Machine Learning-Driven Sentiment Analysis of UK CBDC Tweets Versus Thematic Evaluation of Public and Institutional Perspectives: A Comprehensive Study on the Digital Pound

Guneet Kaur

Professional Doctorate in Data Science

Institute of Computing Science and Mathematics

University of Stirling

February 2025

Abstract

In an era of rapid digital transformation and declining cash usage, central banks worldwide are exploring digital currencies to modernise monetary systems and maintain monetary sovereignty. The proposed digital pound in the UK has ignited intense debate among policymakers and citizens, particularly concerning issues of privacy, surveillance, and economic inclusion. Against this backdrop, this study investigates how public sentiment evolves in response to policy milestones and how it aligns — or diverges — from official Bank of England (BoE) narratives concerning a potential digital pound, addressing a critical knowledge gap in understanding public reception of central bank digital currencies (CBDCs). To meet this objective, the study adopts a novel interdisciplinary approach integrating advanced sentiment analysis using fine-tuned transformer models (DistilBERT, RoBERTa, XLM-RoBERTa), communication theories (e.g., framing theory, agenda-setting theory, and Grunig's two-way symmetrical model), and analysis of policy messaging. A bespoke, domain-specific gold-standard dataset was created and validated, enabling the fine-tuning of these models. RoBERTa, trained for three epochs, emerged as the optimal model for classifying nuanced discussions related to the digital pound.

Longitudinal analysis of public discourse on X (formerly Twitter) across three key periods (2020, 2023, and 2024), corresponding to major BoE policy announcements, revealed an "Exploration–Polarisation–Adaptation" sequence: initial cautious optimism evolved into pronounced negativity, particularly concerning privacy and government control, following major policy announcements, with a partial rebound after official BoE responses. A comparative thematic analysis with official BoE narratives highlighted key discrepancies, notably a "privacy framing gap" where the BoE's technically focused approach to data protection diverged from public concerns over surveillance and government overreach. This mismatch underscores a disconnect between the technocratic framing of policy narratives and public anxieties, pointing to the imperative for two-way symmetrical communication to establish trust.

By illustrating how official narratives can both shape and overlook public views, this study contributes practical insights for policymakers and researchers navigating the complex interplay of technology, policy communication, and public opinion surrounding the digital pound. Recommendations include targeted public engagement on privacy, transparent implementation roadmaps, and a shift towards two-way symmetrical dialogue. While acknowledging limitations related to data source representativeness and the potential influence of external factors, this study provides a comprehensive, empirically grounded understanding of public sentiment dynamics in digital monetary policy. Future research should extend these findings by incorporating broader data modalities and demographic insights, and by applying automated hyperparameter optimisation techniques to further refine understanding of public sentiment dynamics around the digital pound and similar CBDC innovations.

Keywords: *Digital Pound, CBDCs, Sentiment Analysis, Transformer Models, RoBERTa, Temporal Analysis, Communication Theory, Two-Way Symmetrical Communication, Privacy, Government Surveillance, Public Trust, Future Research.*

Declaration

I understand the nature of plagiarism and the University's policy on it. I certify that this thesis reports my original work undertaken as part of the Professional Doctorate Programme, except where I have explicitly cited the work of others, including tweet examples from X *(publicly available)* discussed in Sections 4.3.4, 6.4.3.1, 6.4.4, 7.8.1, and 10.5.1, as well as examples from 2024 Consulting Paper (Section 9.3) and the Technology Paper (Section 9.4), with specific page numbers referenced.

I further confirm that generative artificial intelligence was not used to generate data or compose the final text in this thesis; only Grammarly was utilised for proofreading and grammar corrections.

Stirling, February 2025

Guneet Kaur

Dedication

This thesis is dedicated to the loving memory of my grandparents, Kanwaljeet Singh, Mohinder Kaur, and Gurdial Singh, and my beloved aunt, Manpreet Kaur, whose enduring love, guidance, and unwavering belief in me shaped my life and work. Though they are deeply missed, their presence remains a constant inspiration, and this accomplishment honours their profound legacy.

Acknowledgements

First and foremost, I would like to thank God. Without divine strength, guidance, and grace, none of this would have been possible. My faith has been an unwavering source of hope and perseverance throughout this study's journey.

I wish to express my profound gratitude to my supervisors, **Dr. Saemundur Oskar Harraldson** and **Dr. Andrea Bracciali**, for their invaluable guidance, patience, and insightful feedback throughout the research journey. Their encouragement was instrumental in shaping the direction and depth of my research.

My heartfelt thanks go to my parents — **Jobandeep Kaur**, **Tarwinder Singh**, and **Maninder Singh,** brother **Sehajmeet Singh**, sister **Sukhmeet Kaur** and brother-in-law **Jagjit Singh** for their unwavering love, emotional backing, and confidence in my abilities.

Above all, I am deeply indebted to my husband, **Onkar Singh**, who stood by me through the most challenging moments of this journey, patiently navigating my emotional ups and downs. My beloved daughter, **Asees Eve Kaur**, has been both my inspiration and motivation throughout the study. Pursuing a crucial research degree immediately after giving birth was a formidable challenge, and the three of us have managed to endure and grow through it together.

To my friends, **Dilip Kumar Patairya** and **Suvankar Maity**, thank you for lending a listening ear and offering steadfast support during periods of intense workload and stress. I would also like to thank my colleagues at Cointelegraph for their support during the research period.

Finally, I would like to express gratitude to the broader academic community for the resources, conferences, and open forums that allowed me to learn, collaborate, and share my findings. Through this process, I have grown not only as a researcher but also as an individual, ever mindful that the ability to question, adapt, and navigate the deep waters of complexity is what drives personal growth.

To all who made this work possible: **Thank you**.

List of Publication (s)

The following position paper was published and served as a basis for the literature review in Chapter 2:

- Kaur, G., 2024. Privacy implications of central bank digital currencies (CBDCs): a systematic review of literature. EDPACS, 69(9), pp.87-123.

The following paper has been accepted for publication on May 30, 2025:

- A paper titled '*Comparative Analysis of Transformer Models for Sentiment Classification of UK CBDC Discourse on X*,' based on the work presented in Chapters 4 and 5, has been accepted for publication in the Discover Analytics Journal.

Table of Contents

List of Figures

List of Tables

List of Acronyms

CBDC: Central Bank Digital Currency

BoE: Bank of England

CP: Consultation Paper

TP: Technology Working Paper

NLP: Natural Language Processing

BERT: Bidirectional Encoder Representations from Transformers

RoBERTa: A Robustly Optimized BERT Pretraining Approach

LLM: Large Language Model

LIME: Local Interpretable Model-Agnostic Explanations

EDA: Exploratory Data Analysis

LDA: Latent Dirichlet association

NMF: Non-negative matrix factorisation

TF-IDF: Term Frequency-Inverse Document Frequency

PELT: Pruned Exact Linear Time

EPA: Exploration Polarisation Adaptation

Chapter 1 – Introduction

1.1 Background and Context

The global financial landscape is rapidly transforming, and almost every nation has undergone and participated in a significant digital revolution in the past two decades. Technological advancements alongside evolving societal expectations drive this sweeping transformation. Moreover, the aftermath of the global financial crisis of 2007-2008 acted as a turning point in the history of financial instruments and the broader monetary system, signalling a shift that ultimately resulted in the development of cryptocurrencies, also known as crypto assets [1]. Bitcoin was notably the first decentralised cryptocurrency introduced in 2009 via a white paper titled *"Bitcoin: A Peer-to-Peer Electronic Cash System"* by a person or group of people known by the alias Satoshi Nakamoto [2].

In the early days of cryptocurrencies, neither central banks nor other public institutions paid much attention to Bitcoin or perceived it as an imminent threat. This lack of concern can be attributed to cryptocurrencies' relatively minor influence during their initial years [1]. However, as the ecosystem matured, the number of cryptocurrencies expanded dramatically — reaching 15,882 across 1,192 exchanges as of December 19, 2024, [3], according to data from CoinGecko, a leading cryptocurrency data aggregator. This explosive growth, coupled with the proliferation of scams, fraud, and opaque market practices, prompted central banks and governments worldwide to reconsider their stance on digital assets.

One key response has been the exploration of central bank digital currencies (CBDCs). Unlike decentralised cryptocurrencies like Bitcoin (BTC) and Ether (ETH), CBDCs are digital forms of a nation's sovereign currency, issued and regulated by central banks. In addition, through CBDCs, central banks aim to preserve and reinforce their monetary authority while leveraging technology to improve accessibility to financial systems. The heightened interest in CBDCs also stems from the remarkable rise of private cryptocurrencies - exemplified by Bitcoin's price rise to 14,000 USD in January 2017 [4], which sparked curiosity among central banks, including the central banks of Canada, Japan, Sweden, China, Switzerland, the United Kingdom, and the United States, to investigate the design, implications, and potential adoption of digital fiat currencies. The COVID-19 pandemic is another major factor that led to a substitution effect on various payment forms like cash as customers realised the relevance of digital payment instruments [5]. For instance, cash and coin payments in the United Kingdom saw a significant 35% decline in 2020 [6]. In the same year, there was a 12% increase in contactless payments, constituting more than a quarter (27%) of the total payment transactions [6]. Fast forward to late 2024, only 8% of UK adults reports exclusively using cash, according to NatWest data [7]. While digital payments dominate, cash usage continues to decline, with the volume of cash payments in the UK dropping by 7% in 2023 to six billion, compared to 6.4 billion in 2022 [8]. Cash accounted for just 12% of all payments in 2023, down from 14% in 2022, returning to 2021 levels after a brief rise in 2022 [8]. Despite this decline, cash remains vital for specific needs, with 54% of vulnerable individuals deeming it "essential," up from 47% in 2022. Notably, 76% of digital payment users still carry cash for emergencies, with 48% doing so often or always [7]. These statistics highlight a transitionary phase in consumer payment preferences where cash retains residual importance while digital options predominate.

Thus, CBDCs are a strategic response to the pressures exerted by the rise of private cryptocurrencies, fintech innovations, and changing consumer behaviours. In addition, central banks, such as the Central Bank of Bahamas, which introduced the Sand dollar (CBDC), often mention financial inclusion, improving payment efficiency, strengthening national defence, reducing service delivery costs, modernizing payments and maintaining trust in national currencies as other drivers of CBDCs [9]. Above all, a CBDC is a regulatorily compliant alternative that aligns with contemporary payment preferences, supporting broader monetary policy objectives and everyday financial activities.

1.1.1 The Digital Pound Initiative in the UK

In this evolving monetary landscape, the Bank of England (BoE) in the United Kingdom has considered the digital pound - colloquially referred to by some as “Britcoin,” mirroring the global trend of central banks assessing CBDCs as instruments of payment efficiency and modern monetary frameworks. While the bank has no plans yet to introduce a digital pound, it started exploring the merits and potential design with its 2020 Discussion Paper [10], followed by the 2023 Consultation [11] and Technology Working Papers [12]; these documents have stimulated extensive discourse, featuring contributions from academics, industry experts, financial institutions, policymakers, and the wider public. The bank also published a response to the public feedback to Consultation and Technology Working Papers in January 2024 [13], [14]. However, as the BoE explores next steps, some scepticism or concerns regarding potential government overreach and data privacy have surfaced in mainstream UK news, social media, and public feedback [13], [14]to BoE’s policy documents, complicating the prospective rollout.

To ensure that the digital currency solution aligns with national objectives, public trust, risk management principles, and the evolving demands of a globalised economy, the BoE and HM Treasury are working to future-proof their monetary policy and payments infrastructure, as reflected in the ongoing dialogue. This further serves as a microcosm of a global shift toward understanding how CBDCs might integrate within existing financial systems and what that integration means for retail consumers, businesses, and the central banking mandate. As the UK and other nations weigh with these questions, public sentiment and discourse around such digital currency initiatives become integral to guiding policy design and fostering equitable, well-informed adoption.

1.1.2 The Role of Public Sentiment in Policy Design

Public sentiment has long been recognised as one of the key drivers influencing the direction and emphasis of public policies. The notion of “policy mood”— a measure that reflects the public’s underlying preferences, often guides policymakers to ensure enacted policies resonate with the broader public’s evolving attitudes - has received considerable attention in the literature [15].

In the context of finance and fintech, understanding public sentiment becomes even more crucial as traditional regulatory approaches can be outpaced because of rapid technological innovation and market shifts. Studies have demonstrated that public opinion substantially influences government policy, as observed by [16] and the introduction of new financial instruments and financial regulation often responds to expert recommendations and public demands for transparency, stability, and consumer protection [17]. This implies that sentiment analysis is no longer just a commercial tool or for marketing purposes; it can also be used to assess how public opinion shifts around important monetary policies, which may hinder or expedite reforms.

Moreover, public sentiment shapes how the public receives new technologies and financial solutions. For instance, a mix of scepticism, enthusiasm, and calls for oversight is often seen with decentralised finance, cryptocurrencies, and digital payment platforms [18], [19]. With the rise of government-backed digital currencies, this dynamic has expanded to CBDCs, requiring regulators to weigh the public's trust, comfort level, and perspective about new forms of money, economic benefits and technical feasibility. All these factors feed into the policymaking process and guide decision-makers to design policies that serve intended economic functions and align with societal expectations and normative preferences [20].

In the case of the UK's digital pound initiative, ongoing public discourse on platforms like X (formerly Twitter) offers real-time insights into the shifting policy mood. By analysing the public's perspective using advanced analytical methods, policymakers can transcend top-down expertise and integrate a grounded awareness of how citizens view new financial proposals. Moreover, introducing a digital pound is not merely a technical undertaking; it is a socio-economic transformation with profound implications for various stakeholders, including individuals, businesses, and the financial system as a whole. Public perception and acceptance are crucial for successfully adopting and integrating any new form of currency [21]. Therefore, understanding public sentiment, concerns, and expectations is paramount for policymakers seeking to navigate this complex transition.

1.2 Research Problem and Motivation

1.2.1 Challenges in Understanding Public Discourse on CBDCs

Analysing the social media discourse around complex financial concepts like CBDCs is challenging, as public sentiment is often volatile and influenced by policy announcements and broader macro events, such as interest rate changes, inflation reports, or major financial institution failures. Understanding financial discourse requires distinguishing between genuine concern, informed critique, and misapprehension. Furthermore, traditional sentiment analysis methods, including basic lexicon-based or simple machine learning approaches often struggle or even fail to capture domain-specific jargon (e.g., "CBDCs, stablecoins, cross-border settlement, or privacy-enhancing technologies.")[22]. Hence, more sophisticated analytical frameworks become necessary to handle domain-specific language effectively.

1.2.2 Need for Advanced Analytical Approaches

Advancements in natural language processing (NLP), particularly the emergence of transformer-based models, offer promising avenues for addressing the analytical complexities discussed in Section 1.2.1. Models such as DistilBERT[1], RoBERTa[2], and XLM-RoBERTa[3] have demonstrated state-of-the-art performance in sentiment classification across various domains [23], [24], [25]. By fine-tuning these pre-trained models on a domain-specific, gold-standard dataset, researchers can enhance their capacity to detect context-dependent sentiments, complex evaluative language, and subtle emotional tones. Such approaches enable a more accurate and granular understanding of public discourse, facilitating insights into how sentiments evolve over time and in response to key policy events.

---

1 https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student

2 https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

3 https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

1.2.3 Bridging the Gap Between Public and Official Narratives

A critical dimension of this study involves going beyond sentiment classification and analysis by comparing public discourse to official BoE narrative or communications. Although BoE's motivation and vision for introducing a digital pound are clearly laid out in policy documents, it remains crucial to analyse how the public interprets, emphasises, or questions these proposals. This research aims to identify areas of convergence or alignment and divergence by juxtaposing narratives and themes present in official documents with those emerging from social media. The resulting insights could inform policymakers about prevalent concerns, refine [11], [12], [13], [14]communication and messaging and adjust policy to gain public trust and meet their expectations.

1.3 Research Aims and Objectives

The overarching objective of this research is: *To investigate public sentiment towards the digital pound on X using fine-tuned transformer models,* examine its evolution in response to policy milestones [11], [12], [13], [14], *and compare findings with official narratives or communications from the Bank of England and HM Treasury to* derive actionable guidance for policymakers.

1.3.1 Research Questions (RQs)

To achieve this aim, the following research questions are addressed:

- **RQ1: Model Selection and Justification:** Which transformer-based model (DistilBERT, RoBERTa, or XLM-RoBERTa) performs optimally for sentiment analysis on Twitter/X data related to the digital pound when fine-tuned on a domain-specific gold standard dataset, and what are the theoretical and empirical justifications for using fine-tuning in this context?
- **RQ2: Model Capabilities and Limitations:** What are the capabilities and limitations of the selected transformer model (identified in RQ1) in accurately predicting sentiments in the digital pound discourse, and how can its robustness and explainability be evaluated using techniques like LIME and robustness testing?
- **RQ3: Sentiment Trends and Topics Over Time:** What key themes, topics, and sentiment patterns emerge from multifaceted analysis of Twitter/X data concerning the digital pound across three major timelines linked to BoE policy events, and how do sentiments, emotions, and semantic relationships shift in response to these events? What patterns of change appear from the temporal analysis?
- **RQ4: Analysis of Official Communications/Narratives:** What are the key themes and narratives presented in the official Bank of England policy documents and responses related to the digital pound?
- **RQ5: Comparative Analysis and Communication Theories' Lens:** What alignments and discrepancies exist between public concerns expressed on X and the narratives in BoE policy documents, and how do established communication theories (e.g., framing theory, agenda-setting theory) explain these alignments and discrepancies? What are the implications of these findings for effective policy communication?

1.4 Significance of the Study

1.4.1 Theoretical Significance

This study contributes to the scholarly dialogue on financial policy, sentiment analysis, and NLP by delving deeper than general-purpose sentiment classification and advancing the understanding and use of fine-tuned transformer-based models in specialised, domain-specific contexts, especially in emerging fields like CBDCs. It underscores how transformers can decode intricate financial discourse and track policy mood in real time.

1.4.2 Methodological Significance

The research demonstrates that sentiment analysis of policy-sensitive text benefits from a multi-layered analytic design. Firstly, it demonstrates the utility of combining advanced transformer-based sentiment analysis with temporal trend exploration and established interpretability techniques such as LIME (Local Interpretable Model-agnostic Explanations) through a robust framework, providing a more comprehensive and nuanced understanding of public sentiment dynamics than would be possible with any single method alone. This is achieved by employing rigorous evaluation methods, including creating a domain-specific gold standard dataset, a comprehensive comparison of transformer models, and robustness testing. The gold standard dataset ensures that the models are trained and evaluated on data that is representative of the specific language and context of digital pound discourse. The comparative analysis of different transformer models allows for the selection of the optimal model for this particular task. Finally, the use of robustness testing addresses the critical need for explainability and robustness in NLP applications, especially in sensitive domains like finance and policy. This methodological rigor can serve as a valuable guide for future research endeavours in similar or adjacent fields, promoting best practices in NLP-based sentiment analysis for financial policymaking. Finally, the comparative element of this research, contrasting public sentiment with official narratives, adds another layer of methodological complexity and value, requiring the careful integration of qualitative and quantitative data analysis techniques.

1.4.3 Practical Significance for Policy and Communication

The insights generated from this research can be used in practical contexts such as digital pound policymaking by allowing policymakers to understand public sentiment trends, significant themes in public discourse [20], and communication gaps. The application of communication theories, such as framing theory and Grunig and Hunt’s two-way symmetrical model, further provides a structured approach for officials to refine policy messaging and engagement strategies, thereby fostering greater transparency, trust, and public acceptance of the centralised digital pound [15].

1.5 Key Contributions of the Study

This research provides several novel and rigorous contributions to the study of public sentiment on the UK digital pound, as outlined below:

- **Domain-adapted transformer modelling for policy discourse:** Existing CBDC sentiment studies have largely relied on general-purpose tools, such as lexicon-based models (e.g., VADER) or off-the-shelf transformer checkpoints, with little evidence on how reliably these methods interpret sentiment in complex, policy-sensitive financial discourse. This research systematically fine-tunes and evaluates three transformer architectures (RoBERTa, XLM-RoBERTa, and DistilBERT) on a domain-specific corpus of UK digital pound tweets. The models exhibit statistically significant improvements over non-adapted approaches (e.g., VADER) in identifying sentiment subtleties, demonstrating that domain adaptation is critical for robust performance in this setting.

- **Creation of a gold-standard UK CBDC tweet corpus:** The literature reflects a lack of annotated datasets for UK-specific CBDC discourse due both to the early stage of real-world implementations and the proprietary nature of financial datasets. This study addresses that gap by developing a manually annotated, policy-relevant sentiment dataset, providing a reusable benchmark that supports reproducibility and facilitates rigorous model evaluation in future NLP research on public opinion in emerging monetary policy domains.
- **Transparent and robust sentiment pipeline:** Responding to calls for explainability in applied NLP, this thesis integrates LIME and adversarial robustness checks into the model evaluation process. These techniques offer insight into feature importance and model stability, thereby increasing transparency and accountability in sentiment predictions applied to financial policy debates.
- **Longitudinal sentiment tracking aligned with Bank of England milestones:** Unlike prior studies that provide static sentiment analyses, this research examines sentiment evolution over time by aligning tweet data with key policy events. This longitudinal analysis offers empirical insights into how public attitudes shift in response to official communications. Notably, this study identified an "Exploration–Polarisation–Adaptation" pattern, illustrating the dynamic nature of public sentiment throughout the policy development process. This temporal dimension adds depth to the understanding of sentiment dynamics in relation to policy developments.
- **Theory-grounded comparative narrative analysis:** While earlier work typically isolates either public or institutional narratives, this study directly compares public discourse on X (Twitter) with official communications from the Bank of England and HM Treasury. It applies framing theory, agenda-setting, and two-way symmetrical communication models to interpret areas of convergence and divergence, marking the first theory-informed comparative discourse analysis of the UK's digital pound communication strategy.
- **Methodological synthesis for policy communication research:** By combining transformer-based sentiment analysis with qualitative thematic analysis in a unified framework, the thesis provides a scalable and transferable mixed-methods approach for analysing complex public discourse in other high-stakes financial and regulatory contexts.
- **Practical insights for CBDC stakeholders:** This study reveals a critical disconnect between public concerns and institutional narratives surrounding the digital pound, particularly on issues of privacy, surveillance, and financial inclusion. While official communications tend to frame privacy in technical terms, public discourse reflects broader anxieties about surveillance and state control, a divergence identified as a "privacy framing gap." By conducting a theory-informed comparative analysis of public and official narratives, the research highlights the need for more inclusive, transparent, and dialogic communication. It offers actionable recommendations, including clearer implementation roadmaps and two-way engagement

strategies, to strengthen public trust and improve the alignment of CBDC policy messaging with societal expectations.

1.6 Scope and Delimitations

1.6.1 Data Selection and Context

This study focuses on English-language tweets drawn from the UK context and aligned with critical policy milestones associated with the digital pound's development. While the UK serves as the primary geographic focus, the methodological approach may hold relevance for other jurisdictions contemplating CBDC implementations. Nonetheless, caution must be exercised in generalising the findings beyond the UK's regulatory, economic, and cultural environment.

1.6.2 Model and Method Limitations

The chosen models and analytical techniques, while advanced, have inherent limitations. Transformer models cannot capture every linguistic nuance or cultural reference, and interpretability methods may not fully explain complex internal workings of the models. Regarding the data source, unlike automatic data collection methods, data was collected manually, and the domain-specific gold standard was created through a robust annotation process to minimise biases associated with Twitter/X data. Still, it is important to acknowledge that the platform's user base may not fully represent the broader population, and the dynamic nature of social media discourse can introduce inherent volatility and context-dependent interpretations. Therefore, the findings of this research will be interpreted within the context of these limitations.

1.7 Thesis Structure

This thesis is organised into the following chapters:

- **Chapter 2 (Literature Review):** Reviews the theoretical and empirical literature on CBDCs, sentiment analysis, transformer-based NLP models, and communication theories, highlighting research gaps this study aims to address.
- **Chapter 3 (Research Methodology):** Details the data collection, and summarise annotation procedures, domain adaptation techniques, and evaluation metrics. It provides a rationale for model selection and outlines the steps taken to ensure methodological rigor.
- **Chapter 4 (Domain-specific Gold Standard Development for Digital Pound Sentiment Analysis):** Describes annotation protocols applied to develop the domain-specific, gold-standard dataset essential for fine-tuning and evaluating the transformer models.
- **Chapter 5 (Experimentation with Transformer Models):** Presents a comparative analysis of DistilBERT, RoBERTa, and XLM-RoBERTa to determine the most effective model configuration for sentiment analysis in this domain.
- **Chapter 6 (Evaluating the Robustness and Explainability of RoBERTa for Digital Pound Sentiment Analysis):** Examines the robustness of the chosen model through adversarial testing and interpretability analyses, employing techniques such as LIME to elucidate model decision-making processes.

- **Chapter 7 (Multifaceted Sentiment Analysis of Digital Pound Discourse Beyond EDA):** Offers initial insights into public discourse topics, sentiment distributions, thematic relevance, semantics and relationships among key terms, setting the stage for deeper temporal and comparative examinations.
- **Chapter 8 (Temporal Analysis of Public Discourse on the Digital Pound):** Explores how public sentiments, topics, and relationships evolve over time, correlating changes with major BoE policy announcements and external events.
- **Chapter 9 (Thematic Analysis of the Public Feedback and Bank of England's responses to the Consultation and Technology Papers):** Identifies and analyses themes within BoE's 2024 response papers, establishing a baseline against which public discourse can be compared.
- **Chapter 10 (Comparative Analysis of Public Discourse on X and Bank of England's 2024 Response Papers):** Aligns and contrasts the themes derived from public discourse with those in BoE documents, highlighting potential synergies and misalignments that inform communication strategies.
- **Chapter 11 (Concluding Remarks and Future Research Directions):** Summarises the findings, assesses their theoretical and practical contributions, and proposes future research directions and discusses policy implications.

Chapter 2 – Literature review

2.1 Introduction

This chapter comprehensively reviews academic and professional literature on CBDCs, the digital pound, domain-specific sentiment analysis using fine-tuned transformer models, and relevant communication theories. Critically, it compares these findings with official communications and policy narratives from the Bank of England and HM Treasury, aiming to provide policymakers with valuable insights and inform public discourse on the digital pound.

The review draws on publications from 2018-2024 (including some seminal earlier works) covering CBDCs, digital currencies, sentiment analysis, NLP (especially transformers), communication theories, and policy communication. Sources include peer-reviewed articles, conference proceedings, book chapters, reports from central banks and international organisations, and studies using computational social media analysis in finance, policy and other domains. Studies unrelated to the core themes or published before 2018 were excluded for this review (unless required for historical context).

2.2 Thematic Structure

For the purpose of this review, the chapter is structured around four key thematic areas, providing a comprehensive and coherent overview of the relevant literature. These themes are:

1. CBDCs and the digital pound.
2. Foundations of sentiment analysis and domain-specific challenges.
3. Transformer-based models for domain-specific sentiment analysis.
4. Comparing official and public discourses: Communication theory foundations.

2.2.1 CBDCs and the Digital Pound

2.2.1.1 The Concept, Types, and Global Landscape of CBDCs

Contrary to popular belief, CBDCs are not a new idea; in fact, they have been around for thirty years. For instance, the Bank of Finland launched the Avant smart card in 1993, marking a milestone in developing electronic money [26]. The Avant system served as a crucial precursor to what is now acknowledged as the world's first CBDC, despite being phased out in the early 2000s. However, as stated earlier, due to quick technological developments and a steady drop in cash usage, CBDCs have recently attracted a rise in global research interest. The confluence of these considerations has prompted central banks in every country to actively investigate the many pros and cons of CBDCs *(to be discussed later in the Section 2.2.1.1.1)*. The growing understanding that CBDCs have the potential to change financial landscapes on a global scale is underscored by the expanding scholarly and practical interest in these digital currencies [27]. A thorough comprehension of the complex ramifications of CBDCs is developing as conversations and investigations proceed, adding to the continuing discussion about the future of digital banking.

In essence, a CBDC is a digital version of a nation's legal tender that the central bank issues and controls, unlike decentralised cryptocurrencies like Bitcoin (BTC) or Ether (ETH), which function outside of established financial systems. According to the BIS's Committee on Payments and

Market Infrastructures, Markets Committee [28], a CBDC is characterised as digital currency issued by a central bank, denominated in the national unit of account, and it signifies a liability on the part of the central bank. This type of CBDC is meant for general-purpose use (i.e., for retail customers) rather than wholesale entities.

It is also crucial to note that, at present, central banks are crucial in issuing two different types of money (i.e., cash and electronic central bank reserves), each of which performs crucial roles in the financial ecosystem. Additionally, central banks play an essential role in facilitating the infrastructure required to operate a third category of currency known as private money. This group mostly includes electronic deposits that are readily available to the public and are kept in commercial banks. Crucially, it is vital to note that commercial bank deposits do not constitute liabilities for the central bank, unlike cash and reserves. A new kind of central bank money would be introduced by establishing CBDCs. However, a CBDC should not be confused with a synthetic CBDC issued by a private entity that issues digital tokens backed by a reserve of assets from the central bank, such as government securities or reserves kept at the central bank. These digital tokens are intended to resemble the features and value of a currency issued by a central bank. Stablecoins, or synthetic CBDCs, are often created to maintain a consistent value about a single currency or a basket of currencies. Nonetheless, given their issuing and backing mechanisms, both are different financial instruments.

Bech and Garratt [29] identified two key types of CBDCs, mainly retail and wholesale CBDCs. Retail CBDCs are made for everyday transactions between people and businesses and are intended for general public use. Users have digital accounts directly with the central bank or other approved intermediaries, and they function according to an account-based structure *(similar to most forms of commercial bank money and balances in reserve accounts)*. In an account-based system, CBDCs are connected to particular digital accounts held by people, companies, or institutions. These accounts keep track of transactions, making tracking and conducting regulatory supervision simple. This strategy directly connects CBDCs and account holders, just like the conventional banking system does. Moreover, retail CBDC transactions are frequently transparent, traceable, and governed by regulations. Retail CBDCs strive to offer a safe and effective way to make regular payments, including purchases, remittances, and bill payments.

Wholesale CBDCs, in contrast, are designed for use among businesses in the financial industry, including banks, payment processors, and clearinghouses. These CBDCs make large-scale interbank settlements and wholesale financial transactions possible. Wholesale CBDCs emphasis on boosting the effectiveness and security of the infrastructure supporting the financial markets while lowering settlement risks. Wholesale CBDCs may use an account-based or token-based system *(similar to cash and many other digital currencies)*. A token-based system involves the issuance of digital tokens that represent CBDCs. Without the use of intermediary accounts, these token transactions take place directly between parties. This system functions similarly to the transmission of actual cash, but digitally. Since transactions involving token-based CBDCs are pseudonymous and don’t reveal the parties' identity, they have privacy advantages. The verification mechanism necessary during exchanges is a key distinction between token-based and account-based money [30]. The capacity of the payee to confirm the legitimacy of the payment object is crucial to token-based systems; however, account-based systems heavily rely on the ability to verify the account holder's identity.

2.2.1.1.1 A Comparative Overview of Global CBDC Initiatives: Motivations, Implementations, and Key Concerns

Countries worldwide are not uniformly exploring CBDCs; their pace of adoption and approaching digital versions of native fiat currencies is different. The Atlantic Council's CBDC tracker (as of September 2024) provides a useful framework for categorising these efforts into distinct stages of implementation (from launched to inactive/cancelled), reflecting uneven progress and commitment among different countries [31]. As per the Tracker, at least 134 countries are actively exploring potential CBDC models, with 44 nations progressing to pilot phases involving real-world testing and limited-scale implementations (e.g., Russia, Iran, Norway, and Brazil), underscoring the broad appeal of CBDCs and their perceived role in modernising monetary frameworks [31].

Moreover, 3 nations — The Bahamas, Nigeria, and Jamaica — have already launched live CBDCs, each reflecting distinct policy objectives and contextual nuances. The Bahamas' Sand Dollar, for instance, targets financial inclusion across dispersed island communities [32], whereas Nigeria's eNaira aims to reduce reliance on cash transactions and drive more robust digital financial services [33] and Jamaica also mentioned financial inclusion as its key motivation to introduce the Jamaican Decentralised Exchange or JAM-DEX [34]. In parallel, 20 countries, including Georgia, Canada, and Peru, are currently at the development stage, focusing on designing technical architectures, regulatory frameworks, and stakeholder engagement strategies [35]. Another 20 nations are at a research stage, including the UK, Egypt, Argentina, Nepal, and Hungary, where discussions often revolve around privacy, interoperability, and aligning digital currency solutions with specific national priorities and the public sentiment Some of these research-phase countries, like the UK, also engage in public consultations [11], [12] and face varying degrees of public resistance [13], [14], indicating that consumer attitudes and social acceptance will play a significant role in shaping CBDC outcomes, an issue explored further in this thesis's focus on sentiment analysis. Notably, 21 nations have chosen not to pursue active CBDC development (e.g., Uruguay, Kuwait, Zambia, and North Korea), citing various reasons such as technological hurdles, policy constraints, or concerns over economic impact [35]. Finally, two countries, Ecuador and Senegal, have cancelled their CBDC efforts, highlighting the evolving and sometimes uncertain trajectory of digital currency adoption [35].

On the positive side, CBDCs promise benefits, such as greater payment efficiency, financial inclusion, and potentially lower operational costs. Nevertheless, privacy remains a significant concern, especially in account-based systems where breaches of personal data can escalate to identity theft [29]. Choi et al. [36] found that enhanced privacy features significantly increase willingness to use CBDCs, especially for sensitive purchases, suggesting a strong link between data protection and public acceptance. They found that increased privacy features and information on CBDC's privacy benefits greatly improve willingness to use it, especially for making purchases of privacy-sensitive products, with a possible 60% increase, through a randomised online survey with 3,500 participants.

Some scholars further confirm that countries differ widely in how they approach data privacy and regulatory oversight. Kshetri and Loukoianova [37] highlights several CBDC regulatory frameworks (for data protection) used in Cambodia, China, Japan, the Marshall Islands, South Korea, and Thailand. The authors noted that in Cambodia, institutions manage user data and store it separately for transaction privacy, whereas in China, every transaction detail, including amount and parties' identities, can be recorded by the government. In Japan, commercial banks handle nodes independently, unable to access other institutions' user data, and in the Marshall Islands, SOV blockchain avoids centralising personal information by allowing users to choose trusted verifiers for cryptographically signed SOV IDs [37]. On the contrary, South Korea is piloting privacy features, employing private institutions for users' electronic wallets, and Thailand's Aztec technology ensures participant-only access to transaction details, while concerns persist about the Bank of Thailand's transaction monitoring capability. The authors also noted that in Europe and several Asia-Pacific nations like Australia, Japan, New Zealand, and South Korea, there are strict laws governing data privacy [37]. Nonetheless, Chinese consumers do not have strong privacy rights. Concerns are raised by the Chinese government's use of digital currency electronic payment (DCEP) as a mechanism to manage illicit financial flows [38], [39].

Meanwhile, case-specific developments, such as Nigeria's Naira redesign policy, which was introduced in 2022 to promote digital payments, curb illicit activities, and enhance monetary integrity, faced significant challenges, including cash shortages, inefficient digital infrastructure, and low public trust, compounded by inadequate public sensitisation and tight deadlines [40]. As a result, these issues disrupted Nigeria's economy, with citizens facing financial hardships and businesses suffering reduced productivity. In a different context, Norges Bank has focused on the declining use of cash in Norway as a primary motivation for exploring CBDCs to safeguard essential payment system functionalities in a progressively cashless society [41]. This pursuit reflects a broader concern about ensuring continued public access to central bank money in the digital age. Yet, the technological implications of implementing DLT and programmable money within a centralised framework remain partially understood, necessitating careful consideration of potential risks [42]. A primary concern surrounding CBDCs centers on privacy. In the context of Norway, while CBDC implementation *per se* does not necessarily violate the European Convention on Human Rights (ECHR) Article 8 or Article 102 of the Norwegian Constitution [41], robust privacy safeguards are paramount to protect people's rights regarding personal data processing within a CBDC system.

Against this backdrop, Náñez Alonso et al. [43] investigated optimal conditions for CBDC implementation across various countries, identifying those best suited for adoption based on predefined criteria. They highlighted several promising regions and countries based on similarities to early adopters like the Bahamas and China: the Baltic Sea area (Lithuania, Estonia, and Finland) in Europe; Uruguay and Brazil in South America; Malaysia in Asia; and South Africa in Africa. However, the authors stressed that CBDC success hinges on public acceptance; they argued that while CBDCs contribute to societal digital transformation [43], their value proposition requires further investigation, particularly regarding user benefits compared to existing digital payment methods, paralleling with the findings of Ahiabenu and Olaleye [35]. This raises the critical question of whether the substantial investment in CBDC infrastructure is justified by its advantages in a rapidly evolving payments landscape with shifting consumer preferences.

In the Bahamas, for instance, the above question proves especially relevant given its rollout of the Sand dollar. Branch et al. [27] evaluated Sand dollar's implementation, including key design elements such as wallet-balance caps, user-friendly interoperability (e.g., linking Sand dollar wallets to the Automated Clearing House), and KYC protocols. They found that initial adoption remained low; however, incremental uptake was observed with subsequent educational campaigns and technological enhancements [44]. This demonstrates the importance of robust public awareness and reliable infrastructure. Yet, as with all CBDCs, long-term success depends on how effectively they can address persistent challenges (e.g., data security, user autonomy) in a context where physical cash remains culturally and economically significant.

In essence, CBDCs pose complex challenges related to privacy, governance, and user autonomy, even if they are poised to address issues of efficiency, financial inclusion, and modern payments. Since trust frequently depends on strong data security, user-friendly design, and transparent regulation [45], public sentiment will probably determine whether these digital currencies will eventually flourish or stall. Thoroughly understanding these factors can guide policymakers and stakeholders in designing CBDCs that not only meet technical and economic objectives but also address public concerns.

2.2.1.2 Digital Pound in the UK Context

In the United Kingdom, discussions around a prospective "digital pound" have garnered increasing attention, prompted by both technological shifts in payments and broader concerns about the future of cash usage. The BoE and HM Treasury have released a series of documents outlining preliminary visions, design considerations, and stakeholder consultations. A significant milestone was the 2020 Discussion Paper, which delineated potential models and operational frameworks for a UK CBDC [10]. This document highlighted key motivations such as ensuring continued access to central bank money, bolstering the resilience of payment systems, and fostering innovation within the financial sector. Expanding on these fundamental ideas, the authorities released 2023 Consultation and Technology Papers, which thoroughly examine design architectures, use case scenarios, and possible monetary policy ramifications [11], [12]. These publications reflect a cautious yet progressive approach. Notably, the BoE emphasises sustained access to UK central bank money and promotes innovation, choice, and efficiency in domestic payments as primary motivations for the digital pound. Officials emphasise that any planned CBDC should be a supplement to traditional forms of money rather than a straight replacement, even though suggestions recognize the diminishing relevance of actual currency in daily transactions [13], [14]. In addition to government engagements, industry input has also been crucial. For example, the significance of ensuring interoperability between the digital pound and private-sector payment innovations like open banking efforts has been emphasised by financial market actors and fintech companies [14].

The 2024 response papers from the BoE and HM Treasury to the public consultation (Bank of England and HM Treasury, 2024 provide further insights into the evolving debate [13], [14]. For instance, the response from UK Finance highlights cautious support for investigating a digital pound, but it also mentions the necessity of a comprehensive cost-benefit analysis and strong cooperation with business partners. Members doubt whether a standalone CBDC is the best way to accomplish BoE goals, but they see potential in utilising current infrastructures, such as programmable wallets or multi-asset ledgers. Important areas requiring more research include holding limitations, economic incentives for middlemen, and privacy assurances [46]. The response urges the BoE to work with industry partners on continuing research and iterative design, advocating for a “marathon, not a sprint" approach [46]. On the other hand, Goodell [47] criticises the intentions and the suggested design of the digital pound from an academic standpoint, pointing out flaws in user control and privacy assurances. The paper questions whether a centralised “core ledger” accurately reflects the benefits of real currency, primarily anonymity and direct custody, and advocates for privacy-enhancing technology (such as blind signatures and zero-knowledge proofs). Goodell also raises concerns about custodial wallets, suggesting they risk morphing the digital pound into an “account-like” structure rather than a genuine analog of cash. To allow end-users to store and transfer digital pounds without undue reliance on middlemen, the paper advocates for a more decentralised or permissioned model. To maintain a streamlined narrative, the detailed thematic analysis of these 2024 response papers — covering specific codes, emergent themes, and cross-references with public discourse — will be presented in Chapters 9 and 10 respectively.

Together, these perspectives indicate a tension between regulatory compliance (e.g., user identification, AML/KYC mandates) and individual autonomy (e.g., offline capability, privacy-as-a-right). Regardless of stance, public opinions underscore the importance of transparency, iterative testing, and further public-private discourse to refine the digital pound’s design.

2.2.2 Foundations of Sentiment Analysis and Domain-Specific Challenges

2.2.2.1 Lexicon-Based Methods

Sentiment analysis, also known as opinion mining, is a field of natural language processing (NLP) that trains computers to understand text like humans and aims to identify and extract subjective information from text data [48]. Early approaches to sentiment analysis relied on lexicon or rule-based methods to assign polarity (positive, negative, or neutral) to text data [48], [49], [50]. These methods utilise predefined dictionaries of words and their associated sentiment scores. Although these methods were simple to implement as no labelled training dataset is required and proved effective for certain domains (e.g., product reviews), they often struggle with more complex contexts like domain-specific language and social media feeds, which introduce diverse linguistic styles (e.g., sarcasm, emojis, slang). For instance, general Lexicons like General Inquirer catalog broad sets of words as “positive” or “negative," offering extended coverage and immediate usability. However, they may fail to capture specialised usage or emotive slang [51].

In addition, Min and Zulkarnain [52] provided empirical evidence that VADER (Valence Aware Dictionary and sentiment Reasoner), tuned for social media text, performs better than TextBlob on informal texts like tweets. Yet, it can still underperform if domain-specific financial or policy terms are absent from its dictionary. This is because what’s considered positive in one domain could be classified as negative in another. For instance, words like “huge” and “small” can have different meanings depending on the context under consideration [53]. This disadvantage poses challenges to its usage in specialised fields like finance and policy, where "The pound remains strong," typically conveys positive sentiment, whereas "The BoE supports a strong tightening policy," could indicate negative sentiment.

To bridge this gap, some researchers have utilised additional lexical resources like SentiWordNet [54], [55]. It is built from the WordNet database and assigns each WordNet synset a positive, negative, and objective score distribution. However, this resource could be "too generic" and may not detect domain-specific words, such as in the case of an online textual review of Booking.com data, as discussed in the literature [55]. Conversely, Rutkowska and Szyszko [56] look at how well different sentiment lexicons analyse communications from central banks. They examine four lexicons, concentrating on dictionary-based approaches: two domain-specific (Apel and Grimaldi, 2014; Bennani and Neuenkirch, 2017), an economic and financial (Loughran and McDonald, 2011), and a generic (Minqing and Bing, 2004) lexicon [57], [58], [59], [60]. The study analyses lexicon performance through content alignment, sentiment detection accuracy, and mutual information using both qualitative and empirical testing on monetary policy releases from 15 small open economies. The findings show that all lexicons successfully convey the intentions of central banks. Notwithstanding this, domain-specific lexicons have inherent limitations, such as limited applicability across different central banks or economic environments [61]. Also, they may not incorporate emerging terms and nuanced expressions with the evolution of economic language. Another study analyses sentiment in ECB and Fed communications, finding significant persistence and speaker-specific effects, but also revealing that most variation is unexplained, questioning the reliability of dictionary-based sentiment analysis [62]. Thus, the choice of lexicon is contingent upon particular research objectives and contextual elements.

2.2.2.2 Classical Machine Learning Approaches, Deep Learning Techniques, and Hybrid Models

Consequently, machine-learning classifiers, such as Support Vector Machines (SVM) and Naïve Bayes (NB), which train on annotated corpora of extensive social media or product review datasets have become prevalent, showing high accuracy under controlled conditions. Nevertheless, when transitioning from controlled research settings to real-world deployments, these widely used approaches face notable drawbacks. For instance, Rahat et al. [63] demonstrate that SVM and NB can achieve high accuracy on reasonably well-structured and balanced datasets. However, their performance tends to suffer if the data contains a lot of noise or domain shifts, such as when colloquial language changes or when completely new slang terms appear in social media forums. In these circumstances, the models' robustness may be compromised since they might incorrectly classify reviews because of unrecognised tokens or linguistic formulations that haven't been observed before.

Meanwhile, Guia et al. [64] highlight the substantial tuning costs associated with more complex classifiers, such as random forests. Overfitting on small or unique review sets can achieve optimal accuracy, frequently requiring extensive hyperparameter experimentation (e.g., choosing the number of trees and feature subsets). This is especially evident when working with small, text-heavy e-commerce feedback, as ensemble algorithms may cling onto misleading characteristics due to sparse context and repeated words. If the "real world" introduces new idioms or domain-specific language, the final model may produce impressive accuracy on the training data but poorly on unseen/new data. The study conducted by Banik and Rahman [65] on Bangla textual movie reviews supports the observation that these difficulties are exacerbated in morphologically rich languages. They discovered that unless intensive linguistic preprocessing (such as comprehensive stopword removal or morphological stemming) was incorporated, both SVM and NB had trouble with inconsistent or incomplete token representations. Without these improvements, the relevant signals that classifiers rely on were diluted by word variations caused by complicated inflection, and feature sets exploded with near duplicates. This issue is particularly problematic in fields like movie reviews, where the token space is further complicated by slang or transliterations specific to a certain region.

Yet, recent research by Suasnawa et al. [66] and Yogi et al. [67] emphasises that SVM, NB, and other machine-learning techniques may perform poorly if contextual or semantic subtleties are not fully considered. For example, Suasnawa and colleagues discovered that unless domain adaptation was handled, the performance difference between NB and SVM widened on highly informal Twitter data about online learning — a domain full of newly introduced or altered language during the pandemic. Likewise, Yogi et al. [67] underscore how "comparative studies" of multiple ML techniques are crucial for fine-tuning hyperparameters, dealing with out-of-vocabulary terms, and ultimately enhancing classifier accuracy across diverse social media platforms.

Building upon these challenges, the literature increasingly gravitates toward deep learning techniques like Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM), and Convolutional Neural Networks (CNN), often paired with word embeddings such as Word2Vec and FastText. These methods do, however, have certain advantages and disadvantages. To bridge this gap, Salur and Aydin [68] provide a brand-new hybrid deep learning model that combines various deep learning architectures (CNN and BiLSTM) with word embedding techniques (Word2Vec, FastText, and character-level) in a synergistic manner. Their method outperforms current algorithms on a Turkish Twitter dataset, demonstrating improved feature extraction and classification accuracy.

In a similar vein, Islam et al. [69] thoroughly examine current deep learning architectures for sentiment analysis, carefully weighing their advantages and disadvantages. From a comprehensive comparative investigation, they conclude that capsule-based RNN techniques are better than classic CNN and RNN models, with an impressive accuracy of 98.02%. The researchers also present the novel CRDC (Capsule with Deep CNN and Bi-structured RNN) model, which performs exceptionally well on various benchmarks. In particular, on datasets like IMDB (88.15%), Toxic (98.28%), CrowdFlower (92.34%), and ER (95.48%), the CRDC model performs better than current techniques, highlighting how advanced deep learning models can improve automated sentiment analysis applications' accuracy and efficacy. Nonetheless, Sharma et al. [70] thoroughly examines sentiment analysis as a game-changing NLP tool in domains such as social media, commerce, healthcare, and disaster relief and notes the shortcomings of the deep learning techniques used today, like recurrent neural networks (RNNs), CNN, LSTM, GRU, and BiLSTM, in managing complex emotions and computational effectiveness. This work calls for future research to focus on creating sophisticated hybrid models, enhancing the diversity and quality of data, incorporating explainable AI for more transparency, and using generative models to improve sentiment modelling and dataset augmentation.

Together, these findings show that even with advanced deep learning methods, systematic domain adaptation is necessary for highly shifting social media or domain-specific data. This could mean: (i) continuously extending specialised lexicons to encompass evolving slang using rigorous data preparation (e.g., morphological stemming, synonyms unification), (ii) methodically fine-tuning the classifiers to reduce overfitting.

Current models risk overlooking the nuanced, context-dependent sentiment indicators essential for understanding central bank communications and public discourse around financial policy innovations [62]. Importantly, there remains limited exploration of systematic domain adaptation in contexts like CBDCs or other cutting-edge financial discourse, where policy nuances and jargon evolve continuously, and policy mood plays a critical role in shaping public reception.

Consequently, this study aims to bridge this gap by using advanced NLP techniques, including fine-tuned transformer-based models — ultimately addressing the crucial need to handle context-dependent sentiment indicators in emerging fields such as financial policy communication. Ultimately this approach informs policy design and stakeholder engagement in the context of the digital pound and other similar monetary innovations.

2.2.2.3 Nuances in Financial and Policy Discourse: Public Sentiment, Monetary Innovation, and the Role of Social Media

Unlike traditional domains (like movie or product reviews), financial and policy discourse frequently uses complex grammatical structures and specialised terminology to express emotions [71]. Scholars have long noted that the jargon used in finance and policy is full of acronyms, presumptions about common background, and references to institutional practices, all of which can obfuscate or change meaning for those unfamiliar with the subject [72], [73]. In this context, conventional sentiment classification can be complicated by the frequent use of jargon, acronyms, or references to context-specific terms (such as “interoperability solutions," "privacy-enhancing technologies,” “two-tier architecture,” or “cross-border payments”) by analysts or stakeholders discussing novel monetary policy tools, or financial privacy regulations [74]. As a result, duality in interpretation arises because financial discourse is rarely value-neutral, and terms must be viewed through a prism of economic, institutional, or political objectives.

Furthermore, implicit sentiment is commonly present in these texts. For instance, a phrase that seems “positive,” like “adoption,” might have a negative connotation when used to describe the use of a decentralised cryptocurrency or competitive stablecoin that challenges the function of central bank currency [74]. On the other hand, neutral terms like “interoperability” could be interpreted positively if they indicate that a CBDC will integrate seamlessly with current payment rails, or negatively if they imply reliance on private sector infrastructure that is often considered hazardous. This interplay challenges systems that rely on broad sentiment labels or are purely supervised or generic lexicon-based and may misclassify statements. For example, Siregar et al. [75] demonstrate that policy texts in the monetary and fiscal realms are far from neutral constructs. Instead, any textual component (such as sentence form and phrase choice) may act as a “hegemonic tool,” allowing policymakers or drafters to hide or selectively emphasise particular groups’ interests and ideological positions. This argument supports earlier findings of Section 2.2.2.1 that domain-specific sentiment lexicons and strong domain adaptation techniques are essential; otherwise, standard classifiers risk incorrectly classifying such statements because of the political language they contain [67]

In their exploration of this change, Masciandaro et al. [76] draw attention to the move away from secrecy and toward active social media participation, especially on platforms like X (previously Twitter). They discussed how central banks now use social media to communicate monetary policy pronouncements and shape market expectations, reaching both professional and non-expert audiences. Moreover, according to their research, high-frequency social media data provides useful insights into public opinion, financial market movements, and the efficacy of central bank communication methods, which use computational text analysis and machine learning. However, the researchers also highlight the inherent difficulties in ensuring clarity and sustaining consistent messaging when dealing with varied audiences as social media participation expands outreach. Consequently, sentiment becomes even more complex due to this conflict between precise communication and wide distribution, especially in the context of sentiment analysis of conversations about digital currencies.

In addition, policymakers might frame new monetary instruments or rules to highlight advantages (like technological innovation or financial inclusivity) while downplaying disadvantages (like privacy issues or operational complexity). Although political communication theory has extensively documented such “framing” or “agenda-setting” procedures [77]; however, there is no evidence of their application within central banking discourse. According to framing theory, how a topic is communicated to the general audience can significantly impact how it is understood [77].

2.2.2.3.1 Transition from Technocratic to Public Engagement

Traditionally, central banking was frequently seen as a technocratic field [78], insulated from public opinion and directed toward specialist audiences — particularly financial markets [79]. The necessity to preserve independence and prevent political interference frequently justified the “veil of secrecy” surrounding central bank operations [79]. Nonetheless, it is becoming increasingly clear that public opinion greatly influences monetary policy adoption [80] and eventual effectiveness, particularly regarding important innovations like CBDCs. This is because the legitimacy and efficacy of novel monetary instruments are inextricably linked to public understanding and acceptance.

However, households and non-financial businesses often lack the time, financial literacy, and incentives to look for and analyse information related to monetary policy [78]. This information asymmetry makes it difficult for central banks to interact with the public. Along with addressing public concerns and fostering trust, they must figure out how to clearly and easily convey complicated technological topics. Thus, central banks’ policy communication and its impact on the public is not simply a matter of public relations; it’s about the legitimacy and efficacy of novel monetary instruments.

The influence of stakeholder feedback on policy design and acceptance has been highlighted by several studies. For instance, Korhonen and Newby [81] analysed that some central banks, including the European Central Bank (ECB), the Bank of Canada, and the Federal Reserve System, have held “listening events” to gather public input during strategic reviews and early evidence hints that such events can boost perceived legitimacy among participants [82]. Nevertheless, research on these efforts regarding digital currencies is sparse, and it remains unclear whether they substantially reshape public knowledge or attitudes in the long run.

2.2.2.3.2 Public Trust as a Linchpin of Monetary Innovation

Whether monetary innovations, like CBDC, are accepted or viewed with suspicion depends on public trust in central banks, which is essential for democratic accountability and policy efficiency [79]. The public is typically more receptive to new monetary tools or innovations when a central bank is seen as reliable and trustworthy. On the other hand, regardless of the technical advantages of new policies or technology, poor trust can impede their adoption. While Blinder et al. [79]’s work provides a historical context, more recent research has focused on the specific challenges of communicating about digital currencies. For instance, Auer et al. [83] stress that to increase public confidence in CBDCs, central banks must communicate clearly and consistently. They contend that to ensure broad adoption, central banks must allay public worries about financial stability, privacy, and security. In a speech about the digital euro, Panetta [84] emphasised the value of public participation and consultation during the design phase. The author concludes that the ability of a digital euro to satisfy the demands and aspirations of European residents is what will determine its success.

Broader developments in digital governance also mirror this move toward increased openness and public participation. The way that digital technologies are changing public participation in policymaking has been the subject of several studies. Janssen et al. [85], for instance, look at how citizens might participate in policy making through online platforms. They drew attention to how these platforms might promote more inclusive decision-making, enhance public comprehension of policy concerns, and boost transparency. This is consistent with central banks’ growing focus on public consultation in the digital era. Additionally, public choice theory and behavioural economics studies emphasise how crucial it is to consider public biases and perceptions when designing policies [86]. Recent research has adapted similar findings to the setting of digital currencies. For instance, Shahani and Ahmed [87] highlight how important cognitive biases influence cryptocurrency investment choices, especially in Pakistan. Their results show that the effects of socio-psychological factors like stress, social contacts, and money anxiety are mediated by biases, including herding, overconfidence, and representativeness. Representativeness bias highlights the complex interaction between psychological and social factors in behavioural finance, whereas herding and overconfidence biases influence investing choices [20], [87].

Moreover, research using privacy calculus theory, such as the study by Jabbar et al. [20] investigated how users weigh the potential advantages of CBDC adoption, including convenience, simplicity of use, and credibility against privacy concerns. Even though participants had mixed opinions on data exposure, many were prepared to forgo privacy in exchange for substantial benefits like ease of use, emphasising the necessity for banks to prioritise these considerations to promote CBDC acceptability. This underscores the need for central banks to prioritise understanding public sentiment regarding privacy concerns and incorporate these insights into the design and communication of CBDCs to promote broader acceptance.

2.2.2.4 Social Media as a Mirror of Policy Reception

Social media platforms have emerged as a vital forum for public discourse and opinion formation in today's digital era. They provide a real-time mirror into how the public responds to economic policies, financial rules, and policy pronouncements. This is especially important for financial innovations like CBDCs, which are frequently the focus of heated discussion and scrutiny on social media. Platforms such as X act as a useful source of information for tracking emerging trends, figuring out public sentiment, and monitoring how public opinion changes over time.

Twitter data has been used in numerous studies to understand how the public responds to different economic policies, financial restrictions, and policy announcements. Bollen et al. [88] conducted groundbreaking research showing a relationship between stock market fluctuations and sentiment on Twitter. This study demonstrated how social media data can serve as an early market sentiment indicator. More recently, using Twitter data, scholars have examined public responses to certain policy events, such as Brexit and the COVID-19 epidemic. For instance, Chandio and Sah [89] examine Twitter sentiment regarding Brexit and UK MPs to predict election results. They looked at tweets on Brexit, the EU, Theresa May, and Jeremy Corbyn using the Twitter API and discovered a shift in sentiment that was primarily unfavourable following the UK Parliament's vote in January 2019. Corbyn's positive sentiment remained stronger as May's support decreased. However, the study highlights the significance of clearly presenting data to non-experts using TextBlob for sentiment analysis and Matplotlib for visualisation. This research highlights social media's growing impact on political discourse, aligning with a study [90]that underscores its role in shaping voter behaviour and democratic processes.

Studies examining public sentiment during the COVID-19 pandemic further illustrate the potential and limitations of social media analysis. One study examines fear-sentiment progression and compares machine learning techniques (Naïve Bayes and logistic regression), another study investigates public sentiment regarding COVID-19 using data from Twitter, which achieves high accuracy for short tweets while highlighting performance limitations for longer tweets[91]. Another study employing sentiment and emotion analysis on tweets about COVID-19 in Singapore [92] uncovered trends in public sentiment and emotions, such as delight and terror, linked to significant events like the circuit breaker. Although the study demonstrated effective communication, including frequent and transparent messaging, it recognises the need for further research to improve methods for examining social media patterns in public health emergencies. This is particularly important in the case of CBDCs, where a complex mix of factors may influence public opinion, including privacy concerns, economic concerns, and trust in institutions.

In contrast to relying solely on computational methods for social media sentiment analysis, thematic analysis in social media research has gained traction to provide richer context to computationally derived insights. A compelling example of this combined approach is provided by Andreotta et al. [93], who used topic modelling to condense a sizable corpus of Australian Twitter data on climate change and then thematic analysis to contextualise the generated subjects within the larger sociopolitical context. Their work illustrated the value of this methodological fusion by exposing unique conversational elements and recurrent themes (including climate action, scepticism, and public debate) that just computational studies would have overlooked.

Various scholars have also adopted this approach for understanding the public reception of complicated policy matters. Thematic analysis, for instance, has been successfully applied in research looking at public opinion during the COVID-19 pandemic to comprehend how the public responded to public health initiatives [94]. By using thematic analysis to group UK tweets from 2018 to 2020 into themes and sentiments, this study showed how public opinion of remote healthcare changed in response to evolving regulations. Similarly, Mustafa et al. [95] demonstrated the use of thematic analysis for locating important themes in pandemic-related Twitter conversations, such as humour, politics, and sentiments. Their research shows how social media data may inform infodemiological studies and offer insightful information on how the public responds to public health crises. More evidence of the usefulness of thematic analysis in comprehending the complex nature of online discourse can be found in Noor et al.'s [96] bibliometric analysis of Twitter research from 2009 to 2018 using thematic analysis, which highlighted Twitter's broader role across a variety of disciplines, including sentiment analysis, education, health, politics, crisis management, and risk communication.

In addition to the disciplines mentioned above, some studies have explored qualitative methods, including thematic analysis, within the broader context of financial technology and markets. For instance, Varma et al. [97] uses thematic analysis to explore Fintech's impact on the banking industry, identifying opportunities and challenges from technological disruption. This demonstrates the applicability of thematic analysis to understanding the broader context within which CBDCs operate. Furthermore, Kosari et al. [98] uses thematic analysis to examine financial traders' self-regulation, emphasising the significance of psychological aspects and market knowledge, which are equally pertinent to public opinion about financial innovations. Similarly, Sahoo et al. [99]uses thematic analysis to investigate the difficulties and consequences of Banking 4.0, offering perspectives on the wider technological and regulatory framework of contemporary banking. These studies show the significance of thematic analysis in comprehending intricate processes within the financial industry, even though they do not specifically address public sentiment about CBDCs.

Furthermore, social media use in relation to digital currencies has been the subject of recent studies. For instance, Prodan et al. [100] performed a sentiment analysis of CBDCs using Twitter, concentrating on shifts in public opinion. Their study helps comprehend changes in consumer views since it offers a temporal perspective on CBDC sentiment. However, the study's methodology does not explicitly address the challenges of handling noisy or irrelevant data, a common issue in social media analysis. Investigating the impact of different noise reduction strategies on sentiment analysis accuracy would be valuable to the existing body of literature. Another study investigates the connection between social media sentiment and CBDC communication and how it affects cryptocurrency markets. Their analysis mainly demonstrates correlations rather than causality, even though they offer insightful information about the relationship between CBDC releases and cryptocurrency price changes [74]. This limitation is crucial because it remains unclear whether changes in sentiment drive market movements or simply reflect broader market trends and external events, opening up room for further research to explore causal relationships that might influence both sentiment and market behaviour.

Furthermore, the scholarship on crypto-currency sentiment exhibits a common architecture, including lexicon seeding, shallow classifiers, and short-horizon validation, yet each study foregrounds different limitations that, in combination, motivate the present CBDC pipeline. Pano and Kashef [101] interrogate 13 tweet-cleaning permutations and show that rudimentary edits (sentence splitting, tag removal) inflate VADER–price correlations over intraday windows. The experiment, however, halts at correlation; transformer baselines, sarcasm tests, and out-of-sample forecasts are omitted, leaving predictive utility unverified. Rouhani and Abedin [18]pursue scale rather than nuance: polarity is first assigned by the same generic lexicon and then recycled into an SVM, a procedure that suppresses crypto-specific slang and propagates lexicon bias. In contrast, Gurrib and Kamalov [102] shift to news headlines and couple sentiment with LDA/SVM to predict next-day Bitcoin direction; accuracy improves when sentiment cues are added, yet the one-day horizon and dichotomous labels ignore intraday volatility and magnitude, while non-linear encoders are absent. Aslam et al. [103] advance to deep learning with an LSTM-GRU ensemble and claim 0.99 sentiment accuracy, but the metric rests on silver labels (TextBlob/Text2Emotion) and heavy undersampling, conditions that overstate generalisability; the link to market behaviour is only asserted, not tested. Taken together, these studies confirm that crowd emotion contains market-relevant signals but also illustrate three unresolved gaps: dependence on rule-based or self-labelled ground-truth, lack of domain adaptation for fast-evolving crypto slang, and weak evidence that sentiment improves real-time forecasting. The present thesis addresses these deficits by constructing an expert-annotated digital-pound corpus, fine-tuning RoBERTa, XLM-RoBERTa and DistilBERT with robustness checks, and tracing sentiment trajectories against Bank of England milestones, thereby converting descriptive insights into an audited, policy-ready predictive framework.

Some studies have compared the performance of machine learning and deep learning models, including BERT and RoBERTa, and analysed sentiment analysis of tweets about CBDC [104]. This methodological approach is valuable since it shows how sophisticated models can increase sentiment accuracy. However, the study lacks significant information on model tuning and validation and does not explore the temporal evolution of CBDC sentiment in response to key policy announcements, which limits its ability to offer insights into how sentiment might fluctuate during significant events. Similarly, Astuti and Alamsyah [105] use the BERT and RoBERTa models to compare the public debate surrounding DeFi and CBDCs. This comparative method is useful because it illustrates the disparities in the public perceptions of DeFi and CBDCs. The study does not, however, specify its data sources precisely, raising questions regarding the representativeness and potential biases of the analysed discourse. Additionally, while the research demonstrated the benefits of fine-tuning, further work is required to develop high-quality, domain-specific datasets for fine-tuning and explore rigorous annotation methodologies to ensure data reliability and validity.

This focus on methodological rigor and robust data analysis is further exemplified by the work of Kulakowski and Frasincar [106], who developed two sentiment analysis tools — LUKE (an emoji-based vocabulary) and CryptoBERT (a refined BERTweet model) — especially for the cryptocurrency space. Despite having multilingual capabilities, LUKE limited sentiment to emoji interpretation, possibly overlooking important contextual details. However, using LSTM models on Chinese Weibo data, a study directly addresses price prediction and shows better results than conventional techniques [107]. However, they ignore the longer-term interactions between market stability, policy, and sentiment in favour of short-term forecasts. Notwithstanding this, another study offers an important empirical contrast, discovering a statistically significant but economically negligible correlation between intraday Bitcoin returns and StockTwits emotion, particularly during bubble periods [108]. This raises the crucial question of whether sentiment on social media, although representing public opinion, offers useful data for trading techniques or policy initiatives.

While the integration of thematic analysis with computational methods has been successfully applied in various domains, and as demonstrated in the broader Fintech literature, there remains a notable gap in its specific application to analysing public sentiment towards financial policy, particularly in emerging technologies like CBDCs. As such, this directly connects to RQ5 of this research, which, as detailed in Chapter 11, uses a comparative framework to compare alignment and divergence between themes derived from public discourse on X with those in official Bank of England documents and further explores these alignments and discrepancies using communication theories.

2.2.3 Transformer-Based Models for Domain-Specific Sentiment Analysis

2.2.3.1 Transition from Traditional NLP to Transformers

The advent of transformer models marked a paradigm shift in the evolution of NLP. Conventional NLP techniques treated text as a bag of disconnected tokens, ignoring word order and context. For example, bag-of-words models ignore word order and context and instead generate a vector representation of a document based on the frequency of each word [109]. Although it still lacks contextual awareness, TF-IDF (Term Frequency-Inverse Document Frequency) improves on this by weighting terms according to their relevance within a document relative to a corpus [110]. Moreover, sequential processing was introduced by RNNs, especially LSTMs and GRUs, which enabled the model to consider word order [70]. Nevertheless, RNNs have drawbacks, including disappearing gradients that make identifying long-range dependencies in text challenging [111]. Section 2.2.2 explored various traditional NLP methods and highlighted their limitations in capturing context and long-range dependencies.

The paradigm shifted significantly with the advent of transformer structures, which were first presented by Vaswani et al. [112] in their groundbreaking study "Attention is All You Need," transformed natural language processing by utilising the attention mechanism. Transformers process all words in a parallel manner, which enables them to better grasp long-range dependencies than RNNs, which process text sequentially. The attention mechanism allows the model to capture contextual linkages and subtle meaning transitions by weighing the relative relevance of various words in a sentence. Accurate sentiment analysis requires this contextual awareness, particularly in specialist fields where language usage and vocabulary can be extremely complex.

BERT (Bidirectional Encoder Representations from Transformers), popularised by Devlin et al. [113], expanded the field even further by integrating bidirectional training. BERT is trained to anticipate masked words in a phrase based on both the preceding and following context, in contrast to earlier language models that were trained unidirectionally. This bidirectional training has resulted in notable advancements in several NLP tasks, including sentiment analysis, and enables BERT to gain a deeper understanding of linguistic context [114].

Several transformer-based models, each with unique benefits, have been emerged following BERT:

- **RoBERTa:** By employing a larger training dataset, a longer training duration, and a different training objective, RoBERTa (Robustly Optimized BERT Pretraining Approach) outperforms BERT, as noted by Liu et al. [115].
- **DistilBERT:** It is a knowledge distillation-trained variant of BERT that is 40% smaller, 60% faster, and lighter while maintaining 97% of BERT's language comprehension abilities [116].
- **XLM-RoBERTa:** By training on a large multilingual corpus, XLM-RoBERTa expands RoBERTa to multiple languages, making it especially appropriate for cross-lingual tasks [117].

2.2.3.2 Architecture, Training Methodologies, and Advantages of DistilBERT

DistilBERT is a distilled version of the BERT (Bidirectional Encoder Representations from Transformers) model. DistilBERT attempts to replicate BERT's output distributions through a process called distillation, i.e., using BERT's output as soft targets and training on the same corpus as BERT. According to Hinton et al. [118], distillation is the process by which a smaller "student" model learns to mimic the behaviour of a larger, more well-known "teacher" model, maintaining a sizable amount of its performance while lowering computational complexity.

The distillation loss, which measures the difference between the student and teacher outputs, is combined with the standard cross-entropy loss in the loss function [118]. Cross-entropy loss is a commonly used loss function in classification tasks, which measures the dissimilarity between the predicted probability distribution and the true labels. Cross-entropy loss penalises confident wrong predictions more heavily to ensure the model is optimised to raise the probabilities of correct classes.

Distillation loss is utilised in the knowledge distillation process. The key idea is that the teacher helps the student understand more complex class relationships by providing "soft targets," which are probability distributions over classes (instead of hard labels). A softened version of the cross-entropy between the student's output and the teacher's output distribution is frequently used to compute the distillation loss.

In essence, true labels' accuracy is optimised by cross-entropy loss. In contrast, the task of ensuring that the student learns nuanced knowledge about the data distribution from the teacher's output is ensured by distillation loss.

DistilBERT is 40% smaller and 60% faster than BERT while retaining 97% of BERT's language understanding skills [116]. Its performance and efficiency balance makes it a compelling option for sentiment analysis tasks. DistilBERT's architecture is similar to BERT's, however it has six layers as opposed to 12 in the base version. This thorough reduction produces a lighter model, which is especially useful for deployment in resource-constrained contexts as it allows for quicker inference and lower memory usage, facilitating real-time sentiment analysis on large volumes of X data [116]. Although DistilBERT, the base model, is trained mostly on English data, multilingual versions, like DistilBERT Multilingual, have been developed, expanding its use to datasets with more than one language [117].

The content of social media users, especially on X, often transcends linguistic boundaries because they frequently engage in code-switching- alternating between languages in a single tweet [119]. DistilBERT's architecture can handle sentiment analysis in multilingual data sets because it effectively handles such scenarios, albeit with certain limitations compared to models explicitly designed for multilingual tasks [120]. Moreover, the primary focus of this study — the UK's CBDC — is predominantly discussed in English, so DistilBERT's monolingual strengths are particularly pertinent. Although multilingual capabilities are not a primary requirement, its efficiency and performance make it a viable option for handling large-scale English-language datasets.

2.2.3.3 Architecture, Training Methodologies, and Advantages of RoBERTa

Robustly Optimised BERT Pretraining Approach or RoBERTa is an optimised variant of BERT, designed to improve performance on various NLP tasks, including sentiment analysis. To achieve higher accuracy and a better understanding of language, it modifies BERT's original methodology. It has become a highly effective tool for sentiment classification, especially in monolingual English contexts, due to its ability to eliminate the Next Sentence Prediction (NSP) objective, longer training durations, larger batches, and more data handling [115]. The NSP task involves predicting whether two sentences are sequentially related. RoBERTa eliminates this based on the observation that it contributes minimally to downstream task performance.

RoBERTa's underlying architecture is the same as BERT, employing a multi-layer bidirectional transformer encoder. However, RoBERTa builds upon the foundations of BERT's original architecture proposed by Vaswani et al. [112]. It does so by using extensive training data and utilising a streamlined training process. RoBERTa's architecture consists of 12 layers for the base model, each with feedforward neural networks and self-attention heads. By focusing on various sentence fragments, the self-attention mechanism enables the model to gather contextual information for deciphering sentiment intricacies [121]. The model can comprehend the text's sentiment more thoroughly due to each layer's intermediate input representations.

Pre-trained on a dataset of 160 GB, which includes sources like Common Crawl, RoBERTa's size surpasses that of BERT [115]. Because of this thorough pre-training, RoBERTa can capture various language patterns and semantic links, and generalise better to unseen data, which is crucial for complex tasks like sentiment analysis. Like BERT, Roberta uses the MLM objective to train the model to predict a predetermined percentage of masked tokens. However, Roberta eliminates the NSP objective, making training straightforward and enabling the model to concentrate more on dependencies at the sentence level [115]. In addition, RoBERTa produces more robust and stable representations by using longer training times and larger batch sizes (up to 8,000 tokens). Moreover, during the training process, RoBERTa dynamically masks tokens for each epoch, i.e., exposing the model to a broader range of contextual predictions, thus improving its ability to capture subtle language nuances, unlike BERT, which uses static masking of tokens during pretraining.

As the datasets under consideration for all three timelines are in English, RoBERTa's monolingual language modelling capabilities allow it to capture linguistic subtleties, such as idiomatic expressions, sarcasm, and colloquial language, more effectively than models trained on multilingual data. Social media platforms like X often contain informal language, slang, and abbreviations. RoBERTa's pretraining on large, diverse English datasets helps it understand such informal patterns better than models trained on more structured datasets.

The use of informal languages, such as abbreviations and slang, is common on social media platforms like X. Unlike models trained on structured datasets, RoBERTa, being trained on diverse English datasets, helps it better comprehend such informal patterns. Furthermore, RoBERTa performs effectively even in situations when there is limited fine-tuning data because of the substantial pretraining. This is especially helpful for tasks like sentiment analysis, where there may not be as much annotated data unlike other NLP tasks.

2.2.3.4 Architecture, Training Methodologies, and Advantages of XLM-RoBERTa (XLM-R)

XLM-RoBERTa (XLM-R) is a multilingual variant of RoBERTa, designed to handle text across multiple languages. Pretrained on a large, multilingual dataset, XLM-R can process text in over 100 languages, making it an ideal candidate for tasks that involve multiple languages or code-switching, which are common in social media settings. This ability to understand cross-lingual and multilingual data gives XLM-R an edge in tasks where sentiment is expressed in different languages or where users mix languages in the same sentence [122].

XLM-R or XLM-RoBERTa is a multilingual variant of RoBERTa, capable of handling text across multiple languages. It is pre-trained on a large, multilingual dataset. XLM-R can process text in over 100 languages; this ability to understand cross-lingual and multilingual data makes it an ideal candidate for NLP tasks involving code-switching or multiple languages, which is a common practice in social media settings [123].

XLM-R employs a transformer-based encoder with 12 layers (for the base model), each with multi-head self-attention mechanisms and feedforward layers similar to RoBERTa. However, its ability to handle multiple languages using a single shared vocabulary (Byte-Pair Encoding (BPE) vocabulary trained on a multilingual corpus) and a unified transformer architecture make it different from RoBERTa. Words across different languages into a shared subword vocabulary are tokenized by XLM-R using BPE, and out-of-vocabulary (OOV) words are handled by representing words as combinations of smaller subword units [117]. It breaks OOV words into familiar subwords, thus improving its performance in low-resource or mixed-language environments. This helps XLM-R predict masked tokens in different languages, allowing it to develop a multilingual understanding of text and making it useful in low-resource environments. For instance, the emotion extracted from English text can be applied to other languages with comparable structures.

2.2.3.5 Domain Adaptation and Fine-Tuning

As discussed in Section 2.2.2.1, domain-specific language is a major challenge for sentiment analysis. This limitation applies to pre-trained transformer models even though they perform well on general-purpose NLP tasks; fine-tuning them to fit particular domains can enhance their performance. The underperformance arises because specific language found in highly technical or policy-driven disciplines is frequently absent from mainstream corpora utilised for initial pre-training, derived from open-domain text like Wikipedia and BookCorpus. Consequently, when using off-the-shelf models in these specialised domains, terms such as “quantitative easing,” “privacy enhancing technologies,” or “two-tier CBDC architecture” might not be sufficiently represented in the model’s base vocabulary or learned semantic relationships, resulting in incorrect classifications or weaker contextual embeddings [124]. To mitigate this gap, researchers employ fine-tuning, wherein a pre-trained transformer undergoes additional training using a smaller, domain-specific dataset to recalibrate its language representations toward specialised vocabulary and context [125]. This domain adaptation greatly improves performance in specialist domains such as medicine, finance, or policy [124], [125], [126], [127]

Fine-tuning can take various forms, reflecting different levels of resource availability and domain complexity:

- **Continued pre-training:** Before doing any supervised fine-tuning, some researchers use in-domain unlabelled text to extend the model’s general pre-training [128]. To further train BERT or RoBERTa, for instance, they may compile a sizable corpus of financial news, policy documents, and annual reports to train the model to include domain-specific distributional semantics. When there is a large amount of unlabelled domain text, this two-stage method —continued pre-training followed by supervised fine-tuning — can result in stronger performance [128]. However, this approach is computationally intensive and may not always be feasible.
- **Task-specific fine-tuning:** Practitioners commonly use an annotated dataset aligned with the target task to fine-tune pre-trained models (e.g., sentiment classification of BoE communications). It is less resource-intensive, and the advantage is a straightforward pipeline: from a pretrained model → to domain-labelled data → to the final classifier output [125]. A key limitation here is the reliance on high-quality annotated data, which can be expensive and time-consuming to create, especially for specialised domains.
- **Hybrid techniques (Lexicons + transformers):** Some researchers use hybrid approaches like domain-specific lexicons containing policy and financial acronyms and integrate them into the fine-tuning process to enable the model to simultaneously learn from textual context and domain-tailored dictionaries [56], [57], [58], [59]. While

this can be effective, the creation and maintenance of comprehensive domain-specific lexicons can be challenging, and the integration method can introduce complexities.

A growing body of work demonstrated the superior performance of fine-tuned BERT and RoBERTa models compared to traditional machine learning classifiers (e.g., SVM, Naive Bayes) and non-fine-tuned language models in these fields. This benefit stems from the ability to fine-tune pre-trained language models to the unique nuances of domain-specific language. For instance, FinBERT, a BERT model optimised for financial texts, was introduced by Araci [129] and greatly enhanced the ability to identify minor positive or negative signals associated with corporate earnings reports. The significant advantages of domain adaptation were further demonstrated by Liu et al. [130], who experimented on several financial benchmark datasets and found that FinBERT achieved 91% accuracy. In contrast, the base BERT model achieved 86% accuracy. Legal-BERT [131] and SciBERT [132] echoes these benefits in the legal and scientific spheres, respectively. In addition, domain-specific fine-tuned models are used in public policy to reflect complex sentiments regarding rules or policy modifications [133]. However, accurate sentiment capturing in the context of CBDCs requires refined transformer models due to the intricate economic debates and nuanced views on privacy, technology, and monetary policy. Online debates over CBDCs, for instance, may entail complex discussions regarding privacy trade-offs, financial inclusion, and possible effects on current financial systems [13], [14], [47]. Various aspects of public opinion on these complex issues can be identified with fine-tuned models.

2.2.3.6 FinBERT's Limitations: A Call for Domain-Specific Fine-tuning?

Despite their potential, models such as FinBERT, fine-tuned on financial news data, have drawbacks, mainly when applied to social media data [129]. The training methodology of FinBERT is a major point of criticism. According to Araci [129], FinBERT's training corpus comprises English news items from the LexisNexis database on businesses listed on the OMX Helsinki stock exchange. Although this corpus is useful for recording formal financial language, there are several challenges when assessing popular opinion on platforms like X. First, the style and tone of the source material — formal news articles — differ greatly from the conversations on social media. News stories use intricate sentence structures and specialist financial language while adhering to journalistic standards of neutrality and formality. Social media language, on the other hand, is frequently informal, colloquial, and marked by emojis, misspellings, slang, and an expressive and subjective tone [51]. In addition, Gössi et al. [134] argued that FinBERT struggles with complex sentences that include conjunctions like “but,” “while,” and “though,” which introduce conflicting sentiments, leading to significant misclassifications. Sentence complexity significantly reduces the model's accuracy, making it unable to grasp subtle emotions buried in sophisticated financial terminology. Furthermore, FinBERT struggles with the complex linguistic patterns and specialised financial jargon included in Federal Open Market Committee (FOMC) texts [134].

Although a study on distilled transformer models highlighted FinBERT’s high accuracy, precision, recall, and F1 score in financial sentiment analysis, it does not explore other critical metrics such as robustness to adversarial examples or performance under varying data distributions [135]. As a result, linguistic style variations and domain-mismatch could make it challenging for FinBERT to generalise to social media data. Second, FinBERT's applicability to social media is further limited by the data pre-processing techniques used to create its training data. To concentrate the training on pertinent financial information, Araci [129] noted eliminating sentences that lacked any lexical items. However, this may have unintentionally eliminated words that contained crucial contextual information or more complex sentimental sentiments. Due to this pre-processing phase, the model may have been less sensitive to the implicit or subtle sentiment frequently present in social media posts, which also skewed the model toward more overt emotional expressions. Third, there is a market and regional bias brought about by the emphasis on news from the OMX Helsinki exchange. Although the broad concepts of financial emotion may be universal, local laws, context, and public opinion will likely impact the precise language, issues, and sentiment expressions about CBDCs. As a result, the subtleties of public debate surrounding the digital pound in the UK might not be adequately captured by a model trained on Finnish stock market news. Moreover, FinBERT's emphasis on overall financial sentiment can cause complex discussions about privacy, technology, and monetary policy that are unique to CBDCs to be missed.

In addition to FinBERT's particular drawbacks, other domain-specific BERT modifications may not be directly suitable for CBDC sentiment classification. The absence of high-quality, publicly accessible domain-specific datasets is a major obstacle. Although certain financial datasets are available, they might not be appropriate for capturing the nuances of emotion surrounding CBDCs on social media [105]. This calls for developing custom datasets, which can be costly in terms of both time and resources. Additionally, while transformer models excel at capturing context, they can still struggle with social media data, highlighting the need for careful evaluation and robust testing of domain-adapted models.

Therefore, this research employs a domain-specific gold standard dataset of tweets related to the digital pound — elaborated in Chapter 4 — to fine-tune RoBERTa (and other transformer-based models, see Chapter 5). By creating or utilising a meticulously annotated corpus reflecting public opinion on this new subject, the study lessens the domain mismatch problems that beset larger general-purpose or broad-finance models (e.g., FinBERT) [134], [135]. The model is fine-tuned to acquire distinct language and emotional expressions pertinent to the digital pound by using a CBDC-specific gold standard. These terminology and nuances are less likely to be found in mainstream or even broad financial corpora. As a result, the model can better understand the subtle changes in sentiment on associated themes like financial inclusion, privacy, and policy debates in online discussions of a digital pound.

Additionally, this approach directly addresses the shortcomings of FinBERT and related broad-finance sentiment models, whose corpora usually represent formal news reports or stock market language rather than real-time, informal social media dialogue, by anchoring the entire pipeline in a domain-specific, rigorously annotated dataset. Furthermore, using methods like dynamic masking and a richer feature space, RoBERTa's enhanced training regime reduces the risk of overfitting that is frequently connected with fine-tuning on comparatively small datasets of tweets. This design decision is essential to maintaining generalisability in light of the rapidly changing terminology and perceptions surrounding central bank digital currencies.

Finally, by focusing exclusively on the digital pound–related tweets, the model remains attuned to the most relevant *lexical, thematic, and contextual patterns* shaping public perception of this nascent financial technology. Rather than diluting training data with extraneous financial topics, the dataset in Chapter 4 narrows the scope to reflect the actual discourse policy stakeholders must address. Consequently, the fine-tuned RoBERTa model can produce more accurate, context-aware sentiment classifications, serving as a robust tool for both academic inquiry and practical policymaking guidance regarding the digital pound.

2.2.3.7 Evaluation and Robustness of Transformer models

In addition to using domain-specific datasets, ensuring transformer models' reliability, transparency, and explainability is crucial, mainly when used for informing policy decisions. As important as attaining high accuracy on benchmark datasets is comprehending the reasons behind a model's predictions and ensuring it is resilient to biases and adversarial attacks. This is especially crucial when working with social media data, which can be noisy, biased, and manipulable.

By recognising significant input text elements, techniques such as LIME (Local Interpretable Model-agnostic Explanations) shed light on the model's decision-making process. LIME reveals which features (words or phrases) contributed most significantly to a particular prediction by locally approximating the complex model with a simpler, interpretable model [136]. This is crucial for comprehending the model's logic and spotting any biases. For instance, LIME may show that a model is overly dependent on particular terms linked to specific perspectives on CBDCs, which may be a sign of bias in the training set [137]. Similarly, the process of producing slightly altered inputs intended to "fool" the model is known as adversarial testing [138]. Researchers can find flaws and strengthen robustness by observing how the model reacts to these hostile scenarios. This is especially pertinent regarding exploiting social media through coordinated campaigns or false information. For example, a model may be susceptible to minor typo corrections, or the usage of particular hashtags intended to influence sentiment [139].

Building confidence in model predictions and ensuring their suitability for guiding policy decisions depend on these evaluation and robustness testing techniques. This is in line with RQ2 (Model Capabilities and Limitations), which will be addressed using robustness testing and LIME to assess the explainability and robustness of the chosen transformer model.

2.2.4 Comparing Official and Public Discourses: Communication Theory Foundations

This section explores key communication theories and models that provide a lens for understanding how official bodies attempt to shape public narratives and how these narratives are received and interpreted by the public. It establishes the theoretical framework for analysing the relationship between official communications and public sentiment regarding CBDCs.

2.2.4.1 Theories of Policy Communication and Public Engagement

Effective communication is essential for the public to comprehend, accept, and participate in policy. The dynamics of this process can be better understood by applying several communication theories.

According to the *agenda-setting theory*, the public's opinion of the significance of different topics is influenced by the media [140] and, consequently, official communications such as from the BoE and HM Treasury. Official entities can affect the public's perception of the most noteworthy features of this technology by emphasising particular characteristics of CBDCs (such as technological innovation and financial inclusion) in their messaging. This can influence public opinion and draw attention to specific policy aspects. *This research will analyse official communications to identify the key themes and topics emphasised by the BoE and HM Treasury and compare these with the topics discussed by the public on social media to assess the agenda-setting effect.*

Framing theory by Entman [141] further explains how the frame, i.e., how information is presented, can influence its interpretation. Selecting and emphasizing some aspects of a problem while downplaying others is known as framing. For instance, portraying CBDCs as a means of improving financial stability may cause different public responses than presenting it as a tool for government surveillance. These theories highlight how official language can influence public perception of complicated policy matters. *In the context of the study under consideration, how official communications frame CBDCs and its potential impact on public sentiment will be explored.*

Building upon these theories of influence, Grunig and Hunt's (1984) four models of public relations provide a useful framework for analysing the strategic approaches organisations, including government bodies, take in their communication efforts [142]. These models represent different approaches to communication and relationship management, ranging from one-way dissemination of information to two-way engagement and dialogue:

- **Press agentry/publicity:** This paradigm emphasises one-way communication, mostly through publicity and propaganda, with little regard for audience reaction or veracity. Regardless of the veracity of the material shared, the objective is to attract attention and advance a positive image.
- **Public information:** While one-way communication is still emphasised in this model, the emphasis is on providing the public with correct and truthful information. Although being honest is respected, audience participation and feedback are still not given enough attention.
- **Two-way asymmetrical:** This strategy mostly aims to convince the public to accept the organisation's point of view, but it also emphasises research to understand public views and preferences. The organisation uses feedback to improve its persuasive messages, yet the information flow is inconsistent despite the two-way contact.
- **Two-way symmetrical:** This model highlights reciprocity and strongly emphasises knowing one another and having sincere conversations with the public. Both sides listen to each other, provide feedback, and make adjustments to maintain balanced, two-way communication. This research utilises this model to assess whether BoE engages in meaningful dialogue or primarily attempts to persuade the public without establishing a fully reciprocal exchange (Chapter 10).

While other theoretical frameworks have been applied to the study of public perception of CBDCs, such as the privacy calculus theory applied by Jabbar et al. [20] and the privacy paradox explored by Koziuk et al. [143], there remains a significant gap in research that explicitly examines the communication strategies of official bodies *using established public relations and communication theories and models*. Jabbar et al. [20] investigated the trade-offs individuals make between privacy risks and potential benefits of CBDCs, whereas Koziuk et al. [143] focused on the discrepancy between individuals' stated privacy concerns and their actual online behaviour related to digital currencies. These studies do not directly address the communication dynamics between policymakers and the public while providing valuable insights into individual-level perceptions. *Therefore, understanding official communication narratives through the lens of established communication models is a key supporting aspect of this research (RQ5), providing a crucial context for interpreting public sentiment.*

2.3 Synthesis and Identification of Gaps

This study integrates insights from finance, fintech, economics, communication studies, and NLP to analyse public sentiment toward CBDCs, specifically the digital pound, and its interaction with official communications. The literature on CBDC emphasises issues with public acceptance, privacy problems, and global motivations. The examination of sentiment analysis approaches revealed that traditional sentiment analysis methods struggle with domain-specific nuances in financial discourse. Nonetheless, transformer-based models, enhanced through domain adaptation, address these limitations by capturing context and specialised vocabulary. Finally, communication theories (e.g., Grunig and Hunt's two-way symmetrical model) provide a framework for evaluating how well official strategies foster public awareness and engagement.

This synthesis reveals several critical gaps in the existing literature:

- **Limited application of advanced NLP to CBDC discourse on social media:** While some studies have begun to use transformer models like BERT to explore public perceptions of CBDCs [104], [105], the application of *fine-tuned* transformer models specifically to understand the nuances of CBDC discourse on social media is still in its nascent stages. This gap is significant because social media platforms like X serve as a crucial space for opinion formation and public discourse on complex policy issues like CBDCs. Related work on cryptocurrencies further reinforces this point: although sentiment analysis using NLP has been applied to crypto markets, these studies often rely on lexicon-based tools, self-labelled data, or shallow classifiers with limited robustness and domain adaptation [18], [101], [102], [103]. Consequently, they fall short in capturing semantic nuance, evolving slang, or longer-term narrative dynamics, i.e., shortcomings that are particularly consequential in fast-moving, policy-sensitive domains such as CBDCs.
- **Lack of longitudinal, comparative analysis of public and official narratives within a robust theoretical framework:** Existing studies frequently concentrate only on official communications or user-generated

content (tweets, forums) [20]. Studies that monitor the dynamic evolution of public sentiment over time are essential for understanding if and when negativity or scepticism peaks and whether official communications can (or do) steer sentiment over time. This longitudinal, comparative approach, grounded in communication theory, is crucial to understanding the efficacy of official communication strategies and spotting possible discrepancies between official messaging and public reception.

- **Insufficient integration of communication theories:** The literature lacks computational studies employing a robust theoretical framework for interpreting their findings within the context of communication and policy. This research addresses this critical gap by employing Grunig and Hunt's two-way symmetrical model as a theoretical lens, providing a structured approach to analysing the relationship between official communications and public sentiment *related explicitly to CBDCs*. Similarly, while the importance of how information is presented is highlighted by framing theory, there is limited research specifically examining how official bodies frame certain aspects of CBDCs (in this case, the digital pound), which ultimately influences public sentiment toward such novel monetary tools. This theoretical grounding allows for a deeper understanding of the communication dynamics at play and their impact on public perception of CBDCs.

Moreover, emerging research on cryptocurrency sentiment provides instructive methodological lessons that sharpen the gaps identified above. Empirical work on Bitcoin and altcoins demonstrates that generic lexicon-based pipelines are highly sensitive to tweet-cleaning choices and struggle with crypto-specific slang and sarcasm [101]. Even when headline polarity is fused with market features, the resulting linear models capture only coarse next-day directionality and overlook intraday volatility [102]. Subsequent large-scale studies that recycle lexicon labels into shallow classifiers inherit this bias and under-represent contextual nuance [144]. Deep ensembles improve headline accuracies, yet their reliance on silver-standard labels and undersampled data sets inflates performance and leaves sentiment–price dynamics untested [103]. Collectively, the evidence underscores three imperatives: rigorous domain adaptation, transparent model diagnostics and longitudinal evaluation. Guided by these insights, the present thesis constructs a purpose-annotated digital-pound corpus, fine-tunes multiple transformers with explainability and stress-testing modules, and situates sentiment shifts within a theory-driven, temporal analysis, thereby translating cross-domain learning into a policy-relevant CBDC framework.

The research framework, illustrated in Figure 2.1, outlines the systematic process employed to investigate public sentiment towards the digital pound.

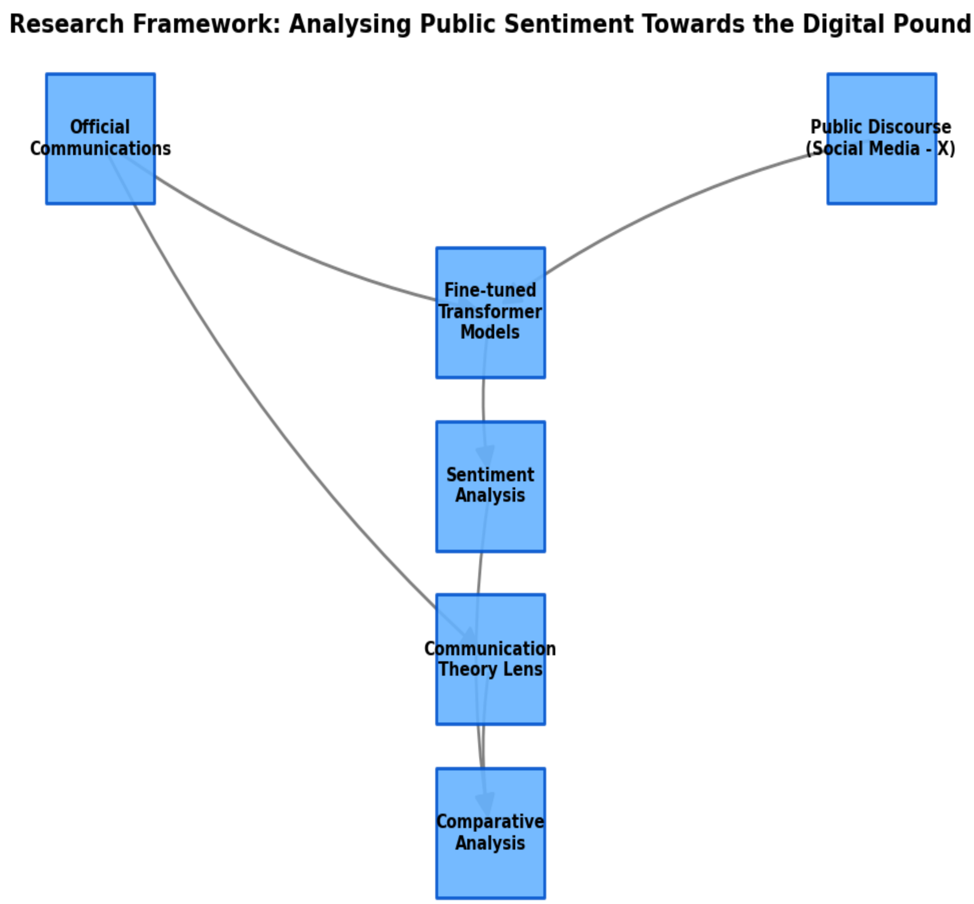


Figure 2.1: Research framework for analysing public sentiment towards the digital pound.

2.4 Conclusion

This literature review has demonstrated the need for a more nuanced and comprehensive approach to understanding public sentiment towards CBDCs, particularly in the context of official communications. By synthesizing advanced NLP techniques, specifically fine-tuned transformer models, with established communication theories, this research aims to bridge the identified gaps in the literature. This interdisciplinary approach allows for a more robust and insightful analysis of the complex interplay between official messaging and public reception, directly addressing the research questions outlined in Chapter 1. Importantly, the research holds practical implications for the digital pound as the UK navigates the design and other complexities associated with its digital currency, public trust and clear communication become critical. Advanced analytics on social media data can support policy development by identifying misunderstanding, scepticism, or areas of heightened interest.

The next chapter details the research design, outlining the methodology used to investigate the interplay between official narratives and public sentiment surrounding the digital pound. This design allows for a nuanced analysis of how public opinion is shaped by, and in turn influences, official communications about emerging financial technologies.

Chapter 3 – Research methodology

3.1 Introduction

This chapter describes the methodological framework used to answer the research questions listed in Section 1.3.1. It incorporates both quantitative and qualitative methodologies, each closely related to one or more of the study's RQs, given the multifaceted nature of CBDC research.

This chapter begins with an overview of the research design and philosophical stance, then discusses data collection methods, preprocessing and gold standard creation, model selection and fine-tuning, robustness checks, sentiment/topic analyses, temporal analysis, and comparisons with BoE documents. Each step is clarified in light of the research questions to demonstrate how the methodology supports the study's objectives.

3.2 Research Design Overview

3.2.1 Mixed-Methods Orientation

Building upon the pragmatic philosophical foundation, which advocates using any research methodologies most appropriate for addressing the central research question [145], this study employs a mixed-methods approach and integrates computational (quantitative) and qualitative research techniques within a pragmatic framework [146]. Prioritising the research issue and using suitable techniques to address it, this paradigm recognises that both quantitative and qualitative data provide insightful information that complements each other. Because it enables the analysis of both broad patterns in public opinion (captured computationally/quantitatively) and the complex meanings and interpretations ingrained in the discourse (explored qualitatively), this method is particularly well-suited for examining public discourse and policy texts. Lukito and Pruden [147] advocate for such methodologies and highlight the need to consider language variations and self-reflexivity in research contexts [1].

- **Computational (quantitative) components:** Transformer-based model training, performance metrics, sentiment classification and analysis (including topic modelling and temporal analysis).
- **Qualitative components:** Thematic analysis of policy documents, interpretive comparison of institutional and public narratives.

This dual approach addresses both *what* the sentiments and trends are *(computationally or quantitatively)* and *why* these sentiments arise or differ from the official narrative *(qualitatively)*, providing a more holistic and nuanced understanding of the research problem [147]. Table 3.1 below explains how RQs link to methodological components:

Research Questions	Methodological Components
RQ1	Data collection (domain corpus), gold standard creation, model experimentation and comparison, theoretical justification for fine-tuning
RQ2	Performance metrics (accuracy, precision, recall, F1-score), robustness checks (adversarial testing and error analysis), LIME-based explainability
RQ3	Multifaceted sentiment and temporal analysis (monthly/event-based sentiment trends)
RQ4	Qualitative thematic coding of BoE and HM Treasury response documents
RQ5	Comparative analysis of public sentiment (RQ3) and official narratives (RQ4), application of communication theories (framing, agenda-setting, two-way symmetrical communication)

Table 3.1: Linking RQs to methodological components.

The integrated nature of the mixed-methods approach is visually represented in Figure 3.1.

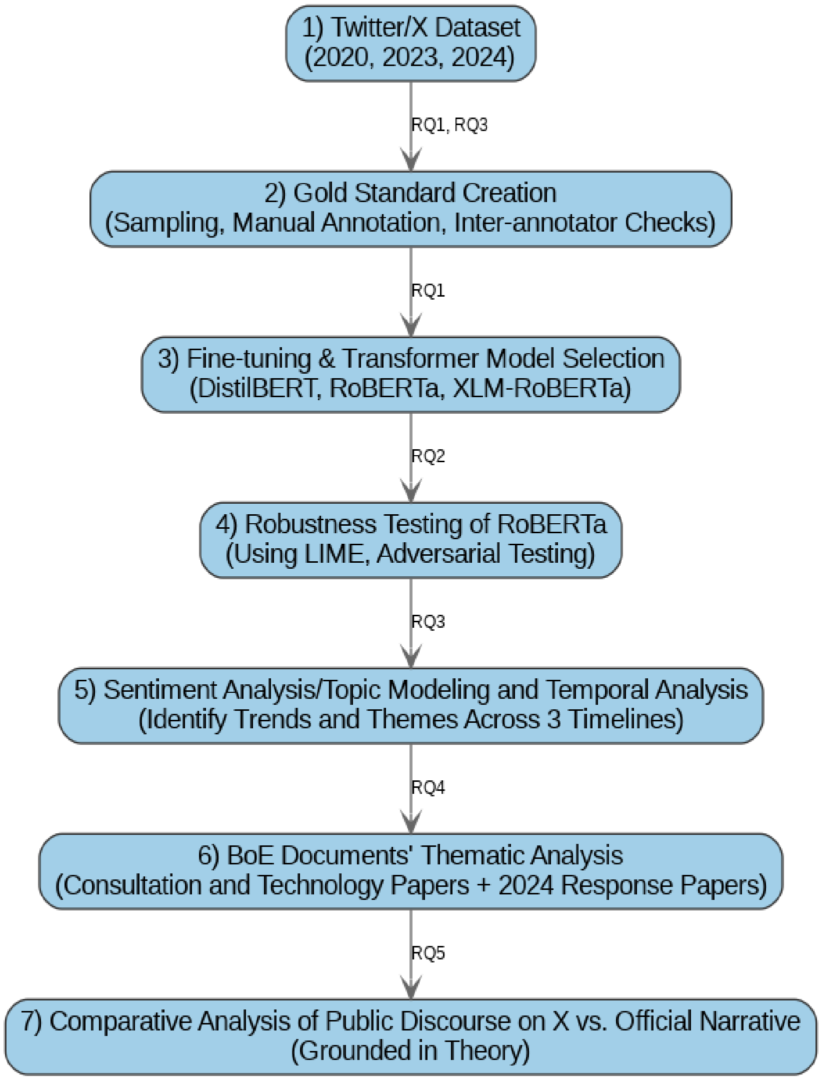


Figure 3.1: Visual representation of the research design.

3.3 Data Collection

3.3.1 Sourcing X Data

The data for the study under consideration was sourced from X (formerly Twitter). The platform's rich repository of user-generated content was leveraged to capture public sentiment on the UK CBDC. Data was collected manually because of X's API restrictions, mainly because it is not available for free for research purposes anymore after Tesla's CEO, Elon Musk, took over in October 2022, and the basic tier does not connect to necessary endpoints. The issue remained unsolved even after connecting with the X's customer support. A major methodological problem brought about by this change in API policy prompts the need to switch to manual data gathering. Although automated data gathering using the API is typically favoured due to its effectiveness and scalability, the limitations imposed rendered it impractical for this study. Manual collection was the most viable alternative to ensure access to the required data.

Although the manual method was labour-intensive, it ensured comprehensive and targeted data retrieval, recording a variety of tweets to represent a spectrum of viewpoints and emotions. This process involved:

- **Identifying relevant keywords**: Relevant keywords were identified and used to search for tweets based on a preliminary search and knowledge of the topic. Finding the right keywords at the outset was essential to concentrating the data collection on pertinent conversations on the UK CBDC. The keywords were chosen based on a review of existing literature, policy documents, and initial X searches to find trending terms and hashtags.
- **Search:** X's advanced search option was used to collect tweets using the identified keywords. This function allowed for precise filtering of tweets based on keywords, date ranges, and language, which was essential for targeting the specific timeframes and topics of interest in this research.
- **Collecting tweets**: Tweets containing the identified keywords were collected manually during the specified periods.
- **Language:** The language was set to English to concentrate on public sentiment within the UK context.
- **Copy and organise:** The tweets containing identified keywords, including hashtags, links, and @mentions, were copied and pasted into the Google sheet. The data was then organised into three columns: 'Keywords,' 'Date,' and 'Tweet.' This structured organisation of the data in a Google Sheet facilitated subsequent preprocessing and analysis.

Despite being a time-consuming manual data collection method provided several advantages. The manual method allowed flexibility in adjusting the search parameters and adding more keywords in case new, pertinent topics surfaced during the data collection periods. Moreover, manual data collection allowed for a deeper understanding of the discourse surrounding the UK CBDC, which influenced later phases of analysis. When combined with the deliberate inclusion of relevant keywords, this manual method helped create a robust and focused dataset.

3.3.1.1 Dates of Paper Publications and Timelines

Three distinct periods were chosen, specifically around the publication of BOE's Discussion paper in 2020, the consulting and technology papers in 2023, and the response to the consulting and technology papers in 2024.

- **January 1, 2020 - June 30, 2020**: This period revolves around the publication of the Bank of England's Discussion Paper titled "central bank digital currency: Opportunities, challenges and design" on March 12, 2020 [10].
- **February 1, 2023 - June 30, 2023**: This period covers the release of the Bank of England's Consultation Paper on the digital pound titled "The digital pound: A new form of money for households and businesses?" and the Technology Working Paper

titled “The digital pound: Technology Working Paper,” both published on February 7, 2023 [11].

- **January 1, 2024 - March 31, 2024**: The data for this period captures public reactions to the publication of the consulting and technology papers on the digital pound as well as the response of the bank and HM Treasury to the public's response, both released on January 25, 2024 [12].

3.3.1.2 Justification for Timelines

The selected timelines aim to address RQ3 on temporal sentiment shifts and coincide with noteworthy publications and events about the UK CBDC, offering a rich contextual framework for analysing public sentiment. The start dates follow with the release of significant Bank of England documents, ensuring that the gathered data corresponds with the initial public responses and the subsequent discussions. Here are the three key dates on which the analysis is based:

- **March 12, 2020:** This marks the release of BOE’s Discussion paper on the digital pound, which covers foundational concepts and considerations. Analysing sentiment during this period could yield insights into public reception to these initial ideas to understand how people reacted to this new form of money.
- **February 7, 2023:** The bank released both Consultation and Technology papers to receive feedback from the public on this date. This period is significant for sentiment analysis because of the possibility of intensified discussion and scrutiny.
- **January 25, 2024**: This date captures the release of the response to the Consultation and Technology papers, which disclose an aggregated summary of the public’s responses to the original papers as well as the bank and HM Treasury’s response to the public on their shared opinions. This period is important to reflect the public’s evolving opinions and reactions to major announcements related to the digital pound.

These timelines allow the author to analyse how public perception evolves over time and in response to major announcements through a comprehensive temporal analysis. This longitudinal approach is crucial for understanding the dynamic relationship between policy announcements and public sentiment.

3.3.1.3 Keywords used

A specific set of keywords were used to ensure the relevance and focus of keywords on the UK CBDC. These include:

- cbdc uk
- digital pound
- digital pound uk
- cbdc anonymity
- cash is king uk

An initial total of 6,283 tweets was gathered across the above keywords. Following the cleaning and deduplication steps (explained in Section 3.3), 6,200 unique tweets remained. Notably, '*digital pound'* saw the highest number of discarded tweets (45), whereas other keywords had minor reductions, and *cbdc anonymity* remained unchanged (see Figure 3.6). A more detailed breakdown — showing the year-wise distribution and keyword-specific reductions — is provided in Section 3.3.1.4.

An advanced search query was used to collect relevant tweets for sentiment analysis, filtering by keywords, language, and date range. For example, the query (as shown in Figure 3.2) is focused on tweets containing "cbdc uk" in English, posted between January 1, 2020, and June 30, 2020. This approach captured public opinion before, during, and after important occasions, such as the publishing dates of the Bank of England's Discussion Paper (March 12, 2020) and Consultation and Technology Papers (February 7, 2023). The same process was applied to other keywords. Manually collected tweets with hashtags, links, and mentions were stored in a Google Sheet with keywords, date, and tweet columns.

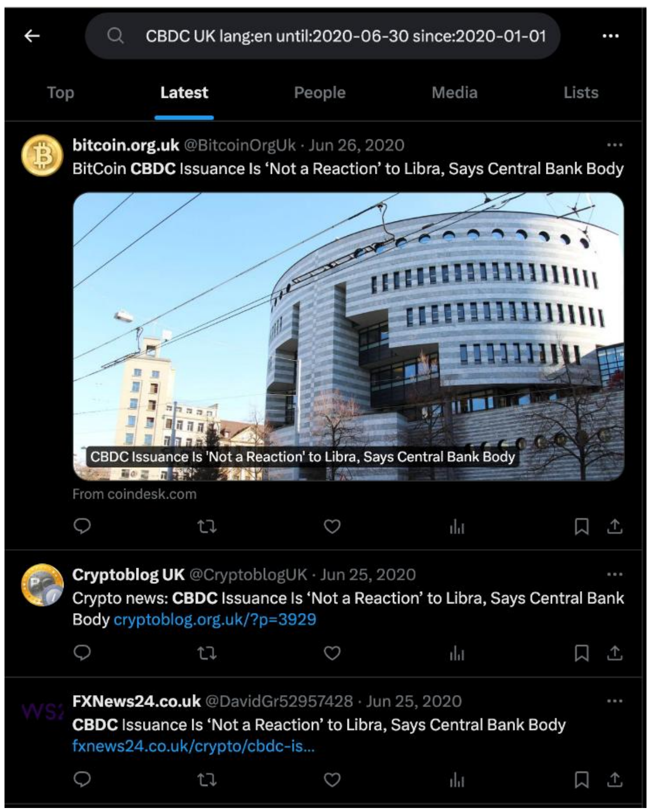


Figure 3.2: Search query used for "cbdc uk" keyword.

Medagoda and Shanmuganathan [148] noted that using keywords ensures that timely sentiment patterns and temporal trends are captured. By efficiently filtering tweets to reveal certain interesting themes, keywords detect sentiment polarity (positive, negative, or neutral) and analyse sentiment patterns over time. They also demonstrated how keyword clustering and sentiment classification can reveal shifts in public sentiment about important events over time. This is in line with one of the critical objectives of the study under consideration. Similarly, Agichtein et al. [149] noted that keywords are frequently more adaptable and can arise naturally within the content of tweets, offering a more thorough and precise representation of the discourse. These studies provide empirical support for the use of keyword-based data collection in social media research and highlight its effectiveness in capturing relevant information and tracking sentiment changes over time.

3.3.1.4 Data Volume

During the data collection phase, the datasets comprised 281 tweets from 2020, 4754 tweets from 2023, and 1248 tweets from 2024, making up a total of 6283 tweets, as shown in Figure 3.3. This variation in data volume across the different timeframes reflects the changing levels of public discussion and interest in the digital pound over time. The larger volume of data in 2023 suggests a significant increase in public engagement with the topic following the release of the Consultation and Technology Papers.

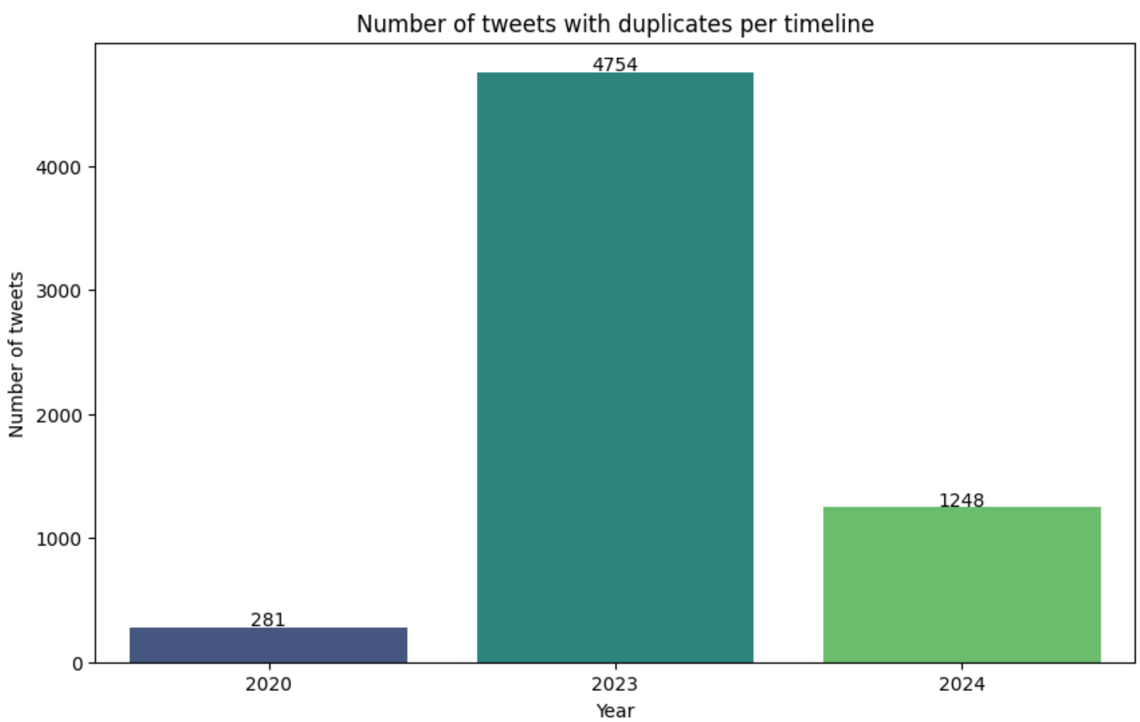


Figure 3.3: Number of tweets per timeline (including duplicates).

Additionally, as different keywords were used to collect the data, the bar chart below (Figure 3.4) shows the number of tweets collected for each keyword associated with the digital pound discourse. With 2,913 tweets, the keyword “cbdc uk” yielded the most, suggesting a substantial public discussion on this subject. The keyword “digital pound” came next with 2,206 tweets, indicating strong interest. Both “digital pound uk” and “cbdc anonymity” attracted 636 and 307 tweets, respectively, indicating moderate engagement. With just 221 tweets, the term “cash is king uk” yielded the fewest, indicating comparatively less public conversation using this keyword. This distribution reveals varying discussion and interest levels regarding particular aspects of the UK CBDC.

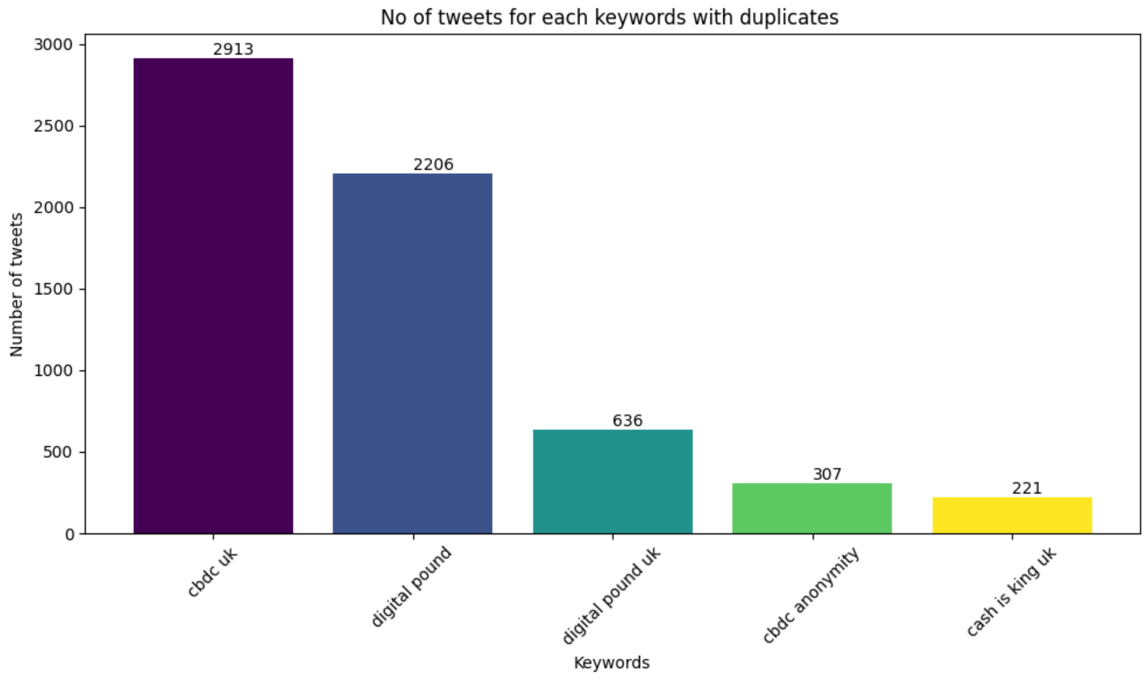


Figure 3.4: No. of tweets collected for each keyword.

3.4 Data Preprocessing

Transforming raw data into a clean and structured format is essential for effective data analysis. This practice is known as data preprocessing, which is crucial in all data science-related projects, particularly in NLP where text data often contains noise, inconsistencies, and irrelevant information. Preprocessing ensures that the data is suitable for subsequent analysis, including sentiment analysis and topic modelling, and improves the performance and reliability of the models used.

3.4.1 Preprocessing Steps for X Data

For the initial sentiment analysis using transformer models (addressing RQ1 and RQ2), the precise date and time stamp within each timeline (e.g., specific day and time) was unnecessary and could introduce unnecessary complexity to the model. Transformer models primarily focus on the textual content for sentiment classification. Therefore, the date information was removed during the initial preprocessing for model training and evaluation. Critically, it was retained in a separate file and reintroduced later in the analysis, specifically for the temporal analysis component (addressing RQ3).

Other preprocessing steps taken to remove noise and standardise tweets include the following:

3.4.1.1 Importing Necessary Libraries and Data

- **Pandas and Numpy:** These libraries are used for data manipulation and analysis. Specifically, Pandas is used to handle structured data in DataFrame format. Pandas provides efficient data structures and functions for data cleaning, transformation, and analysis, while NumPy offers support for numerical operations.
- **Regular Expressions (re)**: This library is used to search and manipulate text. It helps identify patterns such as mentions, URLs, and hashtags. Regular expressions provide a powerful and flexible way to identify and remove or replace specific patterns in text data, which is essential for cleaning social media data.
- **BeautifulSoup**: It is a library used to parse HTML and XML documents. It helps strip HTML tags from tweets.
- **String**: It provides a set of string operations, particularly useful for removing punctuation.
- **NLTK (Natural Language Toolkit)**: A comprehensive library for natural language processing tasks; it includes tools for tokenising, tagging, and parsing text. NLTK is a widely used library in NLP, offering a range of tools for text processing tasks such as tokenisation, stemming, lemmatisation, and stop word removal. Tokenisation is crucial for preparing text data for NLP models.
- **Reading data**: The dataset is read into a pandas DataFrame from an Excel file for further processing.

3.4.1.2 Defining clean_tweet function

This function performs various cleaning operations on individual tweet texts, as explained in Table 3.2.

Various clearing operations	Description
HTML tag removal	HTML tags can add noise to text data; BeautifulSoup helps strip away these tags, leaving only the text content.
Punctuation removal	From a usability perspective, punctuation marks are generally useless for sentiment analysis. Therefore, removing them reduces the complexity of tokenisation and simplifies the text.
URL removal	URLs can skew the analysis and often do not contribute to the sentiment of the text. These are removed using regular expressions.
Number removal	Unless directly relevant, numbers can introduce noise into the sentiment analysis. Regular expressions are used to remove them.

Tokenisation	Tokenisation divides the text into individual words, or tokens. This phase is essential for word-level analysis as it facilitates further processing steps.
Reconstructing text	The text is made clear and consistent by joining the tokens back into a single string.

Table 3.2: Various cleaning operations performed on the data.

3.4.1.3 Defining the 'clean' function

The entire DataFrame is cleaned using the 'clean' function, which performs additional cleaning steps, as explained in table 3.3 below:

Further cleaning operations	**Description**
Lowercasing	To maintain consistency, all data was converted to lower text. This prevents 'Pound' and 'pound' from being treated as different words.
URL removal	Ensures any leftover URLs are removed from the tweets.
HTML tag removal	Ensures any leftover HTML tags are removed.
Mention removal	Removes X mentions (e.g., @username). Mentions do not contribute to sentiment and can add noise.
Hashtag removal	Removes hashtags without changing the meaning of the words. This is important because hashtags can sometimes display sentiments.
Punctuation removal	Ensures any remaining punctuation marks are removed.
New line removal	Replaces new line characters with spaces to keep the text as a single continuous string.
Whitespace normalisation	Replaces numerous spaces with a single space to ensure uniform spacing across the text.

Trimming text	Text length is restricted to 128 characters to maintain consistency, which is important for certain machine learning models.

Table 3.3: Further cleaning operations performed on the data.

The cleaned data is saved back to an Excel file *(preprocessed__full_data.xlsx)* for further analysis.

3.4.1.4 Duplicates Removal

Duplicates were removed during the data preprocessing step to ensure the uniqueness of tweets in each timeline. Using the ‘pd.read_excel function,' data corresponding to the years 2020, 2023, and 2024 was loaded from three separate Excel files into DataFrames. Then, rows containing any missing values were removed using the ‘dropna method’ with the ‘inplace=True argument,' which modifies the DataFrame in place and ensures that no rows with NaN values remain. Handling missing values is an important step to prevent errors during subsequent analysis.

Subsequently, the 'drop_duplicates’ method was utilised on the ‘Tweet’ column for each DataFrame to eliminate redundant data. This method identifies and removes duplicate rows, retaining only the first occurrence of each unique tweet and discarding subsequent duplicates. The ‘subset=[Tweet']’ parameter specifies that the duplicate check should only be performed on the ‘Tweet’ column. Finally, the cleaned DataFrames, now containing only unique tweets, were saved back to Excel files using the ‘to_excel' method, with the ‘index=False’ parameter to exclude row indices from the output files. This thorough process ensures that the datasets are free from missing or null values and duplicates, thus maintaining data integrity and preventing bias in the subsequent analysis, especially the sentiment analysis and topic modelling.

Figure 3.5 shows the number of unique tweets (after removing duplicates) for each year. There are 279 unique tweets for 2020, 4,702 unique tweets for 2023 (the highest number), and 1,219 unique tweets for 2024. This makes a total of 6,200 unique tweets. The smallest dataset in 2020 indicates less activity or interest in a digital pound and related keywords in that year. In contrast, the largest dataset in 2023 suggests a significant increase in discussions or mentions of the keywords. The 2024 dataset with an intermediate number indicates reduced yet sustained activity compared to 2020. This distribution of data across the timelines provides valuable context for the temporal analysis (RQ3), allowing for comparisons of sentiment and topic trends over time.

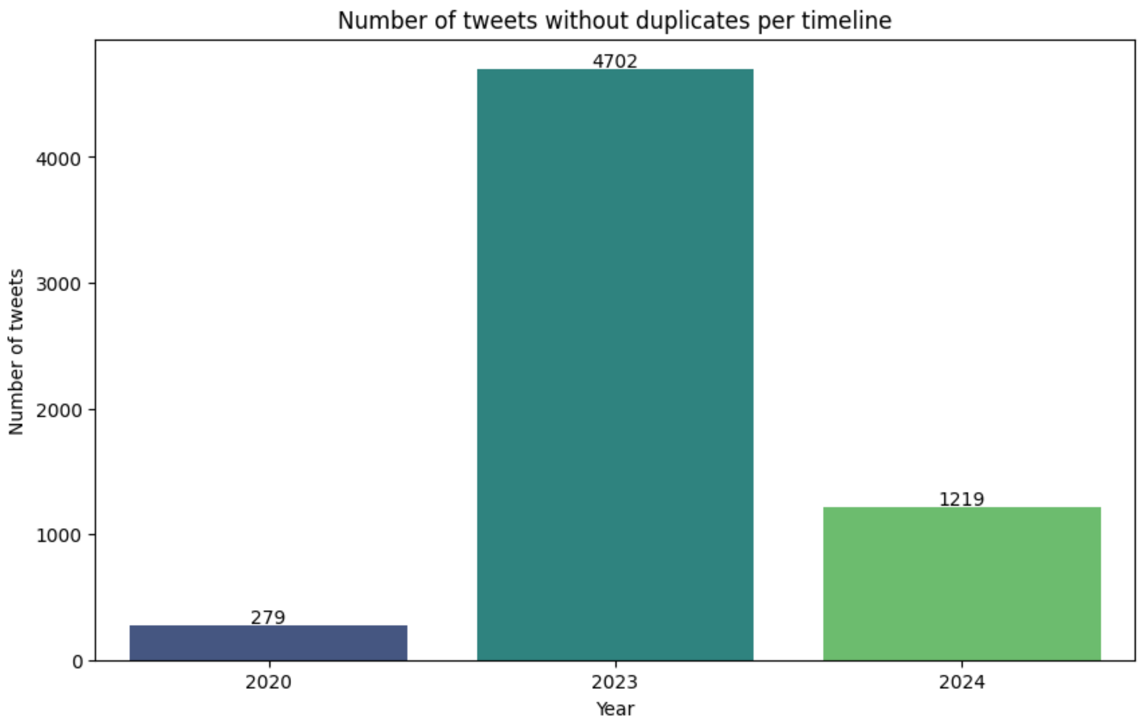


Figure 3.5: Number of tweets without duplicates per timeline.

The dataset's size (6,200 unique tweets and later 5,702 relevant tweets as explained in Chapter 7) is broadly comparable with peer-reviewed resources that have supported credible cryptocurrency-sentiment research. Gadi and Sicilia's CryptoLin corpus contains 2,683 annotated news items yet yields stable model performance across multiple sentiment engines [150], while Divesh et al. [151]examine market mood using 5,000 carefully filtered YouTube comments and obtain 94 % classification accuracy with an ensemble approach [148]. Against this benchmark, the present tweet corpus is sufficiently large for supervised fine-tuning and temporal analysis, especially given its deliberate stratification around three policy milestones and its diversified keyword strategy. Nevertheless, the sample is interpreted as representing the *engaged online discourse* on a UK CBDC rather than the entire population; conclusions are therefore framed as indicative and are cross-checked against official-document themes to mitigate over-generalisation.

The bar chart below visualises the number of unique tweets for each keyword across all timelines (Figure 3.6). The keyword 'cbdc uk' has the highest number of unique tweets, totalling 2,893. Following this, 'digital pound' has 2,161 unique tweets. 'digital pound uk' comes next with 619 unique tweets, while 'cbdc anonymity' has 307 unique tweets. The keyword 'cash is king uk' has the lowest count, with 220 unique tweets. Therefore, it is evident that the 'cbdc uk' keyword is the most discussed, indicating high interest or relevance; 'digital pound' is also highly discussed on the platform, showing significant engagement. On the other hand, 'digital pound uk,' 'cbdc anonymity,' and 'cash is king uk' keywords have fewer mentions, suggesting they are less central to the discussions around the UK CBDC.

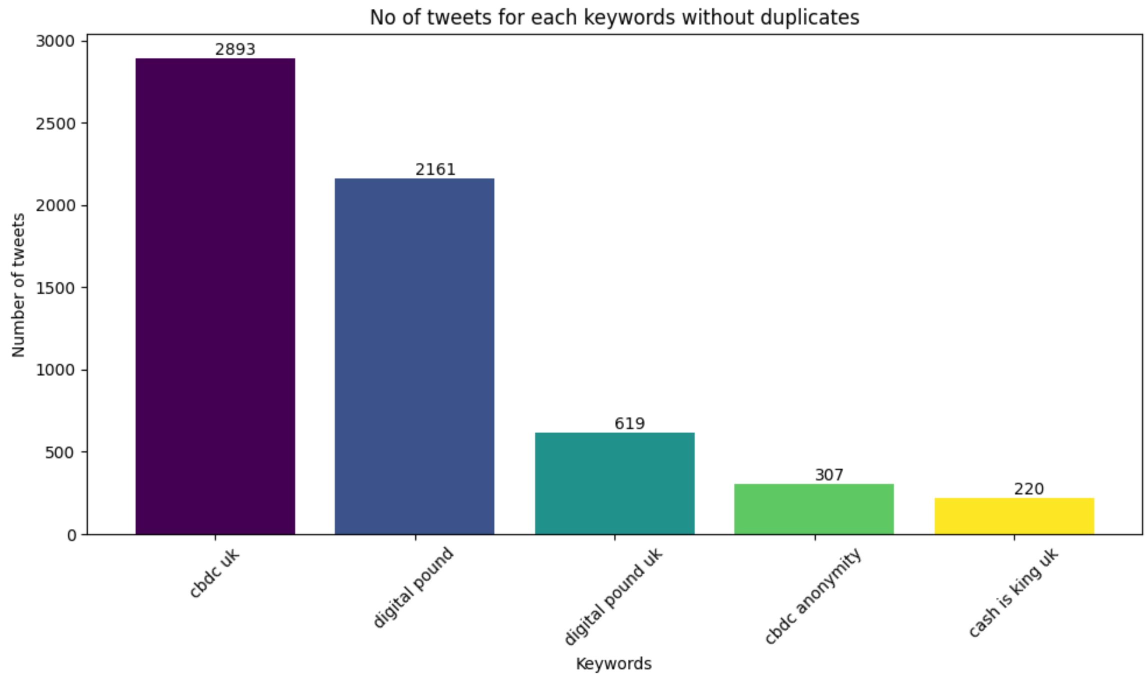


Figure 3.6: Number of tweets for each keyword without duplicates.

3.5 Gold Standard Creation

To address RQ1, a domain-specific gold standard dataset was created to serve as the ground truth for fine-tuning and evaluating candidate transformer models for sentiment analysis within the digital pound discourse.

- **Sampling:** A stratified random sampling approach was employed, ensuring representation from each of the three timeframes (2020, 2023, and 2024). Sample sizes (approximately 162, 355, and 292 tweets, respectively) were determined based on a 95% confidence level and a 5% margin of error.
- **Annotation scheme:** A domain-specific annotation scheme with three sentiment labels (positive, negative, and neutral) was developed, considering the nuances of CBDC discussions. Clear annotation guidelines, including examples of domain-specific sentiment cues (e.g., reactions to "cashless society," "privacy," "anonymity"), were provided to three independent annotators with backgrounds in the NLP field.
- **Annotation process:** Three independent annotators with backgrounds in NLP were recruited to manually label the sampled tweets. This use of multiple annotators (triangulation) helps to mitigate individual biases and improve the reliability of the gold standard.
- **Inter-annotator agreement:** Inter-annotator reliability was measured using Cohen's Kappa (for pairwise agreement) and Fleiss' Kappa (for agreement among all three annotators). Disagreements were resolved through discussion and careful review of the annotation guidelines to reach a consensus; any remaining divergences underwent

adjudication to finalise gold-standard labels, ensuring the quality and validity of the final gold standard dataset.

3.6 Model Section and Finetuning

Three pre-trained transformer models (DistilBERT, RoBERTa, XLM-RoBERTa) were fine-tuned for sentiment analysis. Fine-tuning was performed on the gold standard dataset using an 80/20 train-validation split. Models were tokenised using their respective Hugging Face tokenisers (max sequence length: 128). Hyperparameters used in the process include: Learning Rate: 1e-5, Batch Size: 8, Optimiser: AdamW with weight decay. Each model was trained for 30 runs at epochs 3 and 5. Performance *(of the validation set)* was evaluated using accuracy, precision, recall, and F1-score.

3.7 Evaluation, Robustness, and Explainability

The chosen model's capabilities (RQ2) were evaluated using class-specific F1-scores and confusion matrices. In addition, robustness was assessed using noisy (introducing typos) and adversarial (modifying tweets) examples. Performance degradation was measured by comparing its performance on the clean test data to determine its resilience to real-world data imperfections and potential adversarial attacks. LIME was used for local explainability, identifying influential words/phrases for interpretation, adding interpretive transparency, crucial given policy significance and public trust issues in finance.

3.8 Sentiment Trends, Topics, and Temporal Analysis

The best-performing model (from RQ1) was used to classify the sentiment of all unique tweets. Subsequently, topic modelling was conducted using NMF, LDA, and BERTopic to identify shared themes within the dataset, with BERTopic specifically employed for temporal analysis. To address RQ3, monthly sentiment averages were calculated alongside event-based scores around key policy announcements, enabling the detection of spikes or dips in public sentiment. This analysis aimed to correlate user reactions with each major BoE publication, highlighting possible linkages between policy milestones and changes in sentiment. Topic coherence was also assessed, and variations in both topic prevalence and tweet sentiment were examined over time, offering deeper insights into the evolution of public discourse on the prospective digital pound.

3.9 Thematic Analysis of BoE Response Papers

This section addresses RQ4, which aims to extract key themes and narrative frames from responses to public feedback via 2024 response papers. A qualitative thematic coding approach was employed to systematically read and code each document; a code book was prepared for both papers. Through iterative analysis, recurring priorities and official stances were distilled, revealing the extent to which particular issues — like privacy safeguards or implementation timelines — are foregrounded or downplayed. By synthesising these coded narratives, the study establishes consistent lines of emphasis or omission (e.g., whether privacy is strongly emphasised, whether user experience is portrayed as a design imperative), thereby clarifying the BoE's framing of the digital pound in its official policy communications.

3.10 Comparative Analysis and Communication Theories

This final methodological stage integrates insights from public discourse analysis (addressing RQ3) with the BoE document narratives (RQ4) to determine how official policy documents align or diverge from public concerns. First, an alignment and discrepancy mapping are conducted by juxtaposing the top discussion topics derived from the public's X discourse (e.g., privacy, anonymity, distrust of centralisation) against the coded themes from BoE and HM Treasury publications. This mapping highlights areas of synergy (such as instances where both parties emphasise privacy) or mismatch (e.g., strong public demand for regulatory clarity not explicitly addressed in the official documents).

To interpret these findings, the study draws on communication theories, including framing theory, agenda-setting theory, and Grunig and Hunt's two-way symmetrical communication mode to evaluate whether the BoE's responses genuinely reflect public feedback. By exploring these communication frameworks, the comparative analysis (RQ5) offers nuanced insights into the effectiveness of the BoE's policy messaging, the degree of mutual influence between official narratives and public sentiment, and the broader implications for fostering public trust in the digital pound.

3.11 Ethical Considerations

This research adhered to strict ethical guidelines to ensure responsible data handling and analysis:

- Only publicly available tweets were collected from X. No private or protected data was accessed.
- User identifiers (e.g., usernames) were not collected or stored.
- This research did not involve directly interacting with or recruiting X users. No attempts were made to identify or contact individual users.
- All official BoE and HM Treasury documents used in this research are publicly accessible and were properly cited and referenced.
- All data was stored securely and accessed only by authorised researchers.

3.12 Conclusion

This chapter outlined a rigorous mixed-methods framework designed to address the study's research questions on the UK digital pound. Grounded in a pragmatic philosophy, it combined computational techniques, such as transformer-based sentiment analysis, topic modelling, and temporal analysis, with qualitative thematic coding of official documents. Data was manually collected from X due to API restrictions, targeting three key policy milestones to capture shifts in public sentiment. A domain-specific gold standard enabled accurate model fine-tuning and evaluation. Comparative analysis, informed by communication theories, assessed alignment between public discourse and BoE narratives. Each methodological step was purposefully aligned with specific research objectives, ensuring a holistic and nuanced exploration of how the digital pound is perceived, discussed, and shaped.

Chapter 4 – Domain-specific gold standard development for digital pound sentiment analysis

4.1 Introduction

This chapter addresses the critical first step in answering the overarching aim of this study: *understanding public sentiment towards the digital pound and its evolution over time*. This foundational step directly supports several research questions, most notably RQ1 and RQ2, by providing the necessary domain-specific gold standard dataset for fine-tuning and evaluating the chosen transformer model. Furthermore, this dataset is crucial for addressing RQ3 by enabling valid and generalisable sentiment analysis across different timelines and facilitating the identification of sentiment trends and patterns.

It details the crucial process of creating a gold standard dataset for sentiment analysis of digital pound-related X data. As emphasised in sections 2.2.2 and 2.2.3, a domain-specific dataset is essential for training and evaluating machine learning models in NLP, particularly in sentiment analysis. The methodological steps used to create this benchmark dataset are described in this chapter, ensuring its validity and reliability for further study. It explains annotation rules, how a representative sample of tweets is chosen, the recruitment and training of annotators, the evaluation of inter-annotator agreement (IAA), and the implementation of quality control procedures, including an adjudication process. Finally, the chapter presents the inter-annotator agreement results.

4.2 The Importance of a Gold Standard in Sentiment Analysis

A gold standard dataset plays a critical role in ensuring the precision and reliability of computational models in NLP and sentiment analysis. As discussed in Chapter 2, a gold standard dataset is a carefully selected dataset that has undergone manual annotation and verification for particular attributes, including sentiment. While machine learning models can process vast amounts of data to identify patterns, they *fundamentally* rely on high-quality annotated datasets to learn effectively. In the context of sentiment analysis, the gold standard provides the ground truth against which a model's performance is measured and validated. Without a robust gold standard, the credibility and reliability of sentiment analysis results are significantly compromised. Therefore, the significance of a gold standard in sentiment analysis cannot be overstated [152].

Furthermore, for this research, the creation of a gold standard dataset is particularly important, given the lack of real-world implementation of the digital pound and the limited availability of related datasets. X and other social media platforms offer a wealth of real-time data that reflects the public's thoughts, feelings, and trends [66]. However, reliable sentiment analysis faces substantial hurdles due to tweets' naturally informal and unstructured nature. Slang, acronyms, emojis, and context-dependent terms are common in tweets, which makes it more difficult for algorithms to automatically interpret sentiment. Generic sentiment analysis models, trained on general-purpose text, often struggle to capture the specific nuances and domain-specific vocabulary used in discussions about complex topics like CBDCs. Therefore, a domain-specific gold standard, tailored to the language and context of digital pound discourse, is essential for achieving accurate and reliable sentiment analysis in this specific domain. This necessitates a meticulous and time-consuming manual annotation process, leveraging human expertise to identify and categorise the sentiment expressed in these tweets. The inter-annotator agreement scores (post adjudication process), which will be presented later in this chapter, demonstrate the robustness of this annotation process.

4.3 Gold Standard Construction Methodology

Several methodical steps were taken to create a separate gold standard dataset for each timeline. This provides a reliable benchmark for evaluating the performance of the chosen sentiment analysis model(s) across different contexts and time periods. This methodological approach is essential for addressing RQ1, RQ2, and RQ3. These findings, in turn, provide valuable context for the analysis of official response documents (RQ4) and the comparative analysis of public and policy narratives (RQ5). The following steps were undertaken:

4.3.1 Step1: Annotation Guidelines Development

As rigorous and thorough linguistic analysis is critical in the annotation process [153], the first step in the creation of a gold standard dataset involves the development of comprehensive annotation guidelines. For this study, three commonly used sentiment labels were selected: Positive, Negative, and Neutral. These categories were chosen to capture the range of sentiment expressed in the digital pound discourse, providing sufficient granularity for analysis while maintaining inter-annotator agreement. The following guidelines were provided to annotators to identify hidden sentiments in tweets:

- **Positive sentiment**: Tweets expressing enthusiasm or support for the digital pound, including the use of positive emoticons. For instance, "*a digital pound, that sounds interesting. looks like the bank of england has big plan ahead."* The sentiment expressed in this tweet is positive as it shows favourism towards digital pound.
- **Negative sentiment**: Tweets containing criticism or scepticism about the digital pound, including negative experiences and emoticons. For instance, "*hell no to cbdc. we understand how important our individual anonymity is. we don't need central powers to dictate our decisions."* The sentiment in this tweet is negative as it clearly shows criticism regarding CBDC.
- **Neutral sentiment**: Informational tweets and questions about the digital pound without sentiment or ambiguous feelings. For instance, "*do high street banks want cbdc's, or do they hate the idea?"* The sentiment in the given tweet is neutral since it is a question seeking information without expressing a positive or negative opinion.

The examples (randomly chosen) provided for each labelling category belong to the 2023 timeline data. While the above guidelines ensured that annotators understood the criteria for each sentiment category, it is acknowledged that annotations alone do not thoroughly comprehend linguistic processes; instead, they serve as a means to encode that understanding [154]. In contrast to rule-based systems, which necessitate formal generative rules, annotations provide a more flexible and incremental approach. This research handled annotations by leveraging the annotators' diverse expertise with the digital currency context *(see Section 4.3.3)*.

4.3.2 Step2: Select a Representative Sample

To ensure the annotated subset of tweets accurately reflects the diversity of perspectives, linguistic styles, and themes present in the larger dataset, a statistically significant random sample of tweets was selected from each timeline. This sampling strategy is essential for creating a manageable yet representative dataset for annotation, maximising the generalisability of the sentiment analysis results. This study considered several factors when determining the appropriate sample size for each timeline:

- Firstly, population size was taken into consideration, as a higher sampling percentage is typically used for smaller populations due to the importance of each data point in the overall dataset.
- Secondly, to ensure accuracy, the desired confidence level and margin of error were factored in; a 95% confidence level and a 5% margin of error were chosen, necessitating larger sample sizes.
- Lastly, resource and time constraints were also considered, as the available resources, including time and human annotators, impact the feasible sample size. Specific sample sizes for each timeline were calculated using the finite population correction (FPC)-adjusted formula[4] to address these factors, ensuring statistical significance while balancing resource constraints.

For this study, as noted in Section 3.3.1.4, the tweet populations for the three timelines are as follows:

- **2020**: 279 tweets
- **2023**: 4702 tweets
- **2024**: 1219 tweets

The first step involved loading the datasets containing tweets from the respective timelines. To achieve this, the pandas library was to read these datasets from Excel files and to ensure that the datasets contain unique tweets, duplicate entries (if any) were removed. This is done using the '*drop_duplicates*' function in pandas, specifying the 'Tweet' column to identify duplicates. The number of unique tweets remain the same, as noted above i.e., 279, 4703 and 1219 tweets for all three timelines respectively. Then, the sample size for each timeline was calculated using the FPC-adjusted formula.

Notably, the standard formula assumes an infinitely large population (where removing a few observations does not significantly affect the remaining population). However, this assumption does not hold true for finite populations, such as in the case of the study under consideration. Therefore, a finite population correction or FPC-adjusted formula was used to ensure accurate representation. The FPC factor accounts for the decrease in variability that occurs when sampling

[4] https://www.statisticshowto.com/finite-population-correction-factor/

a significant portion of the population [143]. This adjustment is necessary when dealing with smaller populations. The resulting sample sizes and percentages are shown in Table 4.1.

Timeline	Recommended sample size (approx-rounded up to whole numbers)	Percentage of total tweets
2020	162	58%
2023	355	7.5%
2024	292	23.6%

Table 4.1: The recommended sample sizes and corresponding percentages of the total tweets for each timeline.

A higher percentage of 58% was required for the 2020 timeline due to the smaller population size (279 tweets), where each tweet represents a more significant portion of the total population. On the contrary, the required percentage of the sample was lower (7.5%) despite having the largest population, as the large size of the population dilutes the influence of individual tweets. Similarly, for the 2024 timeline, the need for representativeness and manageability was balanced with the intermediate population size, resulting in 23.6% of total tweets being used as a sample size. Considering the distinct qualities of each population size, these calculations ensured that each timeline's sample was robust and representative of the overall sentiment.

For the 2020 timeline, the 'tweet' column from the *df_2020* DataFrame was isolated. A sample of 162 tweets was selected from this column using the sample method, with '*random_state=0'* to ensure reproducibility. This sample was stored in the '*df_2020_sampling'* DataFrame. The same procedure was applied to the 2023 and 2024 timelines, where 355 and 292 tweets were sampled and stored in the '*df_2023_sampling'* and '*df_2024_sampling'* DataFrames.

4.3.3 Step3: Recruit and Train Annotators

To reduce bias and enhance consistency, three trained annotators independently labelled each random sample of tweets. They were selected for this study based on their proficiency in English and familiarity with the context (e.g., digital currencies). All three annotators are based in the UK and aware of the Bank of England's ongoing efforts to design and implement the digital pound in the UK. The annotators are Tayyub Yaqoob (Annotator1), Ramakrishnan Subramanian (Annotator2), and Jesse Mensah (Annotator3). Tayyub Yaqoob is the Digital Analyst at 8 Million Stories, Ramakrishnan Subramanian is the Director of Data and Analytics at SquareTrade Europe, and Jesse Mensah is a Data Scientist at Places for People. Annotator1 and Annotator2 each have over 4 and 10 years of experience working with data, including NLP, while Annotator3 has 5 years of experience as a Data Scientist, including some NLP projects. All annotators were provided with the guidelines (as noted in Section 4.3.1) during a remote meeting. Each annotator was assigned a random sample from all three timelines to label tweets independently. They were given 10 days to determine their decision points (based on given guidelines) and provide the labelled datasets. A human annotator must make a series of judgments to complete an annotation task, which are called decision points [155].

The decision behind recruiting three annotators is based on the fact that discrepancies between two annotators could lead to a binary disagreement, which may disregard the true nature of the data. Adding a third annotator helps to achieve a majority decision, minimising the impact of a single annotator's subjective perspective on tweets. In cases where two annotators disagree, a third can act as a tiebreaker and establish a more reliable agreement. This triangulation process makes the labelled data more resilient and reliable for further analysis. Furthermore, an additional annotator helps create a pool of expertise and lead to a multifaceted understanding of the data, enriching the overall annotation process [156]. Also, three annotators make finding mistakes that one or two annotators could miss easier. This redundancy contributes to a cleaner and more accurate dataset.

4.3.4 Step4: Inter-Annotator Agreement Assessment

In the NLP field, it is often assumed that the annotations within the gold standard reflect the ground truth of the given domain and are, therefore, accurate [153]. This assumption serves as the foundational premise for gold standards in evaluation processes. Alternatively, the extent of accuracy is expected to be well understood. This is where inter-annotator agreement or IAA plays a key role in validating the trustworthiness of the gold standard by assessing annotation consistency and reliability among different annotators. To understand the level of consensus among multiple annotators, inter-annotator agreement, a statistical measure, ensures that annotations are consistent across different individuals rather than being subjective or arbitrary. As such, high IAA indicates that annotations can be used to train and evaluate NLP models as they are considered reliable and credible sources of truth [157]. Moreover, the importance of IAA is underscored by the annotators' subjective interpretations, which can influence annotations due to differences in annotators' understanding of guidelines, background knowledge, and personal biases [158] Therefore, measuring IAA helps identify discrepancies and resolve them for the subsequent use of this data in machine learning models.

After the annotators completed their individual labelling tasks, the annotated files were collected and compared to assess the level of agreement across the three timelines. The comparison revealed some discrepancies in sentiment labelling across three timelines. The bar chart (see Figure 4.1) illustrates the number of tweets categorised by agreement levels among annotators for 2020, 2023, and 2024. The agreement levels are divided into three categories: "All Agreed," "At least 2 Agreed," and "None Agreed." In 2020, most tweets had at least two annotators in agreement (105), while only a few tweets had no agreement (17). In 2023, the highest number of tweets (245) fell into the "At least 2 Agreed" category, with fewer tweets in the "All Agreed" (67) and "None Agreed" (26) categories. Similarly, in 2024, "At least 2 Agreed" dominated with 204 tweets, followed by "All Agreed" (51) and "None Agreed" (36). This distribution highlights varying levels of consensus across different timelines.

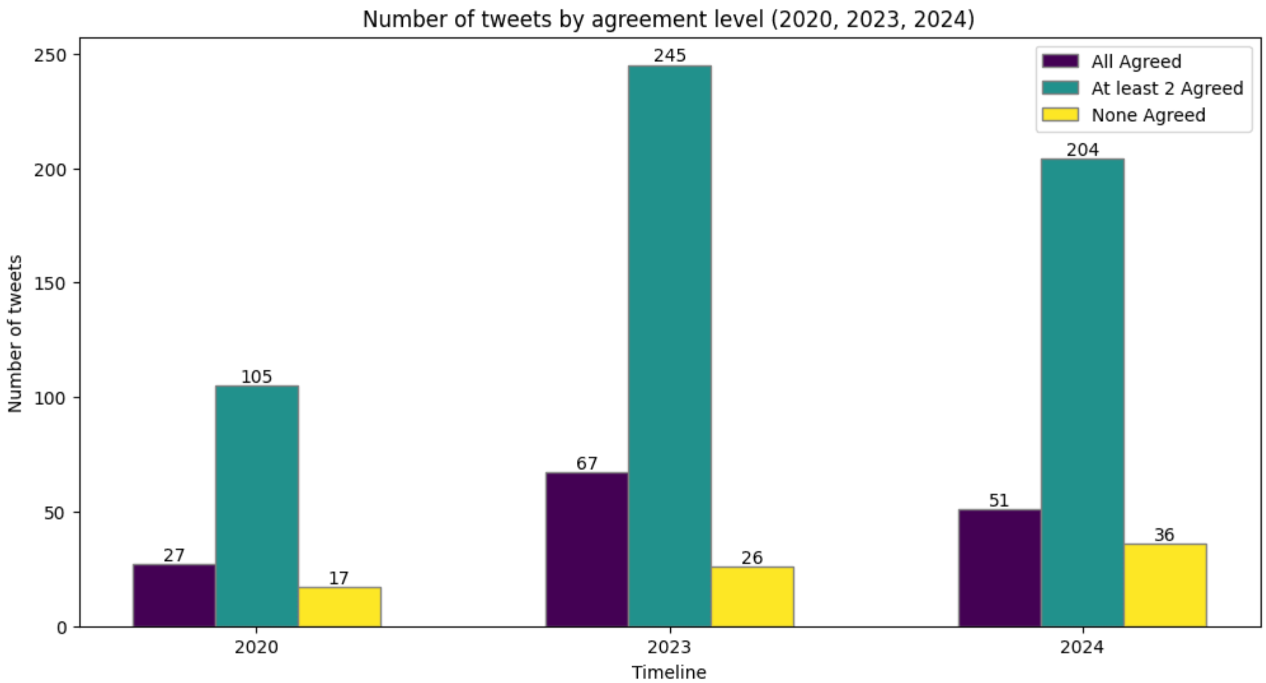


Figure 4.1: Varying agreement levels among annotators across three timelines.

A two-tier approach was adopted to simplify the annotation process and resolve the disagreement among annotators. Tweets, where at least two annotators agreed, were accepted as the final label; this approach is based on the principle that majority agreement among multiple annotators reduces subjectivity and increases the likelihood of accurate and reliable annotations [157]. This majority-rule approach also simplifies the process, allowing for the efficient resolution of most disagreements.

However, tweets where none of the annotators agreed required further attention to determine the most appropriate label. These tweets were subjected to an adjudication process to resolve discrepancies (step 5). The bar chart (Figure 4.2) illustrates the number of tweets where none of the annotators agreed across 2020, 2023, and 2024 timelines. The number of such tweets increased from 17 in 2020 to 26 in 2023 and further to 36 in 2024. This increasing trend suggests a potential increase in the complexity or ambiguity of tweet content over time, possibly due to evolving language, the emergence of new themes, or increased polarisation of the discourse.

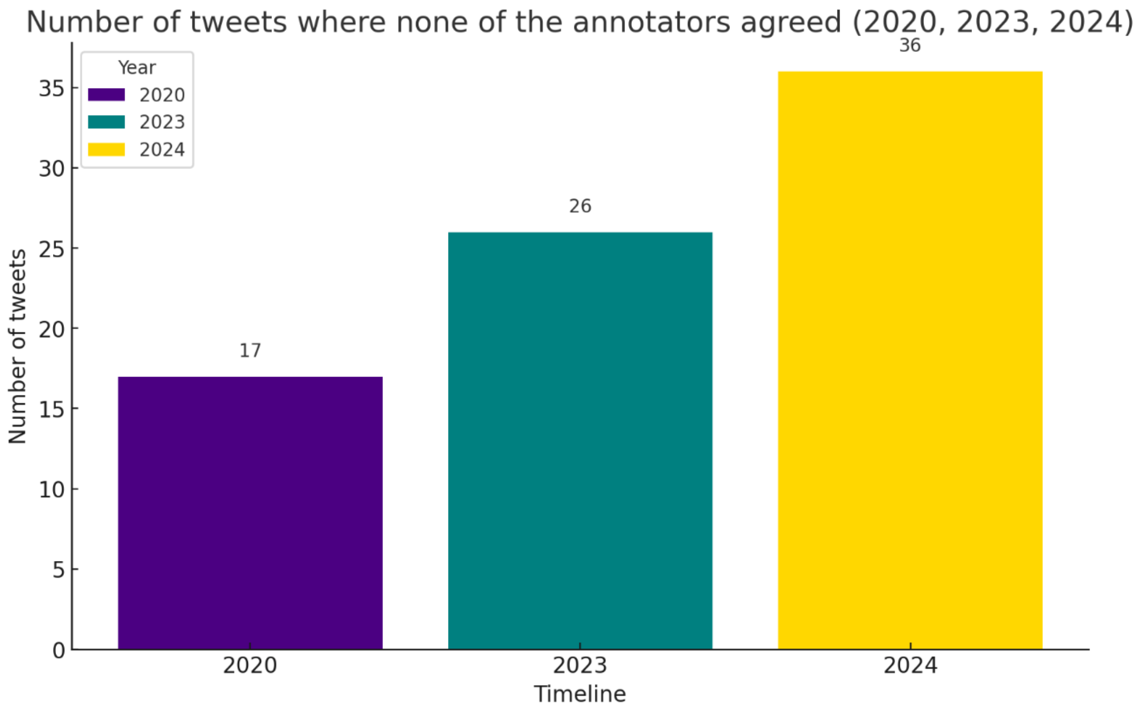


Figure 4.2: No. of tweets where none of the annotators agreed (all timelines).

Moreover, annotators 2 and 3 reported the presence of irrelevant tweets within the dataset during the annotation process. Figure 4.3 shows the number of irrelevant tweets identified across three timelines. The year 2023 exhibits a significantly higher count of irrelevant tweets (431) compared to 2020 (29) and 2024 (38), indicating a substantial increase in noise within the dataset for 2023. Including irrelevant tweets in the IAA calculation could introduce noise and bias, potentially undermining the reliability and validity of the gold standard. Therefore, these irrelevant tweets were removed from the dataset *before* calculating IAA, ensuring that the agreement scores reflect the consistency of sentiment annotation on relevant content. This step was crucial for maintaining the integrity of the gold standard.

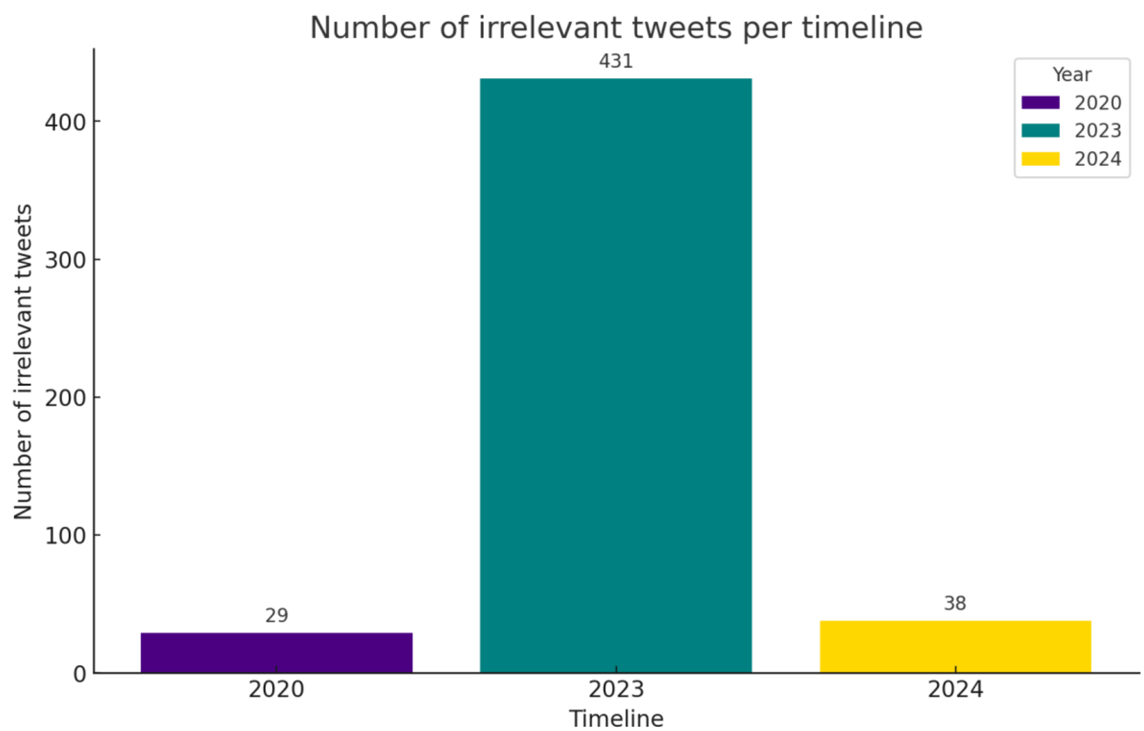


Figure 4.3: Number of irrelevant tweets per timeline.

Table 4.2 provides examples of tweets where disagreement among annotators is evident:

Timeline	Tweet	Annotator	Sentiment	Explanation
2020	"the appropriate degree of anonymity in a cbdc system is a political and social question, rather than a narrow technical question. as discussed above, cbdc would need to be compliant with aml regulations, which rules out truly anonymous payments."	Annotator1	Negative	This tweet discusses the tension between anonymity and regulatory compliance. Annotator 1 may have focused on the negative implications of AML regulations for anonymity, while Annotator 2 might have seen the need for regulation as a positive aspect. Annotator 3 likely focused on the objective nature of the statement, classifying it as neutral.
		Annotator2	Positive	
		Annotator3	Neutral	

2023	“i can't prove the negative. yes, any information captured can be abused, any capability exploited but both are already here through the much more powerful control of smartphones, banking, video and credit cards. i believe the ecb is serious about cbdc privacy, not its anonymity.”	Annotator1	Negative	This tweet acknowledges the potential for abuse but expresses belief in the ECB's commitment to privacy (though not anonymity). Annotator 1 may have focused on the initial negative acknowledgment, while Annotator 2 focused on the positive belief in ECB's intentions. Annotator 3 likely interpreted the mixed sentiment as neutral.
		Annotator2	Positive	
		Annotator3	Neutral	
2024	“don't confuse lack of price movement with lack of development. $qnt is at the center of the uk digital pound and retail bank sTable coin implementation. it will be this year and lots will be annoyed they sold this blue-chip.”	Annotator1	Neutral	This tweet mixes factual statements about development with a speculative prediction and a cautionary note. The different interpretations of this mixed message likely led to the varied sentiment classifications.
		Annotator2	Negative	
		Annotator3	Positive	

Table 4.2: Examples of annotator disagreements.

After reviewing the annotated datasets, all irrelevant tweets were manually identified and removed. This required carefully reviewing every tweet to assess its applicability in light of the predetermined guidelines set forth for this study. Irrelevant tweets that do not pertain to the context of the digital currency landscape, CBDC, or the Bank of England’s digital pound initiative, were excluded from the dataset. For several reasons, tweets were removed manually instead of using automated methods, such as Python scripts or other computational tools. Firstly, the relevance is context-specific and subtle, requiring a level of discernment and understanding that is hard for automated filtering algorithms to fully capture. Due to their experience and contextual knowledge, it was more appropriate for the researcher to make judgments. Secondly, even if the number of tweets was substantial, it was still within a manageable range for manual review, enabling a careful and accurate cleaning process without requiring excessive time investment. The researcher aimed to enhance the quality of the annotated data by removing these irrelevant tweets, thereby ensuring agreement among annotators on all the tweets of the assigned random samples. This step, which removed potential causes of disagreement unrelated to the actual content of interest, was critical to preserving the robustness of the gold standard.

Following the removal of irrelevant tweets, the inter-annotator agreement (IAA) was calculated on the *relevant* tweets, including those where initial disagreement had occurred. This inclusion ensured that the IAA calculation reflected the level of agreement on the tweets that were ultimately deemed pertinent to the study.

4.3.4.1 IAA Analysis and Metrics for Inter-Annotator Reliability

This section will explain the IAA process adopted for this study, including various metrics like Cohen's Kappa, Fleiss' Kappa, and Krippendorff's Alpha.

4.3.4.1.1 Cohen's Kappa

Cohen's Kappa (κ) is a statistical measure used to evaluate the agreement level between two annotators, correcting for chance agreement. It accounts for the possibility that annotators may agree or disagree purely by chance.

A Cohen's Kappa value of 0 indicates no agreement is better than chance, 1 indicates perfect agreement and negative values suggest less than chance agreement. Landis and Koch [159] provide a commonly used interpretation scale for κ values:

- <0: Poor
- 0.01–0.20: Slight
- 0.21–0.40: Fair
- 0.41–0.60: Moderate
- 0.61–0.80: Substantial
- 0.81–1.00: Almost Perfect

To compute Cohen's Kappa, each pair of annotators (Annotator 1 vs. 2, Annotator 1 vs. 3, and Annotator 2 vs. 3) were compared for each sentiment category and each timeline. The 'cohen_kappa_score' function from the 'sklearn.metrics' library in Python was used for this calculation. The results for each year are summarised in Tables 4.3, 4.4, and 4.5:

Sentiment	**Annotator 1 vs. 2**	**Annotator 1 vs. 3**	**Annotator 2 vs. 3**

Positive	0.101	0.064	0.165
Neutral	0.059	0.026	0.266
Negative	0.161	0.169	0.301

Table 4.3: Cohen's Kappa scores for 2020.

Sentiment	**Annotator 1 vs. 2**	**Annotator 1 vs. 3**	**Annotator 2 vs. 3**
Positive	0.061	0.075	0.025
Neutral	0.112	0.180	0.201
Negative	0.153	0.441	0.171

Table 4.4: Cohen's Kappa scores for 2023.

Sentiment	**Annotator 1 vs. 2**	**Annotator 1 vs. 3**	**Annotator 2 vs. 3**
Positive	-0.009	0.143	0.007
Neutral	0.048	0.093	0.169
Negative	0.062	0.418	0.156

Table 4.5: Cohen's Kappa scores for 2024.

The agreement between annotators for positive sentiments is consistently low across all timelines, with κ values mainly in the range of slight agreement (0.01–0.20). In 2024, the agreement seems to be worse than chance due to a negative kappa score between annotators and 2, suggesting that positive sentiments are particularly challenging for annotators to identify consistently, which is common in complex domains like financial discourse [70]. In contrast, agreement on neutral sentiment shows slightly better, but still generally low, levels of agreement. The κ values range from slight to fair agreement across the timelines, with some improvements in 2023 compared to 2020 and 2024. Annotator pairs involving Annotator3 in 2023 displayed fair agreement, indicating some level of consensus improvement. Furthermore, negative sentiments exhibit the highest levels of agreement among annotators. In 2023 and 2024, especially between Annotator1 and 3, the agreement reached moderate levels (κ values above 0.40), indicating that negative sentiments are more consistently identified across annotators, possibly due to their more distinct and recognisable characteristics. Overall, a challenge in achieving high consistency in sentiment analysis is evident due to low to moderate levels of agreement across sentiments and timelines.

4.3.4.1.2 Fleiss' Kappa

Fleiss' Kappa (κ) is suitable for categorical data and adjusts for the agreement occurring by chance [160]. It is an extension of Cohen's Kappa used to evaluate the agreement between more than two annotators.

Like Cohen's Kappa, Fleiss' Kappa ranges from -1 to 1; values closer to 1 indicate high agreement, while values closer to -1 indicate disagreement and values around 0 indicate no agreement better than chance.

This metric is calculated using the 'fleiss_kappa function' from the 'statsmodels.stats.inter_rater' library. The results of the sentiment-specific Fleiss' Kappa scores are shown in the Table 4.6:

Year	Positive	Neutral	Negative
2020	0.077	0.084	0.196
2023	-0.0204	0.1105	0.187
2024	-0.022	0.051	0.224

Table 4.6: Sentiment-specific Fleiss' Kappa scores.

The Fleiss' Kappa scores for positive sentiment show minimal agreement across all years. This suggests that annotators struggled to consistently agree on positive sentiments, with the highest value in 2020 and the lowest in 2023. This reinforces the findings from Cohen's Kappa, indicating that positive sentiment is particularly challenging for annotators to consistently identify. With the maximum score in 2023, Fleiss' Kappa for neutral sentiment suggests a moderate degree of agreement. This shows that annotators were more consistent in identifying neutral sentiments for this year. Fleiss' Kappa indicates moderate agreement for negative sentiment with the highest score in 2024. This confirms the trend observed with Cohen's Kappa, suggesting that negative sentiment is more consistently identified by annotators.

In addition to the above, overall Fleiss' Kappa scores were also calculated to aggregate sentiment-specific information, offering a single, overarching view of annotator reliability. The function 'calculate_fleiss_kappa' computes Fleiss' Kappa for the given annotation data. The results for each year are summarised in the Table 4.7:

Year	Fleiss' Kappa

2020	0.117
2023	0.133
2024	0.093

Table 4.7: Fleiss' Kappa scores for all the timelines.

Across all timelines, the Fleiss' Kappa values are consistently in the "poor agreement" range, indicating little consensus among annotators. These low overall scores, despite the relatively higher agreement on negative sentiment, are likely influenced by the consistently low agreement on positive and neutral sentiment, which constitute a larger portion of the dataset. This highlights the challenges inherent in sentiment annotation, particularly in complex domains.

4.3.4.1.3 Krippendorff's Alpha

Krippendorff's Alpha (α) is particularly useful in content analysis and other fields where data may not be purely nominal or where missing values are present [161]. It is a versatile reliability coefficient that can be used for multiple annotators, different levels of measurement (nominal, ordinal, interval, ratio), and even for incomplete data sets.

Krippendorff's Alpha is computed using the 'krippendorff' module. The sentiment-specific Krippendorff's Alpha scores are shown in the Table 4.8:

Year	Positive	Neutral	Negative
2020	0.079	0.086	0.198
2023	-0.019	0.111	0.188
2024	-0.021	0.052	0.225

Table 4.8: Sentiment-specific Krippendorff's Alpha scores.

The sentiment-specific Krippendorff's Alpha scores largely mirror the trends observed with Cohen's and Fleiss' Kappa. Krippendorff's Alpha scores for positive sentiment reflect minimal agreement, with values dropping to slight disagreement in 2023 and 2024. This further confirms the difficulty in achieving consistent annotation of positive sentiment. Krippendorff's Alpha shows slight agreement for neutral sentiment, peaking in 2023, reinforcing the observation of improved annotator alignment for neutral sentiments in that year. Krippendorff's Alpha scores demonstrate moderate agreement for negative sentiment, with the highest score in 2024, confirming improved reliability in identifying negative sentiments.

Like Fleiss's Kappa scores, overall Krippendorff's Alpha scores were also calculated to provide an aggregated measure of inter-annotator reliability across all sentiment categories. The function 'calculate_krippendorffs_alpha' computes Krippendorff's Alpha for the given annotation data. Table 4.9 summarise the results for each year:

Year	**Krippendorf's Kappa**
2020	0.145
2023	0.127
2024	0.151

Table 4.9: Krippendorf's Kappa scores for all the timelines.

The overall Krippendorff's Alpha scores indicate relatively stable agreement, with a slight dip in 2023 and a recovery in 2024, suggesting that while there were fluctuations, the overall reliability of annotations remained relatively consistent.

4.3.4.1.4 The Need to Address Low IAA

The consistently low IAA scores across all metrics underscore the importance of the adjudication process (explained in Section 4.2.5) in ensuring the quality and reliability of the gold standard. While the IAA scores are not ideal, they are not unexpected in sentiment analysis tasks, especially in complex domains with nuanced language and subjective interpretations [71]. The use of multiple annotators, combined with a rigorous adjudication process, is essential for mitigating the impact of individual annotator bias and improving the overall quality of the gold standard. The adjudication process, which will be described next, plays a critical role in resolving disagreements and creating a more consistent and reliable dataset for subsequent analysis.

4.3.5 Step5: Quality Control Via an Adjudication Process

In this study, the author addressed the challenge of inconsistent annotations through an adjudication process. As the primary researcher and subject matter expert (SME), I served as the adjudicator. Adjudication is critical in resolving discrepancies between annotators to ensure the final dataset is reliable and consistent [162]. The adjudication process involved the following steps:

- **Identify disagreements:** Extract all instances where annotators disagreed on the sentiment labels (Positive, Neutral, Negative).
- **Review disagreements:** Based on predefined annotation guidelines (see Section 4.3.1.) and personal expertise, the author carefully reviewed each tweet where disagreements occurred. Particular attention was paid to the context of the tweet, nuanced language, and potential ambiguities in sentiment expression.
- **Determine final labels:** As the adjudicator, I determined the most appropriate sentiment label for each tweet where disagreements were noted. The adjudicator's decisions were

guided by the goal of aligning the final labels with the overall context and intended sentiment expressed in the tweets.

- **Create an adjudicated dataset:** Then I gathered all the adjudicated labels into a single dataset, ensuring each tweet has one final, agreed-upon sentiment label.

4.3.5.1 Recalculation of IAA Metrics

To adhere to the methodological rigour expected in academic research, the IAA metrics were recalculated. The adjudication process attempts to settle differences and establish a final set of agreed-upon labels by having a SME or prime researcher reconcile divergent annotations. By recalculating IAA metrics post-adjudication, researchers can ensure that reliability scores are based on final, authoritative annotations rather than preliminary, unresolved disagreements, allowing them to obtain a measure that accurately reflects the true level of consensus among annotators [157]. Recalculation is also essential in documenting whether the agreement has been approved over time [163]. In addition, the transparency and validity of their methodological approach strengthen the credibility of any subsequent analyses and conclusions drawn from the data [162], [164].

The Cohen’s Kappa scores after adjudication show varying levels of agreement across the three years and sentiment categories. The scores reflect a range from fair to moderate agreement for the overall year-by-year analysis (see Table 4.10). In 2020, the scores indicated a fair to moderate agreement across all annotator pairs, with Annotator2 vs. Annotator3 showing the highest agreement (0.3977). In 2023, while the agreement between Annotator1 and Annotator3 improved to a moderate level (0.3949), the other pairs showed lower, yet still fair, agreement. By 2024, the agreement between Annotator1 and Annotator3 further improved, reaching a moderate level (0.4552), suggesting that the adjudication process effectively enhanced the consistency between these annotators.

The sentiment-specific Cohen's Kappa scores (Tables 4.11, 4.12, and 4.13) reveal varying levels of agreement across sentiment categories and timelines. In 2020 (Table 4.11), agreement ranges from fair to moderate, with negative sentiment showing the highest agreement (Kappa values between 0.2984 and 0.4204), indicating greater consistency in identifying negative sentiments. Neutral sentiment shows moderate agreement specifically between Annotator2 and Annotator3 (0.4809), while positive sentiment, typically more subjective, exhibits slightly lower agreement. In 2023 (Table 4.12), agreement trends are similar, with neutral sentiment showing moderate agreement, particularly between Annotator2 and Annotator3 (0.3305), and negative sentiment again showing the strongest agreement, reaching a moderate level between Annotator1 and Annotator3 (0.5960). By 2024 (Table 4.13), agreement improves across most categories. Negative sentiment demonstrates the most substantial improvement, with high agreement between Annotator1 and Annotator3 (0.7389). Neutral sentiment also shows moderate agreement, especially between Annotator2 and Annotator3 (0.4077), while positive sentiment continues to exhibit lower agreement, though some improvement is noticeable.

The moderate agreement levels observed, especially for the more distinct sentiment categories and after the adjudication process, suggest that the annotators achieved a reasonable level of consensus.

This level of agreement, while not perfect, supports the validity of the findings and indicates that the adjudicated gold standard dataset is suitable for subsequent analysis. The improved agreement after adjudication highlights the effectiveness of the process in enhancing the reliability and consistency of the sentiment labels.

Year	Annotator 1 vs. 2	Annotator 1 vs. 3	Annotator 2 vs. 3
2020	0.2904	0.2376	0.3977
2023	0.2088	0.3949	0.2299
2024	0.2299	0.4552	0.2662

Table 4.10: Cohen’s Kappa scores after adjudication (year-wise).

Sentiment	Annotator 1 vs. 2	Annotator 1 vs. 3	Annotator 2 vs. 3
Positive	0.2980	0.1515	0.2875
Neutral	0.2737	0.2111	0.4809
Negative	0.2984	0.3615	0.4204

Table 4.11: Sentiment-specific Cohen’s Kappa scores for 2020 after adjudication.

Sentiment	Annotator 1 vs. 2	Annotator 1 vs. 3	Annotator 2 vs. 3
Positive	0.1633	0.1389	0.0841
Neutral	0.2144	0.3111	0.3305
Negative	0.2413	0.5960	0.2500

Table 4.12: Sentiment-specific Cohen’s Kappa scores for 2023 after adjudication.

Sentiment	Annotator 1 vs. 2	Annotator 1 vs. 3	Annotator 2 vs. 3
Positive	0.1979	0.2674	0.1345

Neutral	0.2016	0.2805	0.4077
Negative	0.2800	0.7389	0.2356

Table 4.13: Sentiment-specific Cohen's Kappa scores for 2024 after adjudication.

The overall Fleiss' Kappa scores after adjudication (see Table 4.14) show slight to fair agreement, with 2020 (0.2913) having the highest overall consistency, followed by 2024 (0.2846) and 2023 (0.2494). Sentiment-specific analysis (see Table 4.15) reveals that negative sentiment consistently achieved the highest agreement across all years peaking in 2024 (0.3902). Neutral sentiment also shows moderate agreement, particularly in 2020 (0.3044), while positive sentiment consistently exhibits the lowest agreement, with 2023 (0.0888) showing the greatest difficulty in reaching consensus. These results indicate that, even after adjudication, agreement on negative sentiments was more readily achieved.

Year	Fleiss' Kappa
2020	0.2913
2023	0.2494
2024	0.2846

Table 4.14: Overall Fleiss' Kappa score after adjudication.

Year	Positive	Neutral	Negative
2020	0.2238	0.3044	0.3541
2023	0.0888	0.2805	0.3324
2024	0.1641	0.2737	0.3902

Table 4.15: Sentiment-specific Fleiss' Kappa scores after adjudication.

The overall Krippendorff's Alpha scores after adjudication show slight to fair reliability, with 2024 (0.3133) displaying the highest agreement (see Table 4.16), followed by 2020 (0.2739) and 2023 (0.2405). Sentiment-specific analysis (see Table 4.17) highlights that negative sentiment consistently achieved the highest reliability particularly in 2024 (0.3908). Neutral sentiment shows moderate reliability, especially in 2020 (0.3059), while positive sentiment has the lowest reliability, with the lowest agreement in 2023 (0.0896). These findings align with other statistical tests, consistently showing stronger agreement on negative sentiments after adjudication.

Year	Krippendorff's Alpha
2020	0.2739
2023	0.2405
2024	0.3133

Table 4.16: Overall Krippendorff's Alpha scores after adjudication.

Year	Positive	Neutral	Negative
2020	0.2255	0.3059	0.3555
2023	0.0896	0.2812	0.3331
2024	0.1650	0.2745	0.3908

Table 4.17: Sentiment-specific Krippendorff's Alpha scores after adjudication.

4.3.5.2 Implications for the Research

Despite a rigorous adjudication process that aimed to resolve inconsistencies, the inter-annotator agreement, while improved, remained moderate overall. The moderate agreement achieved, especially for negative and, to a lesser extent, neutral sentiment, combined with the rigorous adjudication process, provides a reasonable level of confidence in the reliability of the adjudicated gold standard dataset. Mozetič et al. [165]report Krippendorff's α values as low as 0.12 ± 0.03 for Albanian and 0.12 ± 0.04 for Spanish Twitter corpora, even after removing inconsistent annotators. Their findings show that model performance tends to converge to these modest agreement levels, indicating that inter-annotator reliability often sets a practical upper bound on predictive accuracy, emphasising the need to monitor annotation consistency rather than expecting models to overcome noisy gold labels [166] . Similarly, Bobicev and Sokolova [166] find average Fleiss' κ and α scores around 0.46 in a multi-label health forum dataset, with many label-specific scores falling below 0.4, confirming that moderate agreement is typical in subjective or domain-specific sentiment tasks [166].

In this context, the present study's agreement scores, i.e., Cohen's κ ranging between 0.21 and 0.39 across sentiment classes and years, and Krippendorff's α between 0.24 and 0.31 overall, fall within the lower but not uncommon bounds for multi-rater sentiment annotation. While "moderate" at best, these figures are above the lower tail observed in prior multilingual and multi-label sentiment studies and reflect the real-world difficulty of assigning sentiment labels to social media discourse, especially for positive sentiment. The use of rigorous adjudication and robustness checks mitigates these limitations and ensures that the annotated dataset supports reliable model training within known performance constraints.

Accordingly, the dataset can be deemed suitable for subsequent analysis, provided that its inherent limitations, particularly the lower agreement observed for positive sentiment, are acknowledged and that findings are interpreted with appropriate caution. This includes recognising that sentiment labels may contain a degree of subjectivity, especially for positive cases, and avoiding overgeneralisation of model performance or downstream analytical claims without referencing these annotation constraints.

4.4 Conclusion

The meticulous process used to produce a robust gold standard dataset for sentiment analysis of tweets about the digital pound is detailed in this chapter. Despite moderate initial inter-annotator agreement, the gold standard dataset's robustness is ensured by rigorous annotation guidelines, a thorough adjudication process, and a focus on relevant content, providing a reliable basis for subsequent analysis.

This rigorous process directly addresses the foundational requirements for investigating the central research questions. Specifically, the resulting gold standard dataset directly provides the essential foundation for addressing RQ1, which aims to compare and fine-tune various transformer models for sentiment analysis in this domain, by providing the necessary labelled data for model training and evaluation. Furthermore, it provides the essential foundation for addressing RQ2 and RQ3, which explore public sentiment towards the digital pound across different timelines, by ensuring the accuracy and reliability of the sentiment analysis performed on the tweets from these periods. While achieving perfect inter-annotator agreement, particularly for nuanced positive sentiments, proved challenging, the final adjudicated dataset exhibits a respectable level of reliability for further analysis presented in the following chapters.

Building on this annotated data, Chapter 5 will detail the experimentation pipeline employed to evaluate three prominent transformer models and the subsequent selection of the most effective model for sentiment analysis within this specific domain. This research tackles the challenge of data scarcity in the context of the *prospective* digital pound, the implementation and piloting of which are yet to occur, by developing a robust, annotated dataset (a gold standard) that enables the analysis of public discourse.

Chapter 5 – Experimentation with transformer models

5.1 Introduction

This chapter presents a comprehensive overview and comparative analysis of three transformer-based models — DistilBERT, RoBERTa, and XLM-RoBERTa — for sentiment analysis of data collected from X about the UK's CBDC. The objective is to identify the ideal model capable of accurately predicting sentiments on unseen data. The chosen transformer models were fine-tuned on the gold standard dataset, finalised in Chapter 4, to facilitate this model selection.

While extensive hyperparameter tuning can often lead to marginal performance gains, it also introduces significant computational overhead. As such, hyperparameter settings for this study were adopted according to recommendations in the established literature; the empirical evidence supporting this approach states that pre-trained transformer models exhibit robust performance with these standard configurations across diverse NLP tasks. *The primary focus of the experiments conducted in this study is to choose a suitable transformer model to classify UK CBDC-related tweets (on the dataset other than the gold standard) rather than maximising predictive accuracy through exhaustive tuning.*

Furthermore, this study emphasises the effectiveness of transformer architectures in capturing complex sentiments within specialised financial discourse through experimentation and evaluation. While other studies have utilized transformer models for financial sentiment analysis, this research combines fine-tuned transformer models with a novel gold standard dataset specific to UK CBDC discourse, providing a more comprehensive and nuanced understanding of public sentiment in this emerging area.

5.1.1 Purpose of Experimenting with Multiple Models

As explored in Section 2.2.3, with the advent of transformer-based models, understanding and generating human language has been improved, ultimately revolutionising the NLP field. Vaswani et al. [112] introduced such models, a type of deep learning architecture for NLP tasks, in their paper titled "Attention is all you need." The key characteristic of these models is their "self-attention" mechanisms to capture complex linguistic patterns and contextual dependencies without requiring the sequential processing of RNNs. Understanding the relationships between words in a sequence guided by a transformer model is crucial for tasks such as sentiment analysis [113].

Selecting the appropriate model is paramount to achieving accurate and reliable results in X's dynamic and informal environment. Moreover, X data is often characterised by brevity, noise, slang, emojis, abbreviations, and code-switching between languages [167]. As a result, the effectiveness of sentiment analysis models in this domain is contingent upon their ability to understand and interpret these complex expressions.

Thus, experimenting with multiple transformer models serves several critical purposes:

- **Understanding model strengths and weaknesses:** Every model has particular architectural elements and pre-training techniques that bestow specific benefits and drawbacks. For example, certain models might prioritise computing efficiency, while others might improve performance in multilingual environments [116].

- **Performance benchmarking:** Transformer models of various types exhibit varying degrees of effectiveness across diverse NLP tasks. Evaluating multiple models helps researchers benchmark their performance and identify the architecture best suited to sentiment classification tasks specific to the datasets at hand [115]
- **Ensuring robustness and reliability:** A comparative analysis of different models helps reduce bias and ensures the researcher did not consciously overlook potential improvements. Moreover, this approach fosters a robust understanding of model behaviours and ensures alignment of research objectives and practical constraints with the performance of the selected model [117].
- **Optimise resource utilisation:** Each transformer model has a different size and computational demands. Evaluating multiple models helps researchers select a model that offers optimal performance without exorbitant computational costs [116], thus balancing performance with resource constraints.

Given the above considerations, this study meticulously evaluates DistilBERT, RoBERTa, and XLM-RoBERTa to effectively classify the sentiments of UK CBDC-related tweets. The subsequent sections provide a detailed overview of each model, explaining its benefits, architectures, and applicability to the task under consideration.

5.2 Experimental Setup

The experimental setup for comparing DistilBERT, RoBERTa, and XLM-RoBERTa was designed to ensure consistency across all models. This includes keeping hyperparameters constant, such as the number of epochs, batch size, learning rate, and optimiser, to evaluate models based on their inherent capabilities rather than different training configurations. This section outlines the training pipeline, hyperparameter choices, and other settings important for evaluating models and selecting an ideal model for further use.

5.2.1 Merging Multi-Timeline Data and Rationale for the Approach

Data from all timelines was merged to fine-tune DistilBERT, BERT, and RoBERTa transformer models to ensure comprehensive and robust sentiment analysis. Goel et al. [168]noted that social media language inherently evolves with emerging slang, is dynamic, and shifts sentiment nuances over time. Also, the datasets (for all timelines) under consideration are not large enough to enhance the models' ability to generalise across different temporal contexts [169]; therefore, merged data exposes them to diverse linguistic patterns and context variations. Moreover, the temporal bias risk is mitigated with temporal diversity in the training dataset. For instance, models trained on data from a single timeframe may underperform when encountering language from different periods [170]. In sentiment analysis, the sentiment conveyed using particular terms may vary across cultures and societies. Furthermore, effective fine-tuning of deep learning models is contingent upon large data volume, and merged datasets contribute to increased data volume to provide ample training examples to improve model accuracy and resilience against overfitting [171].

5.2.2 Consistent Training Procedures Across Models

All models were trained under identical conditions using the same merged dataset of UK CBDC-related tweets to ensure a fair comparison. During the entire training process, a standardised pipeline, including data preprocessing, tokenisation, model training, and evaluation, was followed to ensure that any observed performance differences were not attributed to variations in the training procedure or data handling but instead solely to the model architectures.

- **Data preprocessing**: As noted in Section 3.4, noise such as hashtags, non-alphanumeric symbols, URLs, and special characters were removed from raw tweets. Additionally, text was converted to lowercase to standardise the input format across all models to avoid models focusing on superficial elements like irrelevant symbols of capitalised words.
- **Tokenisation**: As transformer models support formats such as input IDs (numerical representations) and attention masks, as explained in Section 5.2, each model was tokenised using its corresponding tokeniser, including **DistilBERTTokeniser**, **RoBERTaTokeniser**, and **XLM-RTokeniser**. The conversion of tweets into token IDs is managed by the '*AutoTokeniser*' from Hugging Face's Transformers library[5] - a crucial process to ensure compatibility with the model-specific vocabularies and architectures by splitting the text into subword units. It ensures uniform input lengths by applying padding and truncation and generates attention masks that inform the model which tokens should be attended to.
- **Training and validation split**: To eliminate any variability caused by differences in data partitioning, a consistent split across all models was applied: 80:20. This standard approach is frequently used in supervised learning settings [172], [173]. As such, the dataset was split into a training set (80%, comprising 622 tweets) and a validation set (20%, containing 156 tweets) for direct performance comparisons between models.
- **Maximum sequence length set to 128:** The maximum sequence length was set to 128 tokens to standardise input lengths for the transformer models, which have a fixed positional encoding that necessitates uniform input lengths [112]. According to Wolf et al. [174], this length is sufficient for most tweets and allows for the diverse linguistic patterns and subtleties seen in social media conversation without requiring an excessive amount of memory or computational overhead. Setting a maximum sequence length also ensures that the model can handle most input instances without truncation, facilitates fast batch processing, and helps manage the variability in tweet lengths [175].
- **Hardware configuration**: The experiments were conducted using the following hardware and software setup:
  - **Hardware**: MacBook Air M1 (2020 model), 8GB RAM, 256GB SSD.
  - **GPU**: NVIDIA T4 GPU with 16GB DDR6 VRAM.
  - **IDE**: Google Colab was used as the development environment for running all experiments.

[5] https://huggingface.co/docs/transformers/en/index

5.2.3 Hyperparameter Settings

Hyperparameters influence model performance. Therefore, they should be chosen carefully to balance learning efficiency with computational cost. The following hyperparameters were chosen for all models to ensure the comparison is fair and grounded in best practices from the academic literature.

5.2.3.1 Number of Epochs: 3 and 5

An epoch represents one complete pass through the entire training dataset, and the number of epochs determines how often the learning algorithm works through the entire dataset. For this study, the models were trained for both 3 and 5 epochs, with each epoch configuration executed 30 times to ensure reliability and evaluate the impact of training duration on performance metrics. Repeated training is a well-established practice in NLP and machine learning research to account for performance variability introduced by random initialisation, mini-batch sequencing, and other stochastic elements of training [176]. Additionally, fine-tuning transformer models is a computationally intensive task, which contributed to the decision to limit this study to experiments with 3 and 5 training epochs.

Moreover, complex patterns are learned progressively with multiple epochs to improve models' generalisation ability to unseen data; although some scholars advocate for one epoch while dealing with large language models or LLMs [177], [178]. 30 independent runs for each epoch setting help mitigate the inherent variability introduced by factors such as data shuffling, random weight initialisation, and stochastic gradient descent. Furthermore, it provides a clearer picture of the model's true capabilities and variability via average performance and standard deviation [179], [180] and reduces the likelihood of overestimating the model's performance due to favourable conditions in a single run [181]. Additionally, the effect of the randomness caused by transformers' sensitivity to initial conditions and data order is averaged out with multiple runs [180], ensuring that observed performance is not attributed to random data split or seed.

A comprehensive assessment of the models' convergence behaviours and learning curves can be done by training models for 3 and 5 epochs. Unlike training models from scratch, fine-tuning transformer models typically requires fewer epochs because such models already possess extensive pre-trained knowledge [113]. Furthermore, empirical evidence by Srivastava et al. [182] suggests that if transformer models' training is extended beyond a certain point, it may lead to diminishing returns or overfitting, especially when substantial data is unavailable. Given the relatively small size of the gold standard dataset used in this study (778 tweets after merging), limiting the training to 3 and 5 epochs is a particularly prudent approach to prevent overfitting and maintain good generalisation performance. This is consistent with findings in the literature on fine-tuning transformer models for text classification tasks with limited data (e.g., Qasim et al. [25]; Mosbach et al. [183]). Qasim et al. [172], for instance, demonstrated that fine-tuning BERT on relatively small datasets or less training data often achieves optimal performance. Similarly, Mosbach et al. [183] highlighted the importance of careful regularisation and early stopping when fine-tuning pre-trained language models on limited data.

Moreover, the objective of this chapter is not to achieve state-of-the-art performance on the sentiment analysis task but rather to compare the relative effectiveness of different transformer architectures (DistilBERT, RoBERTa, and XLM-RoBERTa) for this specific domain. This approach is aligned with the broader research aim of understanding public sentiment trends, topics, and evolution (RQ3) by establishing a reliable and efficient sentiment analysis pipeline, rather than focusing on maximising individual model performance through exhaustive tuning [171]. *Further, as demonstrated by several studies highlight that transformer models exhibit a degree of robustness to hyperparameter variations [183], [184], particularly when fine-tuned using strong pre-trained checkpoints. This suggests that while fine-tuning can lead to improvements, the core architectural differences between the models are likely to have a more significant impact on performance in this comparative study.* The study aims to identify the optimal training duration by evaluating 3 and 5 epochs, which could help maximise performance without causing overfitting or unnecessary computational costs [171], [185]

5.2.3.2 Learning Rate: 1×10^{-5}

A learning rate influences how quickly a model converges to a minimum of the loss function by determining the step size during optimization. The learning rate, set at 1×10^{-5} (smaller learning rate), reduces the risk of overshooting the optimal parameters by allowing for more precise updates [171]. Empirical studies such as those by Howard and Ruder [186]show that lower learning rates benefit transformer models as they allow task-specific adjustments and preserve the integrity of pre-trained weights. A learning rate 1×10^{-5} also helps minimise the likelihood of degrading model performance via disruptive parameter updates [113]. Furthermore, based on the industry's best practices for fine-tuning large-scale language models, this learning rate helps ensure consistency and reliability in performance outcomes [112].

5.2.3.3 Batch Size

This parameter is crucial in balancing model performance and computational efficiency. It defines the number of training samples processed before the internal parameters of the model are updated. Keskar et al. [187] noted that small batch sizes aid in escaping local minima as they can introduce noise into gradient estimates. In contrast, larger batches provide accurate and stable gradient estimates. A batch size of 8 was chosen to accommodate the memory constraints, especially when utilising GPU resources to train large transformer models. This size ensures that the model receives diverse and representative samples to provide a balance between the granularity of gradient updates and computational efficiency without exceeding hardware limitations [188]. Moreover, smaller batch sizes help the model to explore the parameter space better due to the inherent noise, and such sizes have been associated with improved generalisation performance in a study by Goyal et al. [189].

5.2.3.4 Optimiser: AdamW

An extension of the traditional Adam optimiser, AdamW, was employed. It combines Adam's adaptive learning rate capabilities with a decoupled weight decay mechanism, which prevents overfitting and enhances regularisation (by penalising large weights) without adversely affecting the optimisation process [190]. It takes the first and second moments of the gradients to adjust the learning rates for each parameter individually, facilitating efficient convergence [191]. Due to their extensive parameterisation, transformer architectures are prone to overfitting; AdamW's capability to effectively manage weight decay independently of the learning rate updates makes it useful for transformer architectures. Also, compared to traditional optimisers like standard Adam and SGD, AdamW demonstrated superior performance in fine-tuning large-scale transformer models, such as for NLP tasks [175].

5.3 Evaluation Metrics

Accuracy, precision, recall, and F1-score are the primary evaluation metrics used in this study (see Table 5.1) to evaluate the performance of three transformer-based models. Collectively, these metrics offer a comprehensive evaluation framework to assess each model's ability to generalise across the dataset and accurately classify sentiments of the tweets of all timelines, merged into 1 dataset for fine-tuning purposes.

Metric	Description	Formula
Accuracy	Represents the proportion of correct predictions out of the total predictions made by the model.	(True positives + True negatives) / Total samples
Precision	Measures the proportion of true positive predictions out of all positive predictions made by the model and quantifies the accuracy of positive predictions.	(True positives) / True positives + False positives)
Recall	Also known as sensitivity, measures the model's ability to identify all relevant positive instances within the dataset.	(True positives) / True positives + False negatives)
F1-score	A harmonic mean of precision and recall, providing a single metric that balances both aspects.	2 * (Precision * Recall) / (Precision + Recall)

Table 5.1: Evaluation metrics used.

5.3.1 Why These Metrics Are Necessary for Sentiment Analysis

Accuracy may not effectively capture performance nuances if positive sentiments are overrepresented in the dataset, neglecting the minority classes [192]; therefore, other metrics complement it to ensure a balanced evaluation of model performance across all sentiment classes [193]. **Precision** ensures the trustworthiness of the model's predictions and that the positive classifications made by the model are reliable [194]. There might be some instances where missing positive sentiments could lead to biassed interpretations of public opinion [195]; this is where **Recall** plays a role by ensuring that the model identifies as many positive instances as possible. Similarly, the **F-1 Score** provides a balanced metric that accounts for both the completeness and correctness of positive predictions, a metric critical in sentiment analysis, where both false negatives and false positives can significantly impact the interpretation of public sentiment towards the UK's CBDC [196].

5.4 Results of Model Experiments

This section provides the outcomes of experimental evaluations of three models after they were fine-tuned over 30 runs for 3 and 5 epochs to ensure statistical reliability and mitigate the effects of random initialisation and data shuffling.

A comprehensive statistical analysis is conducted *(following established statistical methodologies)* to determine whether the observed differences in these metrics among the different models and epochs are statistically significant. To facilitate comparisons, the results were organised into tables and visualised through graphs.

5.4.1 Performance Comparison

The data recorded after running epochs 3 and 5 for 30 times each is organised into six groups based on the combination of model and epoch. The complete table is included in Appendix 2.

The average performance and standard deviation of metrics recorded over 30 runs for epoch 3 and epoch 5 are summarised in the Tables 5.2 and 5.3. It then follows a statistical analysis to validate the significance of observed differences.

Model	Epochs	Training Loss	Validation Loss	Accuracy (%)	Precision %	Recall (%)	F1-score (%)
DistilBERT	3	0.72	0.46	82.45	84.20	82.35	80.94
DistilBERT	5	0.26	0.07	98.85	98.91	98.85	98.84
RoBERTa	3	0.37	0.18	94.14	94.31	94.14	94.09
RoBERTa	5	0.16	0.05	98.42	98.51	98.42	98.40
XLM-RoBERTa	3	0.52	0.31	89.98	91.00	89.95	89.65
XLM-RoBERTa	5	0.19	0.11	96.28	96.46	96.28	96.33

Table 5.2: Average performance metrics over 30 runs.

Model	Epochs	Training Loss	Validation Loss	Accuracy (%)	Precision %	Recall (%)	F1-score (%)
DistilBERT	3	±0.02	±0.05	±3.04	±2.16	±2.93	±3.65
DistilBERT	5	±0.03	±0.02	1.16	±1.03	±1.16	±1.17
RoBERTa	3	±0.01	±0.03	±1.61	±1.49	±1.61	±1.66
RoBERTa	5	±0.02	±0.03	±1.59	±1.39	±1.59	±1.66
XLM-RoBERTa	3	±0.14	±0.12	±5.34	±4.36	±5.33	±5.76
XLM-RoBERTa	5	±0.02	±0.02	±1.49	±1.35	±1.49	±1.45

Table 5.3: Standard deviation of performance metrics over 30 runs.

The bar chart (Figure 5.1) below provides a detailed representation of the distribution of various performance metrics across 3 epochs (blue) and 5 epochs (orange) for all metrics. DistilBERT consistently shows higher training and validation losses compared to the other models, indicating that it struggles more to fit the data. Roberta demonstrates a balanced performance with relatively low losses and high accuracy, precision, recall, and F1 scores across both epoch counts, reinforcing its stability and efficiency. XLM-RoBERTa shows improvement at 5 epochs but still exhibits higher losses than RoBERTa, indicating it may need more training or fine-tuning for optimal performance. Overall, RoBERTa outperforms the other models, especially in precision and F1-score, suggesting it's the most effective model for the task under consideration.

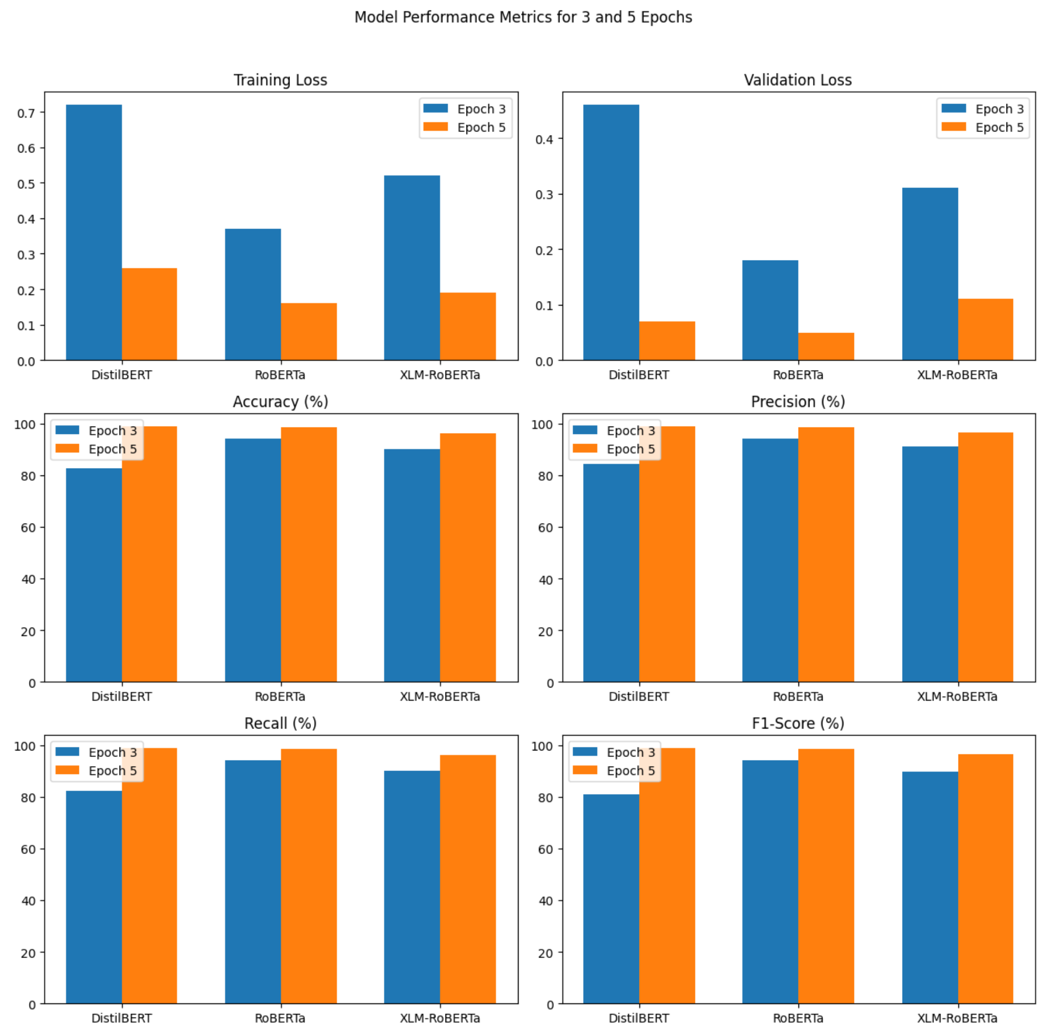


Figure 5.1: Data distribution of all metrics across epochs.

5.4.2 Variability Across Epochs

The results presented in Tables 5.2 and 5.3 show notable variability across different epoch settings. This variability is inherent to deep learning models based on their training process and arises from several factors including random initialisation and stochastic optimisation. As deep learning models initialise their weights and biases randomly at the start of training [197], this can lead to variations in performance metrics across multiple runs due to different convergence paths. Furthermore, Stochastic Gradient Descent (SGD), a type of stochastic optimisation algorithm, introduces randomness via mini-batch sampling and data shuffling [198]. This shuffling can lead to performance fluctuations due to different local minima being reached during training [188].

Also, training data is not processed in the same order by each epoch, which can influence gradient updates and, consequently, the model's learning trajectory and final performance. The number of training epochs significantly affects model convergence, directly impacting metrics such as accuracy, precision, recall, and F1-score, as observed in the performance between 3 and 5 epochs. Furthermore, observed changes in performance metrics could be caused by the non-deterministic behaviour of parallel hardware, such as GPUs, due to the asynchronous nature of computations and floating-point precision limitations [199]. Therefore, it is evident that ensuring reproducibility in deep learning experiments is challenging due to the aforementioned sources of variability. In this study, the effects of random fluctuations were mitigated by conducting 30 runs for each model and epoch combination, which allows for the calculation of mean and standard deviation values, providing a statistical understanding of model performance. This approach is consistent with established practices in experimental machine learning, which emphasise the importance of multiple trials to account for inherent variability and provide statistically sound results.

The observations, as discussed above, will be validated statistically in the next section.

5.5 Statistical Analysis Methodology

This study followed step-by-step statistical procedures for each performance metric (Accuracy, Precision, Recall, F1-Score, Training Loss, Validation Loss), including formulation of hypotheses, identification of dependent and independent variables, assumption checks for normality (Shapiro-Wilk Test) and homogeneity of variances (Levene's Test), selection of appropriate statistical test, such as parametric test like One-Way ANOVA (if assumptions are met) or non-parametric test like Kruskal-Wallis H Test (if assumptions are violated). After selecting appropriate statistical tests, tests were conducted on all metrics, followed by post-hoc analysis, and interpretation of results.

This analysis aims to determine whether there are statistically significant differences in performance between the models and epoch settings. *Specifically, this analysis addresses RQ1 by providing empirical evidence for selecting the optimal model for subsequent analysis. It also contributes to RQ2 by quantifying the performance differences between models under varying different epoch settings.*

5.5.1 Dependent and Independent Variables

- **Dependent variables:** Performance metrics, including Accuracy, Precision, Recall, F1-Score, Training Loss, Validation Loss.
- **Independent variables:** Model type: DistilBERT, RoBERTa, XLM-RoBERTa (categorical); Number of epochs: 3, 5 (categorical).

5.5.2 Hypothesis Formulation

For each performance metric, the following hypotheses were formulated:

- **Null hypothesis (H_0)** = There is no significant difference in the mean [Performance metric] among different models and epochs.
- **Alternative hypothesis (H_1)** = At least one group has a significantly different mean [Performance metric].

Statistical tests were selected based on criteria, as outlined in Section 5.6.

5.5.3 Assumption checks

Before choosing appropriate statistical tests, we verified below two key assumptions for each *model-epoch combination*:

- **Normality**: The data should be approximately normally distributed.
- **Homogeneity of variances**: The variances across groups should be similar.

Model-epoch combination resulted in six distinct groups: DistilBERT trained for 3 epochs (DistilBERT_3), DistilBERT trained for 5 epochs (DistilBERT_5), RoBERTa trained for 3 epochs (RoBERTa_3), RoBERTa trained for 5 epochs (RoBERTa_5), XLM-RoBERTa trained for 3 epochs (XLM-RoBERTa_3), and XLM-RoBERTa trained for 5 epochs (XLM-RoBERTa_5). *This grouping is crucial because it allows examination of the interaction effects between model type and training duration. By creating distinct groups for each combination, it can be determined whether the effect of increasing epochs is consistent across all models or if some models benefit or lose more than others.*

5.5.3.1 Normality Test: Shapiro-Wilk Test

The Shapiro-Wilk test is used to assess whether a dataset follows a normal distribution; it's particularly effective for small to moderate sample sizes. Given 778 tweets were used in the fine-tuning process, this test was applied. *This test is particularly appropriate for this sample size (n=30 for each group, representing 30 runs) and is recommended for assessing normality when the sample size is less than 50.* The test's null hypothesis (Ho) is that the data is normally distributed.

Firstly, data was sorted in ascending order. Using SciPy *(scipy.stats.shapiro)*, the weights were automatically calculated based on the expected values of a normal distribution. The Shapiro-Wilk statistic (W) was then computed, which was then compared to the critical values from the normal distribution, and the corresponding p-value was generated. If the p-value was below the significance threshold (commonly 0.05), the null hypothesis (H_0) was rejected, indicating the data did not follow a normal distribution.

5.5.3.2 Homogeneity of Variances: Levene's Test

Levene's Test helps check if the variances are equal across the six model-epoch groups (homoscedasticity). Given that our normality assumption was not met, Levene's test is a reliable alternative for Bartlett's test since it is less sensitive to deviations from normality. The test's null hypothesis (H_0) is that the variances are equal across the groups.

The test was employed using SciPy (scipy.stats.levene) to test for the equality of variances across groups. First, for each group, the median was calculated, then the absolute deviations of each observation from its group's median were computed and used to perform Levene's test. The test produced a Levene's test statistic (T) and a p-value. If the p-value was less than 0.05, the null hypothesis was rejected, indicating that variances across groups were significantly different (i.e., the assumption of equal variances was violated).

5.6 Statistical Tests and Results

This section details the results for Accuracy metric, and the process used for their calculation. The same process is applied to all other metrics.

5.6.1 Analysis of Accuracy

This section follows the methodology as mentioned in Section 5.6.

5.6.1.1 Hypothesis Formulation

- **H_0:** There is no significant difference in the mean Accuracy among DistilBERT, RoBERTa, and XLM-RoBERTa across 3 and 5 epochs.
- **H_1:** At least one group has a significantly different mean Accuracy.

5.6.1.2 Assumption Checks

a) Shaprio-Wilk test for Normality

The Shapiro-Wilk test was conducted to assess the normality of the accuracy distribution for each group (Model × Epoch combination). The results of the Shapiro-Wilk for Accuracy are presented in Table 5.4.

Group	W Statistic	p-value	Normality Conclusion
DistilBERT_3	0.937	0.0738	Normally distributed
DistilBERT_5	0.813	0.0001	Not normally distributed
RoBERTa_3	0.950	0.1701	Normally distributed
RoBERTa_5	0.742	0.0000	Not normally distributed
XLM-RoBERTa_3	0.873	0.0020	Not normally distributed
XLM-RoBERTa_5	0.963	0.3697	Normally distributed

Table 5.4: Shapiro-Wilk Test results for accuracy.

Interpretation: *The accuracy distributions of DistilBERT (3 epochs), RoBERTa (3 epochs), and XLM-RoBERTa (5 epochs) do not significantly deviate from normality, as they exhibited p-values greater than 0.05. On the other hand, accuracy distributions of DistilBERT (5 epochs), RoBERTa*

(5 epochs), and XLM-RoBERTa (3 epochs) deviate from normality as they showed p-values less than or equal to 0.05.

b) Levene's Test for Homogeneity of Variances

To evaluate the equality of variances across all groups, Levene's Test was performed and results are presented in Table 5.5.

Levene's T statistic	p-value
13.309	0.0000

Table 5.5: Levene's Test for Homogeneity of Variances.

Interpretation: *The null hypothesis (H_0) is rejected as the p-value is significantly less than 0.05. This implies that the variances among the groups violate the assumption of the homogeneity of variances as they are unequal.*

5.6.1.3 Selection of Statistical Tests and Results

Based on the above assumptions' checks, it is evident that the One-Way ANOVA (a parametric test) is **not appropriate** for this analysis. Therefore, a test that does not assume a normal distribution and is robust to unequal variances, the Kruskal-Wallis H Test (a non-parametric test), was selected [200].

Kruskal-Wallis H Test Results:

The H statistic calculation process involves combining and ranking all accuracy observations. Then, each group's average ranks are calculated, followed by the computation of the H statistic. The resulting H value is compared to a chi-squared distribution to determine the p-value. For this task, the SciPy library (scipy.stats.kruskal) was used. The results of this test are presented in Table 5.6.

Statistic	Value
H Statistic	147.539
Degrees of Freedom (df)	5
p-value	0.0000

Table 5.6: Kruskal-Wallis H Test Results for Accuracy.

Interpretation: *The six groups (three models × two epochs) exhibit a statistically significant difference in accuracy, as shown by the p-value less than 0.05, leading to the rejection of the null hypothesis (H_0). This suggests that both the type of model and the number of training epochs significantly influence the model's accuracy.*

Post-hoc analysis:

If a significant difference between the groups is observed with the Kruskal-Wallis's test results, Dunn's post-hoc test is applied to identify which specific groups differ, which adjusts for multiple comparisons using a method like Bonferroni correction while comparing each pair of groups. Given this aligns with the Kruskal-Wallis's test results of Accuracy metric, a post-hoc test called Dunn's Post-Hoc Test was applied with Bonferroni correction to identify which groups differ significantly. Dunn's post-hoc test was conducted using the 'scikit-posthocs' library (scikit_posthocs.posthoc_dunn). The calculation begins by calculating the mean ranks for each group, which are then used to compute Z-scores for each pair of groups to quantify differences between their ranks. Type 1 errors are controlled by applying a multiple comparison correction like Bonferroni to adjust the p-values. Finally, based on the adjusted p-values, the results are interpreted by identifying which groups show significant differences.

Dunn's Post-Hoc Test Results:

The direction and magnitude of the differences between groups are identified via the Z-statistics; positive Z-values mean that the first group ranks higher than the second one, while negative Z-values indicate the opposite. Adjusted p-values provide the significance of pairwise comparisons after controlling for multiple tests. Table 5.7 summarises the significant pairwise differences identified:

Comparison	Adj. p-value	Z-stat	Mean(A)	Mean(B)	Mean Diff (B – A)	% Diff	Interpretation
DistilBERT_3 vs DistilBERT_5	0.0000	-9.7025	82.45	98.85	16.4	19.90%	DistilBERT_5 outperforms DistilBERT_3 by 19.9%.
DistilBERT_3 vs RoBERTa_3	0.0006	-4.0931	82.45	94.14	11.69	14.20%	RoBERTa_3 outperforms DistilBERT_3 by 14.2%.
DistilBERT_3 vs RoBERTa_5	0.0000	-9.1376	82.45	98.42	15.97	19.40%	RoBERTa_5 outperforms DistilBERT_3 by 19.4%.
DistilBERT_3 vs XLM-R_5	0.0000	-6.3936	82.45	96.28	13.83	16.80%	XLM-RoBERTa_5 outperforms DistilBERT_3 by 16.8%.
DistilBERT_5 vs RoBERTa_3	0.0000	5.6094	98.85	94.14	-4.71	-4.80%	DistilBERT_5 is 4.8% higher in accuracy than

							RoBERTa_3.
DistilBERT_5 vs XLM-R_3	0.0000	7.3499	98.85	89.98	-8.87	-9.00%	DistilBERT_5 is 9.0% higher in accuracy than XLM-RoBERTa_3.
DistilBERT_5 vs XLM-R_5	0.0138	3.3089	98.85	96.28	-2.57	-2.60%	DistilBERT_5 is 2.6% higher in accuracy than XLM-RoBERTa_5.
RoBERTa_3 vs RoBERTa_5	0.0000	-5.0445	94.14	98.42	4.28	4.50%	RoBERTa_5 outperforms RoBERTa_3 by 4.5%.
RoBERTa_5 vs XLM-R_3	0.0000	6.785	98.42	89.98	-8.44	-8.60%	RoBERTa_5 is 8.6% higher in accuracy than XLM-RoBERTa_3.
XLM-R_3 vs XLM-R_5	0.0008	-4.041	89.98	96.28	6.3	7.00%	XLM-RoBERTa_5 outperforms XLM-RoBERTa_3 by 7.0%.

Table 5.7: Significant Pairwise Comparisons from Dunn's Post-Hoc Test for Accuracy.

Interpretation: *As seen in Table 5.7, significant differences in accuracy across all group comparisons can be observed. The findings indicate that accuracy for DistilBERT, RoBERTa, and XLM-RoBERTa models is consistently enhanced if training epochs are increased from 3 to 5, as indicated by negative Z-statistics when comparing 3 epochs to 5 epochs. With extended training, DistilBERT_5 significantly outperformed RoBERTa_3 and XLM-RoBERTa_3; RoBERTa_5 and XLM-RoBERTa_5 significantly improved over their 3-epoch counterparts. These results underscore the importance of adequate training duration for optimising model performance in the context of UK CBDC tweet sentiment analysis. The results suggest that increasing training epochs from 3 to 5 leads to statistically significant gains in accuracy across all tested models. Furthermore, the comparisons between different models with the same number of epochs (e.g., DistilBERT_5 vs. RoBERTa_3) provide insights into the relative performance of different architectures. While the table shows several statistically significant differences, the magnitude of the improvement varies. Model selection and training duration are both critical factors in achieving optimal performance.*

5.6.2 Analysis of Precision, Recall and F1-Score

The statistical analysis of Precision, Recall, and F1-Score follows the same process as outlined for Accuracy in Section 5.7.1, including assumption checks and non-parametric testing due to violations of normality and homogeneity of variances. The results are presented in the Table 5.8.

Metric	Group	W Statistic	p-value	Normality Conclusion
Precision	DistilBERT_3	0.977	0.7367	Normally distributed
	DistilBERT_5	0.846	0.0005	Not normally distributed
	RoBERTa_3	0.965	0.4112	Normally distributed
	RoBERTa_5	0.798	0.0001	Not normally distributed
	XLM-RoBERTa_3	0.882	0.0031	Not normally distributed
	XLM-RoBERTa_5	0.974	0.06415	Normally distributed
Recall	DistilBERT_3	0.946	0.1331	Normally distributed
	DistilBERT_5	0.813	0.0001	Not normally distributed
	RoBERTa_3	0.0950	0.1701	Normally distributed
	RoBERTa_5	0.742	0.0000	Not normally distributed
	XLM-RoBERTa_3	0.876	0.0023	Not normally distributed
	XLM-RoBERTa_5	0.961	0.3378	Normally distributed
F1-Score	DistilBERT_3	0.957	0.2659	Normally distributed
	DistilBERT_5	0.814	0.0001	Not normally distributed
	RoBERTa_3	0.953	0.2013	Normally distributed
	RoBERTa_5	0.721	0.0000	Not normally distributed

	XLM-RoBERTa_3	0.864	0.0012	Not normally distributed
	XLM-RoBERTa_5	0.959	0.2900	Normally distributed

Table 5.8: Shapiro-Wilk Test Results for Precision, Recall, and F1-Score.

Interpretation: *From the above analysis, it can be observed that half of the groups for each metric across Precision, Recall, and F1-Score do not follow a normal distribution (see results in Table 5.9). For instance, for Precision, only DistilBERT_3, RoBERTa_3, and XLM-RoBERTa_5 groups followed the normal distribution. That's why the Kruskal-Wallis H Test was employed for further analysis.*

Metric	**Levene's Test Statistic**	**p-value**	**Homogeneity Conclusion**
Precision	10.041	0.0000	Variances are not equal
Recall	13.487	0.0000	Variances are not equal
F1-score	13.814	0.0000	Variances are not equal

Table 5.9: Levene's Test for Homogeneity of Variances for Precision, Recall, and F1-Score.

Interpretation: *Levene's Test results (Table 5.9) conclusively reveal that Precision, Recall, and F1-Score variances are significantly unequal across different models and training epochs, necessitating non-parametric tests. Also, p-values are below the common alpha level of 0.05, rejecting the null hypothesis of equal variances.*

Kruskal-Wallis H Test Results:

Metric	**H Statistic**	**Degrees of Freedom (df)**	**p-value**	**Significant?**
Precision	150.156	5	0.0000	Yes
Recall	147.605	5	0.0000	Yes
F1-score	147.723	5	0.0000	Yes

Table 5.10: Kruskal-Wallis H Test Results for Precision, Recall, and F1-Score.

Interpretation: *For all three metrics, the Kruskal-Wallis H Test results (Table 5.10) yielded highly significant results (p-value = 0.0000), indicating that at least one model-epoch differs significantly from the others. Thus, this confirms that variations in model performance are not due to random chance but are influenced by the model type and the number of training epochs, necessitating further post-hoc analyses to identify specific group differences.*

Dunn's Post-Hoc Test Results:

The Dunn's post-hoc test results (Table 5.11) revealed several statistically significant pairwise differences between the model-epoch combinations for Precision, Recall, and F1-Score.

Metric	Comparison	Adj. p-value	Z-Stat.	Mean (A)	Mean (B)	Mean Diff (B – A)	% Diff	Interpretation
	DistilBERT_3 vs DistilBERT_5	0	-9.8387	84.2	98.91	14.71	17.50%	DistilBERT_5 outperforms DistilBERT_3 by 17.5%
	DistilBERT_3 vs RoBERTa_3	0.0008	-4.0435	84.2	94.31	10.11	12.00%	RoBERTa_3 outperforms DistilBERT_3 by 12.0%
	DistilBERT_3 vs RoBERTa_5	0	-9.2515	84.2	98.51	14.31	17.00%	RoBERTa_5 outperforms DistilBERT_3 by 17.0%
	DistilBERT_3 vs XLM-R_5	0	-6.572	84.2	96.46	12.26	14.60%	XLM-RoBERTa_5 outperforms DistilBERT_3 by 14.6%
	DistilBERT_5 vs RoBERTa_3	0	5.7952	98.91	94.31	-4.6	-4.70%	DistilBERT_5 is 4.7% higher in Precision than RoBERTa_3
	DistilBERT_5 vs XLM-R_3	0	7.345	98.91	91	-7.91	-8.00%	DistilBERT_5 is 8.0% higher in Precision than XLM-RoBERTa_3
	DistilBERT_5 vs XLM-R_5	0.0162	3.1652	98.91	96.46	-2.45	-2.50%	DistilBERT_5 is ~2.5% higher in Precision than XLM-RoBERTa_5
	RoBERTa_3 vs RoBERTa_5	0	-5.208	94.31	98.51	4.2	4.50%	RoBERTa_5 outperforms RoBERTa_3 by 4.5%
Precision	RoBERTa_5 vs XLM-R_3	0	6.7503	98.51	91	-7.51	-7.60%	RoBERTa_5 is 7.6% higher than XLM-RoBERTa_3

	XLM-R_3 vs XLM-R_5	0.0009	-4.0881	91	96.46	5.46	6.00%	XLM-RoBERTa_5 outperforms XLM-RoBERTa_3 by 6.0%
Recall	DistilBERT_3 vs DistilBERT_5	0	-9.7111	82.35	98.85	16.5	20.00%	DistilBERT_5 outperforms DistilBERT_3 by 20.0%
	DistilBERT_3 vs RoBERTa_3	0.0006	-4.1067	82.35	94.14	11.79	14.30%	RoBERTa_3 outperforms DistilBERT_3 by 14.3%
	DistilBERT_3 vs RoBERTa_5	0	-9.1822	82.35	98.42	16.07	19.50%	RoBERTa_5 outperforms DistilBERT_3 by 19.5%
	DistilBERT_3 vs XLM-R_5	0	-6.5199	82.35	96.28	13.93	16.90%	XLM-RoBERTa_5 outperforms DistilBERT_3 by 16.9%
	DistilBERT_5 vs RoBERTa_3	0	5.6763	98.85	94.14	-4.71	-4.80%	DistilBERT_5 has 4.8% higher Recall than RoBERTa_3
	DistilBERT_5 vs XLM-R_3	0	7.2533	98.85	89.95	-8.9	-9.00%	DistilBERT_5 exceeds XLM-RoBERTa_3 by 9.0%
	DistilBERT_5 vs XLM-R_5	0.0134	3.3163	98.85	96.28	-2.57	-2.60%	DistilBERT_5 is 2.6% higher in Recall than XLM-RoBERTa_5
	RoBERTa_3 vs RoBERTa_5	0	-5.1733	94.14	98.42	4.28	4.60%	RoBERTa_5 outperforms RoBERTa_3 by 4.6%
	RoBERTa_5 vs XLM-R_3	0	6.7788	98.42	89.95	-8.47	-8.60%	RoBERTa_5 is 8.6% better than XLM-RoBERTa_3
	XLM-R_3 vs XLM-R_5	0.0008	-4.0881	89.95	96.28	6.33	7.00%	XLM-RoBERTa_5 outperforms XLM-RoBERTa_3 by 7.0%
F1-Score	DistilBERT_3 vs DistilBERT_5	0	-9.6851	80.94	98.84	17.9	22.10%	DistilBERT_5 outperforms DistilBERT_3 by 22.1%

	DistilBERT_3 vs RoBERTa_3	0.0009	-4.0088	80.94	94.09	13.15	16.20%	RoBERTa_3 outperforms DistilBERT_3 by 16.2%
	DistilBERT_3 vs RoBERTa_5	0	-9.1822	80.94	98.4	17.46	21.60%	RoBERTa_5 outperforms DistilBERT_3 by 21.6%
	DistilBERT_3 vs XLM-R_5	0	-6.5199	80.94	96.33	15.39	19.00%	XLM-RoBERTa_5 outperforms DistilBERT_3 by 19.0%
	DistilBERT_5 vs RoBERTa_3	0	5.6763	98.84	94.09	-4.75	-4.80%	DistilBERT_5 is 4.8% higher in F1-score than RoBERTa_3
	DistilBERT_5 vs XLM-R_3	0	7.2533	98.84	89.65	-9.19	-9.30%	DistilBERT_5 is 9.3% higher in F1-score than XLM-RoBERTa_3
	DistilBERT_5 vs XLM-R_5	0.0231	3.1652	98.84	96.33	-2.51	-2.50%	DistilBERT_5 outperforms XLM-RoBERTa_5 by ~2.5%
	RoBERTa_3 vs RoBERTa_5	0	-5.1733	94.09	98.4	4.31	4.60%	RoBERTa_5 outperforms RoBERTa_3 by 4.6%
	RoBERTa_5 vs XLM-R_3	0	6.7503	98.4	89.65	-8.75	-8.90%	RoBERTa_5 is 8.9% higher than XLM-RoBERTa_3
	XLM-R_3 vs XLM-R_5	0.0006	-4.0881	89.65	96.33	6.68	7.40%	XLM-RoBERTa_5 outperforms XLM-RoBERTa_3 by 7.4%

Table 5.11: Significant Pairwise Comparisons from Dunn's Post-Hoc Test for Precision, Recall, and F1-Score.

Interpretation: *The statistical analyses conducted, employing Kruskal-Wallis H tests (Table 5.10) and Dunn's post-hoc tests (Table 5.11), provide strong evidence that both model selection and training duration exert a statistically significant influence on sentiment analysis performance on UK CBDC tweets. These analyses confirm that increasing training epochs from 3 to 5 leads to significant performance gains, as evidenced by the negative Z-statistics observed in comparisons between these training durations. For instance, DistilBERT_5 consistently outperforms other groups with extended training from 3 to 5 epochs. This improvement is also visualised in Figure 5.2, which shows that increasing epochs significantly improves performance, with the yellow box plots (representing 5 epochs) consistently yielding higher and more stable results across all metrics compared to 3 epochs (green box plots). Similarly, RoBERTa_5 and XLM-RoBERTa_5* *demonstrated superior performance over their 3-epoch versions. Positive Z-statistics, particularly in comparisons involving 5-epoch variants or different model architectures, further support the conclusion that both training duration and model selection influence performance. These patterns are consistent across all three metrics, highlighting the critical impact of training duration and model architecture on performance. However, while RoBERTa_5 exhibits the highest performance metrics at this stage of the analysis, a crucial aspect of model evaluation remains: the assessment of potential overfitting. Therefore, before definitively concluding on the optimal model (RQ1) and fully addressing RQ2, this study conducts an analysis of overfitting.*

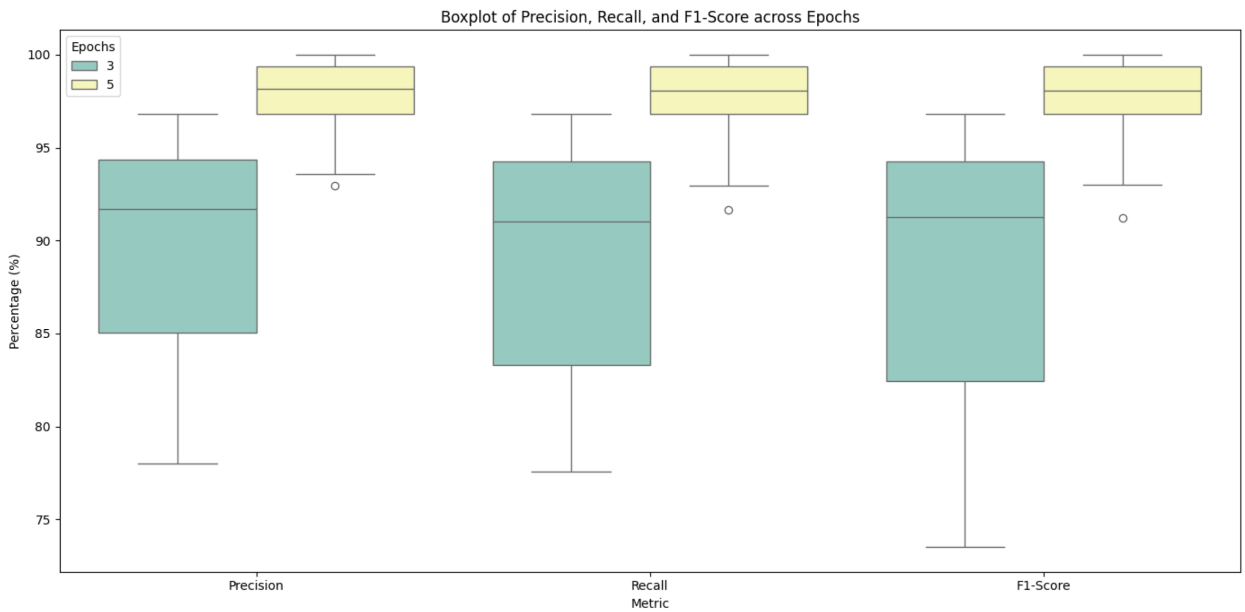


Figure 5.2: Precision, Recall and F1-score across Epochs.

5.6.3 Analysis of Training Loss and Validation Loss

The analysis presented in this section follows the same methodology as outlined in Sections 5.7.1 and 5.7.2.

Metric	Group	W Statistic	p-value	Normality Conclusion
Training Loss	DistilBERT_3	0.928	0.0432	Not normally distributed
	DistilBERT_5	0.978	0.7821	Normally distributed

	RoBERTa_3	0.978	0.7739	Normally distributed
	RoBERTa_5	0.975	0.6808	Normally Distributed
	XLM-RoBERTa_3	0.77	0.0000	Not normally distributed
	XLM-RoBERTa_5	0.92	0.0270	Not normally distributed
Validation Loss	DistilBERT_3	0.946	0.1303	Normally distributed
	DistilBERT_5	0.655	0.0000	Not normally distributed
	RoBERTa_3	0.947	0.1366	Normally distributed
	RoBERTa_5	0.875	0.0022	Not normally distributed
	XLM-RoBERTa_3	0.865	0.0013	Not normally Distributed
	XLM-RoBERTa_5	0.977	0.7338	Normally distributed

Table 5.12: Shapiro-Wilk Test Results for Training Loss and Validation Loss.

Metric	**Levene's Test Statistic**	**p-value**	**Homogeneity Conclusion**

Training Loss	18.933	0.0000	Variances are not equal
Validation Loss	23.245	0.0000	Variances are not equal

Table 5.13: Levene's Test for Homogeneity of Variances for Training Loss and Validation Loss.

Metric	**H Statistic**	**Degrees of Freedom (df)**	**p-value**	**Significant**
Training Loss	167.847	5.0000	0	Yes
Validation Loss	157.195	5.0000	0	Yes

Table 5.14: Kruskal-Wallis H Test Results for Training Loss and Validation Loss.

Metric	**Comparison**	**Adj. p-value**	**Z-Stat**	**Mean(A)**	**Mean(B)**	**Mean Diff (B – A)**	**% Diff**	**Interpretation (Lower = Better)**
	DistilBERT_3 vs DistilBERT_5	0	6.4295	0.72	0.26	0.26 - 0.72 = -0.46	-63.90 %	DistilBERT_3's loss is significantly higher ⇒ DistilBERT_5 is better.
	DistilBERT_3 vs RoBERTa_3	0.0006	4.103	0.72	0.37	0.37 - 0.72 = -0.35	-48.60 %	DistilBERT_3 has higher loss ⇒ RoBERTa_3 is better.
Training Loss	DistilBERT_3 vs RoBERTa_5	0	10.4569	0.72	0.16	0.16 - 0.72 = -0.56	-77.80 %	DistilBERT_3 has higher loss ⇒ RoBERTa_5 is better.

	DistilBERT_3 vs XLM-RoBERTa_5	0	8.9208	0.72	0.19	0.19 - 0.72 = -0.53	-73.60 %	DistilBERT_3 has higher loss ⇒ XLM-R_5 is better.
	DistilBERT_5 vs RoBERTa_5	0.0008	4.0274	0.26	0.16	0.16 - 0.26 = -0.10	-38.50 %	DistilBERT_5 has higher loss ⇒ RoBERTa_5 is better.
	DistilBERT_5 vs XLM-RoBERTa_3	0	-4.7942	0.26	0.52	0.52 - 0.26 = +0.26	1	XLM-R_3’s loss is +100% higher ⇒ DistilBERT_5 is better.
	RoBERTa_3 vs RoBERTa_5	0	6.3539	0.37	0.16	0.16 - 0.37 = -0.21	-56.80 %	RoBERTa_3 has higher loss ⇒ RoBERTa_5 is better.
	RoBERTa_3 vs XLM-RoBERTa_5	0	4.8178	0.37	0.19	0.19 - 0.37 = -0.18	-48.60 %	RoBERTa_3 has higher loss ⇒ XLM-R_5 is better.
	RoBERTa_5 vs XLM-RoBERTa_3	0	-8.8217	0.16	0.52	0.52 - 0.16 = +0.36	2.25	XLM-R_3’s loss is much higher ⇒ RoBERTa_5 is better.
	XLM-R_3 vs XLM-R_5	0	5.0544	0.52	0.19	0.19 - 0.52 = -0.33	-63.50 %	XLM-R_3 has higher loss ⇒ XLM-R_5 is better.
	DistilBERT_3 vs DistilBERT_5	0	8.6235	0.46	0.07	0.07 - 0.46 = -0.39	-84.80 %	DistilBERT_3 has higher val. loss ⇒ DistilBERT_5 is better.
Validation Loss	DistilBERT_3 vs RoBERTa_3	0.0013	3.9209	0.46	0.18	0.18 - 0.46 = -0.28	-60.90 %	DistilBERT_3 has higher val. loss ⇒

								RoBERTa_3 is better.
	DistilBERT_3 vs RoBERTa_5	0	10.0667	0.46	0.05	0.05 - 0.46 = - 0.41	-89.10 %	DistilBERT_3 has higher val. loss ⇒ RoBERTa_5 is better.
	DistilBERT_3 vs XLM-RoBERTa_5	0	6.7937	0.46	0.11	0.11 - 0.46 = - 0.35	-76.10 %	DistilBERT_3 has higher val. loss ⇒ XLM-R_5 is better.
	DistilBERT_5 vs RoBERTa_3	0	-4.7026	0.07	0.18	0.18 - 0.07 = +0.11	1.571	RoBERTa_3's val. loss is higher ⇒ DistilBERT_5 is better.
	DistilBERT_5 vs XLM-RoBERTa_3	0	-6.8841	0.07	0.31	0.31 - 0.07 = +0.24	3.429	XLM-R_3's val. loss is much higher ⇒ DistilBERT_5 is better.
	RoBERTa_3 vs RoBERTa_5	0	6.1458	0.18	0.05	0.05 - 0.18 = - 0.13	-72.20 %	RoBERTa_3 has higher val. loss ⇒ RoBERTa_5 is better.
	RoBERTa_5 vs XLM-RoBERTa_3	0	-8.3274	0.05	0.31	0.31 - 0.05 = +0.26	5.2	XLM-R_3's val. loss is much higher ⇒ RoBERTa_5 is better.
	RoBERTa_5 vs XLM-RoBERTa_5	0.016	-3.273	0.05	0.11	0.11 - 0.05 = +0.06	1.2	XLM-R_5's val. loss is higher ⇒ RoBERTa_5 is better.
	XLM-R_3 vs XLM-R_5	0	5.0544	0.31	0.11	0.11 - 0.31 = - 0.20	-64.50 %	XLM-R_3 has higher val. loss ⇒ XLM-R_5 is better.

Table 5.15: Dunn's Post-Hoc Test for Training Loss and Validation Loss.

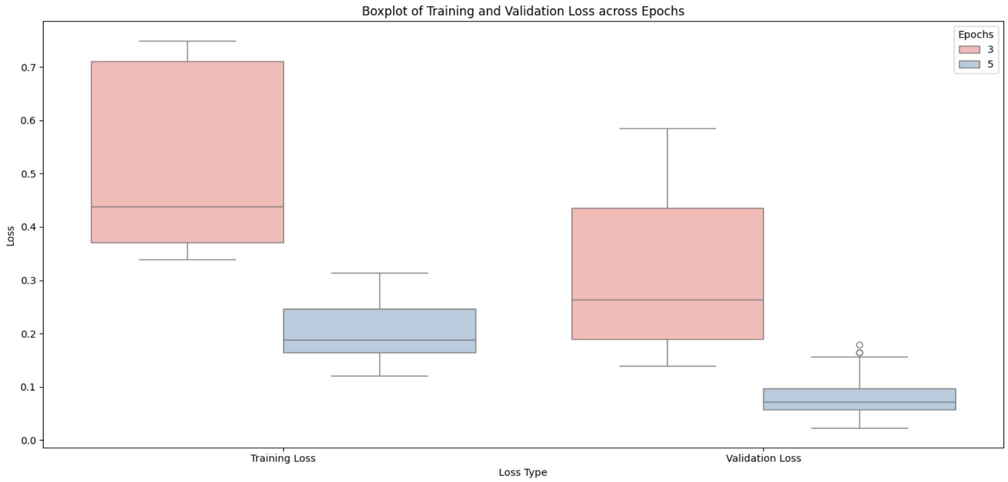


Figure 5.3: Training and Validation Loss across epochs.

Interpretation: *The statistical analysis of training and validation loss revealed that most groups violated assumptions of normality and homogeneity of variance. Therefore, the Kruskal-Wallis H test was used, revealing highly significant differences ($p < 0.0001$) among the six groups for both metrics. Dunn's post-hoc test confirmed that increasing training epochs from 3 to 5 significantly reduced both training and validation loss across all models. This is consistent with the box plots in Figure 5.3, which visually compare training and validation loss at 3 epochs (red) and 5 epochs (blue). The figure clearly shows that increasing the number of epochs significantly reduces both losses across all models. Furthermore, the losses are considerably lower and more tightly distributed at 5 epochs, indicating improved model performance and better generalisation, compared to the greater variability observed at 3 epochs. Specifically, DistilBERT_5 and XLM-RoBERTa_3 exhibited the largest decrease in loss, outperforming their 3-epoch counterparts and other models. RoBERTa_5 also demonstrated exceptional performance, achieving significantly lower losses compared to its 3-epoch counterpart and other models. These findings underscore the importance of extended training duration and appropriate model selection for minimising loss values.*

5.7 Analysis of Findings

Based on the above statistical analysis, distinct performance characteristics of each model have been observed for the CBDC sentiment prediction task. RoBERTa *(specifically, RoBERTa_5)* consistently outperformed both its lower-epoch counterpart and other models across multiple metrics; it exhibits substantial Z-scores and extremely low p-values (< 0.0001) in comparisons conducted using the Dunn's Post-Hoc Test, indicating that the improvements are not due to random chance but are statistically robust.

5.7.1 Performance of Each Model: Strengths vs. Weaknesses

Figure 5.4 shows that RoBERTa's accuracy at 5 epochs is consistently higher, with the median very close to 100%, and a tightly packed box plot, indicating both high accuracy and consistency. While

RoBERTa trained for 3 epochs (RoBERTa_3) also performed well (median accuracy around 94%), its accuracy exhibited greater variability compared to RoBERTa_5.

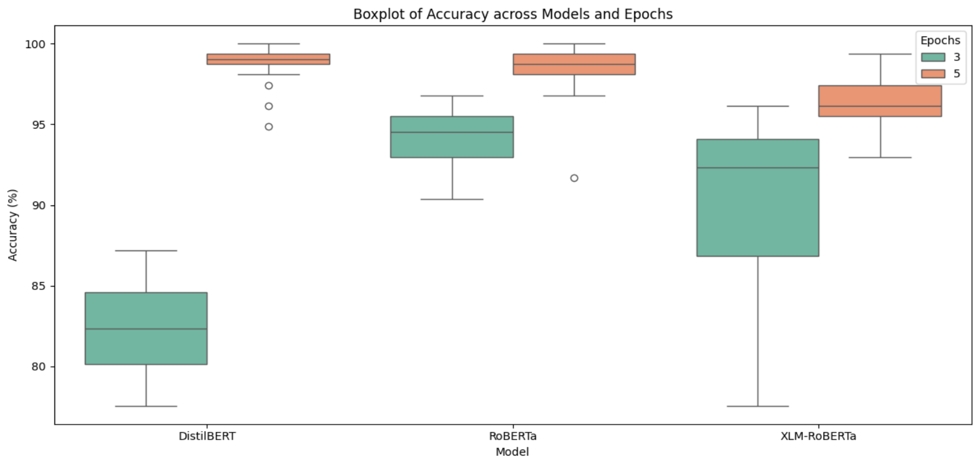


Figure 5.4: Models' accuracy across models and epochs.

In comparison, DistilBERT showed moderate performance, and XLM-RoBERTa underperformed despite increased epochs. However, increasing epochs improved DistilBERT's performance. Notably, the multilingual capabilities of XLM-RoBERTa did not confer advantages on English-only data, suggesting potentially unnecessary computational overhead. These findings suggest that RoBERTa balances performance and task suitability, emerging as the most effective model. While this initial analysis focuses on performance metrics, it lays the groundwork for a more in-depth exploration of RoBERTa's capabilities and limitations. Further investigation, including explainability and robustness testing (as detailed in Chapter 6), is necessary to fully address RQ2 and provide a comprehensive understanding of RoBERTa's strengths and weaknesses in the context of digital pound sentiment analysis.

5.8 Selection of RoBERTa_5 vs. RoBERTA_3

5.8.1 Justification Based on Experimental Results

RoBERTa was selected over DistilBERT and XLM-RoBERTa based on its superior performance across all evaluated metrics. Based on the initial statistical results (Tables 5.11-5.15) and the accuracy boxplots (Figure 5.4), RoBERTa_5 initially appeared to be the superior model. RoBERTa_5 achieved the highest Accuracy score; Dunn's Post-Hoc Test showed significant differences when compared to DistilBERT_3 ($Z = -9.1376$, $p = 0.0000$) and XLM-RoBERTa_3 ($Z = 6.7850$, $p = 0.0000$). Additionally, in Precision, RoBERTa_5 > RoBERTa_3 ($Z = -5.2080$, $p = 0.0000$), suggesting improved ability to correctly identify positive and negative sentiments. RoBERTa_5 also achieved the lowest training and validation losses among all models (Table 5.15), making it an ideal choice for this task, aligning with the findings of Liu et al. [115] regarding RoBERTa's effective learning through removing the next sentence prediction objective and dynamic masking.

5.8.2 Overfitting Analysis and Final Model Selection

Despite RoBERTa_5's superior performance metrics observed through the statistical analysis above, concerns arise regarding potential overfitting. Therefore, evaluating the results through established machine learning principles is essential before choosing between RoBERTA_5 and RoBERTA_3 because the objective is to use the ideal model to classify the sentiment in CBDCs-related tweets.

When a model learns from the underlying patterns in the training data along with the noise and outliers, overfitting occurs, meaning that it is poorly generalising to unseen data [171]. Significantly lower Training and Validation Loss and higher performance metrics in the case of RoBERTa_5 show that the model has become too tailored to the training data. The histogram of Validation Loss (Figure 5.5) reveals a wider distribution of loss values for RoBERTa_3, whereas RoBERTa_5's loss values are tightly clustered around lower values (0.02–0.05). This tight clustering suggests that even though the model has minimised loss on the validation set, it raises concerns about the model's robustness to data variations, a characteristic often associated with overfitting, as highlighted by Hawkins [201].

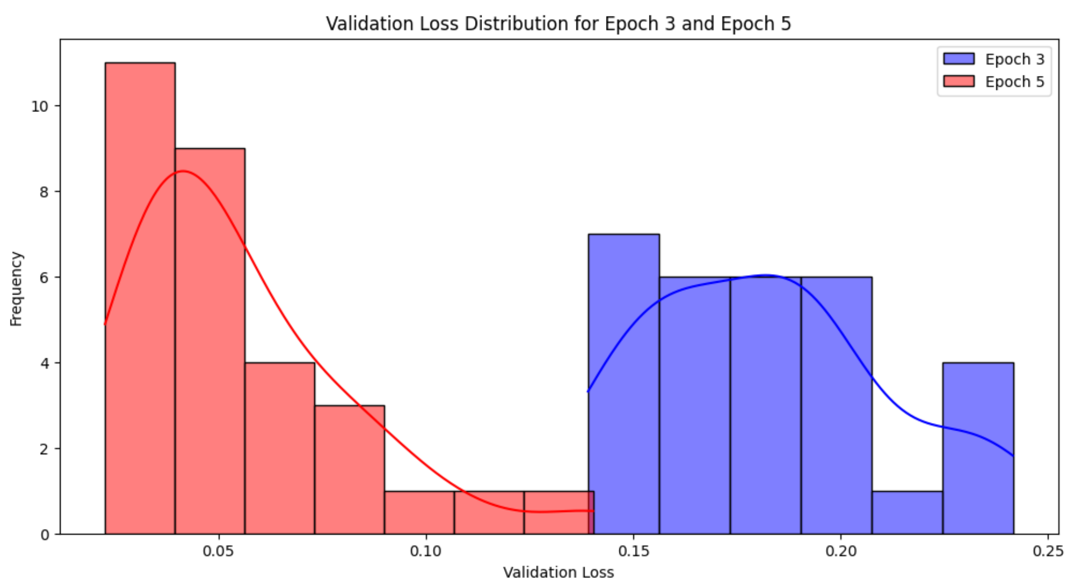


Figure 5.5: The histogram of Validation Loss for RoBERTa_3 and RoBERTa_5.

To investigate this, the correlation between Validation Loss and key performance metrics (Accuracy and F1-Score) was analysed for each epoch. For epoch 3, the correlation between Validation Loss and Accuracy (Figure 5.6) and between Validation Loss and F1-Score (Figure 5.7) is relatively moderate, implying that both Accuracy and F1-Score gradually decline with an increase in Validation Loss. Furthermore, it indicates robustness and lower risk of overfitting as the model's performance degrades steadily with increasing loss [168]. According to Sain and Vapnik [202], such a model is more likely to perform well in real-world scenarios; this aligns with the principles of developing/utilising models that generalise well beyond the training dataset.

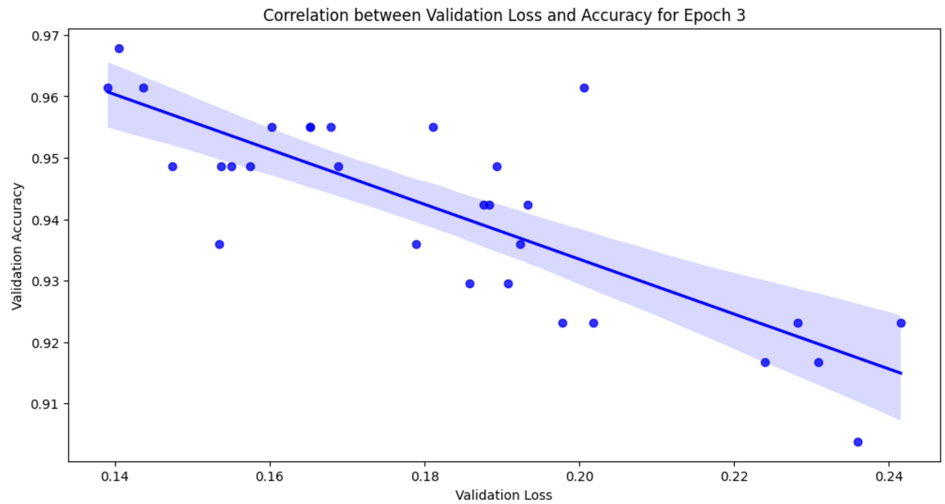


Figure 5.6: The correlation between Validation Loss and Accuracy for epoch 3.

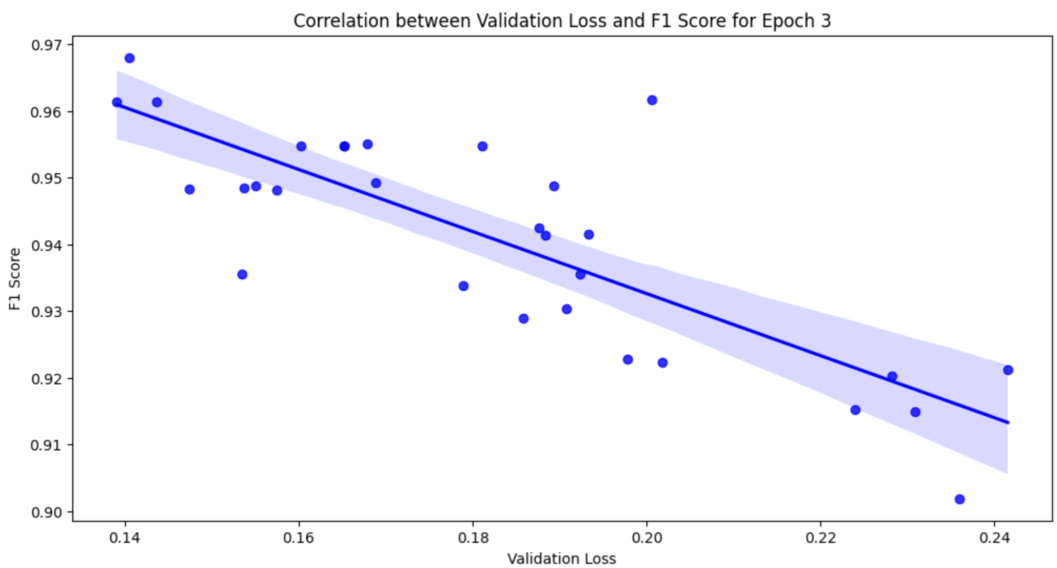


Figure 5.7: The correlation between Validation Loss and F1-score for epoch 3.

In contrast, Figures 5.8 and 5.9 reveal a sharp correlation, showing a significant decline in Accuracy and F1-Score with small increases in Validation Loss. This further suggests that the model is highly sensitive to changes in Validation Loss, which is one of the critical characteristics of overfitting [201]. This sensitivity implies that RoBERTA_5's performance could deteriorate rapidly when presented with slightly different data, as a consequence of learning the training data, including its noise, too effectively.

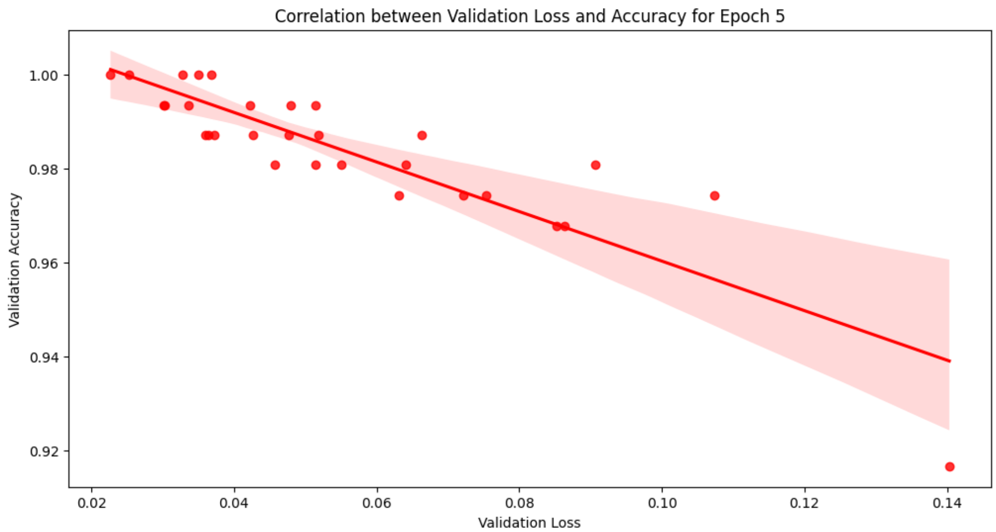


Figure 5.8: The correlation between Validation Loss and Accuracy for epoch 5.

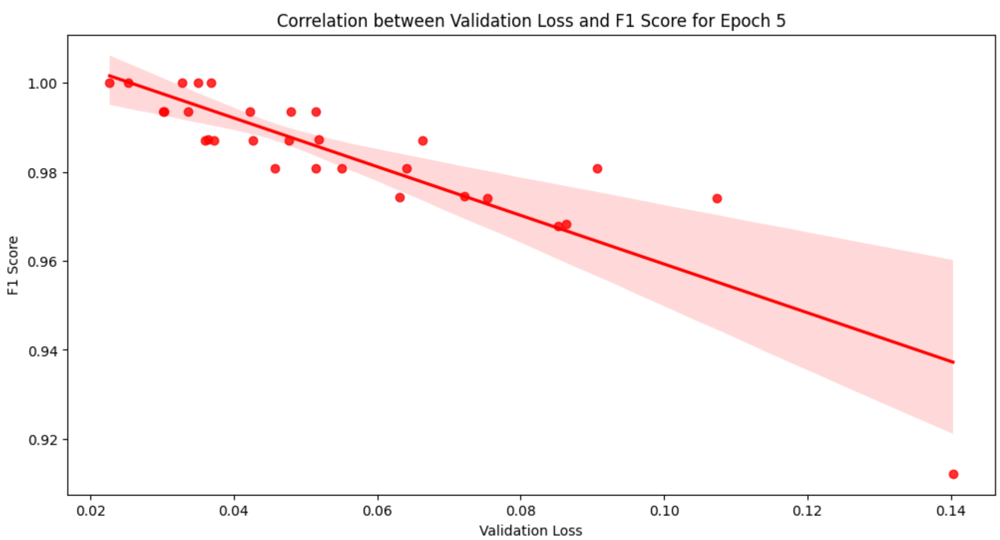


Figure 5.9: The correlation between Validation Loss and F-1score for epoch 5.

This analysis, considered in the light of the bias-variance trade-off principle [203], suggests that while RoBERTa_5 exhibits lower bias (due to increased training), it suffers from higher variance and overfitting. Conversely, RoBERTa_3, with its higher bias but lower variance, demonstrates more stable performance and better generalisation, likely due to a form of regularisation through early stopping [204].

In the context of sentiment analysis for UK CBDC tweets, minimising errors on unseen, real-world data is of paramount importance. An overfit model like RoBERTa_5 is more susceptible to making errors when presented with new, slightly different data. In the context of CBDC discussions, misclassifying public sentiment could have implications for understanding public perception and potentially influencing policy decisions. For example, failing to identify a surge in negative sentiment could lead to missed opportunities for addressing public concerns. Therefore, prioritising generalisation ability, as demonstrated by RoBERTa_3, is crucial for developing a reliable and robust sentiment analysis tool. *Consequently, RoBERTa_3 is selected as the suitable model for classifying CBDC tweets, directly addressing RQ1. Furthermore, this analysis contributes significantly to RQ2 (Model Capabilities and Limitations) by highlighting the trade-off between maximising performance on training data and achieving strong generalisation ability. It demonstrates the limitations of simply selecting a model based on training set performance and emphasises the importance of considering overfitting and its potential consequences in real-world applications.*

5.9 Impact of Hyperparameter Tuning and Computational Cost

The transformer models were fine-tuned with hyperparameter values that are consistently recommended in the text-classification literature (Section 5.2.3). These settings strike a pragmatic balance between convergence stability and computational efficiency, particularly when working with modestly sized corpora [116]. These settings usually bring BERT-type encoders close to their performance plateau while avoiding unstable loss spikes that arise with larger learning rates or longer schedules on limited data [113], [183].

To evaluate the effect of training duration, the models were fine-tuned for three and five epochs, configurations were selected based on common transformer practices. Performance improved notably between the two settings: DistilBERT's F_1 score rose from 80.94% to 98.84%, RoBERTa from 94.09% to 98.40%, and XLM-RoBERTa from 89.65% to 96.33% (Table 5.2). The stability of these improvements across 30 repeated runs (Table 5.3) confirms the reliability of the results and validates the chosen epoch range within the study's computational constraints.

A comprehensive grid or random search was not pursued because each additional hyperparameter configuration would have incurred a full cycle of gradient updates across 622 training tweets and 30 experimental repetitions, a workload that was infeasible given the available GPU time. Consequently, the reported F_1 scores (Table 5.2) reflect performance under a validated yet deliberately narrow hyperparameter corridor and should be interpreted accordingly.

Previous large-scale tuning studies have shown that deviations from these standard configurations seldom raise F_1 by more than one or two percentage points [185]. Given these sharply diminishing returns, the computational burden could not be justified. While an unconstrained search might unearth marginally stronger configurations, the current settings achieve a pragmatic compromise: it aligns with best practices, ensures methodological comparability with prior transformer-based studies, and maintains responsible resource usage, particularly in settings with limited data and compute.

5.10 Conclusion

This chapter makes several key contributions to the burgeoning field of financial sentiment analysis, specifically concerning CBDCs. First, it provides empirical validation of RoBERTa's effectiveness in capturing nuanced sentiment within this specialised financial discourse, demonstrating its suitability for analysing complex topics like CBDCs. Critically, this research highlights the significant impact of training epoch selection on model generalisation, particularly with smaller, specialised datasets. While increasing epochs generally improves performance, this study demonstrates that excessive training can lead to overfitting, as observed with RoBERTa_5.

The findings show that RoBERTa trained for only 3 epochs (RoBERTa_3) achieves a superior balance between performance and generalisation. This challenges the assumption that more training is always optimal, offering practical guidance for fine-tuning models in financial applications. Furthermore, this study demonstrates that for specialised monolingual tasks, RoBERTa (monolingual model) outperforms its multilingual counterpart, XLM-RoBERTa, suggesting that multilingual capacity does not necessarily translate to improved performance in specialised, English-only domains.

Finally, this work contributes to the growing body of research on the role of epoch selection in enhancing RoBERTa's generalisation capabilities for CBDC sentiment analysis, establishing a foundation for future research utilising transformer models to analyse sentiment surrounding emerging financial technologies.

Chapter 6 - Evaluating the robustness and explainability of RoBERTa for digital pound sentiment analysis

6.1 Introduction

This chapter presents a comprehensive *post hoc* evaluation of the fine-tuned RoBERTa model, RoBERTA_3, for financial sentiment analysis tasks [203]. Building upon the experimental findings from Chapter 5, this chapter delves deeper into RoBERTA_3's capabilities and limitations, addressing RQ2 (Model Capabilities and Limitations). This evaluation covers the following:

- Model explainability using Local Interpretable Model-agnostic Explanations (LIME).
- Robustness testing through adversarial examples and noise injection.
- Comprehensive error analysis.
- Statistical validation of the results.

The above methods aim to assess RoBERTA_3's performance and resilience to data perturbation and enhance stakeholders' understanding of the underlying reasons for the model's predictions. It is also important to acknowledge that the robustness testing was conducted using a re-fine-tuned version of RoBERTa_3 due to the unavailability of the initially fine-tuned model. Retraining with identical hyperparameters and the same dataset ensures a functionally equivalent model for these analyses [183], a common practice in NLP research.

6.2 Evaluation Methodology

6.2.1 Data Preparation

To ensure the validity and reliability of the evaluation, a rigorous data preparation process was implemented, including manual labelling, strategic sampling, and accuracy assessments.

6.2.1.1 Data Compilation, Labelling, and Sampling Strategy

The test data, which includes predicted sentiment labels made by RoBERTA_3, comprised tweets (other than the gold standard dataset) collected for sentiment analysis. To evaluate RoBERTA_3's capabilities in generalising on unseen data, the test data was manually labelled for contextually accurate annotations, which automated labelling processes or crowd-sourced annotations might misclassify [205]. The manual labelling follows the guidelines in Chapter 4 *(Section 4.3.1)* for gold standard dataset labelling, providing a foundation for subsequent model evaluation.

A simple random sampling method was employed to evaluate the model's performance on representative subsets of data, and it involves two stages:

- **Initial sample extraction (5% of data):** A random sample of 5% of the original test dataset (247 tweets out of 4924 tweets) was extracted to provide sufficient instances to capture diversity while permitting detailed analysis. Moreover, this approach helps reduce selection bias by ensuring that every data point has an equal probability of being selected.
- **Secondary sample extraction (10% of the initial sample):** A further 10% random sample (24 tweets) was extracted from the initial 5% sample to examine errors on a smaller, manageable dataset and the model's predictions, facilitating manual verification.

6.2.1.2 Accuracy Assessment of Samples

Firstly, the percent agreement on the initial sample was calculated to assess the agreement between the model's predicted sentiments and the true sentiments in the sampled data. The accuracy of the percent agreement was 76.11%, revealing the presence of misclassifications. Then, the accuracy of the secondary sample was evaluated, and the percent agreement was significantly higher, at 91.67%. This difference in agreement between the two samples suggests potential variability in the difficulty of annotating different subsets of the data. The lower agreement on the larger, more diverse initial sample suggests limitations in the model's ability to generalise to the full range of linguistic variations, while the higher agreement on the smaller sample indicates more consistent performance on potentially less complex or more homogenous tweets.

6.2.2 Evaluation Metrics and Tools

The following key metrics were used to evaluate RoBERTa_3 across multiple dimensions:

- **Robustness testing:** The model's resilience was tested against modified or degraded input data. This includes adversarial inputs like word substitutions and noise like typos or random punctuation, providing insight into its reliability in real-world scenarios.
- **Error analysis:** Misclassified tweets were analysed to identify patterns or linguistic nuances that contributed to errors; confusion matrices supported this analysis, highlighting where the model struggled the most (e.g., confusing neutral and negative sentiments).
- **Explainability:** Specific words that influenced the RoBERTa_3's predictions were identified using LIME (Local Interpretable Model-agnostic Explanations) to interpret and understand how the model prioritises and processes certain features in sentiment classification.
- **Statistical tests:** The significance of observed differences was statistically validated using permutation tests, Shapiro-Wilk tests, paired t-tests, and Wilcoxon signed-rank tests.
- **Performance metrics:** To evaluate model's performance in classifying sentiments (negative, neutral, and positive) across the Original, Adversarial, and Noisy datasets, Accuracy, Precision, Recall, and F1-Score were used. Confusion matrices also helped visualise misclassifications and gain insight into the model's prediction capabilities for each sentiment class. In addition, detailed classification reports were produced to offer a granular view of the model's performance.

6.2.2.1 Implementation Details

The model was evaluated using several Python libraries:

- **Transformers:** For loading and utilising the pre-trained RoBERTa model and tokenizer.
- **PyTorch**: As the deep learning framework for model operations.
- **Pandas and NumPy**: For data manipulation and numerical computations.
- **Scikit-learn**: For performance metrics and statistical testing.
- **LIME**: For generating explanations of individual model predictions.
- **Seaborn and Matplotlib**: For data visualisation.

6.3 Model Performance Assessment on Validation Data

Building upon the fine-tuning process described in Chapter 5, this section analyses RoBERTa-3's performance on the validation dataset. The subsequent evaluation of RoBERTa-3's predictive performance on the unseen test dataset is discussed in Section 6.4.2.1.

6.3.1 Training and Validation Loss Across Each Epoch

Figure 6.1 shows that during epoch 1, the model's initial learning phase, both training and validation losses were relatively high (0.7605 and 0.5635, respectively). Effective learning and improved generalisation were observed during epoch 2, as both training and validation losses decreased significantly to 0.5561 and 0.3272, respectively. The losses further reduced to 0.3934 (training) and 0.1986 (validation) during epoch 3, indicating no signs of overfitting with continued learning.

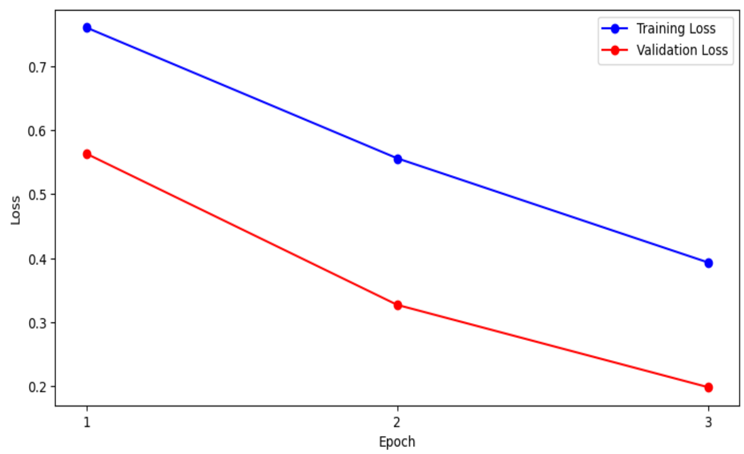


Figure 6.1: Training and Validation Loss of RoBERTa_3 Across Epochs.

6.3.2 Validation Performance Metrics Across Epochs

All metrics substantially improved across all epochs, as seen in Figure 6.2. Accuracy increased from 75.64% to 92.31%; precision, recall, and F1-score all showed similar improvements, indicating balanced performance across all sentiment classes. Achieving over 92% across all metrics by epoch 3 suggests that the model effectively learned the data distribution.

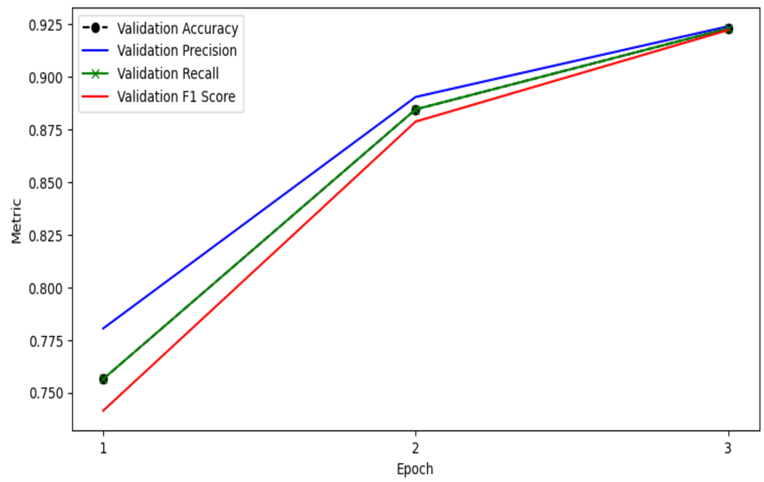


Figure 6.2: Validation Performance Metrics Across Epochs.

6.3.3 Confusion Matrix and Classification Report on Validation Data

The confusion matrix's (Figure 6.3) rows represent the actual labels (ground truth) and predicted labels could be observed in columns.

Figure 6.3: Confusion Matrix for RoBERTa_3 on Validation Data.

Figure 6.3 shows that the model performs very well on the negative class, with 71 out of 74 actual negative instances correctly predicted. There are only a few false positives (predicted neutral or positive when actually negative). The model also performs reasonably well on the neutral class, with 38 out of 46 actual neutral instances correctly classified. However, there are more misclassifications compared to the negative class, with some neutral instances being confused with positive or negative. The model's performance on the positive class is decent, with 35 out of 38 actual positive instances correctly classified. There are a few false negatives (actual positive classified as neutral or negative). The model struggles to distinguish between neutral and positive sentiment, as evidenced by misclassifications in both directions, suggesting difficulty with subtle nuances in this spectrum.

As shown in the classification report (Figure 6.4), RoBERTa_3 achieves high precision (96%) and recall (99%) for negative sentiments, effectively capturing signals such as scepticism and concerns. For neutral sentiments, the model achieves 93% precision and 83% recall, enabling the identification of informative discussions. For positive sentiments, the model achieves 85% precision and 92% recall, indicating its ability to identify supportive views. The high F1-scores (97% for negative, 87% for neutral, and 89% for positive) demonstrate a strong balance between precision and recall across all classes. These results are particularly important in the context of CBDC sentiment analysis, where accurate identification of public opinion is crucial for policymakers [20].

```
Classification Report:
              precision    recall  f1-score   support

    negative       0.96      0.99      0.97        72
     neutral       0.93      0.83      0.87        46
    positive       0.85      0.92      0.89        38

    accuracy                           0.92       156
   macro avg       0.91      0.91      0.91       156
weighted avg       0.92      0.92      0.92       156
```

Figure 6.4: Classification Report for RoBERTa_3 on Validation Data.

6.4 Model Performance Assessment on Original Test Data, Adversarial, and Noisy Datasets

6.4.1 Robustness Testing

RoBERTa_3's robustness was evaluated by conducting experiments designed to simulate potential real-world challenges, including testing the model's ability to predict sentiment labels when faced with adversarial inputs and noisy data, which are common in NLP applications.

6.4.1.1 Adversarial Testing

Adversarial testing was conducted to evaluate how the model responds to intentional perturbations (designed to deceive or confuse) in the input text. The below steps were taken:

Step 1: Define replacement mapping: A set of word substitutions was defined to systematically alter specific words within the tweets. The replacements were as follows:

- "and" → "&"
- "is" → "was"
- "the" → (removed)

- “in” → “inside”
- “on” → “upon”
- “of” → (removed)
- “for” → “4”

The above approach is grounded in empirical studies introducing syntactic and lexical variations commonly found in informal digital communication to challenge NLP models effectively. For instance, the use of symbols in social media abbreviations is mirrored by replacing “and” with “&.” Similarly, altering “is” to “was” aligns with Jia and Liang’s [206] observation that such grammatical changes can significantly impact reading comprehension systems and tense shifts could challenge a model’s temporal understanding. In addition, removing function words like “the” and “of” mimics social media users’ informal writing style or common typing errors; this could help assess RoBERTa_3’s ability to interpret meaning primarily from content words. Substituting “in” with “inside” and “on” with “upon” introduces adversarial examples to uncover how the model generalises to paraphrased inputs, as noted by Ribeiro et al. [207]. Similarly, substituting “for” with “4” simulates phonetic and numeric variations, often prevalent in user-generated content [208] and could help understand whether the model is robust to character-level perturbations.

Step 2: Applying substitutions: Each tweet was processed by splitting the text into individual words, and the defined substitutions were applied case-insensitively. Unspecified words remained unchanged.

Step 3: Generating adversarial tweets: The adversarial version of each tweet was reconstructed by recombining the modified words, and such tweets were stored in a new column titled *‘adversarial_tweet’* within the dataset’s Pandas DataFrame.

6.4.1.2 Noise Injection Testing

Noise injection simulated the informal language, typographical errors, and random punctuation common in real-world tweets. The following steps were involved:

Step 1: Random word selection: To ensure reproducibility, a single word was randomly selected for each tweet using a controlled randomisation process.

Step 2: Appending random punctuation: To introduce noise into the tweet, a random punctuation mark — either “.,” “!” or “?” — was appended to the selected word. The punctuation choice reflects common stylistic emphases or typing errors seen in social media communications. The selection of the punctuation marks was informed by their common occurrence on social media texts due to hurried typing or autocorrect features [209], as highlighted by Pruthi et al. [210], who note that due to such perturbations, models may misclassify inputs. Furthermore, Belinkov and Bisk [211], demonstrate that misplaced punctuation, natural noise, and even synthetic noise can degrade neural machine translation system performance, justifying the use of these punctuation errors to assess model robustness.

Step 3: Generating noisy tweets: The altered tweets were stored in another new column named *'noisy_tweet'* within the DataFrame to facilitate direct comparison between the original and noisy versions of each tweet during analysis.

6.4.2 Model Prediction Process and Results

The RoBERTa_3 model generated the sentiment predictions for the original, adversarial, and noisy tweets to assess the impact of adversarial and noise-injected modifications on the model's performance. The first step involved loading the fine-tuned RoBERTa_3 model to ensure consistency with the training configuration, as described in Chapter 5. Then, using the RoBERTa tokeniser, each tweet was tokenised and encoded; this step adheres to the model's input requirements regarding sequence length and padding. Following this, to obtain logits, the encoded tweets were passed through the model; logits were converted into probability distributions via the softmax function, and the sentiment class with the highest probability was selected as the predicted label. Furthermore, to facilitate a comprehensive comparison of the model's performance across different input conditions, these predictions were systematically stored in three new columns within the DataFrame:

- *'original_pred'* for the original tweets.
- *'adversarial_pred'* for the adversarial tweets.
- *'noisy_pred'* for the noisy tweets.

6.4.2.1 Results Summary

Several evaluation metrics were used to quantify the model's performance under each testing condition, as noted in Section 6.2.2.

6.4.2.1.1 Accuracy Scores

Accuracy was calculated as the proportion of correct predictions to the total number of samples for each dataset, using this formula: (Number of correct predictions) / (Total samples) * 100%

Dataset under consideration	No. of correct predictions	No. of incorrect predictions	Total samples	Accuracy (%)
Original tweets	173	74	247	70.04
Adversarial tweets	179	68	247	72.47
Noisy tweets	169	78	247	68.42

Table 6.1: Accuracy scores on different datasets.

Table 6.1 shows that the RoBERTa_3 model performs consistently across different datasets, with lower accuracy on the noisy dataset (68.42%) and highest on the adversarial dataset (72.47%).

These results indicate that the model is relatively robust to adversarial modification and performs reasonably well when exposed to the noisy dataset.

6.4.2.1.2 Confusion Matrices

The confusion matrix below (Figure 6.5) shows a high misclassification rate between neutral and positive tweets. For example, 20 tweets out of 88 neutral ones were misclassified as positive, and 18 negative tweets were misclassified as neutral. The high misclassification rate (4 false negatives) and low number of true positives (17) indicate that the model struggles with the positive sentiment (a common challenge in sentiment analysis due to the nuanced nature of these sentiments).

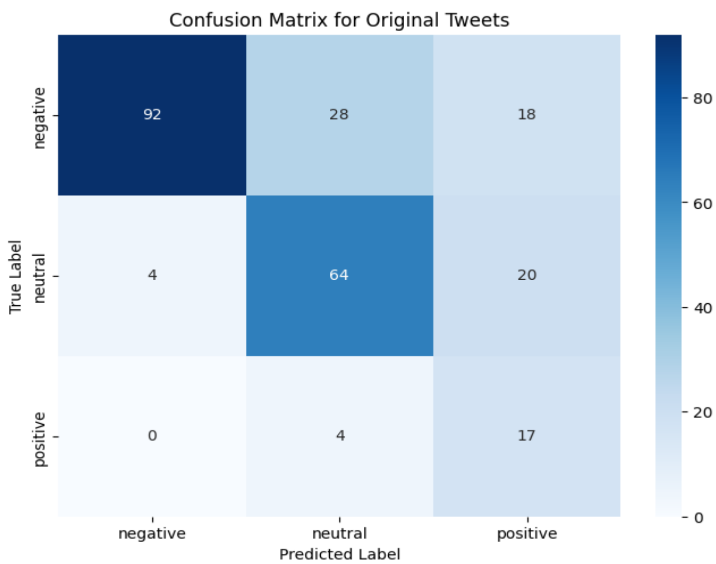


Figure 6.5: Confusion matrix for original tweets.

In the adversarial dataset, classifications between neutral and positive remain challenging, as shown in the confusion matrix (Figure 6.6). 22 neutral tweets were incorrectly classified as positive; however, slight improvements are observed in predicting negative and positive tweets (18 correctly identified), suggesting some robustness to the applied word substitutions.

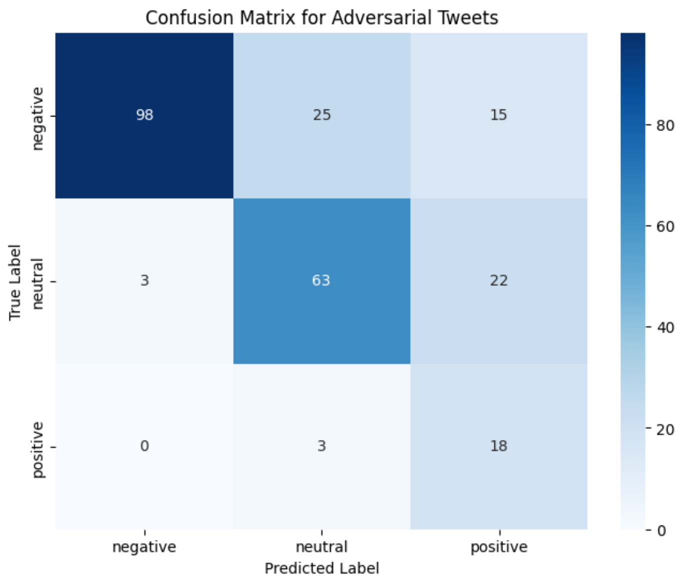


Figure 6.6: Confusion matrix for adversarial tweets.

When exposed to the noisy dataset, the model struggles to separate neutral from positive and negative tweets (Figure 6.7). Only 57 neutral tweets were correctly classified, with 6 positive tweets misclassified as neutral. While neutral tweet performance experiences a slight drop, the model still correctly classified a substantial number of neutral tweets despite the noise.

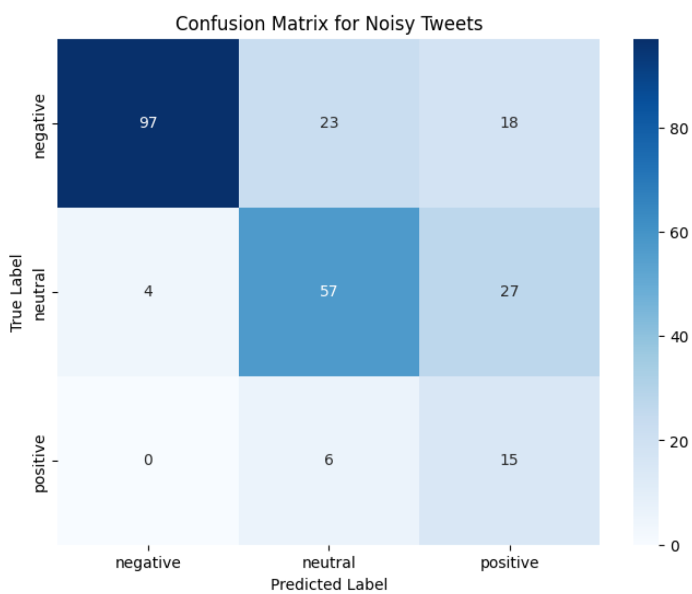


Figure 6.7: Confusion matrix for noisy tweets.

6.4.2.1.3 Classification Reports

The classification report results of all datasets are presented in the Table 6.2. The model demonstrates high precision for the negative class across all datasets (above 0.95), indicating effective identification of negative tweets. Recall varies, with the highest value in the adversarial dataset (0.71) and the lowest in the original dataset (0.67). For the neutral class, performance decreases under noisy conditions, although precision and recall remain relatively consistent across datasets. The lower F1-score (0.66) for neutral tweets in the noisy dataset suggests difficulty distinguishing neutral sentiments in noisy text. The positive sentiment class exhibits the weakest performance across all datasets, particularly in the noisy dataset, with the lowest precision (0.25) and F1-score (0.37). However, commendable recall is still achieved, especially under adversarial conditions.

Dataset	**Class**	**Precision**	**Recall**	**F1-Score**
Original	Negative	0.96	0.67	0.79
	Neutral	0.67	0.73	0.70
	Positive	0.31	0.81	0.45
Adversarial	Negative	0.97	0.71	0.82
	Neutral	0.69	0.72	0.70
	Positive	0.33	0.86	0.47
Noisy	Negative	0.96	0.70	0.81
	Neutral	0.66	0.65	0.66
	Positive	0.25	0.71	0.37

Table 6.2: The classification reports for each sentiment class.

6.4.2.2 Analysis of Results

- **Impact of adversarial modifications:** A slight improvement in overall accuracy from 70.04% in the original dataset to 72.47% in the adversarial dataset indicates RoBERTA_3's resilience to adversarial modifications. Moreover, the classification metrics, where both precision and recall for negative and positive sentiments showed slight improvements, show that the model can maintain its performance by leveraging contextual cues beyond individual word substitutions, owing to its transformer-based architecture and extensive pre-training on diverse corpora. These experimental results align with studies, such as [208], [212], [213], which indicate that pre-trained transformer models are generally more resilient to adversarial examples than previous NLP models. Specifically, the findings of Moradi and Samwald [212] highlight that pre-trained transformers like RoBERTa maintain higher stability under perturbations. As such, RoBERTa_3's performance on the adversarial dataset ensures reliable sentiment analysis in dynamic social media environments where intentional or unintentional word modifications are prevalent.

- **Effect of noise injection:** When exposed to noisy inputs, the RoBERTa_3 model's accuracy slightly declined to 68.42%. However, this represents a marginal drop compared to the accuracy of 70.04% achieved by the original dataset, indicating a commendable ability to handle random punctuation errors commonly found in social media texts. Such irregularities in text are effectively managed due to RoBERTa's BPE tokeniser, which helps it to learn noise without significantly disrupting the tokenisation process. The model's generalisation ability across different text variations is further improved by the extensive pre-training on vast and varied datasets it had gone through. This performance on noisy data makes RoBERTa_3 relevant for analysing CBDC sentiment, where accurate interpretation of informal tweets is crucial.

6.4.2.3 Implications and Limitations for Sentiment Analysis Task

This study demonstrates the RoBERTa_3 model's effectiveness in handling real-world social media data by testing its robustness against adversarial modifications and noise injections — implemented through systematic word substitutions and the introduction of random punctuation errors. User-generated texts are often informal and contain dynamic language; the model's resilience against such attacks underscores that it can capture contextual meanings beyond surface-level word changes. Moreover, accurate sentiment analysis of CBDC-related discussions holds policy relevance in understanding public opinion and guiding communication strategies. Additionally, the study contributes to NLP research by providing empirical evidence of RoBERTa_3's robustness to adversarial and noisy inputs mentioned above.

However, the model's performance under untested conditions, such as abbreviations, complex grammatical errors, and slang, remains uncertain, and the present findings only represent a subset of the linguistic variations' users may employ. Future studies could incorporate human annotator(s) to label adversarial and noisy datasets and then compare model performance under perturbed conditions. This could explain whether the model accurately captured such nuances and help validate its robustness from a human-centric perspective.

6.4.3 Error Analysis

Analysing where RoBERTa_3's predictions diverged from the true sentiments requires identifying and understanding misclassifications across the Original, Adversarial, and Noisy datasets. Error analysis helps uncover underlying patterns or features that may contribute to misclassifications.

6.4.3.1 Misclassification Overview

The number of misclassified tweets for each dataset is as follows:

- **Original dataset**: 74

- **Adversarial dataset**: 68
- **Noisy dataset**: 78

The relatively small variation in misclassifications across datasets suggests consistent model performance despite the perturbations.

Examples of Misclassified Tweets

Two examples are provided from each dataset to illustrate the nature of misclassifications:

Original Dataset

Index 1

- **Tweet**: "our system is not being overwhelmed it is being changed by step changes relating to the great reset, this is why it is accompanied by the gradual but steady removal of personal freedoms in successive reactionary legislation and will usher in cbdc and social control."
- **True sentiment**: Negative
- **Predicted sentiment**: Positive

Index 15

- **Tweet**: "no to #cbdc, say a uk think tank as well as – you guessed it – the uk #bitcoin community."
- **True Sentiment**: Negative
- **Predicted Sentiment**: Neutral

Adversarial Dataset

Index 1

- **Tweet**: "our system is not being overwhelmed it is being changed by step changes relating to the great reset, this is why it is accompanied by the gradual but steady removal of personal freedoms in successive reactionary legislation and will usher in cbdc and social control."
- **True Sentiment**: Negative
- **Predicted Sentiment**: Positive

Index 10

- **Tweet**: "central banks don't care 'how much people spend on sandwiches', says cbdc developer quant quant ceo said that cbdcs will still provide a degree of anonymity, adding that values but not identifiable info would likely be available. #crypto #bitcoin #ethereum #twitter #nft #cbdc"
- **True sentiment**: Neutral
- **Predicted sentiment**: Positive

Noisy Dataset

Index 10

- **Tweet**: "central banks don't care 'how much people spend on sandwiches', says cbdc developer quant quant ceo said that cbdcs will still provide a degree of anonymity, adding that values but not identifiable info would likely be available. #crypto #bitcoin #ethereum #twitter #nft #cbdc"
- **True sentiment**: Neutral
- **Predicted sentiment**: Positive

Index 19

- **Tweet**: "john f kennedy jr says freedom of money is just as important as freedom of speech in a free and democratic society. tell that to the uk government who are going to introduce digital currency and digital id to end free country status of the uk #cbdc #digitalid #bitcoin"
- **True Sentiment**: Negative
- **Predicted Sentiment**: Positive

6.4.3.2 Analysis of Misclassifications

Analysis of these examples reveals challenges in RoBERTa_3's sentiment interpretation:

- **Complex sentences and negative sentiment detection:** Index 1 is a lengthy, complex sentence, leading to misclassifications in both the original and adversarial datasets. The tweet expresses a strong negative expression toward government control and CBDCs, whereas the model misclassifies it as positive. It means that the model struggles while processing complex syntactic structures, which aligns with findings that long sentences containing sophisticated vocabulary and multiple clauses may confuse the model [214].
- **Implicit negativity and sarcasm:** The misclassification of Index 15 in the original dataset indicates difficulty detecting implicit negativity and sarcasm, as exemplified by the phrase "you guessed it," a known challenge in sentiment analysis [215].
- **Quotations and neutral reporting:** The model incorrectly predicts a positive sentiment of the tweet in Index 10 in the adversarial and noisy datasets, possibly due to positive connotations highlighted by phrases like "provide a degree of anonymity." This suggests that informative or neutral content is often confused with positive sentiment by RoBERTA_3, a limitation observed in sentiment analysis models [216], [217]
- **Contextual negativity overlooked:** The misclassification of Index 19 in the noisy dataset, where positive phrases like "freedom of money" are used to emphasise a negative point, highlights the model's difficulty interpreting negative context when surrounded by positive language [218].

6.4.3.2.1 Impact of Adversarial and Noisy Inputs

The consistent misclassification of specific tweets results from inherent challenges in the model's understanding of specific linguistic features and input perturbations (i.e., this is not the sole factor). It is evident from the model's robustness to controlled word substitutions, as indicated by the decrease in misclassifications in the adversarial dataset (68), aligning with the literature's findings that pre-trained transformer models can maintain performance under certain adversarial conditions [206]. Nonetheless, random punctuation errors led to increased misclassification due to disruption in tokenisation and parsing [211], thereby impacting the model's accuracy.

6.4.3.3 Misclassifications Summary

Tables 6.3 summarise the distribution of misclassified tweets across true and predicted sentiment labels for each dataset. This analysis helps identify patterns in the model's errors.

The misclassified tweets from each dataset (Original, Adversarial, and Noisy) were grouped based on their true sentiment labels and incorrectly predicted sentiments. Using the '*misclassification_summary*' function, a tabular summary for each dataset displaying the counts of misclassifications between each pair of true and predicted sentiment classes was created.

True Sentiment	**Predicted Sentiment**	**Original Dataset Count**	**Adversarial Dataset Count**	**Noisy Dataset Count**
Negative	Neutral	28	25	23
Negative	Positive	18	15	18
Neutral	Negative	4	3	4
Neutral	Positive	20	22	27
Positive	Neutral	4	3	6

Table 6.3: Original, Adversarial and Noisy datasets' misclassification summary.

6.4.3.4 Analysis and Implications:

The misclassification summaries reveal patterns in the model's errors across different datasets, as discussed below:

- **Negative sentiment misclassified as neutral or positive:** RoBERTa_3 tends to confuse negative sentiments with neutral and positive ones (the most common type of misclassification across all datasets.). The indirect or subtle language might have impacted the model's understanding of negative tweets. For instance, the tweet represented by Index 1 expresses a strong negative sentiment toward government control and CBDCs. Despite this, phrases like "gradual but steady" and "step changes" or lack of explicit negative words could have misled the model due to positive connotations. However, the model's robustness to word substitutions was observed with a slight reduction in misclassifications in the Adversarial dataset. Additionally, sensitivity to noise (indicated by the consistency) does not significantly impact this error type.

- **Neutral sentiment misclassified as positive:** In the case of uncertainty, the model frequently misclassified neutral tweets as positive, and this tendency is exacerbated in the Noisy dataset due to the presence of elements that the model interprets as positive cues. For instance, in a tweet represented by Index 10, phrases like "provide a degree of anonymity" and "values but not identifiable info would likely be available" contain positive language even if the tweet just reports information without expressing a personal sentiment. This might have led the model to predict a positive sentiment.
- **Positive sentiment misclassified as neutral:** This is the less frequent misclassification error, as the model occasionally confuses positive sentiments with neutral ones. The model struggles to recognize positive sentiments, especially when exposed to the Noisy dataset. For instance, tweets represented by Index 19 suggest positive phrases like "freedom of money" and "free and democratic society" are used to frame a negative sentiment, which might have disrupted the model's ability to interpret the negative context correctly.

The analysis of misclassifications and specific examples indicates some robustness to the applied word substitutions, particularly for negative tweets misclassified as positive. However, the increased misclassification of neutral tweets as positive in the noisy dataset suggests sensitivity to noise. Overall, the model demonstrates reasonable effectiveness for sentiment classification tasks.

6.4.3.5 Class-Wise Analysis of Misclassified Tweets

To gain further insights into the types of errors RoBERTa-3 made, quantitative analyses of the misclassified tweets were conducted. Using the *'get_misclassified_df'* function, the misclassified tweets were extracted from each dataset. Finally, class-wise misclassification rates were calculated to quantify the distribution of errors across sentiment categories.

6.4.3.5.1 Class-Wise Misclassification Rates

Class-wise misclassification rates were calculated for each dataset (Figures 6.8-6.10) to quantify how frequently the model incorrectly classified tweets into different sentiment categories.

The confusion matrix for the original dataset (Figure 6.5) reveals that 60.87% of the misclassified negative tweets were predicted as neutral, and 39.13% were predicted as positive. Additionally, the model incorrectly identified 83.33% of neutral tweets as positive, demonstrating a persistent bias toward positive sentiment prediction even in the presence of neutral emotion.

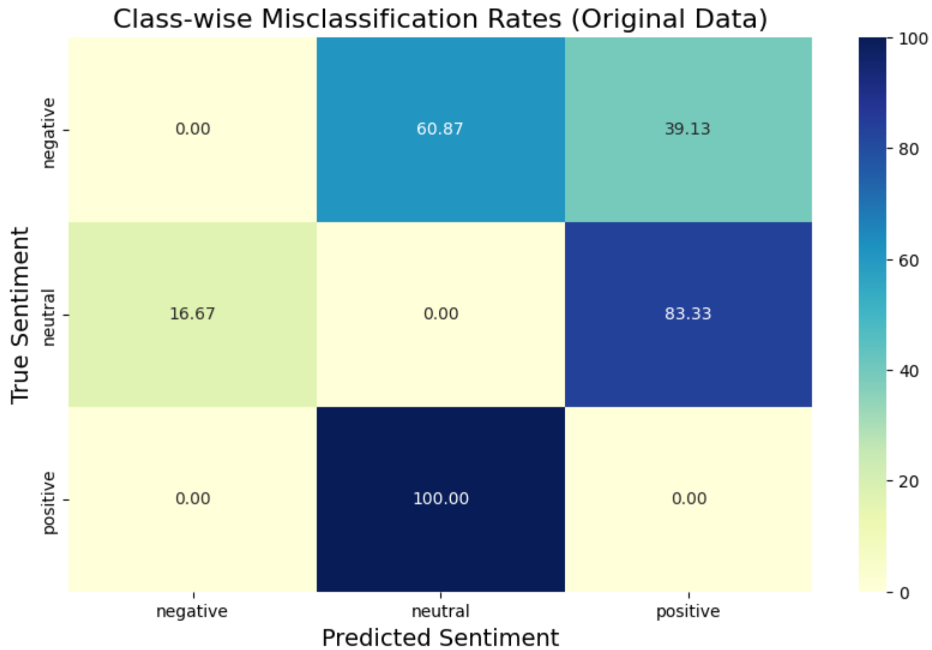


Figure 6.8: Class-wise misclassification rates in original dataset.

The model performed reasonably well in the adversarial dataset (Figure 6.6), correctly classifying 37.50% of misclassified positive tweets and 62.50% of misclassified negative tweets as neutral. The large percentage of neutral tweets mistakenly identified as positive (88%) indicates that the model favoured positive sentiment even when minor word modifications occur.

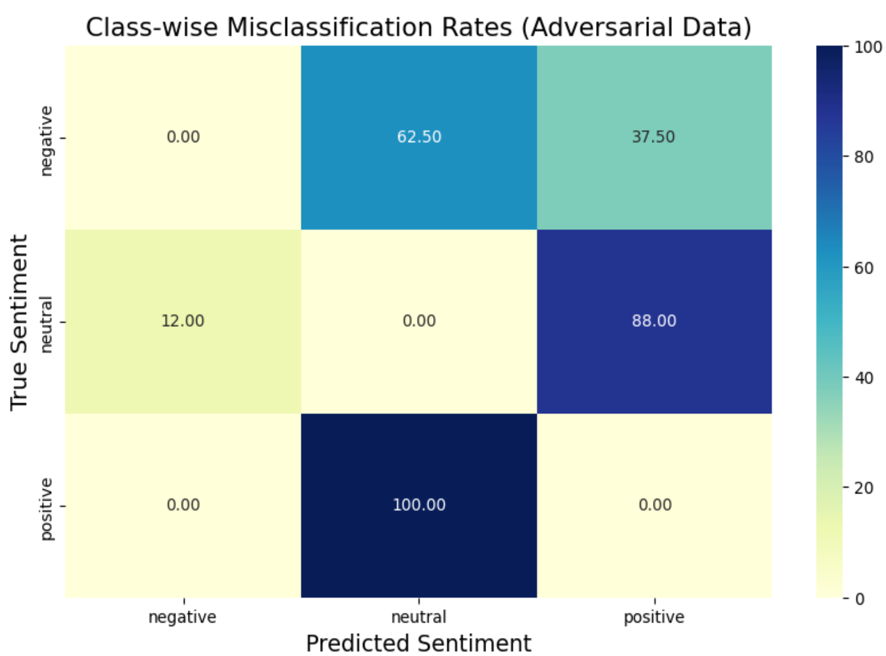


Figure 6.9: Class-wise misclassification rates in adversarial dataset.

An incremental decline in performance is evident in the noisy dataset (Figure 6.7), where 56.10% of incorrectly identified negative tweets are projected as neutral and 43.90% as positive. Neutral tweets had the largest error rate, with 87.10% of them being incorrectly categorised as positive. This suggests that the noise added to the model had a more noticeable impact on its accuracy, particularly when handling neutral tweets.

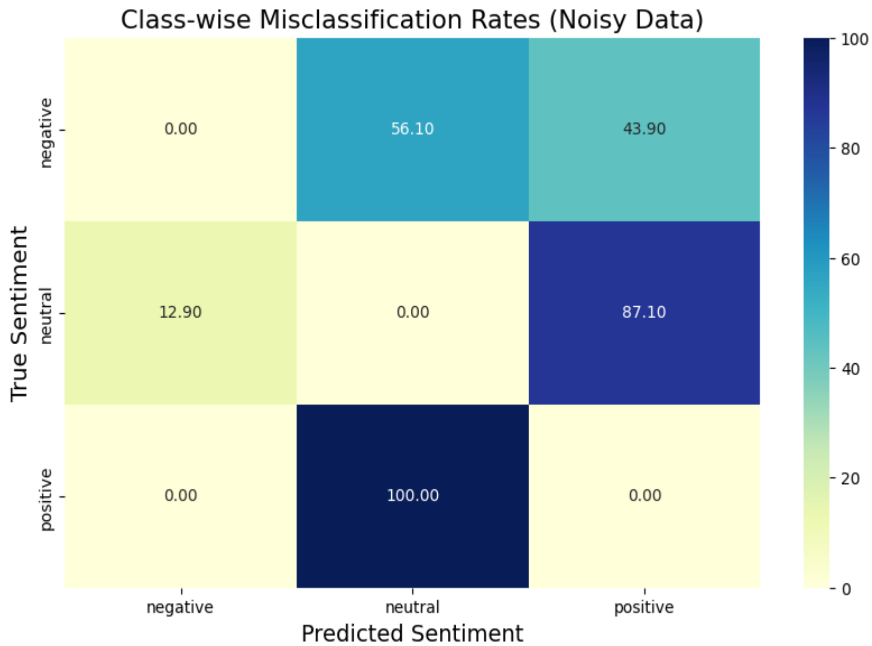


Figure 6.10: Class-wise misclassification rates in noisy dataset.

6.4.3.5.2 Implications for Sentiment Classification

The preceding analysis reveals that RoBERTa_3 is a robust model suitable for classifying UK CBDC-related tweets. The model maintained consistent performance across all three datasets, effectively handling variations without significant loss of accuracy. While a minor bias toward positive sentiment was observed for neutral tweets, this did not significantly impede the model's ability to capture overall sentiment trends.

RoBERTa_3's robustness to noise and adversarial inputs underscores its suitability for real-world applications where data imperfections are commonplace. This resilience ensures that sentiments in test data are reliably classified, thereby facilitating downstream analyses such as identifying temporal trends (Chapter 8) and comparing social media sentiment with official narratives (Chapter 10).

Future work may focus on refining the model to further mitigate biases and enhance its handling of neutral sentiments. Still, the current performance meets the study's objectives.

6.4.4 Explainability and Interpretability with LIME

It is important to understand the decision-making processes of complex models like RoBERTa (at 3 epochs), especially when working with social media data. Therefore, to elucidate the factors influencing RoBERTa_3's predictions, the Local interpretable Model-agnostic Explanations (LIME) technique was used. By approximating the model locally with an interpretable surrogate model, LIME facilitates the interpretation of individual predictions [136], [137] and highlights the most influential features — here, words — in determining sentiment classifications.

LIME was applied to a subset of tweets from the test dataset; explanations were generated for ten specific tweets: one initially selected tweet (Index 0) and nine additional randomly sampled tweets (Indices: 163, 28, 6, 189, 70, 62, 57, 35, 188). These selections encompassed various correctly classified and misclassified instances across all sentiment classes — negative, neutral, and positive.

The LIME application process involved the following steps:

- The *'LimeTextExplainer'* was configured with the class names ['negative', 'neutral', 'positive'] to align with the sentiment labels.
- A custom prediction function compatible with LIME was defined. It utilises the RoBERTa_3 tokeniser and model to output probability distributions over sentiment classes for given texts.
- For each selected tweet, LIME perturbed the text and analysed the impact of individual words on sentiment prediction, identifying the top ten contributing words for each class.
- Explanations were visualised using Matplotlib, where words contributing positively to a sentiment were highlighted in green and those contributing negatively to red.

Example Analyses:

The following examples (Figure 6.11-6.13) illustrate how LIME provides insights into RoBERTa-3's predictions:

Example 1: Correctly Classified Negative Tweet (Index 163)

- **Tweet:** "the public consultation is symbolic act with no real meaning. the public is unlikely to be able to vote on whether cbdc is desirable for uk. there is no debate or discourse, and this silence is replicated across most media outlets and print journalism. if you search for the word"
- **True Sentiment:** Negative
- **Predicted Sentiment:** Negative

Figure 6.11: Lime explanation for index 163.

LIME Explanation: *Words such as "symbolic," "unlikely," "no debate," and "silence" significantly contributed to the negative sentiment prediction (Figure 6.11), indicating that RoBERTa_3 effectively identified key phrases that convey dissatisfaction and lack of genuine engagement.*

Example 2: Misclassified Neutral Tweet Predicted as Positive (Index 62)

- **Tweet:** "these insights will have a significant impact on the design choices for cbdc, and the future of digital currency in the uk."
- **True Sentiment:** Neutral
- **Predicted Sentiment:** Positive

Figure 6.12: Lime explanation for index 62.

LIME Explanation: *Contextually neutral Words, such as "significant impact," "design choices," and "future" heavily influenced the positive prediction because of their association with forward-looking statements (Figure 6.12). This may have led the model to interpret the sentiment as positive, revealing a bias towards optimistic language.*

Example 3: Correctly Classified Neutral Tweet (Index 28)

- **Tweet:** "#digitalpound by 2030? here's what r3's head of govrel emea, chris ford, had to say to @bleadernews on what it means for the #blockchain and #fintech sector, and how the @bankofengland can leverage the industry."
- **True Sentiment:** Neutral
- **Predicted Sentiment:** Neutral

Figure 6.13: Lime explanation for index 28.

LIME Explanation: *The model identified informational terms (Figure 6.13) like "digitalpound," "2030," "blockchain," "fintech," and "leverage" as influential in predicting a neutral sentiment (even these words could be associated with positive sentiment). This demonstrates RoBERTa_3's ability to interpret context and not solely rely on individual keywords.*

6.4.4.1 Key Findings on Model Explainability and Interpretability

In most cases, influential words, as identified by RoBERTa_3, align with what a human annotator would label significant for sentiment classification. As highlighted in example 1 above, negative connotations were correctly highlighted, meaning the model focuses well on contextually relevant terms. Even LIME provided reasons behind misclassifications, such as in Example 2, where positive words lead to incorrect labelling of neutral sentiment. In contrast, in example 3, the model correctly classified the tweet as neutral despite potentially positive keywords, highlighting that RoBERTa_3 can interpret tweets beyond surface-level keyword associations. However, it occasionally overemphasised specific keywords, leading to biassed predictions. But overall, the model generally performed well.

Based on the above analysis, it is evident that the model is robust, providing a solid foundation for identifying genuine patterns and trends in public opinion on CBDCs, which is essential for accurate sentiment labelling. While further accuracy improvements might be possible, they were beyond the scope of this study, which prioritised establishing a reliable and robust baseline for downstream analyses. Moreover, LIME further increases confidence in the labelled data by revealing key factors influencing the model's decisions. Specifically, this validated reliability and interpretability justify the use of the classified data for multifaceted sentiment analysis and temporal sentiment tracking. These downstream analyses, which are crucial for understanding public opinion trends and informing policy development, can now be conducted with greater confidence, as the LIME analysis confirms that the model's predictions are based on contextually relevant features, further enhancing the trustworthiness of the insights derived from the subsequent analysis in Chapters 7 and 8.

6.5 Statistical Validation of Model's Robustness

A series of hypothesis tests were conducted to determine whether the differences in the model's performance across all three datasets and confidence scores are statistically significant.

6.5.1 Objectives and Hypotheses

This statistical analysis assesses RoBERTa_3's resilience to adversarial word substitutions and typographical noise. Specifically,

- **Accuracy differences:** To assess if there is a significant difference in the model's accuracy between the original and modified datasets (Adversarial and Noisy).
- **Confidence scores:** To assess how the model's confidence in its predictions for the true class varies between the datasets.

Dependent and independent variables:

- **Independent variables:** The dataset type (Original, Adversarial, Noisy) and the modification condition (presence or absence of word substitutions / random noise).
- **Dependent variables:** (a) Model accuracy — i.e., the proportion of correct predictions across tweets, and (b) Model confidence — the probability assigned by the model to the predicted class.

Hypothesis statements:

- **Null Hypothesis (H_0):** The model's performance (accuracy and confidence scores) between the original and modified datasets is not significantly different.
- **Alternative Hypothesis (H_1):** There is a significant difference in the model's performance between the original and modified datasets.

6.5.2 Statistical Tests

The below statistical methods were considered to test the above hypotheses:

- **Permutation test for accuracy differences:** The permutation test is a nonparametric method for determining the significance of the difference between two sample means or accuracies. It involves calculating and comparing the observed difference to a distribution of differences obtained by randomly permuting the labels. The test statistic used in this analysis is the difference in accuracy between the two datasets. It is suitable because the impact of adversarial or noisy data on model accuracy might not follow a normal distribution.
- **Shapiro-Wilk test for normality:** The Shapiro-Wilk test assesses whether a sample comes from a normally distributed population. This test was applied to verify this assumption, as discussed in Section 5.6.3).
- **Paired t-test:** The paired t-test compares the means of two related groups to determine if there is a statistically significant difference between these means. This test assumes that the differences between the paired observations are normally distributed.
- **Wilcoxon Signed-Rank Test (used when normality assumption is violated):** It is a non-parametric test used as an alternative to the paired t-test when the differences between pairs are not normally distributed. This test assumes that differences around the medians are symmetric.

Then, the model's predictions (the predicted sentiment classes for each tweet) and confidence scores (the probability assigned by the model to the true sentiment class for each tweet) were obtained for the true class across all datasets. The difference in confidence scores between the original and modified datasets was calculated by subtracting the modified dataset's confidence scores ($Confidence_{Modified}$) from the original dataset's confidence scores ($Confidence_{Original}$).

6.5.3 Results and Interpretation

The results of the statistical tests are summarised in Table 6.4. These tests indicate that RoBERTa's performance, in terms of both accuracy and confidence scores, is not significantly affected by either adversarial word substitutions or noise injection. While the observed differences exist, the high p-values from both the permutation test (for accuracy) and the Wilcoxon signed-rank test (for confidence scores) suggest that these differences are likely due to chance and not a systematic impact of the perturbations. In essence, RoBERTa demonstrates resilience under modified input conditions, reinforcing its resilience in real-world applications where noisy or adversarial modifications may occur.

Comparison	Test	Observed statistic	p-Value	Interpretation
Original vs. Adversarial	Permutation Test (accuracy)	$D_{obs} = -0.0243$	0.2547	Fail to reject H_0; no significant difference in

				accuracy.
	Shapiro-Wilk (Conf. score diff.)	Differences tested for normality	0.0000	Differences not normally distributed; paired t-test not suitable.
	Wilcoxon Signed-Rank (Conf. score)	Ranks of paired differences	0.4489	Fail to reject H_0; no significant difference in model's confidence scores.
Original vs. Noisy	Permutation Test (accuracy)	$D_{obs} = +0.0162$	0.5455	Fail to reject H_0; no significant difference in accuracy.
	Shapiro-Wilk (Conf. Score Diff.)	Differences tested for normality	0.0000	Differences not normally distributed; paired t-test not suitable.
	Wilcoxon Signed-Rank (Conf. Score)	Ranks of paired differences	0.1712	Fail to reject H_0; no significant difference in model's confidence scores.

Table 6.4: Summary of statistical results of original vs. modified datasets.

6.6 Impact of Model Re-Fine Tuning

Due to a technical oversight, the initial fine-tuned model was not saved. A re-fine-tuned model, exhibiting slightly altered performance metrics (shown in Table 6.5), was used for robustness testing. To assess the impact of this change on robustness, results of initial fine-tuned and re-fine-tuned model were compared.

Dependent and independent variables:

- **Independent variable:** Model fine-tuning condition (Initial Fine-Tuning vs. Re-Fine-Tuning).
- **Dependent variables:** Model performance, measured by a) accuracy — proportion of correctly classified sentiment labels; b) confusion matrix distributions — number of correctly/incorrectly classified Negative, Neutral, and Positive predictions.

Actual	Predicted Negative	Predicted Neutral	Predicted Positive
Initial Fine-Tuning			
Negative	72	1	0
Neutral	2	43	3
Positive	0	4	31
Re-Fine-Tuning			
Negative	71	1	0
Neutral	2	38	6
Positive	1	2	35

Table 6.5: Results of initially fine-tuned vs. re-fine-tuned model.

The overall accuracies of the models were comparable despite these discrepancies:

- **Initial fine-tuning accuracy:** (72 + 43 + 31) / 156 = 93.59%
- **Re fine-tuning accuracy:** (71 + 38 + 35) / 156 = 92.31%

In addition to the above, a Two-Proportion Z-Test was conducted to determine whether the observed differences in model performance are statistically significant. This test compares the accuracies of the initial and re-fine-tuned models.

Hypotheses statements:

- **Null Hypothesis (H_0)**: There is no significant difference in the accuracies of the initial and re-fine-tuned models ($p_1 = p_2$).
- **Alternative Hypothesis (H_1)**: There is a significant difference in the accuracies of the initial and re-fine-tuned models ($p_1 \neq p_2$).

Using the correct prediction and total predication values above (Table 6.5), the Two-Proportion Z-Test yielded the following results:

- **Z-Statistic:** $z = 0.4423$
- **P-Value:** $p = 0.6583$

The null hypothesis is not rejected given that the p-value exceeds the conventional significance level (α=0.05), indicating no statistically significant difference is observed in the accuracies of the two models. Therefore, the minor discrepancies in performance metrics are likely due to random variations inherent in the training process. This finding contrasts with Dodge et al. [185] who reported performance instability with small fine-tuned datasets. However, in the present study, the use of 622 tweets for training did not result in substantial performance variance.

To further assess the equivalence of the models' performance, 95% confidence intervals for the accuracies were calculated (Figure 6.14). The Wilson score interval method was used for this purpose. The results are as follows:

- **Initial fine-tuning accuracy 95% CI:** (88.60%, 96.48%)
- **Re-fine-tuning accuracy 95% CI:** (87.04%, 95.55%)

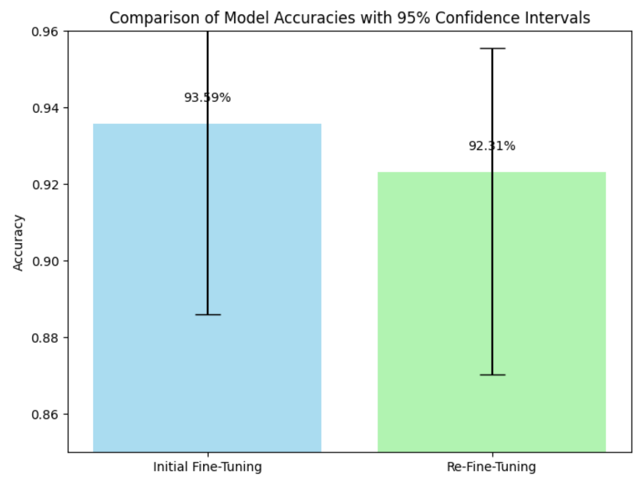


Figure 6.14: Model accuracies' comparison with 95% confidence intervals.

The substantial overlap between these confidence intervals, as shown in the Figure 6.17, suggests that the two models' true accuracies are statistically indistinguishable, further supporting the conclusion that the differences in performance are not significant.

Implications: The statistical analysis indicates that the re-fine-tuned model's performance is effectively equivalent to the initial model's. Consequently, the downstream analyses remain valid and reliable. The robustness testing performed with the re-fine-tuned model provides additional evidence of the model's resilience to adversarial and noisy inputs without undermining the integrity of the earlier analyses.

6.7 Conclusion

The comprehensive evaluation of RoBERTa_3's robustness shows that it is an interpretable and statistically reliable model for sentiment analysis in the context of CBDC discussions on social media, directly addressing RQ2. Specifically, this chapter investigated the capabilities and limitations of the transformer model chosen in Chapter 5, thus implicitly addressing RQ1 in accurately predicting sentiments within the digital pound discourse.

Despite the minor methodological inconsistency encountered, statistical validation confirmed the validity and reliability of the analyses conducted. The results demonstrate RoBERTa-3's robustness in handling variations in social media data without significant loss of accuracy, meeting the project's objectives for robust and reliable sentiment classification. However, error analysis revealed limitations in handling nuanced language, specific CBDC-related terminology implicit sentiment, sarcasm, and distinguishing between neutral and positive sentiment.

This work contributes novel findings on the comparative analysis between initial and re-fine-tuned models, emphasising the negligible impact of re-fine-tuning on model performance and the validity of downstream analyses based on initial predictions. The findings further demonstrate that transformers-based models like RoBERTa_3 effectively capture sentiment from complex, real-world social media data and are robust against common data perturbations. This robustness is essential for practical applications where data quality cannot always be controlled. Moreover, the model's inner workings were explained using LIME, offering stakeholders the trust to rely on the model's insights. This addresses the explainability aspect of RQ2. Finally, the statistical validation reinforces the reliability of the model's performance, providing a solid foundation for its use in downstream tasks in Chapters 7 and 8, which will then be compared with official communications in Chapter 8 and 9 (addressing RQ4: Analysis of Official Communications and RQ5: Comparative Analysis and Communication Theories' Lens). The methodologies employed in this evaluation can serve as a framework for future research aiming to assess and deploy NLP models in similar contexts.

Chapter 7 – Multifaceted Analysis of Digital Pound Discourse Beyond EDA

7.1 Introduction

This chapter offers a multifaceted analysis of public discourse on the digital pound, leveraging X data to explore a range of perspectives beyond traditional exploratory data analysis (EDA). The analysis focuses on three distinct timelines, as noted in Section 3.3.1.1 and aims to provide a comprehensive understanding of public perceptions and discussions surrounding the digital pound by examining sentiments, emotions, topics, semantic relationships, and other relevant patterns within the tweets.

This foundational analysis sets the stage for a deeper exploration of temporal trends in Chapter 8, which will examine the evolution of these patterns over time. The insights presented here directly address RQ3, uncovering the key themes and topics that emerge within the digital pound conversation across different policy-relevant timelines.

7.2 Sentiment Analysis

This section presents the overall sentiment analysis using RoBERTa_3 (justified in Chapter 5 and validated for robustness in Chapter 6) and compares it with the results obtained using VADER for each timeline.

7.2.1 RoBERTa-Derived Sentiment Distribution

RoBERTa_3 was selected to classify the sentiments across all three timelines. The timeline-specific data was merged to form one dataset to ensure sufficient linguistic and sentiment variability across all timelines. The sentiment distribution is presented in Table 7.1.

Timeline	Positive	Negative	Neutral	Total Tweets
2020	71 (28.4%)	86 (34.4%)	93 (37.2%)	250
2023	624 (14.6%)	1,944 (45.5%)	1,703 (39.9%)	4,271
2024	209 (17.7%)	599 (50.7%)	373 (31.6%)	1,181

Table 7.1: Sentiment distribution (RoBERTa).

As the Table 7.1 shows, negative sentiment towards the digital pound is predominant and has become increasingly critical, particularly in 2023 and 2024. This suggests that as the consultation process advanced, there seems to be a growing public concern, dissatisfaction, or scepticism about the digital pound. This observation contributes to RQ3 by addressing the "sentiment trends" aspect of RQ3 by presenting quantitative sentiment distributions across the three timelines.

7.2.2 Evaluating RoBERTa_3 Against a Lexicon-Based Approach

VADER (Valence Aware Dictionary and Sentiment Reasoner) is used as a comparative tool for sentiment classification, specifically to provide per-timeline sentiment analysis and compare it with

RoBERTa (the primary model validated in Chapter 6). VADER is a rule-based sentiment analysis tool that is efficient for small datasets but has limitations in handling domain-specific language [219]. Based on a pre-defined lexicon (a dictionary of words associated with sentiment values), it assigns sentiment scores to words and then sums the scores to predict the overall sentiment for a given tweet.

Confusion matrices (Figures 7.1, 7.2, and 7.3) were generated to compare RoBERTa_3's results (rows) with VADER's predictions (columns). A diagonal element shows agreement, whereas an off-diagonal one indicates disagreement.

a) 2020:

For 2020 timeline, both RoBERTa and VADER accurately classified 31 tweets as negative. However, 18 tweets that RoBERTa labelled as negative were labelled neutral by VADER, and 37 tweets labelled negative by RoBERTa were misclassified as positive by VADER.

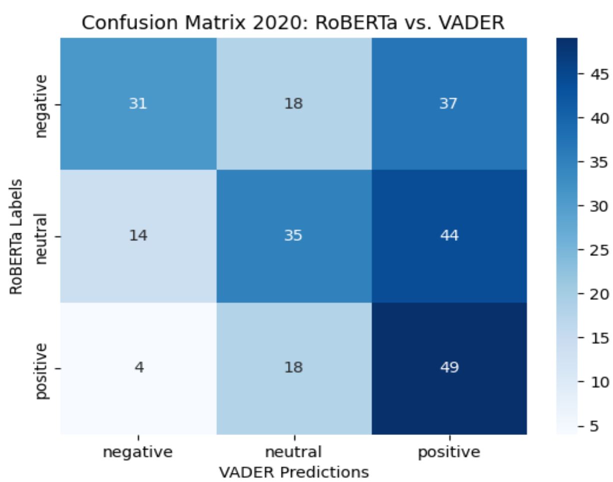


Figure 7.1: Confusion Matrix 2020: RoBERTa vs. VADER.

b) 2023:

VADER correctly classified 812 negative tweets but misclassified 810 negative tweets as positive for 2023 timeline. Additionally, VADER accurately labelled 799 tweets as neutral but misclassified 743 neutral tweets as positive.

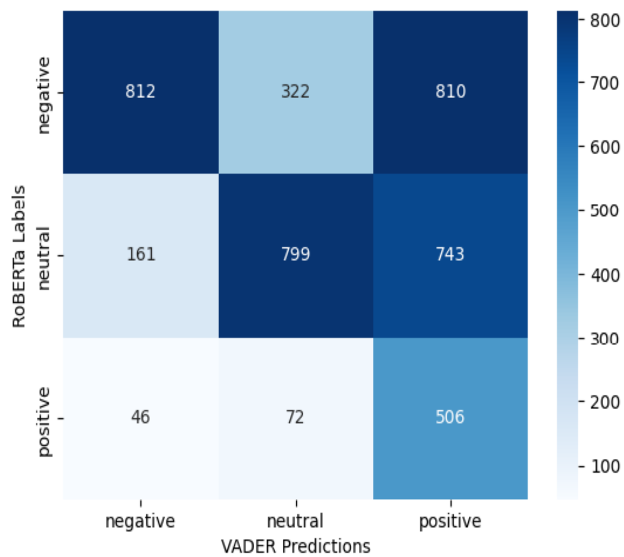


Figure 7.2: Confusion Matrix 2023: RoBERTa vs. VADER.

c) 2024:

For 2024 timeline, VADER accurately classified 256 negative tweets but misclassified 252 negative tweets as positive. It correctly labelled 172 positive tweets but misclassified 202 neutral tweets as positive.

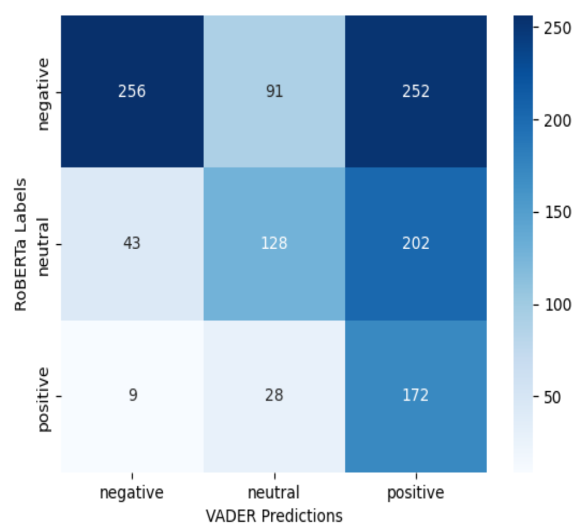


Figure 7.3: Confusion Matrix 2024: RoBERTa vs. VADER.

The above comparison highlights VADER’s limitations [61] and underscores the need for domain-specific models to understand evolving sentiment nuances in specialised domains [124], [126]. VADER's reliance on a fixed lexicon appears insufficient to capture the dynamic and context-dependent nature of sentiment in this evolving domain.

7.2.3 Statistical Validation of RoBERTa_3 vs. VADER Discrepancies

To provide quantitative evidence for the observed discrepancies between RoBERTa and VADER, Cohen's Kappa and the Chi-Square Test of independence tests were conducted.

Dependent and independent variables:

- **Independent variable:** Sentiment classification model (RoBERTa_3 vs. VADER).
- **Dependent variables:** a) Sentiment class labels (Positive, Negative, Neutral); b) Agreement/disagreement levels between models (measured via Cohen's Kappa).

Hypotheses statements:

a) Hypothesis for Cohen's Kappa-

- **Null Hypothesis (H_0):** RoBERTa_3 and VADER exhibit substantial agreement in sentiment classification (Kappa > 0.6).
- **Alternative Hypothesis (H_1):** RoBERTa_3 and VADER exhibit only fair or low agreement (Kappa < 0.6), suggesting inconsistency.

b) Hypothesis for Chi-Square Test-

- **Null Hypothesis (H_0):** There is no statistically significant difference between RoBERTa_3 and VADER sentiment distributions.
- **Alternative Hypothesis (H_1):** There is a statistically significant difference between RoBERTa_3 and VADER sentiment distributions.

The results are presented in Table 7.2.

Timeline	Cohen's Kappa	Interpretation	Chi-Square Statistic	p-value	Test Result
2020	0.2050	Fair agreement	30.307	4.24×10^{-6}	Statistically significant difference between RoBERTa and VADER distributions (reject H_0)
2023	0.2893	Fair agreement	1014.939	2.07×10^{-218}	
2024	0.2524	Fair agreement	226.508	7.45×10^{-48}	

Table 7.2: Comparative statistical analysis of VADER vs. RoBERTa.

The Kappa values for all timelines indicate a fair level of agreement between RoBERTa_3 and VADER, suggesting some overlap in their classifications but also significant discrepancies.

However, this level of agreement doesn't place VADER in a position to be relied upon for precise sentiment analysis, i.e., models are not interchangeable. Similarly, the extremely low p-values (below the conventional significance level of 0.05) in all timelines lead to the rejection of the null hypothesis (H_0) that no significant difference exists between the sentiment distributions of RoBERTa_3 and VADER. This means that both models' sentiment classification differences are statistically significant, beyond what could be attributed to chance.

With an almost infinitesimal p-value (2.07×10−218) and the exceptionally high Chi-Square Statistic of 1014.939, the divergence was particularly pronounced in 2023 potentially due to advanced discussions around digital pound. This analysis reinforces RQ2 regarding the superiority of advanced, domain-specific models (RoBERTa) over rule-based methods like VADER [61], [126].

7.2.4 Implications for Statistical Analysis in Specialised Domains like CBDCs

The above results indicate that VADER's rule-based method has limitations in handling nuanced and context-dependent language [48], [49], [50], which can explain the discrepancies between its predictions and RoBERTa's results. Additionally, the confusion matrices reveal that VADER's rule-based method is not efficient enough to handle nuanced and context-dependent language [219]. It tends to misclassify many neutral or negative tweets as positive, especially in 2020 and 2023 timelines; this could result from its inability to understand subtle negativity, sarcasm, or domain-specific terminology [220], [221]. Moreover, it struggles with correctly identifying neutral tweets, most noticeable in 2023 and 2024, where it often misclassifies them as either positive or negative [222]. As VADER is not trained on the gold standard dataset like RoBERTa, it relies on its built-in lexicon and rules to label the sentiment of each tweet and failed to capture the evolving public sentiment and discourse complexity (the increasing Chi-Square Statistic over time, particularly peaking in 2023 reflects this) as the digital pound became a more prominent topic in 2023.

In contrast, RoBERTa's ability to adapt to emerging language trends over time highlights its robustness [112], [223]. It makes it a more reliable tool for sentiment analysis in specialised contexts where language usage deviates from general patterns. Furthermore, the study highlighted that policymakers and researchers relying upon VADER to gauge public opinion on critical financial innovations could yield misinterpreted results, potentially leading to misguided decisions. This finding is directly relevant to RQ2 and RQ3, highlighting the importance of model selection in accurately capturing public sentiment in complex financial domains.

7.3 Sentiment Distribution

Sentiment distribution across all years is visualised using a bar chart as shown in the Figure 7.4.

7.3.1 Sentiment Distribution Across Timelines

The trend reveals challenges faced in digital pound implementation or on-going public concerns as discussions moved from conceptual discussions to more practical considerations.

- **2020**: Neutral sentiment was slightly dominant in this year, while sentiment distribution is relatively balanced, suggesting that people were likely beginning to understand CBDCs as discussions were exploratory, with a mix of cautious optimism and some critique.
- **2023**: Both negative and neutral sentiments saw a significant increase, with negative sentiment being the highest, indicating growing scepticism or concerns around CBDCs in 2023, potentially due to technical or economic challenges that came to light. Neutral discussions also highlight cautious yet balanced interests.
- **2024**: Negative sentiment decreased compared to 2023 but remains dominant. Neutral sentiment has also dropped, while positive sentiment remains low, reflecting a maturing public opinion where criticisms persist. However, a saturation point in the debate around CBDCs could have arisen as the discussion volume may have been reduced.

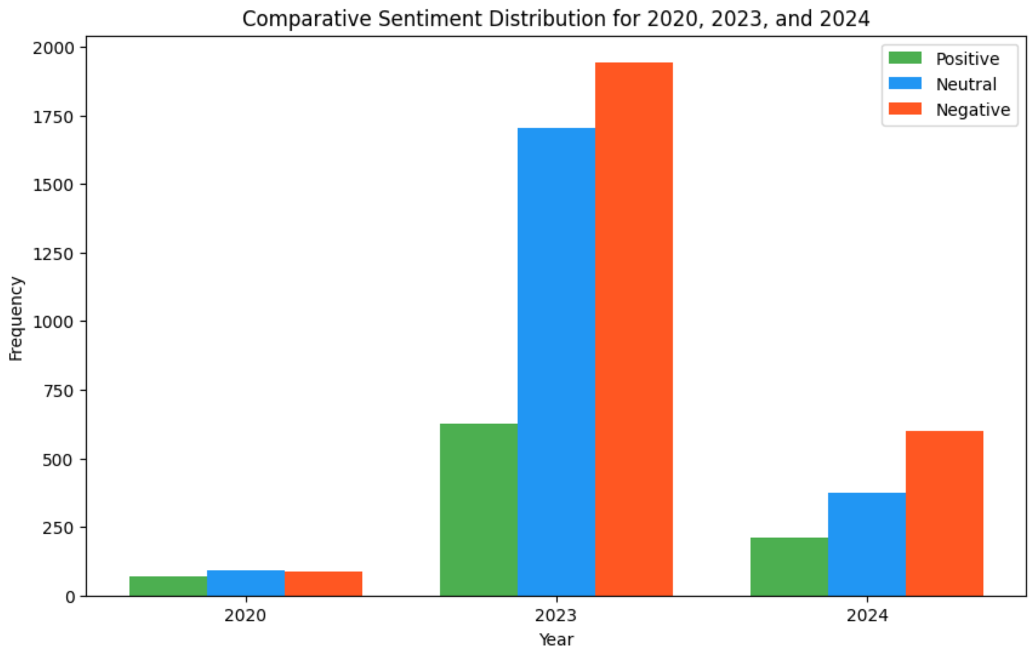


Figure 7.4: Sentiment distribution across all years.

For further analysis of public discourse, natural class distribution (Figure 7.4) was retained because manipulating class frequencies using undersampling or oversampling techniques could introduce artificial proportions not representative of how sentiments occur on X discussions across 2020, 2023, and 2024. Class-rebalancing could also risk undermining external validity, because policy decisions and communication strategies rely on knowing the actual prevalence of each sentiment category, even if it is skewed or heavily weighted toward negativity[224]. Also, the analysis of RoBERTa and VADER in Section 7.2.2 shows that both models classify a large share of tweets as Negative in 2023. The confusion matrix (Figure 7.2) indicates that Neutral (799 vs. 743 or 322 vs. 161, etc.) and Positive (506 vs. 46 or 72) classes also appear in substantial but smaller numbers; forcing data to “balance” these classes artificially might misrepresent underlying sentiment trends. Notably, RoBERTa’s contextual learning and attention mechanism [112] tend to mitigate the impact of class imbalance [225].

7.4 Descriptive Analysis of CBDC-Related Term Prevalence

An analysis was conducted to assess the prevalence and sentiment of 31 predefined CBDC-related terms across all timelines. This list included 31 terms and phrases closely associated with CBDCs, such as:

- **Technological terms**: 'blockchain,' 'cryptocurrency', 'digital,' 'ledger,' 'ethereum,' 'bitcoin,' 'wallet', 'transactions.'
- **Institutional and policy terms**: 'central', 'bank', 'regulation', 'policy', 'monetary', 'financial', 'system', 'economy', 'decentralised', 'central bank digital currency.'
- **Security and privacy terms**: 'security', 'risk', 'privacy', 'anonymity', 'surveillance.'
- **Innovation and adoption terms**: 'innovation', 'technology', 'cashless', 'stablecoin', 'digital pound.'

The analysis of CBDC-related content involved parsing each tweet, identifying individual words, and tallying the number of matches with a pre-established list of CBDC terms. Then, tweets containing one or more of the above terms were classified as CBDC-related. The percentage of CBDC-related tweets was calculated by dividing the number of CBDC-related tweets by the total number of tweets for each timeline. Finally, the sentiments of the CBDC-related tweets were analysed using the RoBERTa_3 model's sentiment labels and the distribution of sentiments was calculated as a percentage of the CBDC-related tweets.

7.4.1 Results

The percentage of tweets containing CBDC-related terms increased from 79.60% in 2020 to 92.98% in 2023, then decreased slightly to 87.13% in 2024 (Table 7.3), indicating a growing focus on explicit CBDC/digital pound topics followed by a slight shift in public interest, possibly due to public interest in other areas of digital finance.

Year	Total tweets	Tweets containing one or more defined CBDC-related terms	Percentage of CBDC-related tweets (%)	Sentiment distribution for tweets containing CBDC-related terms (%)
2020	250	199	79.60	Positive: 27.64 Neutral: 38.69 Negative: 33.67
2023	4,271	3,971	92.98	Positive: 14.33 Neutral: 40.82

				Negative: 44.85
2024	1,181	1,029	87.13	Positive: 19.34 Neutral: 32.85 Negative: 47.81

Table 7.3: Prevalence of CBDC-related terms across all years.

Insights: *In 2020, the sentiment was relatively balanced, with neutral tweets being the most common. Example neutral tweets include "Either CBDC form deposit account ppl somehow still prefer cash in circulation," and "Digital Monetary Institute grows CBDC think tank." The positive sentiment about the digital pound and CBDCs also signalled early optimism and curiosity. In contrast, positive sentiment decreased to 14.33%, while negative sentiment increased substantially to 44.85%; this shift likely corresponds with the release of the Bank of England's consultation and technology papers, which may have raised public concerns regarding privacy, government control, and the overall impact on the existing financial system. An example of a negative tweet includes, "Spot central bank digital currency would enormously increase government control over money and economy." Similarly, negative sentiment continued to rise to 47.81%, with a slight increase (19.34%) in positive sentiment. An example of a negative tweet from this year includes "Yeah, digital pound tracked currency people expect us to use." The persistence of high negative sentiment suggests unresolved issues or scepticism about the implementation and implications of the digital pound. Thus, effective communication from the Bank of England and other stakeholders is crucial to addressing and mitigating issues related to privacy, security, and the potential for increased government control. Moreover, monitoring public sentiment through social media analysis can provide insights into public perceptions of CBDCs, allowing for adaptive policy implementation and communication strategies. This section contributes to RQ3 by quantifying how discussions explicitly tied to CBDCs evolve, thereby revealing changes in thematic focus over time.*

7.5 Emotion Analysis Using NRC Emotion Lexicon

The NRC Emotion Lexicon was applied to each timeline's dataset to detect and quantify key emotions — such as joy, trust, fear, and anger, among others. The NRC Emotion Lexicon categorises emotions into states based on their usage in large datasets and is a widely used resource in emotion detection [226]. This approach aligns with established methods for identifying emotional states in texts by leveraging emotion lexicons [221].

Timeline 1 (2020): Initial optimism

During early-stage discussions about the digital pound at the time of the discussion paper, *trust* and *positive* emotions dominated (Appendix 3), meaning there was an optimistic view of the Bank of England's discussion paper. The key focus was looking at the benefits of the digital pound, while *anticipation, fear* and *anger* were also present, but they didn't reflect widespread concerns. This implies that discussions were exploratory in nature during 2020.

Timeline 2 (2023): Growing scepticism

In 2023, the emotional tone shifted toward *fear* (growing anxiety) and *anger* (rising frustration), a common phenomenon found in the literature regarding discourse becoming critical as implementation details of new technologies emerge [227]. As a result, there was a sharp rise in *negative* emotion, scoring nearly 6,000 (see Figure 7.5), implying that with the launch of Consulting and Technology Papers, the public became more aware of the potential risks (e.g., financial surveillance and data privacy) concerning the digital pound. However, *trust* and *positive* emotions remained high, reflecting the public's confidence in BoE's effective management of the digital pound.

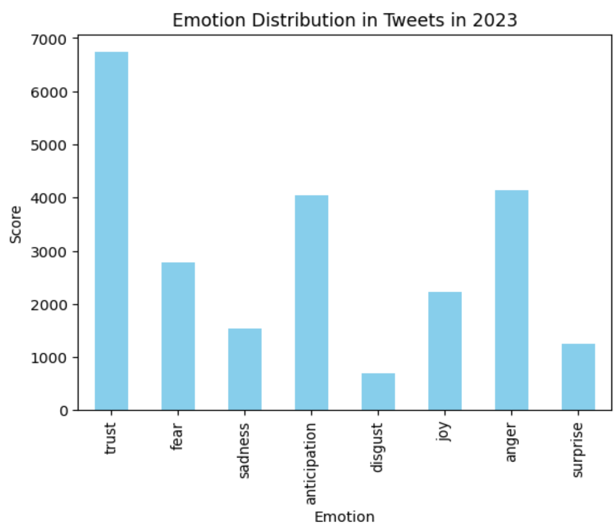


Figure 7.5: Emotion distribution in tweets in 2023.

The 2023 timeline represents the most significant shift in public sentiment and emotional response, coinciding with the release of key policy documents. The focus on 2023 allows for a more detailed examination of the pivotal moment in the digital pound discussions. Emotional distribution data for 2020 and 2024 are presented in Appendices 4 and 5, respectively.

Timeline 3 (2024): Opposition and concern

By 2024, the rise in negative sentiment, anger, and sadness highlighted a growing opposition to the digital pound. *Fear* was still present (around 1,100 instances) but lower than in 2023, which co-exists with resigned emotions like *sadness* (around 7,00 instances). It had become a polarising topic where positive and trusting emotions remained, indicating continued optimism among some; on the contrary, negative emotions reflect more profound public dissatisfaction, possibly driven by unresolved privacy and government control concerns or fears about economic disruption. The growing disillusionment is consistent with the timeline's context, where the public was reacting to the BoE and HM Treasury's feedback to the public's response to the 2023 papers.

Insights: *The emotional analysis reveals that optimism in 2020 was fuelled by exploratory discussions surrounding 2020, while public concerns increase with the launch of Consulting and Technology papers. By 2024, there seems to be a deepening divide between supporters and critics of the digital pound. The above analysis indicates that to ensure widespread public support to the digital pound, public concerns — particularly regarding privacy and government control — should be addressed. By comparing the emotional responses across 2020, 2023, and 2024 (Figures 7.14–7.16), this analysis answers the "emotional responses" component of RQ3***.

7.6 N-gram Analysis

N-gram analysis focuses on uncovering common themes and how sentiment shifts from one period to another based on recurring word sequences. A contiguous sequence of 'n' items — typically words — from a given text sample is called N-gram [228]. In textual analysis, n-grams help identify frequent word combinations representing key themes, concepts, or sentiments within a corpus. A bigram represents a sequence of two words like "digital pound" frequently appearing together, whereas a trigram refers to a sequence of three words like "central bank digital." This technique has been widely utilised in fields such as linguistics and computational linguistics and, more recently, in financial discourse analysis [229] to gain insights into the underlying themes, concerns, and sentiments expressed within large bodies of text.

Tool Used: For this analysis, the 'CountVectoriser' tool from the '*sklearn.feature_extraction.text'* library was employed to identify the most frequent bigrams and trigrams across the specified periods. The vectorizer was configured to remove stop words (common words such as "and," "the," "of") to ensure that only meaningful word sequences were analysed. This process generated a list of the most common word pairs and triples, sorted by frequency of occurrence.

7.6.1 Bigram Analysis

a) 2020

As illustrated in Table 7.4, bigrams such as "central bank" (62 occurrences) and "digital currency" (46 occurrences) were prevalent, reflecting public and institutional focus on the role of central banks in implementing digital currencies. The term "cash king" appeared 45 times, showcasing public concerns about a CBDC potentially displacing cash. Additionally, the mention of former BoE governor "mark carney" indicates that his public statements were influential in early-stage discussions around government-backed digital currency implementation in the UK.

Bigrams	Count
central bank	62
digital currency	46
Cash king	45

bank digital	27
digital pound	19
Bank digital	19
central banks	20
says mark	16
Pound present	14
Present challenge	14

Table 7.4: Top 10 bigrams (2020).

b) 2023

The strong emphasis on the "digital pound" (1871 occurrences, as shown in Table 7.5) reflects widespread engagement with the BoE's consulting and technology papers. Additionally, "bank england" signals the role of the Bank of England in these discussions.

Bigrams	**Count**
digital pound	1871
bank england	853
central bank	736
digital currency	648
bank digital	469
currency cbdc	158
england treasury	104
pound cbdc	100
Cbdc digital	98
Social credit	89

Table 7.5: Top 10 bigrams (2023).

c) 2024

Table 7.6 shows that the three most frequent bigrams of 2024 include “digital pound” (537 occurrences), followed by “bank england” (182 occurrences) and “privacy concern” (52 occurrences). Notably, public apprehensions about the implications of a digital currency on personal privacy were reflected in this year.

Bigrams	Count
digital pound	537
bank england	182
digital currency	97
central bank	91
bank digital	57
privacy concern	52
england treasury	51
digital cbdc	46
pound cbdc	37
big opportunity	29

Table 7.6: Top 10 bigrams (2024).

7.6.2 Trigram Analysis

a) 2020

As shown in Table 7.7, the prominence of “central bank digital” (26 occurrences) and “bank digital currency” (26 occurrences) underscores the foundational discourse surrounding the introduction and implications of CBDCs. The frequent occurrence of “says mark carney” (15 occurrences) highlights the influential role of the former Governor of the Bank of England, whose statements significantly shaped early-stage discussions around the digital pound.

Trigrams	Count
central bank digital	26

bank digital currency	26
says mark carney	15
digital pound present	14
pound present challenge	14
digital currency cbdc	14
present challenge say	13
challenges say mark	13
japan sweden switzerland	7
canada japan sweden	6

Table 7.7: Trigram analysis (2020).

b) 2023

In 2023, a significant increase in the frequency of "central bank digital" and "bank digital currency" reflects the intensified focus on CBDCs within institutional discussions and policy formulations, as illustrated in Table 7.8. The emergence of "digital currency cbdc" highlights a consolidation of terminology. Additionally, trigrams such as "digital pound cbdc" and "digital pound project" suggest foundational efforts to develop the digital pound, highlighting a transition from conceptual debates in 2020 to actionable initiatives in 2023.

Trigrams	**Count**
central bank digital	463
bank digital currency	456
digital currency cbdc	153
bank england treasury	102
digital pound cbdc	97
digital pound project	77

digital pound foundation	73
new form money	71
digital pound bank	61
digital pound bank	56

Table 7.8: Trigram analysis (2023).

c) 2024

Table 7.9 shows the ongoing discourse on "central bank digital" and "bank digital currency." Trigrams such as "digital pound consultation," digital pound foundation," and "england uk treasury" highlight active stakeholder engagement and interdepartmental collaboration between the Bank of England and the UK Treasury. The presence of "persist digital pound" (16 occurrences) suggests ongoing efforts to address challenges and persist in the digital currency initiatives.

Trigrams	Count
central bank digital	55
bank england treasury	51
bank digital currency	51
digital pound cbdc	37
digital pound consultation	27
Digital currency cbdc	25
digital pound foundation	25
treasury bank england	24
say digital pound	21
persist digital pound	16

Table 7.9: Trigram analysis (2024).

Insights: *2020 marks exploratory discussions on the digital pound, whereas, in 2023 and 2024, the discourse shifted toward implementation and practical considerations, including actionable initiatives like setting up the Digital Pound Foundation. Also, it is evident that the discussions on the digital pound occur within a broader international context, as shown by the mention of "japan sweden switzerland" in 2020. Moreover, the introduction and growing frequency of terms like "privacy concerns" indicate public concerns about the impact of UK CBDC on their personal privacy and identity security. The identification of common bigrams and trigrams (Tables 7.4–7.9) directly addresses the "key topics" part of RQ3 by revealing emerging phrases such as "digital pound" and "privacy concerns."*

7.7 Topic Modelling with LDA and NMF

Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) were employed for topic modelling purposes to uncover the underlying topics within the collected tweets for each timeline. Several Python libraries were utilised for this, including:

- **NLTK**: Used for NLP tasks such as stopword removal, tokenisation, and lemmatisation.
- **Gensim**: Provided the LDA modelling capabilities.
- **Scikit-learn**: Used for NMF modelling and TF-IDF vectorisation.
- **PyLDAvis**: Enabled visualisation of LDA topics.
- **Matplotlib**: Assisted in plotting and visualising results.

Alternative visualisation techniques, such as bar charts were also used to represent the topics derived from the models. In addition to the pre-processing described in Section 3.4, an additional step was performed specifically for NMF implementation. This involved transforming pre-processed text into a TF-IDF weighted document-term matrix using Scikit-learn's '`TfidfVectorizer`' before applying the NMF algorithm. This weighting emphasises words important to specific tweets but less common overall.

7.7.1 Latent Dirichlet Allocation (LDA) Implementation

To ensure optimal topic extraction, the implementation of LDA followed a structured approach:

- **Corpus and dictionary creation**: A dictionary mapping of unique words to their corresponding IDs and a corpus representing the frequency of each word in each document was created using Gensim [230]. Prior to this, the text data was pre-processed by removing stop words using the NLTK stop word list and lemmatising words using WordNetLemmatiser. This preprocessing step is crucial for effective topic modelling.
- **Model training**: The Gensim 'LdaModel' implementation was used. The configuration (optimal) with the highest coherence score was identified by applying LDA to each timeline's dataset independently, experimenting with varying numbers of topics (e.g., 5, 10, 15, 20, 25). The model was trained with a random_state of 42 for reproducibility, and alpha and eta were initially set to 'auto', allowing Gensim to learn these parameters. This

approach leverages Gensim's built-in heuristic optimisation for these hyperparameters and was chosen due to the computational demands of training LDA models on a corpus of 5,702 tweets. This work could be extended by investigating specific alpha and eta values using more exhaustive search methods.

- **Hyperparameter tuning**: To fine-tune the distribution of topics within documents and words within topics, different values for the alpha and eta parameters, including 'symmetric,' 'asymmetric,' 0.01, and 0.1, were explored.
- **Coherence score evaluation**: Gensim's *CoherenceModel* with the 'c_v' metric was used to calculate coherence scores to assess the semantic consistency of the topics, thereby guiding the selection of the optimal number of topics.
- **Topic interpretation and labelling**: Top words from each topic were extracted to facilitate a clear and interpretable representation of the discovered topics, and descriptive labels based on thematic relevance were assigned.
- **Visualisation with PyLDAvis**: Leveraged PyLDAvis to create interactive visualisations of the LDA models; this tool provided a comprehensive overview of the topic landscape, allowing for identifying dominant themes and their interrelationships.

7.7.2 Non-negative Matrix Factorization (NMF) Implementation

NMF was employed to provide a comparative analysis of topic modelling techniques:

- **TF-IDF vectorisation**: Using Scikit-learn's *TfidfVectorizer,* the preprocessed text data was transformed into a TF-IDF weighted document-term matrix, emphasising the importance of significant words while mitigating the impact of frequently occurring terms. The TfidfVectorizer was configured with max_df=0.95, min_df=2, and stop_words='english'. max_df=0.95 was chosen to remove words that appear in nearly all documents (95% or more), as these words are likely to be common across the entire corpus and therefore less informative for distinguishing between topics. min_df=2 was used to exclude words that appear in only one document. Such rare words are often typos or very specific to a single document, and their presence can add noise to the model. Using stop_words='english' removed common English words (e.g., "the," "a," "is") that do not typically carry much semantic weight in topic modelling. These standard preprocessing techniques are commonly employed to improve the quality and interpretability of topic models.
- **Model training**: Applied NMF to the TF-IDF matrix, systematically varying the number of components (topics) and the l1_ratio parameter (0.0, 0.5, 1.0) to control the sparsity and optimise topic discovery. A manual grid search approach was employed and the range of topics was chosen to explore a variety of granularities in the topic structure, from a relatively coarse-grained view (15 topics) to a more fine-grained view (25 topics). The l1_ratio parameter controls the balance between L1 and L2 regularisation, influencing the sparsity of the resulting topic-term matrix. An l1_ratio of 0.0 corresponds to no L1 regularisation (only L2), 0.5 balances L1 and L2, and 1.0 uses only L1 regularisation. Exploring these values allows for an assessment of how sparsity affects topic coherence. The maximum number of iterations was set to 400, which was deemed sufficient for convergence based on initial experimentation.

- **Coherence score evaluation**: Using Gensim's *CoherenceModel*, the semantic coherence of the NMF-derived topics was assessed to ensure that the topics were both meaningful and distinct.
- **Visualisation**: Generated bar charts to visually represent the top 10 words and their associated weights within each topic. These top words were then used to represent and interpret the discovered topics.

7.7.3 Evaluation Metrics

Three evaluation metrics were used to compare LDA vs. NMF model results:

- Coherence score, which measures the semantic similarity of words within topics, indicating interpretability.
- Perplexity (LDA only) that assesses how well the model predicts a sample; lower perplexity suggests better model fit.
- Topic diversity, which measures the proportion of unique words across all topics, reflecting the range of themes captured.

7.7.4 Results and Analysis

The key metrics and extracted topics for both LDA and NMF models across the three years are presented in consolidated tables to facilitate a clear and organised comparison.

7.7.4.1 LDA Model Performance Across All Years

The best hyperparameters for LDA are listed in Table 7.10. Complete hyperparameter tuning results can be found in Appendix 5.

Year	Best hyperparameters
2020	Alpha = 'asymmetric', Eta = 'auto'
2023	Alpha = 'asymmetric', Eta = '0.1'
2024	Alpha = 'asymmetric', Eta = 'auto'

Table 7.10: LDA best hyperparameters across years.

LDA coherence scores can be found in Table 7.11.

Year	Optimal number of topics	Coherence score
2020	20	0.4728
2023	25	0.4556
2024	20	0.4381

Table 7.11: LDA coherence scores across years.

LDA perplexity and topic diversity are presented in Table 7.12.

Year	Perplexity	Topic diversity
2020	-5.5323	0.61
2023	-15.0099	0.83
2024	-11.4171	0.81

Table 7.12: LDA perplexity and topic diversity across years.

For brevity, only the top 5 topics from the optimal number of topics for each year have been included in Tables 7.13, 7.14. And 7.15. Full Tables detailing all topics are available in Appendices 6, 7, and 8 for 2020, 2023, and 2024, respectively. The topics presented in these tables were manually labelled to provide a concise and interpretable summary of the key themes emerging from the LDA models. This manual labelling process involved reviewing the top words associated with each topic and assigning a descriptive label that captured the core concept or theme.

Topic no.	Topic label	Top words
1	European News and Cryptocurrency	health, get, market, world, cbdc, news, boe, issue, dont, offer
2	Future Digital Currencies and Economic Considerations	digital, say, cbdc, mark, carney, risk, currency, monetary, planned, banking
3	CBDC Design and Cash Discussions	cbdc, anonymity, bank, system, time, account, would, deposit, rather, design
4	International Crypto Developments	like, france, many, cbdc, sector, crypto, work, even, way, look
5	CBDCs in Global Central Banks	bank, england, central, potential, digital, paper, currency, cbdc, discussion, government

Table 7.13: LDA topics for 2020.

Topic no.	Topic label	Top words

1	UK Bank and Digital Pound Developments	pound, new, plan, launch, cbdc, via, news, cap, call, briton
2	Britcoin and Leadership in Digital Currency	pound, could, britcoin, stablecoins, would, access, cbdc, foundation, part, cunliffe
3	Policy Concerns and Societal Impact	bitcoin, way, work, find, policy, even, industry, every, live, general
4	Tokens and Crypto Challenges	want, use, one, launched, transaction, tax, many, council, cbdc, bring
5	Blockchain Adoption and Private Sector Involvement	private, know, blockchain, open, here, coexist, recommends, adopting, egbp, mixed

Table 7.14: LDA topics for 2023.

Topic no.	Topic label	Top words
1	Government Adoption and Economic Considerations	cbdc, pound, need, people, time, get, way, coming, know, one
2	BoE, Legislation, and Parliamentary Control	cbdc, plan, government, control, boe, legislation, week, proposed, whether, security
3	CBDC Design, Privacy, and Banking Concerns	pound, privacy, concern, design, future, labour, approach, party, issue, banking
4	Freedom, Economic Control, and Public Opinion	use, still, yet, undecided, freedom, solution, via, economic, used, create
5	UK Digital Pound and Financial Governance	pound, bank, england, treasury, currency, central, consultation, cbdc, privacy, potential

Table 7.15: LDA topics for 2024.

Example: PyLDAvis Visualisation for Topic 5 from the year 2024

To illustrate the interactive exploration capabilities of this tool, a representative PyLDAvis visualisation is presented (Figure 7.6), which shows the distribution, relevance, and relationship of topics generated by an LDA model (specifically highlights Topic 5 for the year 2024), revealing its key terms, prevalence, and relationship to other topics. The left side of the visualisation displays the Intertopic Distance Map, which illustrates the relationships between different topics identified by the LDA model. Each circle represents a topic, and the size of each circle indicates how prevalent that topic is within the dataset; larger circles mean the topic appears more frequently. The proximity of circles reflects the similarity between topics, with closer circles sharing more common terms, while topics farther apart are less related. On the right side, a bar plot shows the Top-30 most relevant terms for the currently selected topic (Topic 5 in this case). The red bars represent the frequency of each term within Topic 5, while the blue bars show the overall frequency of these terms across all topics. This contrast helps distinguish terms uniquely relevant to the selected topic from those more common across the dataset. At the top of this section, there's a relevance metric slider (λ) that adjusts the balance between term frequency and term uniqueness for the selected topic. At the bottom left, a small chart shows the marginal topic distribution, indicating the general prevalence of each topic across the entire dataset. In the visualisation, for example, Topic 5 has a noticeably larger circle, suggesting it is one of the more dominant topics in the dataset. This means that the terms associated with Topic 5 appear more frequently in tweets than terms from smaller circles (other topics), making it a significant theme in the sentiment analysis data. Specifically, it represents 8.3% of the tokens and centres around concerns and discussions regarding the potential impact of a digital pound on various aspects of society and the economy. This is evidenced by the high frequency of terms like 'pound,' 'digital,' 'bank,' 'england,' 'future,' 'labour,' 'approach,' 'party,' 'banking,' and 'concerns.' The prominence of 'bank' and 'england' suggests a focus on the role of the BoE in the development and implementation of the digital pound. The terms 'future,' 'labour,' 'approach,' and 'party' point to discussions about the long-term implications and political perspectives on the digital currency.

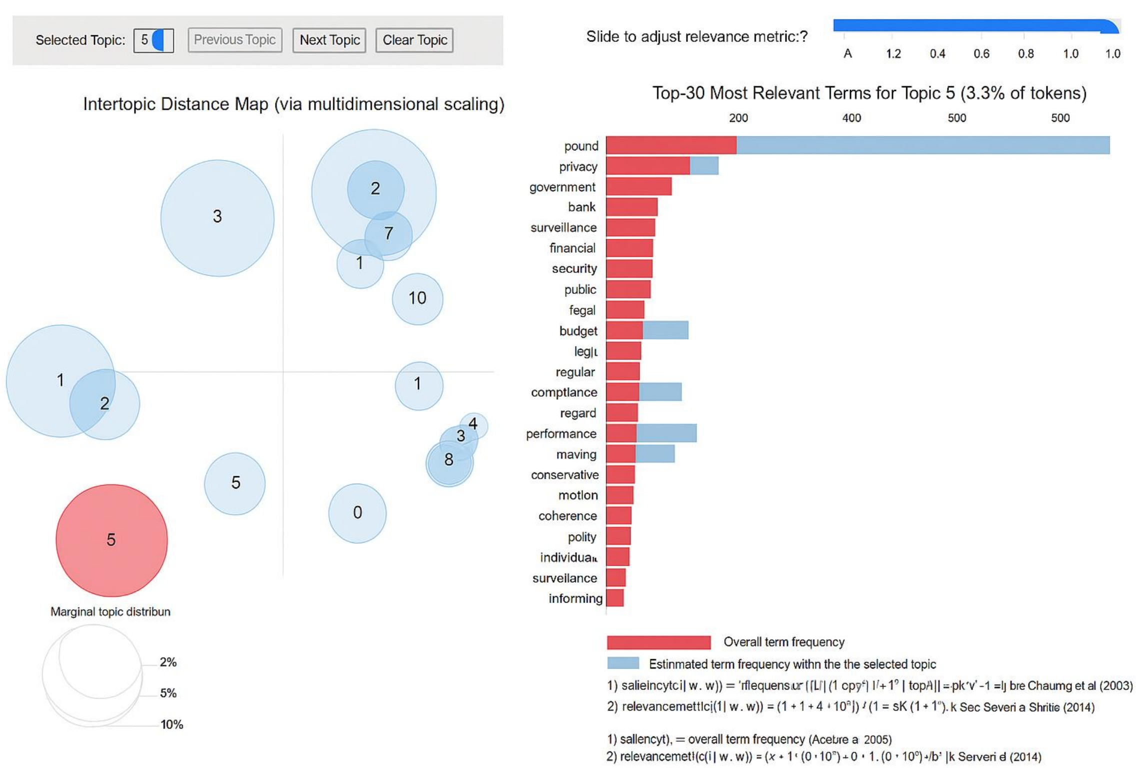


Figure 7.6: PyLDAvis visualisation for topic 5 in 2024.

Insights: *LDA uncovered that digital pound discussions in 2020 were foundational, progressing to regulatory frameworks and privacy concerns by 2024. Most structured thematic clustering was seen in 2020, with a coherence score peaking at 0.4728 in 2020, while as themes broadened in 2023 and 2024, coherence scores started decreasing (0.4556 in 2023) and (0.4381 in 2024), meaning that maintaining topic-specific cohesion became challenging. Moreover, LDA's ability to model complexity in extensive text corpora was observed with the lowest (most negative) perplexity scores in larger datasets (Table 7.12). Hyperparameter tuning, especially with asymmetric alpha, played a key role in optimising coherence, allowing LDA to reveal nuanced, probabilistically-driven themes effectively within CBDC discourse. The asymmetric alpha suggests that certain topics were inherently more prevalent or dominant within the corpus, which aligns with the observed evolution of the discussion from foundational concepts to specific policy and implementation details.*

7.7.4.2 NMF Model Performance Across Years

The best hyperparameters for the NMF model are presented in Table 7.16. Complete hyperparameter tuning results can be found in Appendix 9.

Year	Best hyperparameters
2020	n_components = 15, l1_ratio = 0.0
2023	n_components = 15, l1_ratio = 0.0
2024	n_components = 20, l1_ratio = 0.0

Table 7.16: NMF best hyperparameters across all years.

Table 7.17 shows NMF optimal topics, topic diversity and coherence scores:

Year	Optimal number of topics	Topic diversity	Coherence score
2020	15	0.86	0.7616
2023	15	0.8	0.5739
2024	20	0.755	0.5046

Table 7.17: Optimal no. of topics, topic diversity and coherence scores across each year.

For brevity, only the top 5 topics from the optimal number of topics for each year have been included in Tables 7.18, 7.19, and 7.20. Full Tables detailing all topics are available in Appendices 10, 11, and 12 for 2020, 2023, and 2024, respectively. The topic label was generated manually based on top words, as presented in Section 7.7.4.1.

Topic no.	Topic label	Top words
1	CBDC-Based Retail Systems	based, retail, central, allow, similar, token, manner, ecb, circulate, anonymity
2	Cash and Economic Flow	cash, king, market, profit, economy, pandemic, flow, care, world, going
3	Challenges and Leadership in Digital Currency	present, challenge, mark, say, carney, pound, digital, bitcoin, crypto, coindesk
4	Digital Pound Services	pound, digital, service, year, billion, new, multimillion, british, health, money
5	CBDC Anonymity and Payments	cbdc, anonymity, payment, account, need, like, deposit, idea, anonymous, look

Table 7.18: NMF topics for 2020.

Topic no.	Topic label	Top words
1	Bank of England's Plans	england, bank, plan, briton, cap, face, treasury, governor, deputy, create
2	Support and Project Needs	need, likely, treasury, support, say, project, england, bank, pound, cbdc
3	Digital Pound Foundations	pound, digital, foundation, limit, britcoin, consumer, case, pay, mean, work
4	Programmable CBDCs and Privacy	cbdc, programmable, petition, coming, anonymity, introduction, year, implement, prevent, soon

5	Everyday Digital Payments	new, form, household, money, business, payment, digital, everyday, pound, news

Table 7.19: NMF topics for 2023.

Figure 7.7 shows a representative bar chart for a randomly chosen topic (Topic 10) from the 2023 NMF model. This example is provided for context and to avoid adding numerous topic visualisations. It reveals public concerns in 2023 about privacy in the context of the Bank of England's work on CBDCs, as analysed through Topic 10 in an NMF model. Key terms like "privacy," "boe," and "pseudonymous" indicate a focus on how CBDC technology might impact user anonymity and security, highlighting apprehension about the balance between transparency and privacy in digital currencies.

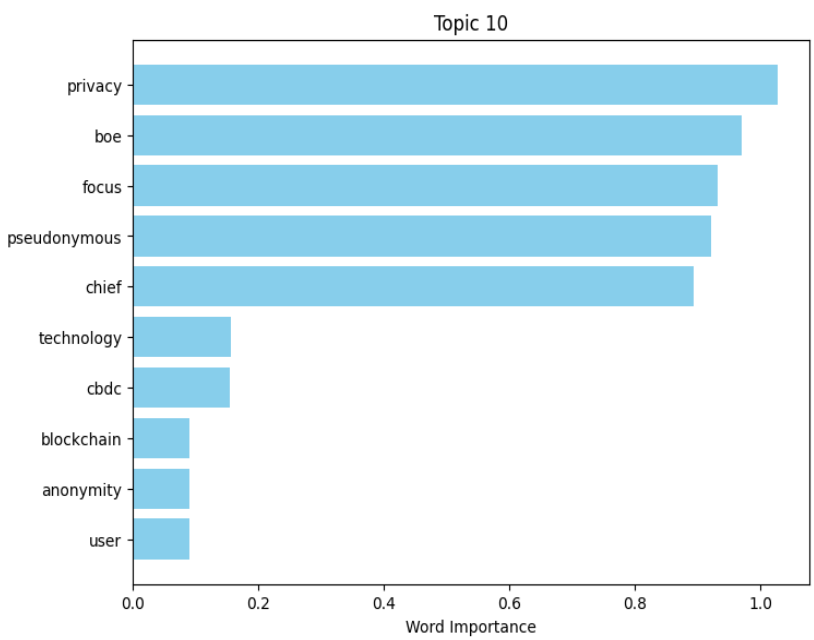


Figure 7.7: Bar chart for topic 10 captured by NMF in 2023.

Topic no.	Topic label	Top words
1	Future of the Digital Pound	pound, digital, foundation, future, decision, money, working, launch, legislation, finance
2	Opportunities and Official Statements	opportunity, big, present, england, bank, cbdc, pound, digital, say, breaking
3	Advancements and	feasibility, advancing, exploring, cbdc, news, crypto,

	Stability	january, cryptocurrency, britcoin, stability
4	Public Opinion and Freedom	wef, people, net, zero, vote, cbdc, reform, party, freedom, cbdcs
5	Privacy and Progress	persist, worry, progress, plan, privacy, cbdc, pound, digital, learn, feed

Table 7.20: NMF topics for 2024.

Insights: *NMF's strength in producing semantically unified topics can be observed from high coherence scores across all years; NMF's straightforward hyperparameter tuning enabled efficient extraction of specific, thematically distinct topics, with topic diversity scores slightly decreasing over time as core themes solidified. Moreover, the coherence score in 2020 (0.7616) reflects its ability to capture clear, structured themes, whereas in 2023 (0.5739) and 2024 (0.5046), the coherence scores decreased as thematic variety and text volume expanded. The consistent selection of l1_ratio = 0.0 across all years indicates that L2 regularisation was sufficient for these models. Also, document-level variations were adapted well by NMF due to its matrix factorisation approach, making it a reliable method for consistent, high-coherence topic modelling in digital finance (in this case, CBDCs) data.*

7.7.5 Model comparison: LDA vs. NMF

Table 7.21 shows perplexity, topic diversity, and coherence score comparison between LDA and NMF across three timelines.

Year	**LDA perplexity**	**LDA topic diversity**	**NMF topic diversity**	**LDA coherence score**	**NMF coherence score**	**Superior model**
2020	-5.5323	0.61	0.86	0.4728	0.7616	NMF
2023	-15.0099	0.83	0.8	0.4556	0.5739	NMF
2024	-11.4171	0.81	0.755	0.4381	0.5046	NMF

Table 7.21: Perplexity, topic diversity, and coherence score comparison between LDA and NMF.

7.7.5.1 Comparative Insights

The analysis reveals a preference for NMF over LDA in this study. LDA's coherence scores were consistently lower than NMF's, suggesting that LDA topics were less semantically cohesive. In contrast, NMF consistently achieved higher coherence scores across all years, producing more semantically meaningful and interpretable topics. This aligns with NMF's known strength in identifying distinct, unified themes [93]. While LDA exhibited decreasing perplexity (better statistical fit) with increasing dataset size, particularly in 2023, this did not translate to higher coherence. This highlights a crucial point: in sentiment analysis, a better statistical fit in terms of likelihood does not necessarily translate to more interpretable topics, i.e., lower perplexity did not correlate with higher coherence. To address RQ3 – understanding public sentiment and its evolution – semantic interpretability is paramount. Therefore, NMF's superior coherence makes it a more suitable model for capturing and modelling UK CBDC sentiments.

Regarding diversity, NMF showed higher diversity in 2020 and decreased slightly in later years, whereas LDA exhibited higher diversity in 2023 and 2024. However, this higher diversity in LDA did not compensate for its lower coherence, as the additional topics may have been less semantically consistent. Given its consistent ability to produce coherent and interpretable topics, NMF has proved to be better than LDA for capturing and modelling UK CBDC sentiments.

The topic modelling results, particularly from the NMF models, reveal prevalent themes of government control, privacy concerns, and regulatory challenges. These themes are consistent with emotion analysis findings (Section 7.5), which showed a rise in fear and anger, emotions linked to surveillance and loss of privacy. The n-gram analysis (discussed in Section 7.6), which identified frequent phrases like "privacy concerns" and "government control," provides further corroborating evidence. This convergence of findings across multiple analytical methods strengthens the validity of the identified themes and provides a robust and multifaceted understanding of public discourse surrounding the digital pound. This contributes substantially to RQ3 by providing a quantitative and qualitative mapping of themes that evolve over time, specifically demonstrating the shift from exploratory discussions to concrete concerns regarding privacy and regulation. The analysis reveals a clear pattern: as discussions around the digital pound progress, public anxieties regarding privacy and government control become increasingly prominent.

7.8 Word Embeddings and Clustering

To gain deeper insights into public discourse on the digital pound, two clustering methods, as discussed below, were applied to X datasets across all years. These methods provide valuable insights into the thematic structure of tweets by grouping semantically similar tweets and identifying prevalent topics. By revealing latent themes and their evolution over time, this analysis complements sentiment and n-gram analyses (Section 7.6) and offers a deeper understanding of public concerns and opinions.

Word2Vec Embeddings with KMeans Clustering: It represents tweets in a high-dimensional vector space based on word contexts, capturing semantic relationships between words.

- **Word2Vec model**: Trained on tokenized tweets to generate word embeddings that capture semantic meanings.
- **Tweet vectorisation**: Each tweet is represented by averaging the embeddings of its constituent words.
- **KMeans clustering**: Applied to the tweet vectors to group semantically similar tweets into five clusters.

TF-IDF Vectorization with KMeans Clustering: This method emphasises important words in the corpus by weighting terms based on their frequency and inverse document frequency.

- **TF-IDF vectorisation**: Converts tweets into feature vectors where each word's weight reflects its importance in the tweet relative to the corpus.
- **Optimal cluster determination**: The elbow method was used to identify the optimal number of clusters (k=5). Figure 7.17 shows a representative elbow plot for 2023. The same process was used for the other two years.
- **KMeans clustering**: Applied to the TF-IDF vectors to cluster tweets based on term importance.

Figure 7.8 shows a "bend" or "elbow" point where adding more clusters does not significantly reduce inertia. The x-axis represents the number of clusters (k), and the y-axis shows the sum of squared distances (inertia), which measures how tightly the data points in each cluster are grouped. In this plot, the elbow appears around k = 5, suggesting that 5 clusters are an optimal choice.

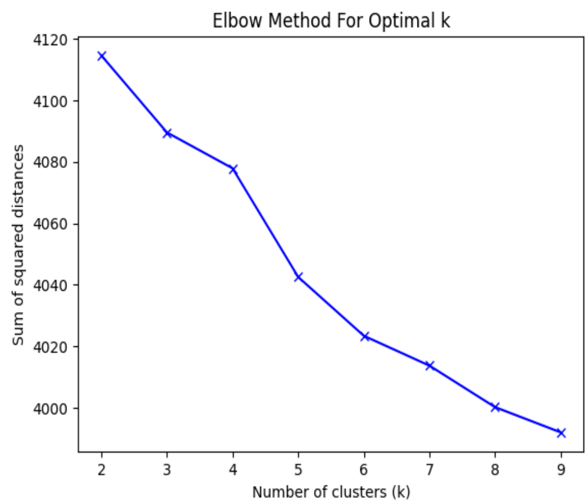


Figure 7.8: Elbow plot for the year 2023.

7.8.1 Result and Key Insights

Tables 7.22, 7.23, and 7.24 show clusters identified using both methods, including tweet excerpts for Word2Vec + KMeans method. Notably, TF-IDF clusters are often driven by specific words rather than overall tweet meaning. Consequently, TF-IDF clusters tend to have top words rather than excerpts because they group tweets based on prominent terms rather than deeper semantic similarity.

Cluster no.	Word2Vec + KMeans	Tweet excerpts for Word2Vec + KMeans	TF-IDF + KMeans

Cluster 0	International central bank collaborations and CBDC design	"ecb mention anonymity potential central bank..." "central bank switzerland sweden canada japan..."	bank, central, currency, digital, cbdc, sweden
Cluster 1	Digital tools and economic initiatives	"either cbdc form deposit account..." "landmark deal provide access digital tool..."	cash, king, economy, market, profit, flow
Cluster 2	Economic concerns and the role of cash	"cash king asmussen issue bigmoney..." "may already guessed economy screwed..."	coinbase, cbdcs, future, ceo, financial, benefit
Cluster 3	Financial security and cash preservation	"bust anyway passenger becomes unsecured..." "cash king plc"	cbdc, digital, pound, china, anonymity, bank
Cluster 4	Digital pound prospects and challenges	"digital monetary institute grows cddc..." "digital pound could present challenge say mark..."	mark, carney, say, present, challenge, pound

Table 7.22: Clusters identified for 2020.

Cluster no.	**Word2Vec + KMeans**	**Tweet excerpts for Word2Vec + KMeans**	**TF-IDF + KMeans**
Cluster 0	Public consultation and awareness efforts	"thought might helpful flag time running..." "consultation creating digital pound bank england..."	cash, king, digital, cbdc, replace, pound

Cluster 1	Technological innovations and critiques	"thinking thing heading direction billionaire..." "gbp ethereum spur innovation..."	likely, treasury, support, say, need, project
Cluster 2	Official announcements and privacy concerns	"head central bank england rule need launch..." "bank england treasury support digital pound..."	cbdc, digital, pound, cbdcs, control, bank
Cluster 3	Regulatory challenges and inclusion	"sec play rumsfeld speech sound like..." "treasury bank england announced statebacked digital..."	pound, digital, england, bank, new, consultation
Cluster 4	Scepticism towards CBDCs	"spot central bank digital currency would enormous..." "central bank digital currency slavery"	central, bank, currency, digital, cbdc, pound

Table 7.23: Clusters identified for 2023.

Cluster no.	**Word2Vec + KMeans**	**Tweet excerpts for Word2Vec + KMeans**	**TF-IDF + KMeans**
Cluster 0	Official positions and privacy assurances	"yeah digital pound tracked currency people expect..." "bank england official say digital pound provide..."	people, digital, cbdc, financial, freedom, plan

Cluster 1	Global control concerns and conspiracy theories	"need look deeper cbdc trial philippine..." "won't happen million died world war freedom..."	opportunity, big, present, england, say, pound
Cluster 2	Geopolitical and economic pessimism	"continues charge germany turn nuclear power..." "everything worldwide allowed cause problem..."	labour, party, want, tokenization, hub, digital
Cluster 3	Business opportunities and economic growth	"bank england say digital pound cbdc present big..." "digital pound cbdc big opportunity business..."	cbdc, digital, bank, pound, government, like
Cluster 4	Privacy concerns and government surveillance	"govt focus privacy control digital pound..." "cbdc linked carbon credit limit way they've..."	pound, digital, privacy, bank, treasury, england

Table 7.24: Clusters identified for 2024.

Combining both methods provides a comprehensive understanding of the digital pound discourse. Word2Vec clusters reveal the context and sentiment behind the topics, while TF-IDF clusters pinpoint the prevalence of specific terms. Moreover, the consistent identification of core themes across both methods strengthens the reliability of the findings. The following prominent themes emerged from the analysis:

- **Official announcements and support:** Discussions around official positions and support for the digital pound are evident across all timelines, particularly in 2020 (Cluster 0), 2023 (Cluster 2), and 2024 (Cluster 0) using Word2Vec, and through frequent mentions of "bank," "england," "treasury," and "governor" in TF-IDF clusters.
- **Privacy and anonymity concerns:** Privacy consistently ranks as a major concern, appearing prominently in Word2Vec clusters across all years, notably in 2020 (Cluster 3)

and 2024 (Clusters 0 and 4). TF-IDF clusters reinforce this with frequent appearances of "privacy," "anonymity," and "concern," especially in 2024 (Cluster 4).

- **Economic stability and cash preference:** Concerns about economic stability and a preference for cash are highlighted in Word2Vec clusters for 2020 (Clusters 2 and 3). TF-IDF clusters corroborate this with "cash," "king," and "economy" as top terms in 2020 (Cluster 1) and 2023 (Cluster 0).
- **Government control and Scepticism:** Scepticism towards government control is apparent in Word2Vec clusters for 2023 (Cluster 4) and 2024 (Cluster 1). TF-IDF clusters echo this with terms like "control," "government," and "freedom" in 2023 (Cluster 2) and 2024 (Cluster 0).
- **Technological innovation and opportunities:** Discussions of technological innovation and opportunities related to the digital pound are present in Word2Vec clusters for 2023 (Cluster 1) and 2024 (Cluster 3). TF-IDF clusters support this with terms like "innovation," "blockchain," and "opportunity" in 2023 (Clusters 1 and 4) and 2024 (Cluster 1).

Insights: *Privacy and anonymity concerns, government control, and economic stability were found to be the consistent themes in the clustering analysis, and these findings echo insights yielded from the sentiment distribution (Section 7.3.1) and emotion analysis (Section 7.5), which showed increasing fear and anger, corresponding with clusters related to concerns over government surveillance and skepticism toward CBDCs. The consistent and significant themes in the public discourse on the digital pound suggest that addressing "privacy and government control" concerns is crucial for public acceptance of the digital pound. Additionally, the need for clear communication about the benefits and security of CBDCs is highlighted by the preference for cash and concerns about economic stability. However, potential support from tech-savvy demographics and industry stakeholders is evident from positive sentiments toward innovation. Thus, from a policy development point of view, to alleviate public fears, there is a need to incorporate strong privacy protections, transparent governance and enhanced communication strategies to educate the public about CBDCs. Finally, continued monitoring of public discourse to adapt policies and communication strategies is crucial. The clustering analysis further addresses RQ3 by illustrating how semantic relationships and thematic clusters evolve over the chosen study periods.*

7.9 Semantic Network Analysis

Semantic network analysis explored relationships between key terms in digital pound discourse, revealing the structure and focus of conversations. It was performed by vectorizing cleaned tweets using CountVectoriser to create a co-occurrence matrix of word pairs. This matrix was used to construct a network graph with words as nodes and co-occurrence frequencies as edge weights. A threshold was applied to the edge weights to retain only the most significant relationships for visualisation and analysis.

7.9.1 Result and Key Insights

As observed, in 2020, high co-occurrence between "bank" and "central" indicates discussions centred on central banks' roles in digital currencies (Table 7.25, Figure 7.9). Moreover, pairs like "bank – digital" and "digital – pound" suggest early conversations about exploring CBDCs and digitising national currency. The frequent pairing of "bank" with "cbdc" and "currency" reflects an interest in how traditional banking institutions are approaching digital transformations. Similarly, Figure 7.9 shows that in 2020, discussions around the digital pound were exploratory, as observed from the clusters around terms like "economy," "blockchain," "reserve," and "launch." Each node in the semantic network represents a frequently mentioned term, and the edges show co-occurrences, reflecting early-stage conversations on the foundational concepts and possibilities of the digital pound, including potential design, implications, and feasibility of CBDCs. This aligns with the 2020 sentiment analysis, which revealed a generally optimistic and exploratory tone (Section 7.2.1).

2020-word pair	**Count**	**2023-word pair**	**Count**	**2024-word pair**	**Count**
bank – central	109	digital – pound	2911	digital – pound	1365
central – bank	109	pound – digital	2911	pound – digital	1365
bank – digital	90	bank – digital	2285	bank – digital	813
digital – bank	90	digital – bank	2285	digital – bank	813
bank – cbdc	77	cbdc – digital	1344	bank – pound	564
cbdc – bank	77	digital – cbdc	1344	pound – bank	564
digital – pound	70	currency – digital	1262	cbdc – digital	483
pound – digital	70	digital – currency	1262	digital – cbdc	483
bank – currency	69	bank – pound	1191	digital – privacy	458
currency – bank	69	pound – bank	1191	privacy – digital	458

Table 7.25: Top 10 Co-occurring word pairs for each year.

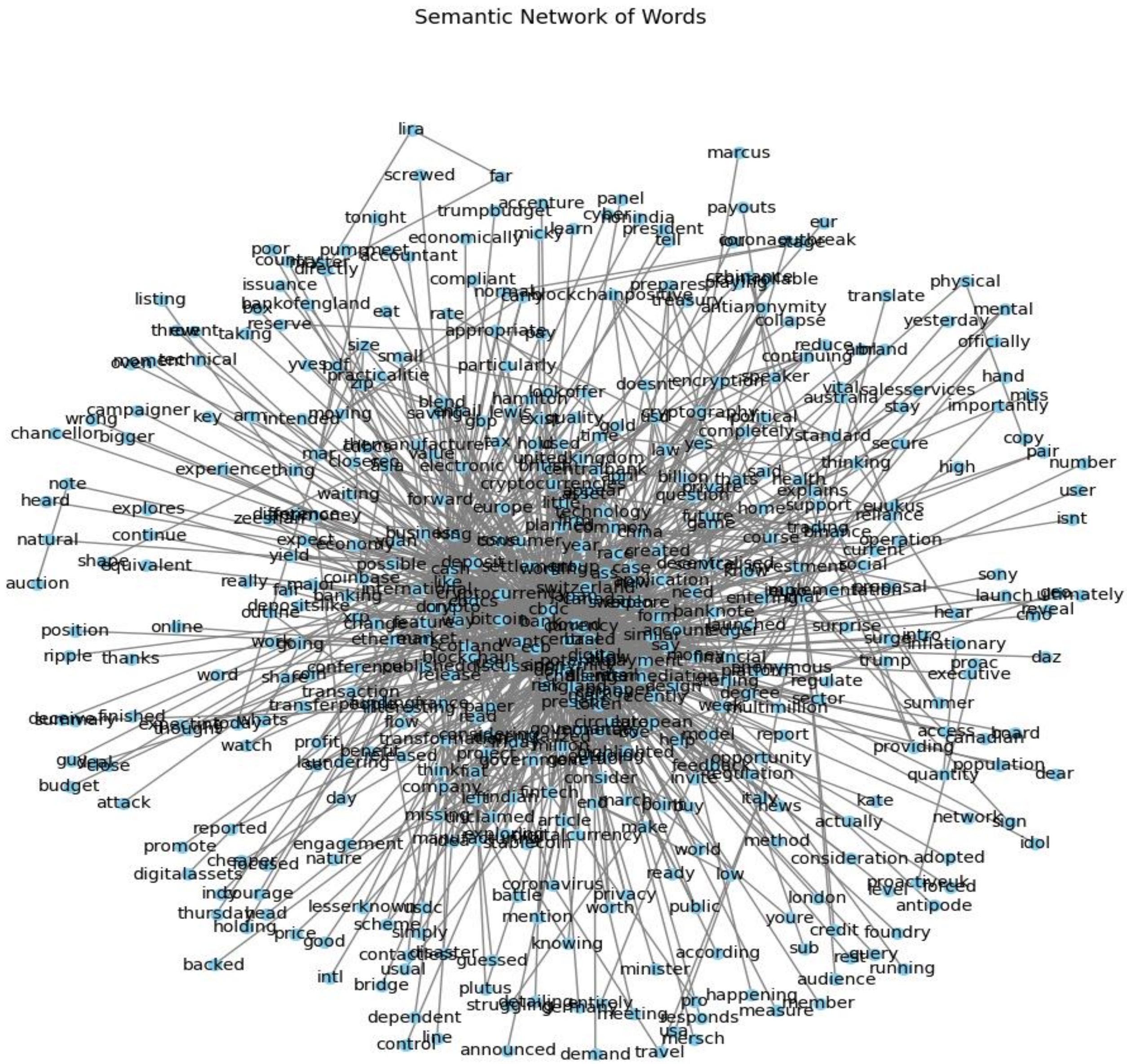


Figure 7.9: Semantic network of words for 2020 timeline.

By 2023, there was a significant increase in “digital–pound” co-occurrences, highlighting intensified public discourse on the UK CBDC, possibly due to official announcements regarding consultation and technology papers. Furthermore, persistent high co-occurrence between “bank” and “digital” suggests ongoing discussions about the role of banks in the digital currency landscape and pairs like “currency – digital” indicate expanding conversations to include digital currencies in general or perhaps CBDCs being explored by other countries, not just the pound. This increased activity mirrors the surge in tweets observed in the data preprocessing stage (in Chapter 3) and the shift towards more sceptical sentiment identified in the sentiment analysis.

While "digital – pound" remains the top pair in 2024, the introduction of "digital – privacy" into the top co-occurring pairs signals growing public concern over privacy issues. Similarly, the prominence of "privacy – digital" reflects increasing apprehension about data surveillance and security related to digital currencies as discussions around the digital pound expand. This heightened focus on privacy aligns strongly with the emotion analysis (Section 7.5), which showed increased fear and anger in 2024, as well as the n-gram analysis (Section 7.6), which identified phrases like "privacy concerns" as increasingly prevalent. Continued co-occurrence of "bank" with "digital" and "pound," alongside rising "privacy" mentions, suggests that discussions are now balancing institutional roles with privacy considerations. This analysis contributes to RQ3 by revealing changes in semantic relationships over time, such as the increased focus on "digital – privacy" in 2024. The increasing co-occurrence of "privacy" with "digital" and "pound" specifically underscores the growing public apprehension about data surveillance and security related to digital currencies as discussions around the digital pound progress.

7.10 Polarity and Subjectivity Analysis

This section explores polarity and subjectivity dimensions of public discourse on the digital pound. Understanding the degree of sentiment (polarity) and the certainty with which it is expressed (subjectivity) provides valuable additional insights beyond categorising sentiment into positive, negative, or neutral.

Two methods were employed to calculate polarity and subjectivity scores:

- **TextBlob:** TextBlob, a lexicon-based sentiment analysis tool, was used to obtain polarity and subjectivity scores directly from its built-in sentiment analysis function. However, as discussed in Chapter 2, TextBlob's reliance on pre-defined lexicons restricts its adaptability to context-specific nuances [52].
- **RoBERTa:** A fine-tuned RoBERTa model, validated for its robustness in Chapter 6, was also utilised to provide a more nuanced analysis. For RoBERTa, probabilities for each sentiment class were derived, and these probabilities were then used to calculate polarity (Positive Score - Negative Score) and subjectivity (1 - Neutral Score) scores.

For both methods, mean, median, and range were calculated. Finally, the distribution and correlation of the polarity and subjectivity scores were visualised using histograms and scatter plots.

7.10.1 TextBlob Results

Table 7.26 presents the polarity and subjectivity statistics derived from TextBlob.

Year	Mean Polarity	Median Polarity	Polarity Range	Mean Subjectivit	Median Subjectivit	Subjectivit y Range	Correlatio n

				y	**y**		
2020	0.04	0	-0.80 to 1.00	0.29	0.25	0.00 to 1.00	0.16
2023	0.04	0	-1.00 to 1.00	0.26	0.23	0.00 to 1.00	0.22
2024	0.04	0	-0.80 to 1.00	0.26	0.23	0.00 to 1.00	0.17

Table 7.26: TextBlob Results.

Insights: *The polarity range spans from negative to positive, but the mean remains slightly positive, as indicated by mean polarity scores of 0.04 across all years. In contrast, the median polarity is 0.00, suggesting that most tweets are neutral. Mean subjectivity scores, ranging from 0.26 to 0.29, indicate that tweets are generally more objective than subjective. Median subjectivity scores reinforce this objectivity as they are close to the mean. Furthermore, the weak positive correlation between polarity and subjectivity (0.16 to 0.22) indicates a slight tendency for more subjective tweets to lean positive. However, the overall picture painted by TextBlob is one of neutrality, with limited shifts in sentiment or subjectivity over time.*

7.10.2 RoBERTA Results

Table 7.27 presents the polarity and subjectivity statistics derived from RoBERTa.

Year	Mean Polarity	Median Polarity	Polarity Range	Mean Subjectivit y	Median Subjectivit y	Subjectivit y Range	Correlatio n
2020	-0.11	0.02	-0.99 to 0.98	0.49	0.41	0.03 to 1.00	-0.34
2023	-0.24	-0.01	-0.99 to 0.99	0.52	0.48	0.02 to 1.00	-0.54
2024	-0.3	-0.16	-0.99 to 0.99	0.61	0.78	0.02 to 1.00	-0.54

Table 7.27: RoBERTa Results.

Insights: *Polarity scores indicate an increasing negative sentiment toward the digital pound and CBDCs, as indicated by mean polarity scores from -0.11 in 2020 to -0.30 in 2024. Similarly, median polarity shifts from slightly positive (0.02) in 2020 to more negative values in subsequent years. Mean subjectivity scores increase from 0.49 to 0.61 over time, suggesting tweets are becoming more subjective and median subjectivity rises significantly to 0.78 in 2024, indicating a higher prevalence of personal opinions. The correlation between polarity and subjectivity (-0.34 to -0.54) shows that tweets become more polarised (positive or negative) as they tend to be more objective* *(but in this case, predominantly negative). These findings are consistent with the trends observed in the sentiment analysis (Section 7.2.1), emotion analysis (Section 7.5), and n-gram analysis (Section 7.6) presented earlier, particularly the increasing negativity and focus on privacy concerns.*

7.10.3 Visualisations

7.10.3.1 Polarity Distribution Comparison

Figure 7.10 shows that TextBlob consistently shows a neutral polarity around 0, while Roberta captures a slight negative skew, especially in 2023 and 2024. This difference likely stems from RoBERTa's fine-tuning on CBDC-related data, making it more sensitive to the subtle negative sentiments prevalent in discussions about the digital pound. In contrast, TextBlob misses such nuances, likely due to its simpler model structure. Given the sentiment-laden discussions on the digital pound, RoBERTa is more suitable for capturing complex sentiment. This divergence in polarity detection between the two models further underscores the importance of using domain-specific models like RoBERTa, as highlighted in Chapters 2 and 6.

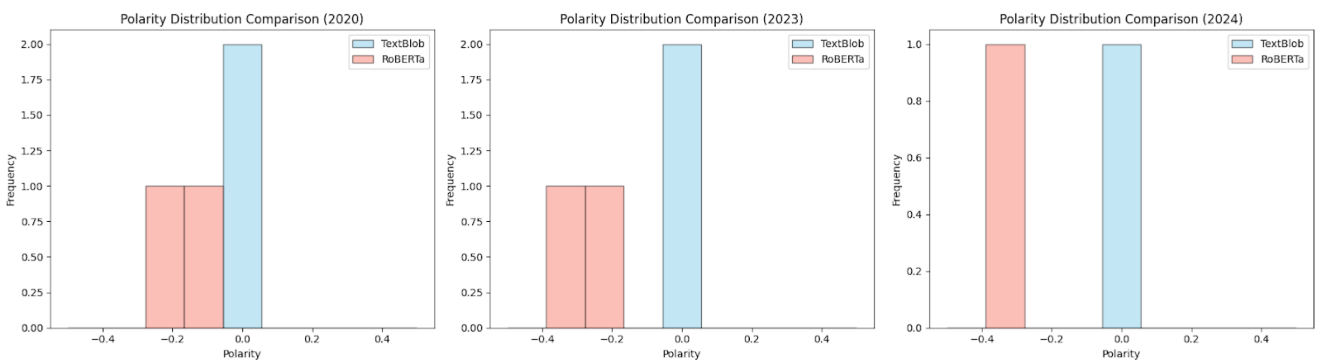


Figure 7.10: Polarity distribution comparison across all years.

7.10.3.2 Subjectivity Distribution Comparison

Figure 7.11 shows that TextBlob indicates more objective assessment as it maintains lower subjectivity levels, centred around 0.2, whereas RoBERTa consistently captures higher subjectivity values. This suggests that TextBlob's more objective assessments might miss capturing subjective language and nuanced sentiment (rising subjectivity evident in social media discussions), making RoBERTa a better model to assess public opinion and complex discussions around the digital pound.

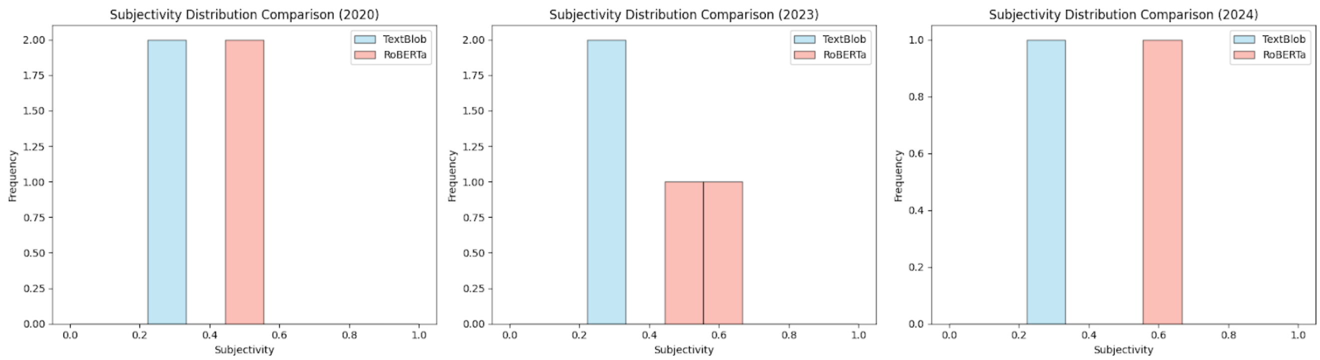


Figure 7.11: Subjectivity distribution comparison across all years.

7.10.3.3 Correlation Scatter Plots

Figure 7.12 shows RoBERTa's strong negative correlation, suggesting that it is capable of detecting nuanced negative sentiments associated with subjective content more effectively, which may provide deeper insights into public sentiment on complex issues like the digital pound. On the contrary, TextBlob's weak correlation and tendency toward neutrality may make it less suitable for capturing these nuanced domain-specific relationships.

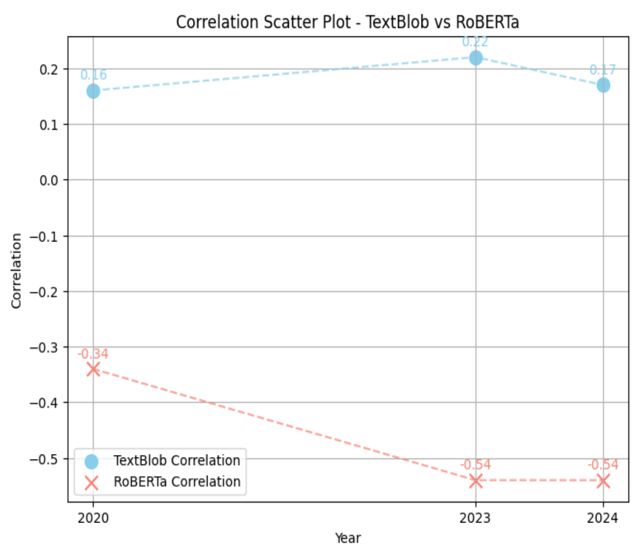


Figure 7.12: Correlation comparison across all years.

7.10.3.4 Key Insights on Public Discourse

The comparative analysis, which reveals RoBERTa's effectiveness in identifying increasing negative sentiments, as evidenced by the rising mean polarity score from -0.11 to -0.30), echoes the trends identified in Section 6.4.2. Specifically, the increasing negativity observed here aligns with the growing difficulty annotators faced in reaching consensus, suggesting a potential rise in polarized and complex expressions within the discourse.

However, TextBlob underestimated negative sentiments as shown by slightly positive mean polarity. Moreover, the increase in mean subjectivity from 0.49 to 0.61 suggests that tweets began becoming more opinionated as discussions around the digital pound expanded, possibly due to the release of Consultation and Technology papers [11], [12]. The visualisations (Figure 7.19-7.21) reinforce the findings from the earlier sentiment analysis (Section 7.2), emotion analysis (Section 7.5), n-gram analysis (Section 7.6), and topic modelling (Section 7.7), demonstrating the increasing negativity and rising subjectivity in public discourse as the digital pound discussions evolved. These insights also address RQ2 by demonstrating the superior performance of a fine-tuned RoBERTa model in this specialised domain.

Overall, relying on TextBlob could lead to incomplete or misinformed conclusions about public sentiment, while RoBERTa's contextual understanding could help stakeholders interpret complex language, slang, and sarcasm, leading to more accurate sentiment analysis. These insights provide strong evidence supporting the argument for the use of advanced NLP models for sentiment analysis in specialised domains like CBDCs, as explored in RQ2, and reveals initial temporal trends, laying the groundwork for Chapter 8's in-depth analysis of their evolution in addressing RQ3. Moreover, the comparative polarity analysis directly contributes to RQ3 by quantifying the shift toward more negative sentiment (from a mean of –0.11 in 2020 to –0.30 in 2024) and increased subjectivity over time.

7.11 Analysis of Sentiment Association with Predefined Aspects

This section explores the association between overall tweet sentiment and a set of predefined aspects related to the digital pound and CBDCs. While not a true aspect-based sentiment analysis (ABSA), which would analyse sentiment *within the context* of each aspect [231], this analysis provides a valuable high-level overview of how general sentiment aligns with key themes of public discourse. It complements the broader analyses presented in this chapter by offering a focused perspective on sentiment related to specific areas of concern or interest. The aspects selected for this analysis are Privacy, Anonymity, Stability, Surveillance, Regulation, Security, and Innovation. These aspects were chosen based on their relevance to CBDCs, and supporting evidence from the other analyses presented in this chapter.

- **Privacy:** Privacy is important because digital transactions increase the risk of data collection and possibly violate individual privacy rights. Topic modelling identified privacy as a key theme, and the emotion analysis revealed increased fear and anger, possibly linked to privacy issues. The N-gram analysis (Section 7.6) also showed frequent mentions of "privacy concerns."
- **Anonymity:** It represents a feature of cash transactions where one can conduct transactions without revealing their identity, but this kind of feature could be potentially compromised in digital systems. Word embeddings and clustering (Section 7.8) concerns over loss of anonymity.
- **Stability:** The introduction of a CBDC could impact monetary policy and other aspects of the banking system; thus, financial stability is crucial; there is apprehension that CBDCs could restrict users from transacting without the risk of enabling enhanced government surveillance. Topics related to "economic considerations" emerged in topic modelling (Section 7.7), making them crucial.

- **Surveillance:** Effective regulatory frameworks are essential to ensure that CBDCs do not violate privacy rights and prevent misuse of central bank-backed digital currency. This is important because emotion analysis highlighted fear associated with government control, and semantic network analysis (Section 7.9) showed increased co-occurrence of terms related to surveillance.
- **Security:** Cybersecurity threats pose significant risks to digital currencies; thus, security is paramount to protect against hacking and unauthorised access. Semantic network analysis (Section 7.9) revealed pairs like "digital – privacy" and "privacy – digital" appearing frequently.
- **Innovation:** Innovation represents the positive potential of CBDCs and highlights their potential for economic growth through technological advancement. Clustering analysis (Section 7.8) revealed groups discussing technological innovations and opportunities despite negative sentiments around the digital pound.

The following steps were undertaken to perform the analysis:

- **Aspect identification**: The chosen aspects were defined and used to detect mentions within tweets.
- **Sentiment association**: For each tweet, if any keyword associated with a specific aspect was present (case-insensitive), the *overall* sentiment of the tweet (as determined by RoBERTa) was *associated* with that aspect. If no keyword for a given aspect was found, the association for that aspect in the tweet was recorded as 'neutral.' This approach, while computationally efficient, assumes that the overall sentiment of the tweet is indicative of the sentiment related to each aspect mentioned. Therefore, this analysis should be interpreted as showing *broad trends* in the *association* of overall sentiment with predefined aspects, rather than precise sentiment towards each individual aspect. It provides a high-level overview of which aspects tend to co-occur with positive, negative, or neutral sentiment in the tweets. In future research, this limitation could be addressed by applying ABSA, which extracts aspects from the tweets and then performs further analysis [231].
- **Data processing and sentiment counting**: This process was applied to the datasets for each timeline, and the results were stored in new columns corresponding to each aspect's sentiment. Then, sentiments were counted for the number of positive, negative, and neutral sentiments for each aspect across all tweets.
- **Visualisation**: Radar charts were created to visualise each year's sentiment distribution across different aspects.

7.11.1 Results of Sentiment Association Analysis

Figure 7.13 presents a visualisation of the association between overall tweet sentiment and predefined aspects for 2020, 2023, and 2024. A radar chart (or spider chart) is a type of chart used to show how different items compare across multiple categories [232]. It looks like a web or spider's web where each category is represented by a spoke or axis radiating from a central point, with equal spacing between them and values for each category are plotted along these axes and connected to form a polygonal shape (a larger spread toward certain axes indicates higher values in those categories). Importantly, this analysis reflects how *overall* tweet sentiment aligns with the *presence* of aspect keywords, not necessarily sentiment *towards* each individual aspect.

Figure 7.13 shows sentiment association with aspects for 2020 (green line), 2023 (blue line), and 2024 (green line). Visually, most aspects appear to register values below approximately 500. It indicates early-stage discussions or less evolved concerns in 2020, as indicated by the smallest values across all aspects. This aligns with the n-gram and topic modelling analyses (Sections 7.6 and 7.7), which indicated that discussions were primarily centred on foundational CBDC concepts.

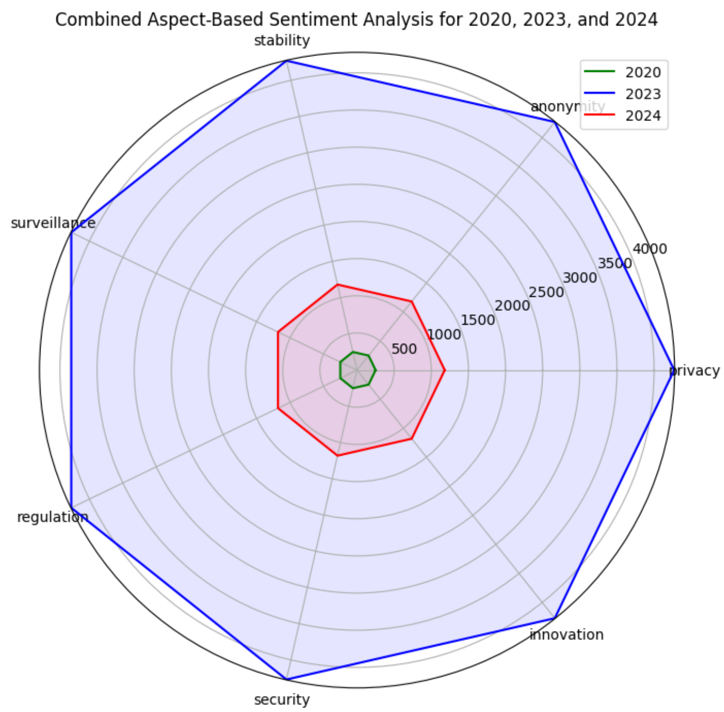


Figure 7.13: Combined Radar Chart of Aspect-Based Sentiment Analysis for 2020, 2023, and 2024.

There is a noticeable increase in the strength of association across most aspects in 2023. Specifically, it exhibits the largest values across 'privacy,' ‘anonymity,' ‘surveillance,' and ‘regulation' aspects, suggesting that with the introduction of Consultation and Technology papers [11], [12], public discussions were extended to privacy and regulatory implications of the digital pound. The prominence of ‘surveillance’ and ‘privacy’ indicates heightened sensitivity around potential government control and data security. For instance, visually, the association with privacy jumps from below 500 in 2020 to somewhere around 3000-3500 in 2023. Surveillance shows a similar, though perhaps slightly less pronounced, increase, also reaching the 2500-3000 range. Anonymity and regulation also show substantial increases, appearing to reach values between 2000 and 2500. While innovation and security *might* show a slight positive association, these are dwarfed by the surge in negative associations with the aforementioned aspects. This analysis corresponds with the 2023 polarity and subjectivity analysis (Section 7.10), as observed by the increase in negative sentiment and the emergence of privacy-related themes in the topic modelling (Section 7.7).

By 2024, while the association remains strong for several aspects, there is a general decrease in intensity compared to 2023. 'Privacy' and 'anonymity' continue to be prominent, though with reduced intensity (in 2000-2500 range). This could result from BOE's response to public feedback on both the papers and the discussion around measures to address public concerns. The continued emphasis on ‘regulation’ and ‘surveillance,' though less pronounced, indicates ongoing public interest in how these aspects will be managed. These trends are consistent with the 2024 semantic network analysis, which showed a shift in discussions around institutional roles with privacy considerations, the less dominant privacy themes in the topic modelling results (Section 7.7), the slight decrease in negative sentiment observed in the polarity and subjectivity analysis (Section 7.10), and the shift in topics towards potential solutions and opportunities observed in the word embeddings and clustering analysis (Section 7.8).

This analysis identifies key sentiment trends related to the digital pound to address RQ3, specifically highlighting a growing negative association with privacy, anonymity, and surveillance in 2023, followed by a slight moderation in 2024. These trends suggest potential shifts in public perception that will be further explored in Chapter 8's temporal analysis, which will examine their relationship to specific policy events.

7.12 Conclusion

The empirical findings from this multifaceted analysis of X discourse on the digital pound offer crucial insights for policymakers and stakeholders navigating CBDC implementation.

Firstly, as the concept transitions from theoretical discussions to tangible proposals, the years 2023 and 2024 saw an increase in negative sentiments toward the digital pound. Policy policymakers must recognize and address the roots of this negativity to avoid resistance to adopting the digital pound. Secondly, significant increases in anger and fear in 2023 and 2024, as observed via the emotion analysis, reveal that emotional responses can profoundly impact public behaviour and acceptance. This means that mere emphasising robust security measures and privacy protections is not enough; wider adoption requires transparent communication to clarify misconceptions and provide reassurances.

Predominant and consistent negative themes, such as privacy, anonymity, and surveillance, as revealed by topic modelling and aspect-based sentiment analysis, highlight the public’s concerns over increased government surveillance and potential infringements on privacy. This implies that ensuring transaction anonymity where appropriate is essential to building public confidence in implementing a digital pound. Moreover, clear accountability mechanisms explaining how the digital pound will not compromise individual freedoms should be established and effectively communicated to the public.

In addition, it is evident that the public is concerned that those reliant on cash transactions could be disadvantaged by the introduction of the digital pound, which could disrupt the existing financial system. This requires policymakers to devise strategies to support those who lack access to digital technologies or prefer cash to ensure that the transition to a digital pound does not exclude marginalised population segments and is inclusive. Despite the criticism of the digital pound, some acknowledge the innovation potential of CBDCs, which needs to be highlighted (with justification), including the need for a digital pound and its benefits for the UK economy.

These insights, derived from a rigorous analysis of public discourse, are critical for responsible innovation in the digital economy and offer valuable lessons for any nation considering CBDC implementation. This research has demonstrated the value of advanced NLP models, particularly fine-tuned transformer models, in capturing nuanced public sentiment in specialised domains, and has illuminated the complex interplay between public discourse, policy announcements, and evolving public perceptions of the digital pound.

Chapter 8 - Temporal analysis of public discourse on the digital pound

8.1 Introduction

Building upon the foundational analysis presented in Chapter7, this chapter delves into the *temporal dynamics* of public discourse surrounding the digital pound. While Chapter7 identified key themes, sentiments, and emotions prevalent across specific timelines, this chapter addresses *'evolution over time'* part of RQ3 by exploring how these elements *evolve* over time in response to key events. Temporal analysis will provide a longitudinal perspective [100], offering crucial insights into the factors influencing public perception and acceptance of the digital pound. By capturing these dynamics, the chapter contributes to understanding how institutional communications and policy announcements shape public opinion.

By analysing data across three periods under consideration, key objective include:

- Identify sentiment trends, volatility, and shifts in response to major events.
- Examine how emotional responses correlate with policy announcements.
- Explore topic dynamics using advanced topic modelling techniques.
- Detect significant change points in sentiment over time.

8.2 Data Preparation for Temporal Analysis

8.2.1 Data Consolidation

To facilitate temporal analysis, the datasets collected for each period were consolidated into a single dataset. This allowed for a longitudinal examination of sentiment shifts, topic evolution, and other key trends. The below preprocessing steps were taken to ensure data integrity and prepare the data for analysis:

- **Date parsing and standardisation**: All date entries were converted to a uniform format (`YYYY-MM-DD`) to ensure consistency. This step is crucial for correctly ordering tweets chronologically and aggregating data within specific time windows. All dates were parsed successfully without errors to ensure there is no variation with respect to different time zones.
- **Data integrity checks**: To ensure all tweets had valid timestamps, the merged dataset was checked, and no missing or invalid values were found in the 'date' column.
- **Sorting and indexing**: The dataset was sorted and indexed by the standardised date column. This chronological ordering is *essential* for temporal analysis, enabling the identification of trends and shifts in public opinion over time. The resulting dataset comprises 250 tweets for 2020, 4,271 for 2023, and 1,181 for 2024.
- **Event window definition:** For each significant event (as mentioned in Section 3.3.1.1, specific time windows were defined to capture pre-event and post-event sentiments. These events are:

  - ➔ **2020 event window**: January 1, 2020 – June 30, 2020.
  - ➔ **2023 event window**: February 1, 2023 – June 30, 2023.

➔ **2024 event window**: January 1, 2024 – March 31, 2024.

An 'Event_Period' column was added to label tweets according to these windows. This categorization allows for the analysis of public sentiment and discourse *in relation* to these specific events.

8.3 Sentiment Analysis Over Time

8.3.1 Sentiment Mapping

For quantitative analysis, sentiment labels (negative, neutral, positive) were mapped to numerical values (0, 1, and 2, respectively). This numerical representation facilitates statistical computations and visualisation of sentiment trends over time. This mapping also aligns with the approach used in the robustness testing discussed in Chapter 6 and sets the foundation for tracking sentiment trends over time, a key element of RQ3.

8.3.2 Sentiment Trends Visualisation with Event Markers

Sentiment scores were plotted as discrete points, trend lines indicating general sentiment movement over time, and markers indicating significant events. The visualisation (Figure 8.1) highlighted how public sentiment fluctuated in response to key announcements. The red line aligns with the release of the BoE's discussion paper in March 2020, the green line marks February 2023 when the BoE released its Consultation, and Technology Working Papers, and the blue line signifies the January 2024 response from the BoE and HM Treasury, addressing the feedback received on prior papers.

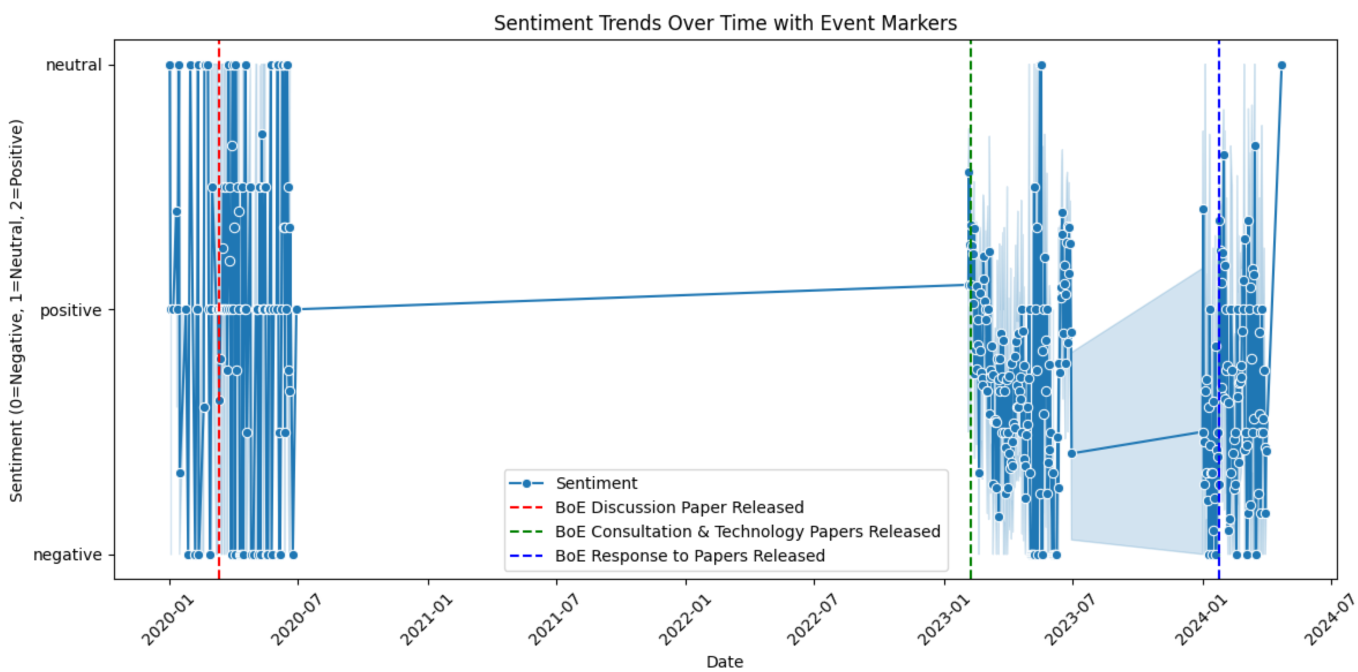


Figure 8.1: Sentiment trends over time with event markers.

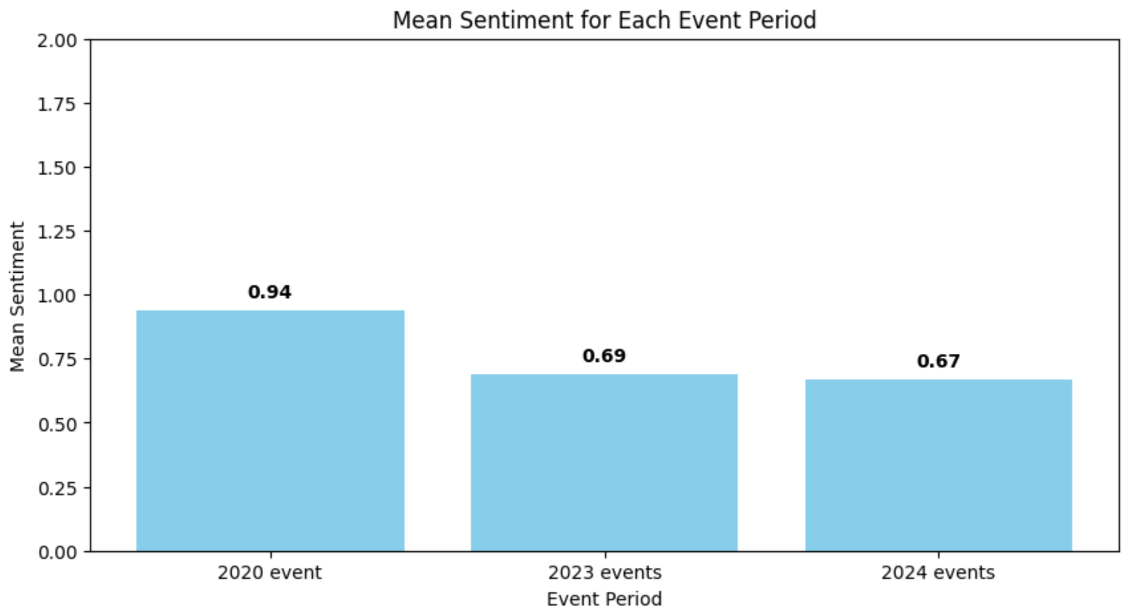


Figure 8.2: Mean sentiment for each event period.

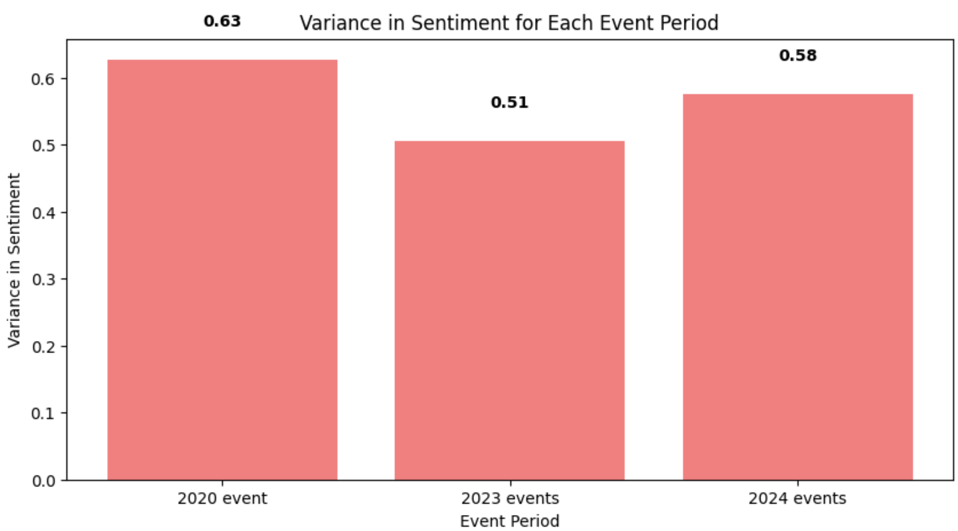


Figure 8.3: Variance in sentiment for each event period.

The sentiment trends, visualised in Figures 8.1, 8.2, and 8.3, reveal distinct patterns across the three time periods.

- **2020:** Sentiment trends in 2020 led to a more fragmented sentiment distribution as the discussions were likely at the exploratory phase. The sentiment distribution in this period shows mixed responses, with both negative (0) and positive (2) sentiments visible but a predominance of neutral (1) responses, probably because the public was still trying to understand the concept of CBDCs and their relevance to the UK economy. The mean sentiment score for this period is 0.94, with a relatively high variance of 0.63, indicating fragmented yet slightly optimistic sentiment, consistent with the early, exploratory phase typical of new technologies. This aligns with the findings of Prodan et al. [100], who reported that public sentiment often begins sceptically without clear regulatory or technical frameworks when new financial technologies are introduced.

- **2023:** With the release of Consultation and Technology Working Papers, there appears to be a more substantial presence of negative sentiments compared to 2020, with a noticeable cluster of tweets labelled as negative (0). The mean sentiment score for 2023 is 0.69, with a variance of 0.5, highlighting the public's analytical stance as they evaluated the pros and cons of a digital pound, including privacy, financial control, innovation, and the potential impacts on existing financial structures.
- **2024:** The year 2024 also shows mixed positive and negative reactions; however, positive sentiment is more pronounced than in previous years. The mean sentiment score for 2024 is 0.67 (Figure 8.2), with a variance of 0.58 (Figure 8.3) indicates positive yet cautious stance of the public as response provided by the BoE and HM Treasury may have addressed some of the public's concerns or clarified points that were previously contentious. Auer et al. [233] noted that when central banks actively respond to public concerns, they can mitigate surveillance or control fears and foster trust. Nonetheless, ongoing debate and concern among some individuals are indicated by prevailing negative sentiments.

As the public's trust and acceptance hinge on transparent and inclusive engagement, CBDCs must balance innovation with traditional values of security and privacy. As Bordo and Levin [234] emphasise, effective communication is essential for balancing innovation with traditional values of security and privacy, which are critical for public trust. This analysis provides valuable insights for policymakers seeking to navigate the complex landscape of public opinion surrounding CBDCs. Overall, these results suggest that public sentiment evolves in line with policy announcements, providing valuable insights for policymakers who must balance innovation with public concerns regarding security and privacy.

8.3.3 Sentiment Trends Following Key Events Across Different Time Windows

To understand the immediate and short-term impact of key announcements on public sentiment, average sentiment scores within specific time windows following each event were analysed. This analysis provides insights into the public's initial reactions and how sentiment evolved in the weeks and months following each announcement. For each event, the average sentiment within the following time windows (chosen based on data availability) were calculated and results are presented in Table 8.4:

- On the day of the event
- 2 weeks after the event
- 1 month after the event
- 45 days after the event
- 2 months after the event

Table 8.1 shows that the average sentiment in 2020 increased from 0.63 on the day of the event to 0.84 after two months, indicating a growing positive reception as the public became more familiar with the digital pound concept. In contrast, average sentiment remained relatively stable in 2023, around 0.78 in the first month, but gradually declined to 0.72 after two months, suggesting that as the public had more time to digest the details of the proposals, emerging concerns tempered the initial positive reaction. However, a high initial sentiment of 0.99 on the day of the event in 2024 reflects optimism following the authorities' responses. The sentiment decreased to 0.75 after two months, indicating that while some concerns were addressed, others persisted. This analysis implies that major announcements generate positive sentiment spikes but maintaining this requires ongoing engagement. The differing patterns observed across the three events highlight the importance of understanding not only *what* is communicated, but *how* it is communicated and how it is received by the public. This section directly addresses RQ3 by quantifying immediate and short-term sentiment shifts following policy announcements.

Event	Time window	Data points	Average sentiment
BoE Discussion Paper Released (2020)	On Day	19	0.63
	2 Weeks After	64	0.92
	1 Month After	107	0.87
	45 Days After	123	0.85
	2 Months After	141	0.84
BoE Consultation & Tech Papers Released -2023	On Day	10	0.70
	2 Weeks After	1,349	0.78
	1 Month After	1,802	0.78
	45 Days After	2,077	0.75
	2 Months After	2,331	0.72
BoE Response to Papers Released (2024)	On Day	121	0.99
	2 Weeks After	424	0.82
	1 Month After	583	0.73

	45 Days After	801	0.77
	2 Months After	888	0.75

Table 8.1: Average sentiment scores after key events.

8.4 Sentiment Volatility Analysis

Sentiment volatility reflects the degree of variation in public emotions and opinions over time, especially in response to significant events related to the digital pound, indicating periods of consensus or divergence among the populace. Volatility analysis helps understand how stable the public sentiment is, how policy announcements impact the public sentiment and highlights potential areas of public concern. Volatility analysis directly addresses RQ3 by measuring how sentiment changes over time due to policy events and identifying patterns of public reaction (stability vs. instability).

To quantify sentiment volatility, several complementary analytical approaches were employed:

- **Day-to-day sentiment change:** Daily changes in sentiment scores were calculated to capture sudden shifts that may correspond to specific events.
- **Aggregated volatility (weekly and monthly):** The sentiment changes over weekly and monthly intervals were aggregated to calculate the mean sentiment change per week and month, recognising that daily sentiment data can be noisy due to user activity patterns and fluctuations in tweet volume. This approach helps smooth out short-term fluctuations and reveal broader trends in volatility.
- **Rolling statistics with a 7-day window:** Using a 7-day window, sentiment scores' rolling mean and rolling variance were computed to further analyse sentiment trends and volatility. This window size aligns with weekly patterns often observed in social media activity due to news cycles and user behaviour [235]. It also provides a sufficient number of data points for each period, enhancing the statistical reliability of the results, while avoiding the potential to obscure significant short-term changes that might occur with larger window sizes. Rolling statistics, also known as moving averages and moving variances, are statistical measures calculated over a fixed-size window to capture local trends and fluctuations by moving through the data sequentially. Rolling variances, in particular, highlight periods of increased or decreased sentiment variability, which can be associated with specific events or evolving public discourse.

8.4.1 Daily Sentiment Volatility Over Time

Figure 8.4 illustrates initial public uncertainty as the concept of CBDCs was introduced, which led to significant sentiment fluctuations in 2020. This aligns with the "alarmed discovery" stage as per the Issue-Attention Cycle, where, in response to a new issue, public interest spikes, often accompanied by uncertainty and diverse opinions [236]. This period also corresponds to the *introduction* stage of the Diffusion of Innovations Theory, where early adopters begin to discuss and evaluate the digital pound, leading to varying opinions [140].

In 2023, volatility spiked again, likely due to the release of BoE Consultation and Technology papers. This increased volatility can be explained through the lens of agenda-setting theory [140], as these policy announcements likely served to focus public attention on the digital pound, leading to heightened scrutiny and a wider range of opinions. This public reaction could also be explained through the Social Amplification of Risk Framework, which explains media coverage of these policy announcements coupled with public discourse may have amplified concerns over privacy, security, and governmental control, leading to increased volatility in sentiment [237].

By early 2024, volatility moderated, leading to a more stable sentiment trend, suggesting that some public concerns may have been addressed by the BoE or they became aware about the concept of central-bank backed digital currencies, aligning with the later stages of the Diffusion of Innovations Theory, where the innovation gains wider acceptance among the majority [238]. However, it is important to note that this does not necessarily indicate a complete resolution of concerns.

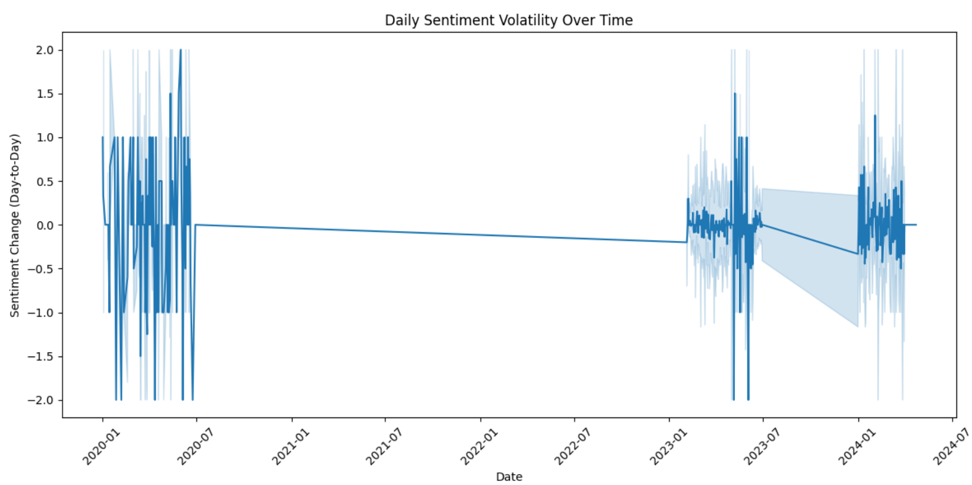


Figure 8.4: Daily sentiment volatility over time.

8.4.2 Weekly and Monthly Average Sentiment Volatility

Figure 8.5 displays that the weekly average sentiment volatility in early 2020 indicates mixed reactions as the concept was still being explored, consistent with the initial engagement phase of the Issue-Attention Cycle [236], where public interest in a new issue begins to rise, but opinions are still forming.

In early 2023, slight variations reappear following this period, likely due to renewed discussions and releases from the BoE. The pattern could be explained by the Social Amplification of Risk Framework [237], which emphasises that perceptions of risk can be amplified beyond inherent danger of a hazard itself. In this case, with new policy details, risk perceptions increase, ultimately leading to amplified emotional responses and increased volatility. By early 2024, volatility is mild, indicating a more tempered public response to ongoing digital pound updates. This suggests that some public concerns may have been addressed or that public understanding of the digital pound has increased.

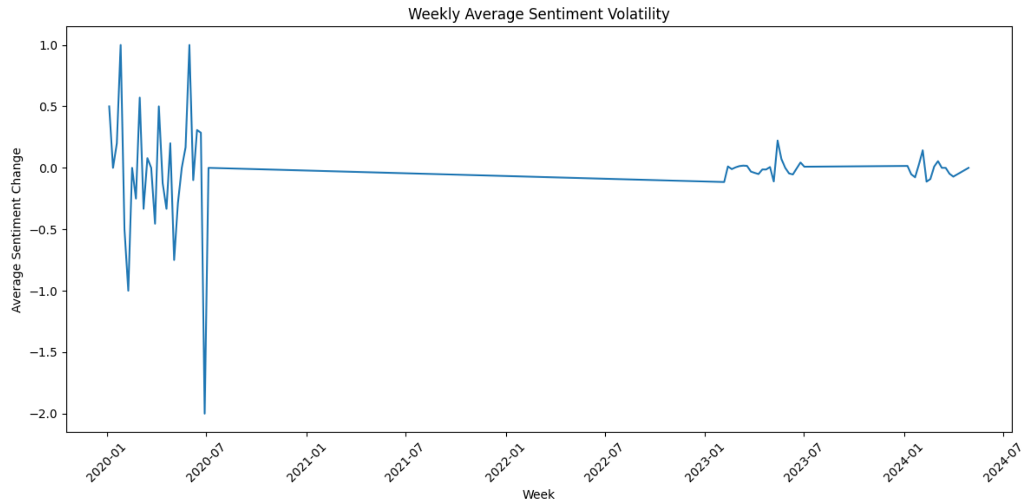


Figure 8.5: Weekly average sentiment volatility.

Figure 8.6 illustrates that early 2020 shows substantial monthly volatility, peaking positively at the start before sharply declining, likely due to initial reactions to the concept of a digital pound - a phenomenon explained by the Hype Cycle Model [239]. The initial positive peak may represent the 'inflated expectations' stage, where the potential of a new technology is often overhyped. The subsequent sharp decline could correspond to the 'disillusionment' stage, as the initial hype fades and the challenges of implementation become more apparent [239].

The average monthly volatility stabilised between 2023 and early 2024 (Figure 8.6), with only minor fluctuations observed. This suggests that the public discourse may have moved into the "slope of enlightenment" and "plateau of productivity" phases of the Hype Cycle, where understanding of the technology improves, and practical benefits become clearer [239], leading to a more balanced and stable public opinion. The increased engagement observed during significant publication periods further supports this interpretation.

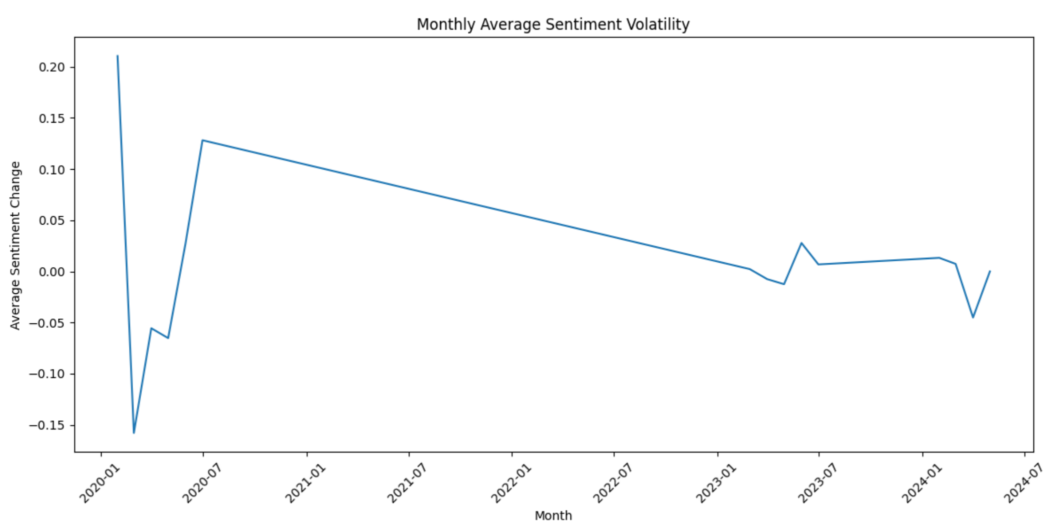


Figure 8.6: Monthly average sentiment volatility.

8.4.3 7-Day Rolling Mean & Variance

The rolling mean illustrates the *direction* of sentiment change, while the rolling variance indicates the *intensity* of public debate. Together they reveal *how much* sentiment is fluctuating, showing the dynamic shifts in public opinion over time. Figures 8.7 and 8.8 show 7-day rolling mean and variance of digital pound-related sentiment. The trends observed in these figures align closely with the patterns observed in the weekly and daily sentiment volatility analysis. The upward trend in the rolling mean in early 2023 corresponds to heightened daily and weekly sentiment volatility (Sections 8.4.1 and 8.4.2), and a downward trend in the rolling mean indicates a shift toward more negative sentiment following the release of the Consultation and Technology working papers. Furthermore, the positive trend into 2024 points to an increasing sense of hope or acceptance as public opinions steadily shift in favour of the CBDC concept while being less volatile.

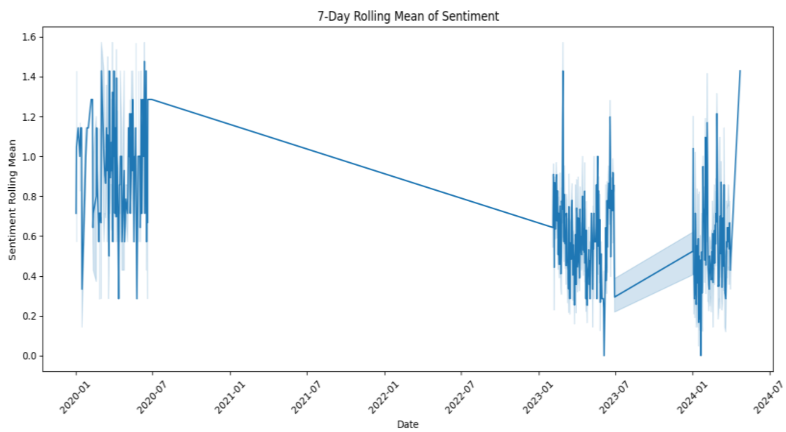


Figure 8.7: 7-day rolling mean of sentiment.

The 7-day rolling variance (Figure 8.8) highlights periods of intensified public debate and complements the monthly average sentiment volatility findings. The high variance in 2020 and 2023 indicates divided opinions as the BoE introduced key papers. By early 2024, moderate weekly fluctuations accompanied by reduced variance suggest that public reactions started to become more balanced while public interest remains high, underscoring the importance of critical events in shaping both the *stability* and *direction* of sentiment over time.

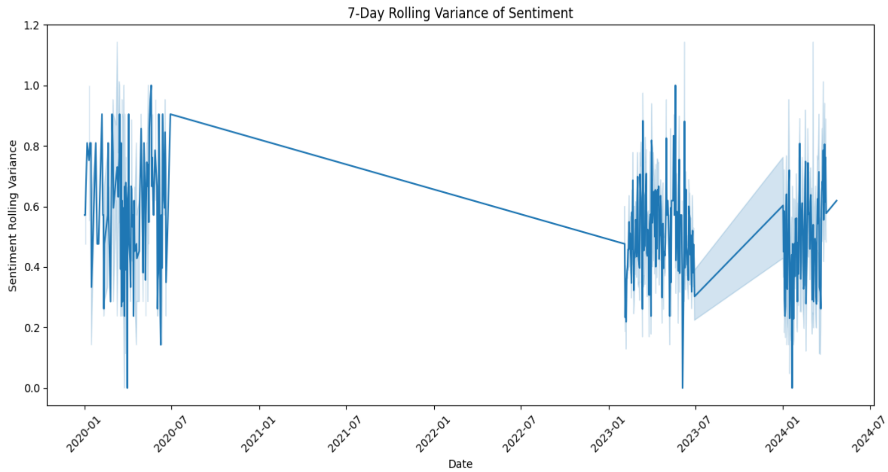


Figure 8.8: 7-day rolling variance of sentiment.

This analysis contributes to RQ3 by demonstrating the direct impact of these events on public sentiment and the level of public discourse.

8.4.4 Key Insights for Policymakers

The volatility analysis answers the "patterns of change" aspect of RQ3 by showing how stable or fluctuating public sentiment is in response to events. The sentiment volatility analysis reveals that the adoption and acceptance of the digital pound concept follow a diffusion curve [226], with significant policy announcements by the BoE leading to initial volatility reflecting early adopters' influence and later stabilisation, indicating wider acceptance among the majority. These findings offer several key insights for policymakers:

- Understanding the tone and volatility of public discourse could help understand how risks are perceived to tailor messages that address such concerns effectively.
- Careful framing of policy communications where potential public concerns are transparently addressed can mitigate negative sentiment shifts.
- Proactive engagement following significant announcements can help stabilise the volatility in public sentiment and inform communication strategies as well as policy adjustments to align with public perceptions of the digital pound.
- Policymakers can maximise the impact of their communication strategies by identifying specific periods of high volatility (heightened public debate) to target their engagement efforts more effectively.

8.4.5 Sentiment Distribution 30 Days Before and After Key Events

To gain a granular perspective on how public opinion shifts in response to policy developments, the distribution of sentiments 30 days before and after each significant event was analysed. This approach offered insights into the immediate impact of these events on public discourse beyond what average sentiment scores can reveal.

Kernel Density Estimation (KDE) plots were used to visualise the sentiment distributions to estimate the probability density function of sentiment scores in a non-probabilistic way, facilitating a smooth comparison between distributions. The blue KDE curve represents sentiment in the 30 days preceding the event, while the red KDE curve shows sentiment in the 30 days following the event.

8.4.5.1 BoE Discussion Paper Release (2020)

Figure 8.9 illustrates the sentiment distribution before and after the release of the BoE's Discussion Paper on March 12, 2020.

Key insights include:

- The public sentiment largely remained balanced, as indicated by a noticeable peak around neutral sentiment (1) in both periods.
- The "before" curve exhibits a cautious engagement, as highlighted by a slightly higher density at the neutral sentiment peak.
- The "after" curve shows a mild increase in optimism following the paper's release, as a slight shift toward positive sentiment can be observed.

As the Discussion paper emphasised the opportunities and innovations associated with the digital pound, it influenced public sentiment positively. This framing and the modest increase in positive sentiment post-event align with the issue-attention cycle described by Downs [236], where a new issue enters public discourse and initially garners optimism and interest.

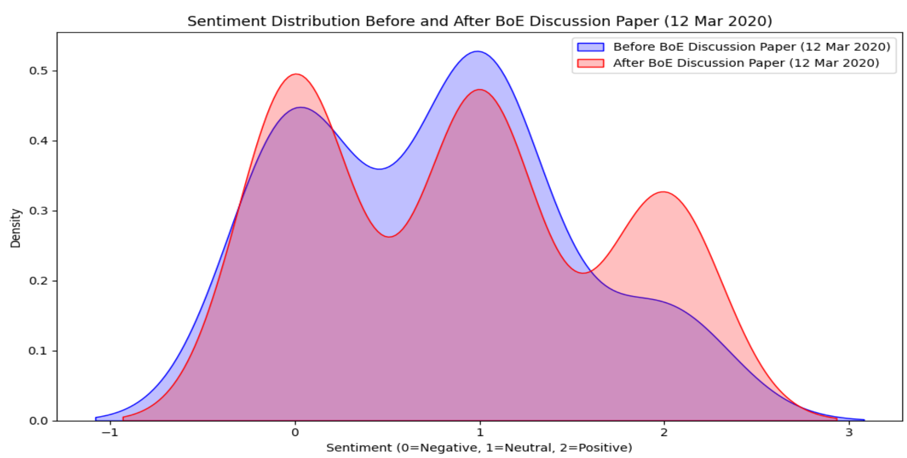


Figure 8.9: Sentiment distribution 30 days before and after BoE Discussion Paper release.

8.4.5.2 BoE Consultation & Technology Papers Release (2023)

Figure 8.10 presents the sentiment distribution surrounding the BoE's Consultation and Technology Working Papers release on Feb. 7, 2023.

Key insights include:

- Like 2020, public sentiment remained largely balanced before and after the publication, as indicated by a prominent peak at neutral sentiment (1) in both periods.
- There was a noticeable increase in negative sentiment (around 0) after the release. This means that the information provided in the consultation and technical documents raised some public concerns. Predominant and consistent negative themes, such as privacy, anonymity, and surveillance, as highlighted during EDA, align with this observation.
- The positive sentiment (2) remains minimal but slightly increases after the event. This suggests that some aspects of the documents may have resonated positively with a segment of the public, perhaps related to the potential benefits of CBDCs.

The framing effect (which posits that the way information is presented (framed) can significantly influence public perception and interpretation) helps understand the shift toward negative sentiment post-event [141]. Public apprehension was heightened based on how risks associated with digital pound were presented in the papers. For instance, discussions of surveillance capabilities or data control without sufficient emphasis on privacy safeguards could have inadvertently amplified negative sentiment. Moreover, the agenda-setting theory [140] suggests that the BoE may have unintentionally amplified public concerns by bringing these issues to the forefront, which could overshadow the perceived benefits, affecting the overall sentiment distribution. According to the agenda-setting theory, media and institutional communications are crucial in determining the public's perception of the issue's importance [140].

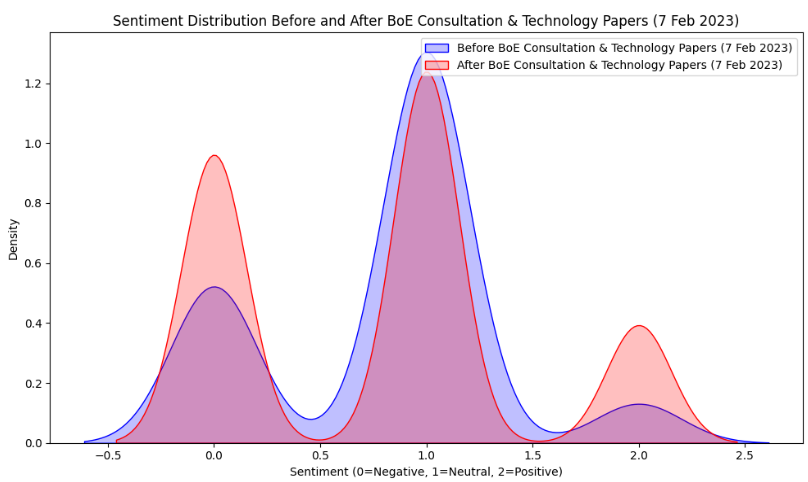


Figure 8.10: Sentiment distribution 30 days before and after BoE Consultation and Technology Working papers release.

8.4.5.3 BoE & HM Treasury Responses to 2023 Policy Papers

Figure 8.11illustrates the sentiment distribution before and after the BoE's Response Papers release on Jan. 25, 2024.

Key observations include:

- Prior to the BoE's clarifications, a prominent negative sentiment peak (around 0) before the response can be observed.

- Following the response, negative sentiment appears to be reduced, as indicated by the red curve and an increased density in the neutral-to-positive range (around 1 and above), suggesting that the BoE's response may have alleviated some concerns.
- An improved public perception following the release could be seen from a moderate increase in positive sentiment (2) after the event.

The positive shift in sentiment distribution post-response reflects that the BoE and HM Treasury may have reframed the narrative around the digital pound based on the public's feedback. This involves addressing privacy concerns, and highlighting benefits. This align aligns with the framing effect, which emphasis that the presentation of information can alter public perception [141].

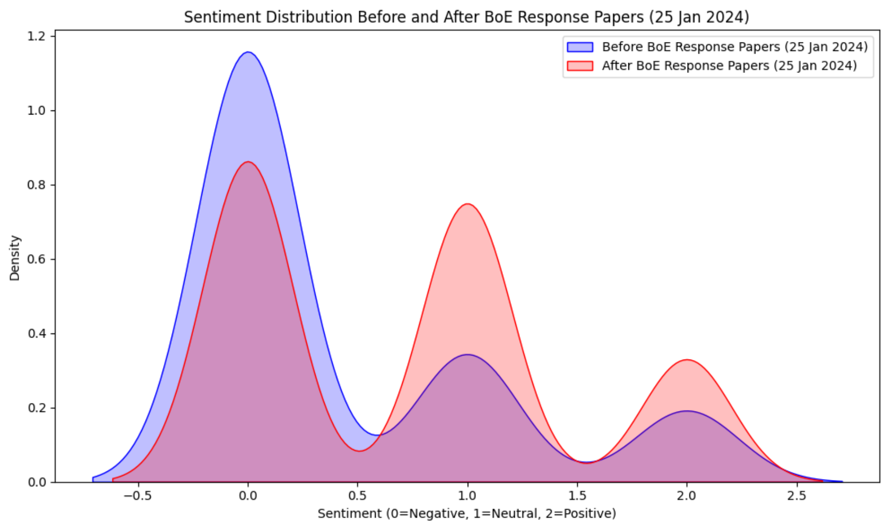


Figure 8.11: Sentiment distribution 30 days before and after BoE Response papers release.

8.4.5.4 Monthly Sentiment Distribution

A month-on-month analysis of sentiment distribution percentages during the study periods was also conducted to examine how public sentiment changes every month. Pie charts (Figure 8.12) were used to illustrate the sentiment distribution for each month.

Across these periods, there is a noticeable variation in sentiment trends. In 2020, the sentiment showed relatively balanced proportions of positive, neutral, and negative sentiments (Figure 8.12), with a peak of positive sentiment in June (41%) and the lowest negative sentiment in the same month (23%).

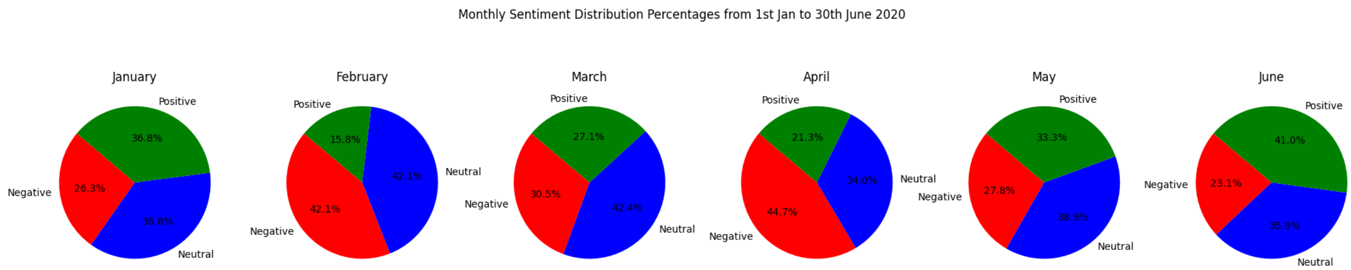


Figure 8.12: Monthly sentiment distribution percentages from 1st Jan to 30th June 2020.

In contrast, 2023 exhibits a more polarised distribution (Figure 8.13), where negative sentiment dominates, especially in April (63.2%), the month with the lowest neutral sentiment (22.2%). This trend in 2023 highlights a more critical public response during this period.

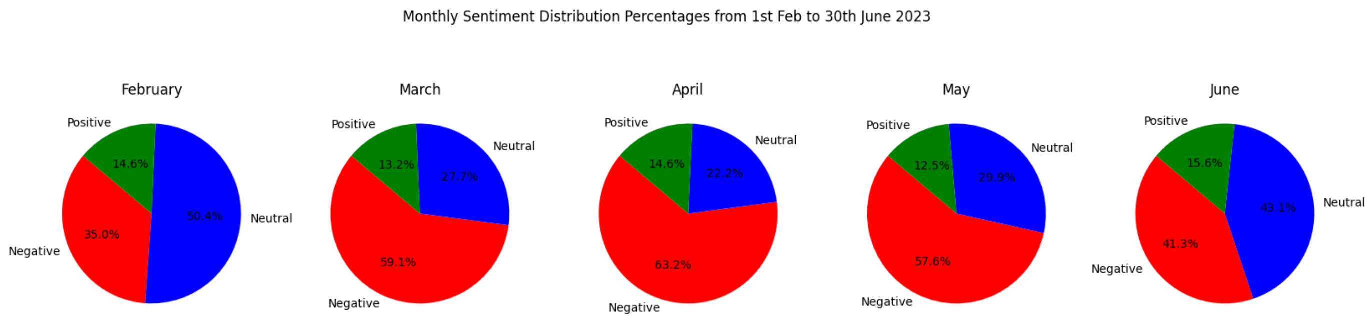


Figure 8.13: Monthly sentiment distribution percentages from 1st Feb to 30th June 2023.

Meanwhile, in early 2024, the sentiment distribution showed a high negative sentiment (particularly in March at 62.3%, as shown in Figure 8.14) and a comparatively lower positive sentiment.

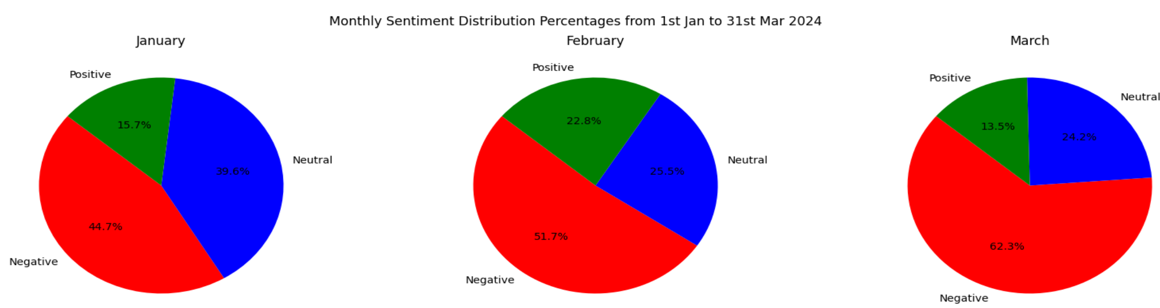


Figure 8.14: Monthly sentiment distribution percentages from 1st Jan to 31st Mar 2024.

Table 8.2 presents the monthly maximum and minimum percentages of positive, negative, and neutral sentiment.

Metric	Month-Year	Value

Highest Negative Sentiment (Monthly)	April 2023	63.20%
Lowest Negative Sentiment (Monthly)	June 2020	23.10%
Highest Positive Sentiment (Monthly)	June 2020	41.00%
Lowest Positive Sentiment (Monthly)	May 2023	12.50%
Highest Neutral Sentiment (Monthly)	February 2023	50.40%
Lowest Neutral Sentiment (Monthly)	April 2023	22.20%

Table 8.2: Monthly maximum and minimum percentages of all sentiment classes.

The above findings suggest a decline in positive reception over time, with sentiment skewing more damaging, especially during key publication dates in 2023 and 2024, reflecting growing concerns or apprehensions.

8.4.5.5 A Distribution-Based Perspective on The Digital Pound Discourse

This review of sentiment distribution presents a unique and important viewpoint, enhancing and broadening the insights obtained from the earlier analyses regarding sentiment trends and volatility. While the exploration of overall sentiment trends (Section 8.3.2) indicated general changes in public perception — demonstrating a transition from a slightly optimistic outlook in 2020 (mean sentiment 0.94) to a more negative stance in 2023 (mean sentiment 0.69), followed by a minor rise in positivity in 2024 (mean sentiment 0.67) — the distribution analysis offers a more detailed comprehension of how these changes unfolded. Specifically, while the average sentiment scores indicate an overarching trend, the distribution analysis uncovers the intricacies of public opinion. For example, during periods characterised by overall negative sentiment (such as in 2023), the distribution analysis indicates that some segments of the population still expressed positive opinions.

Likewise, the volatility analysis (Section 8.4) pointed out times of intensified public discourse (like the spikes in daily volatility in 2020 and 2023), but the distribution analysis demonstrates how these discussion periods led to distinct shifts in the balance of positive, negative, and neutral sentiments. For instance, the spike in volatility in 2023 correlated with a significant rise in the number of negative tweets (centred around a sentiment score of 0), suggesting that the increased debate was associated with a noticeable rise in expressed negative sentiment.

The examination of sentiment distribution (Section 8.4.5.1 to Section 8.4.5.4), which includes assessments before and after significant events as well as monthly breakdowns, provides a detailed view of how these changes and variations occurred among the population. For instance, although the overall pattern indicated a drop in positive sentiment throughout 2023, the distribution analysis indicated that a portion of the public continued to hold positive opinions during this timeframe of general negativity. In a similar vein, while the volatility analysis pointed to heightened discussion in 2023, the distribution analysis illustrated how this conversation resulted in a notable increase in negative sentiment. Additionally, the monthly analysis further tracked these changes, revealing, for example, that negative sentiment peaked in April 2023 at 63.2%. By integrating these various analytical perspectives — trends, volatility, and distribution — stakeholders could gain a broader and more detailed understanding of the changing public discourse regarding the digital pound, moving past mere averages and delving into the complexities of public opinion. This detailed comparison further refines the understanding of immediate public reaction, reinforcing the temporal trends central to RQ3.

However, there are certain limitations to the above analysis, including limited data availability before the 2023 event, which restricts the ability to fully capture pre-event sentiment distributions during that period. The above analysis excludes external factors not related to the digital pound, longer-term sentiment shifts or the delayed impact of events on public opinion. Finally, this analysis is based on X data, which may not be fully representative of the entire population's views on the digital pound. The demographics of X users and their engagement with financial topics could introduce biases into the data. Therefore, the findings of this analysis should be interpreted within the context of these limitations.

8.5 Emotion Analysis Across Events

Building upon the emotion analysis conducted using the NRC Emotion Lexicon (detailed in Section 7.5), this section quantifies the prevalence of specific emotional responses during different event periods, providing a deeper understanding of the affective dimension of public opinion toward the digital pound.

8.5.1 Emotion Extraction Using NRCLex

Using the NRCLex library, each tweet was processed to extract emotion scores across ten categories: trust, fear, joy, positive, anger, anticipation, negative, sadness, surprise, and disgust. This results in a comprehensive profile of emotional responses associated with the discourse on the digital pound.

The emotional distribution representing aggregated emotion counts across these events was visualised using stacked bar charts. In addition, the percentage of each emotion within each event period was calculated to further understand the trends. Line graphs were plotted to illustrate the changes in emotion proportions across events.

8.5.2 Emotion Distribution and Trends Visualisation and Interpretation

Each colour in the stacked bars (Figure 8.15) represents a different emotion with the height of each section corresponding to the count of that emotion. Compared to 2020 and 2024, the 2023 events have a significantly higher volume of emotional responses, likely triggered by the BoE's Consultation and Technology papers release, which provided concrete details on the digital pound, stirring both anticipation and concern.

Moderate positive and negative sentiments in 2024 show a more balanced distribution of emotions, reflecting a more informed and less polarised public response (aligns with polarity and subjectivity analysis discussed in chapter 7). In terms of emotions, Trust and anticipation appear prominently in all events, suggesting that the public generally regard CBDC updates with interest and a degree of confidence. The dominance of these emotions can be explained through the lens of the Curiosity Drive Theory, which posits that novel stimuli elicit curiosity and exploratory behaviour[240]. The fluctuation of trust, however, suggests that public confidence is contingent on how well the BoE addresses public concerns.

Negative emotions (like fear, sadness, and anger) are also present (particularly in 2023) and positive emotions such as joy and trust are prevalent but are relatively smaller in 2024, suggesting a tempered response following the BoE' and HM Treasury's response to the public feedback on 2023 papers. The 2023 public reaction corresponds to the Appraisal Theory of Emotions, which explains that people assess events according to how important they are to their own well-being [241]. It's possible that the thorough policy documents heightened unfavourable feelings concerning individuals' privacy and freedom to use cash. The Risk Perception Theory further explains how people's perceptions of risk might intensify negative emotions like fear and anger [242]. For instance, the sharp increase in anger in 2023 specifically points to public frustration and perhaps a feeling of disempowerment.

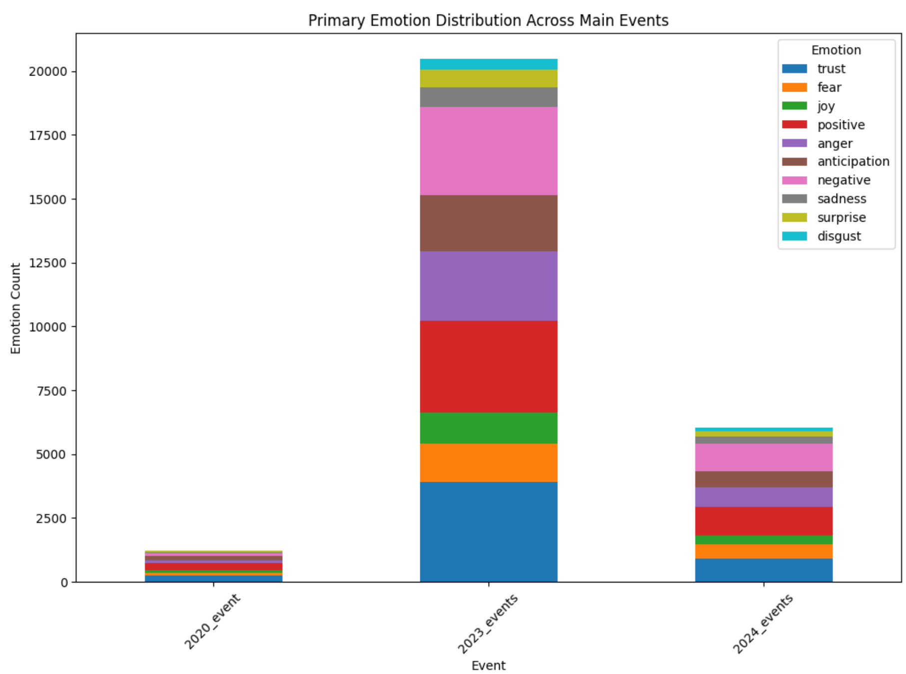


Figure 8.15: Emotion distribution across events.

Here is a breakdown of emotion trends observed in Figure 8.16:

- **Trust**: Initially high, drops in 2023, then recovers slightly in 2024, indicating regained confidence.
- **Anticipation**: Remains steady, showing sustained public interest in CBDC developments.
- **Negative**: Peaks in 2023, reflecting significant public concern, then reduces slightly by 2024.

- **Anger**: Sharp increase in 2023, suggesting frustration with perceived risks.
- **Sadness**: Rises in 2023 but declines somewhat by 2024, indicating reduced pessimism.
- **Joy**: Low overall but sees a slight rise in 2024, showing a small increase in optimism.
- **Positive**: Similarly low but increases slightly by 2024, reflecting a modest boost in public confidence.
- **Fear**: Gradually rises, highlighting ongoing unease about CBDCs' potential impacts.
- **Surprise**: Minimal change, but increases slightly by 2024, possibly due to unexpected reassurances.
- **Disgust**: Low but steadily rising, indicating persistent discomfort with CBDC concepts.

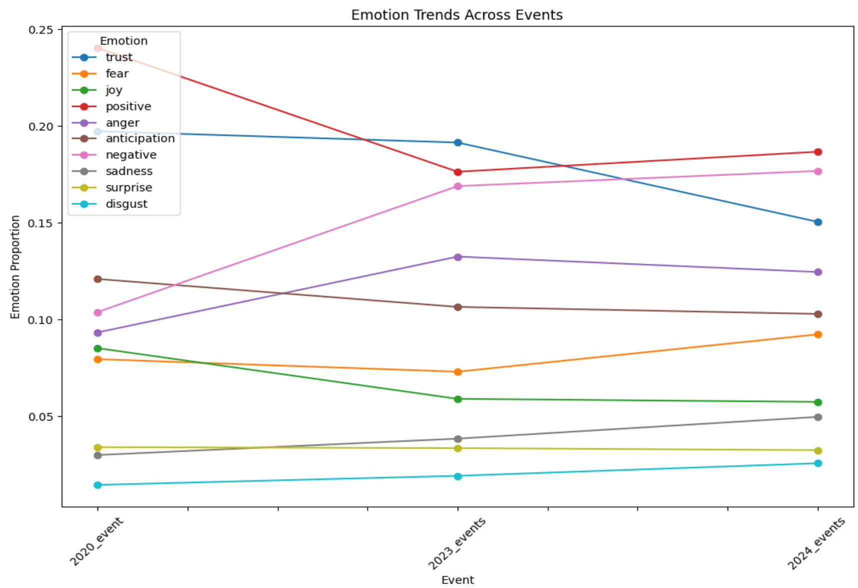


Figure 8.16: Emotion trends across events.

The Social Amplification of Risk Framework helps explain the rise in negative emotions like fear, anger, and sadness in 2023 [237]. The public may have perceived 2023 policy announcements as threatening, increasing negative sentiment. The subsequent reduction in negative emotions in 2024 suggests that the BoE's clarifications could have positively influenced emotional responses. Overall, the analysis revealed that emotions play a critical role in public perception, and policymakers' engagement with the public on an emotional level can enhance the effectiveness of policy communications. Furthermore, regular analysis of public emotions can provide insights into the impact of policy announcements and highlight areas where further clarification or engagement is needed and inform the development of communication strategies that resonate with the public's emotional responses. The fluctuations in trust underscore the importance of transparent and responsive communication to maintain public confidence in the development of the digital pound.

8.6 Topic Modelling with BERTopic

An advanced topic modelling technique, BERTopic, which combines transformer-based embeddings with clustering algorithms was used to analyse the nuanced linguistic patterns and evolving public discourse surrounding the digital pound. While improved interpretability was observed with NMF over LDA (as discussed in Section 7.7.5.1), both traditional methods face inherent limitations in handling context and domain-specific, event-driven text from social media platforms like X [243]. Thus, to capture nuanced linguistic patterns, BERTopic, an advanced approach to understanding the temporal dynamics and evolving sentiments was utilised.

8.6.1 Implementation Steps

To generate coherent topics, BERTopic was implemented in three stages:

1. **Embedding model:** The "all-MiniLM-L6-v2" model from SentenceTransformers was used to generate embeddings for the tweets. This model is efficient and well-suited for creating dense vector representations of short texts like tweets. Prior studies indicate that sentence-level embeddings generally outperform classical vectorisation schemes in capturing contextual similarities and subtle sentiment cues [243], [244].
2. **Clustering and topic extraction:** BERTopic employs HDBSCAN after generating embeddings to cluster the high-dimensional vector space [234]. HBDSCAN can identify clusters of varying densities without requiring the number of clusters to be specified a priori [245], which is advantageous for event-driven, noisy, and highly heterogeneous corpora like social media data [246], [247]. Additionally, to enhance representative capacity to capture both individual terms and short phrases, the model utilises a CountVectoriser with an n-gram range of (1, 2), as recommended for domain-specific corpora [248]. Moreover, English stop words was used within BERTopic to create a document-term matrix and to retain document-specific words.
3. **Temporal segmentation:** The analysis divided the dataset into approximately 20-time bins spanning the study period to capture how topics evolve over time; this segmentation aligns with established practices in temporal text analysis [249], allowing researchers to observe topic emergence, growth, and decline correlated with key events and policy announcements. The public's thematic focus may shift due to new central bank communications, regulatory changes, or economic developments, necessitating a method that can adapt to and dynamically capture these changing themes.

8.6.2 Parameter Selection and Model Tuning

Hyperparameter tuning is critical to ensuring that BERTopic surfaces coherent and meaningful topics to reflect genuine changes in discourse, thus addressing RQ3's thematic evolution component. Key parameters [243]that were tested include the following:

- min_topic_size: Minimum number of documents for a topic.
- - n_neighbors: Number of neighbors for UMAP.
- - n_components: Number of dimensions for UMAP.
- - min_dist: Minimum distance parameter for UMAP.
- - min_df_value: Minimum document frequency for CountVectorizer.

The coherence score (using the c_v metric) guided the selection of the optimal configuration.
Table 8.3 shows a subset of these tuning results.

min_topic_size	n_neighbors	n_components	min_dist	min_df	Coherence_score
5	5	2	0.0	0.01	0.3100
5	5	2	0.1	0.005	0.3356
5	5	2	0.1	0.01	0.3119
5	5	5	0.0	0.005	0.3519
5	5	5	0.0	0.01	0.3300
5	5	5	0.1	0.005	0.3319
5	5	5	0.1	0.01	0.3239
5	10	2	0.0	0.01	0.3387
5	10	2	0.1	0.01	0.3295
5	10	5	0.0	0.01	0.3519
5	10	5	0.1	0.01	0.3334
10	5	2	0.0	0.01	0.3364
10	5	2	0.1	0.01	0.3289
10	**5**	**5**	**0.0**	**0.01**	**0.3633**
10	5	5	0.1	0.01	0.3555
10	10	2	0.0	0.01	0.3397
10	10	5	0.0	0.01	0.3354

Table 8.3: Hyperparameter tuning results of BERTopic.

The optimal configuration, achieving a coherence score of 0.3633, represents a balance between identifying sufficiently large topic clusters and maintaining topical specificity. While the coherence score might appear modest, it’s important to note that achieving high coherence scores with short social media texts is often challenging due to their inherent noise and brevity. This score, in conjunction with qualitative assessment of the resulting topics, guided the final parameter selection.

8.6.3 Topic Assignment, Outliers, and Distribution

The optimal model configuration was applied to the dataset of 5,702 tweets. The results indicated no rows were removed due to missing or null processed text, confirming the robustness of data cleaning steps, discussed in chapter 3. The model flagged 1,476 as outliers (Topic = -1), a typical occurrence in noisy social media datasets when data points do not fit into any coherent cluster [250]. The remaining 4,226 tweets were assigned to valid topics. The thematic structure becomes clear when such outliers are filtered out.

The Table 8.4 provides a sample of the distribution of documents across the identified topics, showing the number of tweets assigned to each topic.

Topic	Document Count
0	257
1	187
2	161
3	143
4	107
...	...
137	10
138	10
139	10
140	10

Table 8.4: Sample distribution of documents per topic.

8.6.4 Temporal Dynamics and Evolving Themes

Studies in sentiment and topic analysis indicate that BERTopic allows for uncovering dynamic insights [243] and modelling the evolution of topics over time and the extent to which topic representations reflect that [244]. The global topic representation, encompassing 139 topics generated by the BERTopic model (Figure 8.17), reveals the breadth of public discussions and concerns across the study period.

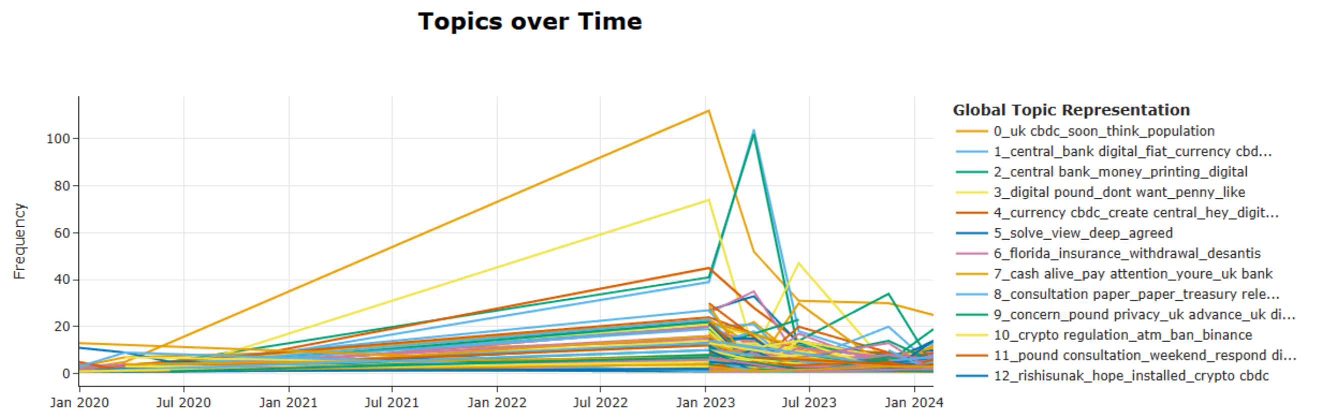


Figure 8.17: Topic evolution over time.

Initially, early phases were dominated by foundational and conceptual discussions, while later periods saw heightened concerns about privacy, identity verification, and regulatory oversight — trends mirroring findings in global financial innovation discourse. The model's granularity in capturing temporal fluctuations offers empirical evidence of evolving public sentiment. For example, the increased prominence of topics related to "privacy and ID" following the release of the Consultation and Technology Working Papers (as seen in the surge of negative sentiment around this theme in Figure 8.18) clearly demonstrates the impact of these documents on public concerns. Notably, "central bank money printing digital" dominates negative and neutral sentiments, suggesting anxieties around monetary policy and digitalisation. In contrast, the dominance of negative sentiment towards central bank actions underscores public apprehension, while the spread in positive sentiment reflects a more varied, albeit less pronounced, optimistic discourse. Such temporal patterns underscore the value of dynamic modelling approaches, as static topic modelling might obscure these subtle shifts. This analysis contributes to RQ3 by demonstrating that topics (sentiment) evolve over time in response to key policy events.

Figure 8.18: Topic per class.

8.6.5 Prevailing and Declining Topics

Topic frequencies were analysed to identify prevailing (increasingly frequent) and declining (diminishing) across the defined time bins. Table 8.5 shows examples of topics that rose in prominence and those that waned over time.

Topic	Change in frequency	Trend
0	23	Increasing
37	19	Increasing
68	14	Increasing
77	-12	Decreasing
19	-8	Decreasing

Table 8.5: Examples of prevailing and declining topics.

Manual inspection of representative tweets associated with these topics revealed that rising topics often corresponded to emerging regulatory debates, privacy issues, and consultation events. In contrast, early speculative themes, such as those focused on the *potential* benefits of the digital pound in abstract terms, gradually lost relevance as concrete policy details and implementation plans were released. This observation aligns with literature indicating that as policy discussions mature, initial speculation gives way to more grounded, domain-specific debates [16], [17], [124], [126].

From the above analysis, notable findings include:

- **Shift from theoretical benefits to practical challenges:** As the project progressed, a maturation in public discourse was observed as discussions moved from the potential benefits of the digital pound to implementation challenges and regulatory implications.
- **Emergence and peaks of privacy and surveillance concerns:** During the BoE's policy releases in 2023 and early 2024, topics related to privacy issues, data security, and government surveillance showed noticeable peaks. The increased public discourse on privacy concerns after the release of policy documents in February 2023 suggests that detailed policy proposals heightened public awareness and apprehension about the potential privacy implications of the digital pound.
- **Impact of policy responses on public engagement:** The BoE and HM Treasury's responses to public feedback in January 2024 renewed discussions on policy responses and public consultation, underscoring the importance of two-way communication in policy processes.
- **Temporal dynamics of niche topics:** External factors or prominent Figures can influence public attention even if they do not have a lasting impact, as shown by mentions of "Ripple" and "Rishi Sunak," which exhibited momentary spikes but remained relatively minor overall. Global economic changes, international debates on digital currencies, or noteworthy occurrences in the broader crypto market may have influenced public discussions in the UK context. This suggests that global trends, events, and local policies influence public opinion on the digital pound.
- **Sustained discussions post-event:** Even after peaks linked to significant events decreased, several themes, like "currency concerns" and "central bank digital currency," continued to be discussed at a baseline level. Regardless of the news cycle, this ongoing conversation indicates that some interests and concerns endure beyond the immediate responses to events, suggesting a deeper level of public involvement with the core elements of the digital pound.

These findings address the "key topics" part of RQ3 by identifying which themes are becoming more or less significant.

8.6.6 Hierarchical Clustering and Intertopic Distances

To further explore the relationships between the extracted topics and gain a deeper understanding of the thematic structure of public discourse surrounding the digital pound, two complementary visualisation techniques: Intertopic Distance Maps and Hierarchical Clustering, were employed. This analysis further addresses the "semantic relationships" element of RQ3.

8.6.6.1 Intertopic Distance Map

The relationships and distances between the extracted topics are visualised using the Intertopic Distance Map. The D1 and D2 axes are abstract dimensions derived from reducing the topic model's high-dimensional space, making it easier to see similarities and distinctions between topics in a 2D space. For example, the highlighted circle (indicating the prevalence or frequency of that topic) represents Topic 9 (Figure 8.19), with keywords such as "concern," "pound privacy," "uk advance," "uk digital,"and "balance" The size of the circle, with 58 occurrences, indicates its relative prevalence in the dataset compared to other topics. The proximity of Topic 9 to Topic 35 (centered around "digital age" and "hmtreasury" with a size of 32) and Topic 91 ("pound referred " with a size of 16) on the map suggests a thematic connection between these topics; this spatial closeness indicates that discussions about digital pound and its implications for privacy are interconnected, with shared concerns regarding advancements in digital policies and their governance.

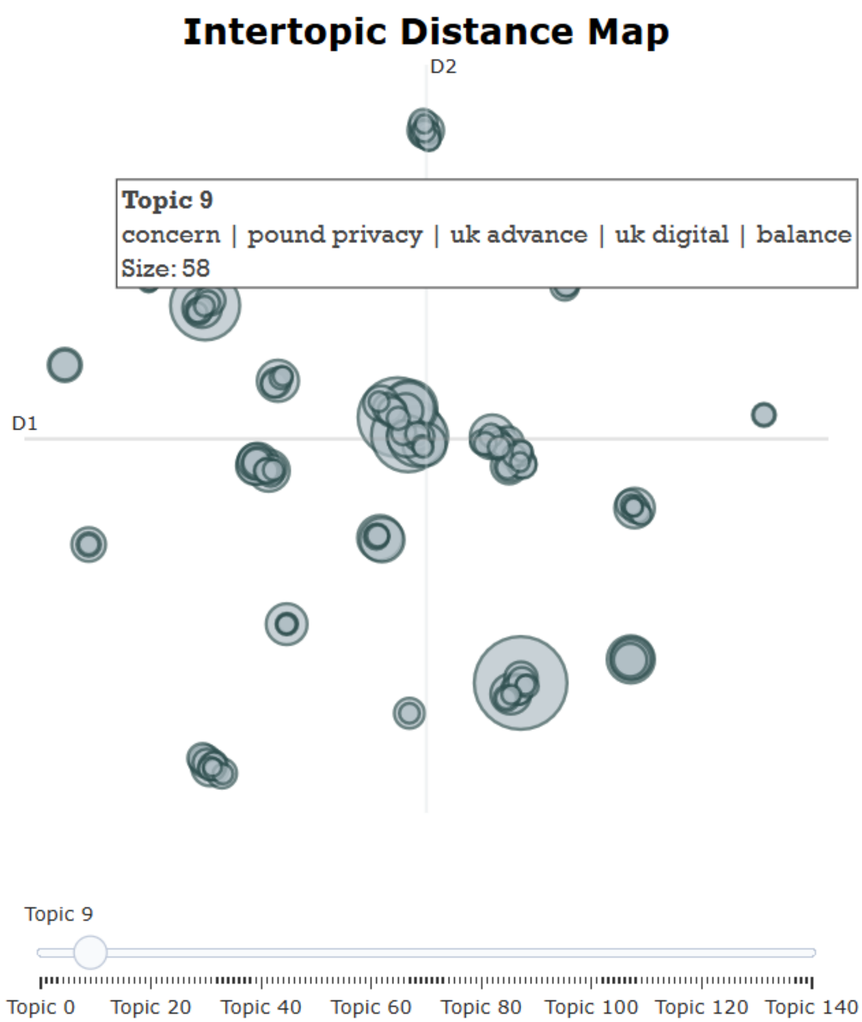


Figure 8.19: Intertopic distance map of topic 9.

8.6.6.2 Hierarchical Clustering

Hierarchical clustering provides another perspective on the relationships between topics, displaying their associations based on thematic similarity in a dendrogram. Each branch in this dendrogram represents a distinct topic, and the distance between them indicates the degree of similarity between branches. Closely related topics exhibit substantial contextual overlap, indicating recurring ideas or opinions in the discourse.

For instance, a cluster focusing on privacy and pseudonymity (topics 134 and 124) suggests these themes are tightly connected in the dataset, reflecting shared concerns around data security and anonymity. Similarly, another cluster captures regulatory and policy discussions (e.g., topics 52, 122, and 108), emphasising structured debates on the governance of the digital pound (Figure 8.20). The distance and hierarchy in the dendrogram illustrate the degree of similarity between clusters, with closely linked topics like those on UK digital policy (topic 58) showing less thematic overlap with clusters focused on blockchain technology or Ripple partnerships (topic 86). This visualisation provides valuable insights into major themes within the dataset, illustrating how topics evolve and diverge over time, aiding in the analysis of public discourse and institutional priorities regarding the UK CBDC.

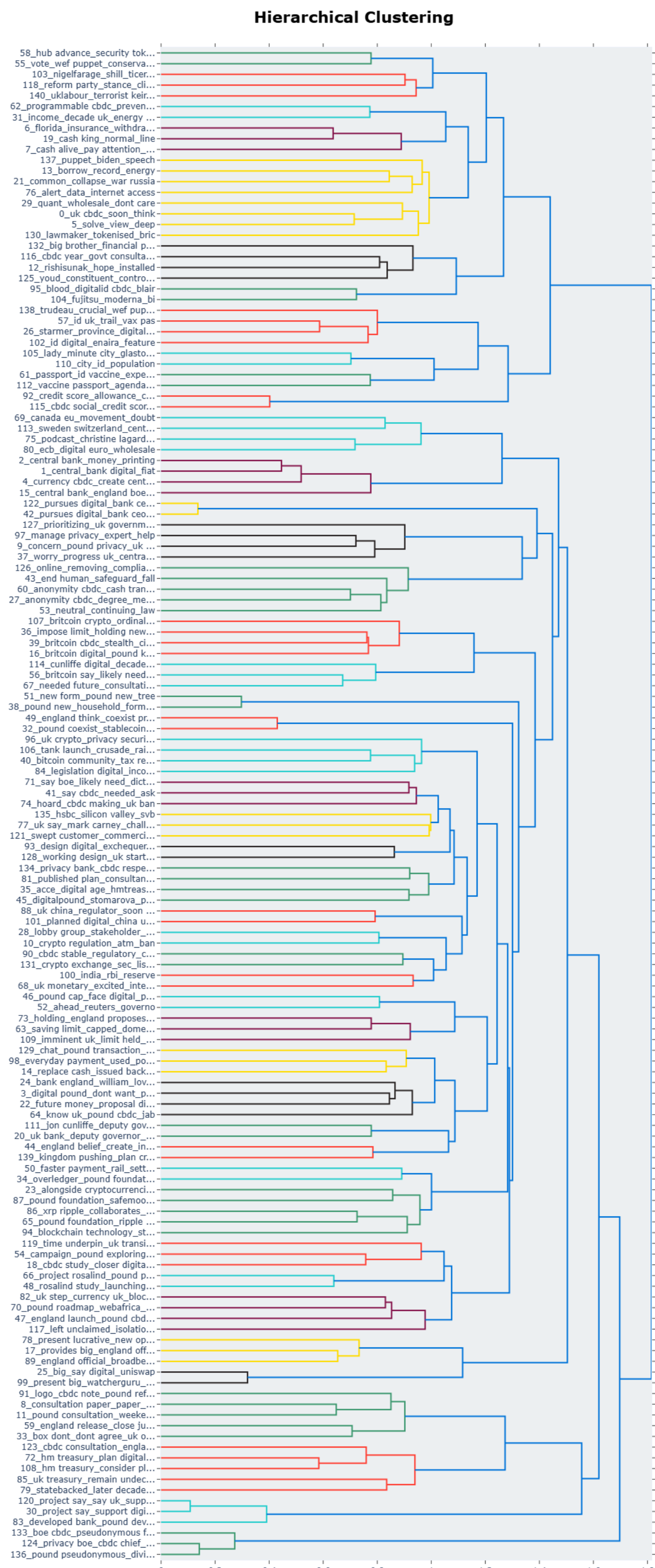

Hierarchical Clustering
58_hub advance_security tok...
55_vote_wef puppet_conserva...
103_nigelfarage_shill_ticer...
118_reform party_stance_cli...
140_uklabour_terrorist keir...
62_programmable cbdc_preven...
31_income_decade uk_energy ...
6_florida_insurance_withdra...
19_cash king_normal_line
7_cash alive_pay attention_...
137_puppet_biden_speech
13_borrow_record_energy
21_common_collapse_war russia
76_alert_data_internet access
29_quant_wholesale_dont care
0_uk cbdc_soon_think
5_solve_view_deep
130_lawmaker_tokenised_bric
132_big brother_financial p...
116_cbdc year_govt consulta...
12_rishisunak_hope_installed
125_youd_constituent_contro...
95_blood_digitalid cbdc_blair
104_fujitsu_moderna_bi
138_trudeau_crucial_wef pup...
57_id uk_trail_vax pas
26_starmer_province_digital...
102_id digital_enaira_feature
105_lady_minute city_glasto...
110_city_id_population
61_passport_id vaccine_expe...
112_vaccine passport_agenda...
92_credit score_allowance_c...
115_cbdc social_credit scor...
69_canada eu_movement_doubt
113_sweden switzerland_cent...
75_podcast_christine lagard...
80_ecb_digital euro_wholesale
2_central bank_money_printing
1_central_bank digital_fiat
4_currency cbdc_create cent...
15_central bank_england boe...
122_pursues digital_bank ce...
42_pursues digital_bank ceo...
127_prioritizing_uk governm...
97_manage privacy_expert_help
9_concern_pound privacy_uk ...
37_worry_progress uk_centra...
126_online_removing_complia...
43_end human_safeguard_fall
60_anonymity cbdc_cash tran...
27_anonymity cbdc_degree_me...
53_neutral_continuing_law
107_britcoin crypto_ordinal...
36_impose limit_holding new...
39_britcoin cbdc_stealth_ci...
16_britcoin digital_pound k...
114_cunliffe digital_decade...
56_britcoin say_likely need...
67_needed future_consultati...
51_new form_pound new_tree
38_pound new_household_form...
49_england think_coexist pr...
32_pound coexist_stablecoin...
96_uk crypto_privacy securi...
106_tank launch_crusade_rai...
40_bitcoin community_tax re...
84_legislation digital_inco...
71_say boe_likely need_dict...
41_say cbdc_needed_ask
74_hoard_cbdc making_uk ban
135_hsbc_silicon valley_svb
77_uk say_mark carney_chall...
121_swept customer_commerci...
93_design digital_exchequer...
128_working design_uk start...
134_privacy bank_cbdc respe...
81_published plan_consultan...
35_acce_digital age_hmtreas...
45_digitalpound_stomarova_p...
88_uk china_regulator_soon ...
101_planned digital_china u...
28_lobby group_stakeholder_...
10_crypto regulation_atm_ban
90_cbdc stable_regulatory_c...
131_crypto exchange_sec_lis...
100_india_rbi_reserve
68_uk monetary_excited_inte...
46_pound cap_face digital_p...
52_ahead_reuters_governo
73_holding_england proposes...
63_saving limit_capped_dome...
109_imminent uk_limit held_...
129_chat_pound transaction_...
98_everyday payment_used_po...
14_replace cash_issued back...
24_bank england_william_lov...
3_digital pound_dont want_p...
22_future money_proposal di...
64_know uk_pound cbdc_jab
111_jon cunliffe_deputy gov...
20_uk bank_deputy governor_...
44_england belief_create_in...
139_kingdom pushing_plan cr...
50_faster payment_rail_sett...
34_overledger_pound foundat...
23_alongside cryptocurrenci...
87_pound foundation_safemoo...
86_xrp ripple_collaborates_...
65_pound foundation_ripple ...
94_blockchain technology_st...
119_time underpin_uk transi...
54_campaign_pound exploring...
18_cbdc study_closer digita...
66_project rosalind_pound p...
48_rosalind study_launching...
82_uk step_currency uk_bloc...
70_pound roadmap_webafrica_...
47_england launch_pound cbd...
117_left unclaimed_isolatio...
78_present lucrative_new op...
17_provides big_england off...
89_england official_broadbe...
25_big_say digital_uniswap
99_present big_watcherguru_...
91_logo_cbdc note_pound ref...
8_consultation paper_paper_...
11_pound consultation_weeke...
59_england release_close ju...
33_box dont_dont agree_uk o...
123_cbdc consultation_engla...
72_hm treasury_plan digital...
108_hm treasury_consider pl...
85_uk treasury_remain undec...
79_statebacked_later decade...
120_project say_say uk_supp...
30_project say_support digi...
83_developed bank_pound dev...
133_boe cbdc_pseudonymous f...
124_privacy boe_cbdc chief_...
136_pound pseudonymous_divi...
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8

Figure 8.20: Hierarchical clustering showing topic association based on thematic similarity.

8.7 Change Point Detection in Sentiment

Change point detection indicates potential shifts in underlying processes or behaviours by providing a statistical approach to identify times when the probability distribution of a time series changes. Change point detection was performed using the Pruned Exact Linear Time (PELT) algorithm implemented in the *ruptures* library to detect significant shifts in sentiment over time. The PELT algorithm has a linear computational cost concerning the time series length and efficiently detects multiple change points in time series data [251]. Moreover, it can capture various change points simultaneously over the study period. Also, it ensures that the detected change points are statistically significant by providing exact segmentation under certain conditions [251].

For the change point detection, the Radial Basis Function (RBF) cost model was selected due to the following reasons:

- Sentiment scores may not follow standard statistical distributions, and the RBF model does not assume a specific distribution for the data, which makes it suitable for sentiment scores.
- RBF effectively detects changes in the time series mean and variance, capturing subtle shifts in sentiment patterns.
- Social media data is often noisy; RB can model non-linear relationships within such data.

Implementation steps include the following:

- Sentiment scores (0 for negative, 1 for neutral, and 2 for positive) and corresponding dates were extracted from the filtered dataset to handle missing values appropriately.
- The PELT algorithm was applied to the sentiment time series with the RBF cost function.
- The penalty parameter (pen=10) was selected to balance missing important shifts and detect too many insignificant changes based on prior experimentation to control the sensitivity of the change point detection.
- The sentiment scores over time were plotted with vertical lines indicating the detected change points. Additionally, sentiment change magnitude was calculated for each detected change point, quantifying the shift in sentiment between consecutive points. This followed a histogram of sentiment change magnitudes to illustrate the distribution of change intensities across the dataset and identify patterns in how frequently small versus large sentiment shifts occurred, contributing to insights on public sensitivity to developments around the digital pound.

8.7.1 Interpretation of Change Points

Figure 8.21 illustrates the detected change points marked in the sentiment time series plot, which shows the sentiment scores over time, with vertical dashed lines indicating the points where significant changes in the statistical properties occurred.

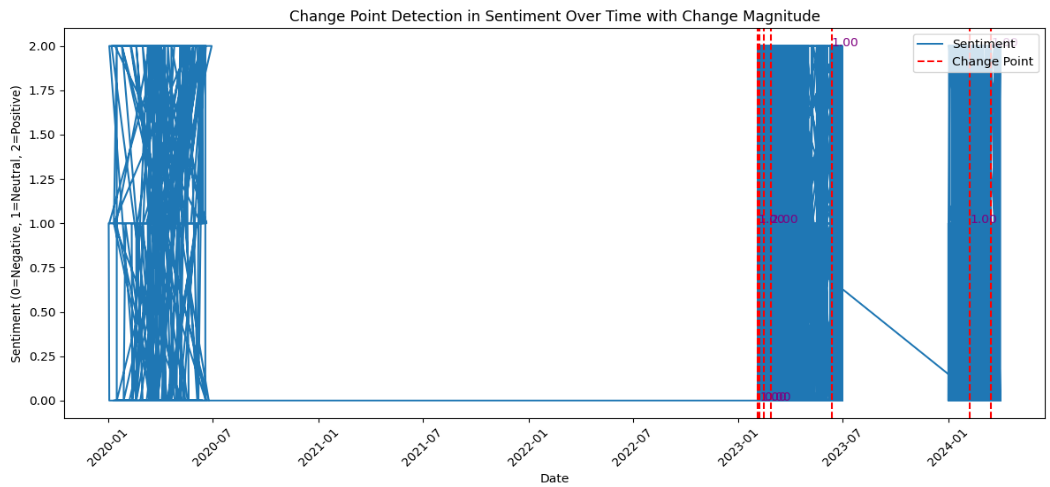


Figure 8.21: Change points detected in sentiment over time.

As an example, tweets surrounding the first detected sentiment change point, dated 2023-06-18, were examined. This change point was further investigated by analysing tweets within a 7-day window (2023-06-11 to 2023-06-25). This yielded 576 tweets, and thematic analysis of these tweets revealed several key drivers behind the observed sentiment shift. Discussions surrounding the *practical introduction* of the digital pound intensified, with many tweets expressing anxieties about the implementation timeline, integration with existing financial systems, and accessibility for various segments of the population. Public discourse also focused heavily on the *BoE's role* in managing and overseeing the digital pound, raising questions about the central bank's level of control, transparency, and potential for government overreach. Furthermore, concerns about the *potential financial impacts* of the digital pound contributed significantly to the sentiment shift, with tweets discussing risks to the existing financial institutions, the impact on interest rates and inflation, and implications for individual savings and investments. These interwoven discussions about the introduction, the BoE's role, and the financial impacts appear to have collectively driven the sentiment shift observed on 2023-06-18.

In addition to the above, the magnitude of change was calculated, as noted in Table 8.6, to provide insight into the strength of these shifts, with higher values indicating more substantial changes in public sentiment.

Date	**Change magnitude**
2023-02-04	1
2023-02-07	1
2023-02-14	0

2023-02-27	2
2023-06-12	1
2024-02-07	1
2024-03-15	1

Table 8.6: The detected change points and their magnitudes.

From Figure 8.21 and Table 8.6, multiple change points were observed with varying magnitude as follows:

- **Feb. 4 and Feb. 7, 2023**: Both dates indicate moderate sentiment shifts and exhibited a change magnitude of 1.00, meaning the public began engaging with the detailed proposals.
- **Feb. 14, 2023**: A stabilisation in sentiment was observed, with a change magnitude of 0.00, suggesting that as the public processed the details, initial reactions were settling.
- **Feb. 27, 2023**: A significant change magnitude of 2.00 was recorded, reflecting a pronounced shift, aligning with the polarity and subjectivity analysis (Section 7.10), highlighting increased scrutiny following continued discussions around the potential drawbacks or societal impacts of the digital pound.
- **Feb. 7, 2024**: A change magnitude of 1.00 indicates renewed public reaction to the authorities' clarifications and responses.
- **March 15, 2024**: Another change in magnitude of 1.00 suggested ongoing sentiment adjustments as the public continued to process the BoE' and HM Treasury's position on the digital pound.

Tweet volumes were also examined alongside detected change points to provide additional context (Figure 8.22). Key observations include:

- **March 2020**: A slight increase in tweet volume corresponds with the BoE's Discussion Paper release, indicating initial but modest public attention.
- **February 2023**: A surge in tweet volume aligns with multiple change points, reflecting heightened public interest and active discussions prompted by the Consultation and Technology papers.
- **January 2024**: Increased tweet volume following the BoE and HM Treasury's responses to public feedback illustrates renewed engagement as the public reacted to official clarifications.

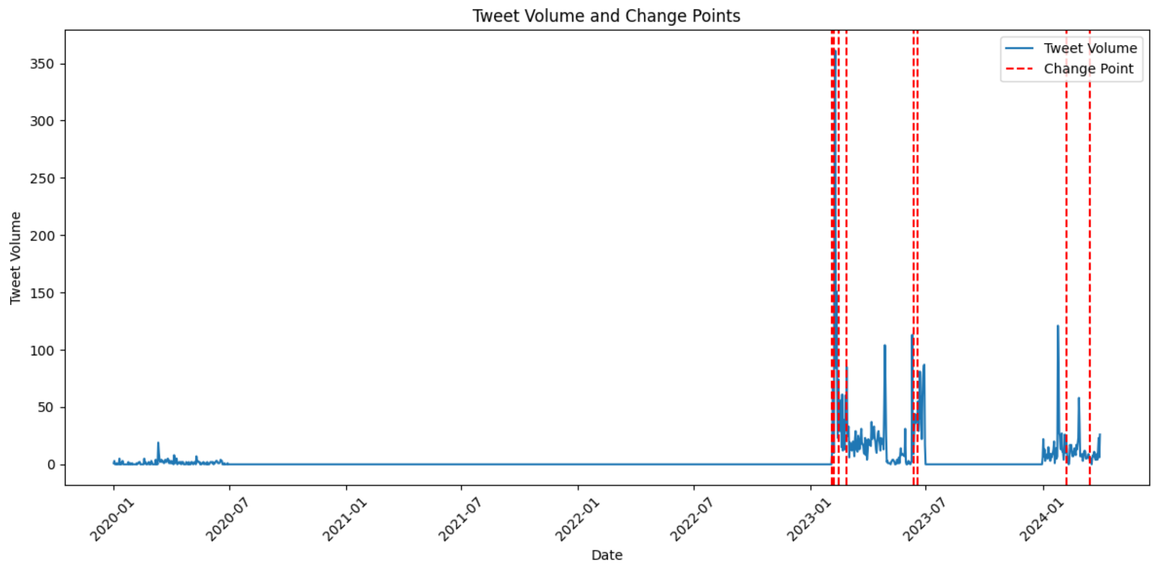


Figure 8.22: Tweet volume and change points.

Overall, the change point detection analysis revealed noticeable shifts:

- **Concentrated change points around February and June 2023 and early 2024:** Shortly after the release of the Bank of England's Consultation and Technology working papers on Feb. 7, 2023, a significant change point was detected, which aligns with the increase in negative emotions identified in the emotion analysis (Section 7.5), and underscore the substantial impact of these policy announcements on public sentiment, as previously observed in Chapter 7. Primary concerns driving these shifts included privacy, security, and the implications of the digital pound's implementation.
- **Divergence in sentiment in early 2020**: The issues surrounding the digital pound became more contentious or dynamic from 2023 onward, whereas sentiment appeared stable before that period (generating curiosity and optimism).
- **Changepoint in early 2024**: A change point was detected around January 2024, coinciding with the release of responses from the Bank of England and HM Treasury to public feedback to 2023 papers. This period saw renewed discussions and a slight improvement in sentiment, suggesting that the authorities' engagement may have clarified some concerns, positively influencing public opinion. It further highlights the importance of two-way communication and public engagement in shaping perceptions of complex economic topics.

The alignment of the detected change points with major policy announcements and events validates the influence of these occurrences on public sentiment dynamics. It underscores the responsiveness of public opinion to institutional communications and policy developments around major economic topics like the implementation of a digital pound. This section answers the "patterns of change" part of RQ3 by statistically identifying when significant shifts in sentiment occur in response to policy events.

8.8 Statistical Analysis of Sentiment Shifts

A series of statistical tests were conducted to evaluate the statistical significance of observed changes in public sentiment to determine whether shifts in sentiment before and after each event were significant (evolved) or could be attributed to random variation.

Dependent and independent variables:

- **Independent variable:** The event period (2020, 2023, 2024), which represents the policy milestone or discussion phase related to the digital pound.
- **Dependent variable:** The sentiment scores assigned to tweets before and after the event, which reflect changes in public sentiment.

Hypotheses statements:

- **Null Hypothesis (H_0):** There is no difference in sentiment before and after the event.
- **Alternative Hypothesis (H_1):** There is a significant difference in sentiment before and after the event.

Underlying assumptions of normality and homogeneity of variances were checked before selecting appropriate statistical tests. This followed a similar approach to that described in Chapter 5.

Normality Tests:

The Shapiro-Wilk Test was used to check normality and results for each event are summarised in Table 8.7:

Period	Statistic	p-value	Normality Conclusion
2020 Event Period	0.8002	<0.0001	Not normally distributed (reject H_0)
2023 Events Period	0.7736	<0.0001	Not normally distributed (reject H_0)
2024 Events Period	0.7537	<0.0001	Not normally distributed (reject H_0)

Table 8.7: Shapiro-Wilk normality test results for sentiment scores.

The Shapiro-Wilk test results ($p < 0.0001$ for all periods) demonstrate that sentiment scores for each event period are *not* normally distributed. This violates the assumption of normality required for parametric tests like ANOVA.

Homogeneity of Variance Test:

Levene’s test was conducted to see if the variances across three groups are equal (null hypothesis) to evaluate the assumption of equal variances. Table 8.8 presents the results of this test for each event.

Test Statistic	p-value	Variance Conclusion
7.0163	0.0009	Variances are not equal (reject H_0)

Table 8.8: Levene's test for homogeneity of variances.

As observed, the p-value is less than 0.05, indicating that variances are not equal across the three periods. This further supports the relevance of non-parametric tests.

8.8.1 Significance Testing

To compare sentiment scores across three periods, the Kruskal-Wallis H test (a non-parametric equivalent of ANOVA) was conducted to assess differences. Given the non-normal distribution and unequal variances, this non-parametric test is suitable for comparing more than two independent groups when the assumptions of ANOVA are not met [241]. The Kruskal-Wallis H test is a rank-based test, meaning it does not assume any specific distribution of the data. This makes it a robust choice for analysing sentiment data, which, as shown in Table 8.7, is not normally distributed.

Test Statistic	p-value	Significance Conclusion
28.1625	0.0001	Significant difference in sentiment across periods (reject H_0)

Table 8.9: Results of Kruskal-Wallis H test.

Given that p-value is less than 0.05 (Table 8.9), null hypothesis is rejected, indicating that there is a statistically significant difference in sentiment across the three periods i.e., policy announcements influence public sentiment.

Post-Hoc Analysis:

Dunn's test with Bonferroni correction was conducted as a post-hoc analysis (Table 8.10) to identify which periods differ significantly. The Bonferroni correction is applied to control the family-wise error rate, reducing the risk of false positives when conducting multiple comparisons. This followed a similar approach to that described in Chapter 5.

Comparison	p-value	Significance interpretation
2020 Event Period vs. 2023 Events Period	0.000003	Significant difference ($p < 0.05$)
2020 Event Period vs. 2024 Events Period	3.66E-07	Significant difference ($p < 0.05$)

2023 Events Period vs. 2024 Events Period	0.346878	No significant difference (p > 0.05)

Table 8.10: Dunn's test with Bonferroni Correction Post-Hoc results.

The post-hoc analysis reveals that there are significant differences in sentiment between the 2020 Event Period and both the 2023 Events Period and the 2024 Events Period. However, no statistically significant difference in sentiment was found between the 2023 Events Period and the 2024 Events Period ($p = 0.347$). This suggests that while sentiment changed significantly from 2020 to later periods, the sentiment in 2023 and 2024 was not statistically different from each other, despite observed fluctuations.

8.8.1.1 Additional Statistical Analysis to Compare 30-Day Windows

An additional analysis was conducted using 30-day windows before and after each key event to provide a more granular understanding of sentiment shifts, allowing for the examination of immediate reactions and short-term trends surrounding each event. For this analysis, the Mann-Whitney U test was employed. It is a non-parametric test suitable for comparing two independent groups [252] when the data are not normally distributed.

Event	Test Statistic	p-value	Significance interpretation
2020 Event	1088.5000	0.3233	Fail to reject the null hypothesis: No significant difference in sentiment.
2023 Events	125136.0000	0.4624	Fail to reject the null hypothesis: No significant difference in sentiment.
2024 Events	40444.0000	0.0000	Reject the null hypothesis: Significant difference in sentiment before and after the event.

Table 8.11: Mann-Whitney U Test results for 30-day window sentiment changes.

Table 8.11 results show that immediate responses to events were muted in 2020 and 2023 (i.e., no statistically significant difference in sentiment before and after the events), while sentiment broadly evolved across all periods. Moreover, the lack of significant short-term change in 2023 within the 30-day window aligns with the post-hoc Dunn's test, indicating that sentiment had largely stabilised by 2023. This implies that both short- and long-term trends are important to capture a comprehensive view of public reaction to policy announcements and changes over time.

Short-term insights are crucial for assessing the immediate impact of policy releases. Proactive engagement and addressing public concerns can mitigate misinterpretations and prevent unintended backlash. As public preferences evolve, evidenced by sentiment shifts from 2020 to 2024 (statistically significant difference in sentiment before and after the event), public opinion on CBDCs will likely continue to change. Policymakers must therefore adopt an interactive approach, incorporating feedback and addressing concerns transparently at each stage of CBDC policy development.

8.9 Conclusion

This chapter complements the findings of Chapter 7 and addressed ‘temporal element’ of RQ3. The findings demonstrate that public sentiment is not static but evolves significantly in tandem with policy developments and institutional communications. The statistical tests confirm that public sentiment shifts are not random but are significantly influenced by policy events, reinforcing the findings of Chapter 7.

- **Key themes, topics, and sentiment trends:** This research indicates a distinct progression in public sentiment. Initial reactions in 2020, after the release of the Discussion Paper, were characterised by a sense of curiosity and cautious hope. However, with the publication of the Consultation and Technology Papers in 2023, there was a noticeable shift towards negative sentiment, largely fuelled by apprehensions regarding privacy, data security, and the possibility of government overreach. By 2024, in light of the responses from the Bank of England and HM Treasury to public feedback, a slight adjustment in sentiment was noted, as some concerns were alleviated, resulting in a more even sentiment distribution. Additionally, topic modelling analysis shed light on the key themes driving these sentiment transitions, highlighting a focus on practical implementation issues, regulatory frameworks, and the potential effects on current financial systems.
- **Shifts in sentiment, emotions, and semantic relationships:** This analysis demonstrates that sentiment, emotions, and semantic relationships within the X discourse shifted significantly over time and in direct response to key policy events. The increase in negative sentiment in 2023, coupled with heightened expressions of fear, anger, and sadness in the emotion analysis, underscores the public’s sensitivity to policy details, particularly those related to privacy and control. The semantic analysis, through topic modelling, revealed how these emotional responses were connected to specific themes, such as surveillance, anonymity, and the role of central authorities. The change point detection analysis further pinpointed the temporal connection between these shifts and specific policy announcements, providing evidence of the immediate impact of institutional communications on public perception.
- **Patterns of change identified through temporal analysis:** Temporal analysis revealed distinct patterns of change, which can be referred to as the ‘Exploration-Polarisation-Adaptation’ (EPA) sequence. The initial phase (2020) was characterised by a more exploratory and less polarised discourse. The subsequent phase (2023) saw a surge in negative sentiment and heightened debate, reflecting increased public scrutiny and apprehension. The BERTopic modelling in this phase indicated a notable shift in the thematic focus toward issues surrounding privacy, data protection, and the risk of government surveillance. In the final phase (2024), there was evidence of some adaptation and adjustment, marked by a slight enhancement in overall sentiment and a more intricate conversation regarding the issues. BERTopic analysis during this period revealed a balanced discourse, as it included discussions on both the concerns and potential solutions.

> The patterns identified through this longitudinal study, supported by BERTopic's capability to track topic evolution, underscore the ever-changing nature of public opinion and the necessity for continuous dialogue and interaction between policymakers and the public. Future research could explore demographic influences on sentiment, the role of traditional media, and comparative analyses across countries.

Nonetheless, understanding public sentiment using social media data is just one aspect of the issue. To achieve a thorough understanding of the conversation around the digital pound, it is equally important to investigate how institutions address public concerns and influence the narrative related to this developing technology. Consequently, Chapter 9 will focus on this other facet of the dynamic, examining the BoE and HM Treasury's reactions to the feedback it received for its Consultation and Technology Documentation. This study of institutional communication will offer valuable insights into how authorities are responding to public concerns and shaping the trajectory of the digital pound.

Chapter 9 - Thematic analysis of the public feedback and Bank of England's responses to the Consultation and Technology papers

9.1 Introduction

Understanding the themes and narratives emerging from BOE's responses to the public feedback on Consultation and Technology Papers is essential for aligning policy objectives with public sentiment and ensuring the successful adoption of the digital pound. The objective of this chapter is to perform a comprehensive thematic analysis of the BoE's response papers to the Consultation and Technology Papers, addressing RQ4.

By extracting key themes and narratives, the analysis serves as a foundation for comparing official communications with public discourse on X, addressing RQ4. The chapter integrates relevant literature and theoretical frameworks to contextualise the findings, offering a robust understanding of stakeholder perspectives on the digital pound. The methodology is designed to be reproducible, employing systematic coding and analysis techniques.

9.2 Methodology

9.2.1 Research Design

A qualitative approach was used to systematically identify and interpret patterns of meaning (themes) across the BoE's response papers. The thematic analysis method, as outlined by Braun and Clarke [253], is suitable for exploring complex textual data and uncovering nuanced themes relevant to the research questions. This qualitative analysis complements the sentiment, emotion, and topic modelling analyses presented in Chapters 7 and 8.

While a RoBERTa-based pipeline was developed and applied to sentiment analysis of X data, as described in Chapters 5, 6, 7, and 8, this chapter analyses structured public feedback submissions to the Bank of England. These submissions, distinct from the short-form, spontaneous nature of tweets, necessitate a different analytical approach. A qualitative thematic analysis is employed here to best capture the nuanced and considered opinions expressed in these formal responses [254].

A strategic decision was made to focus on key questions for detailed analysis while providing synthesised findings for the remaining questions, given the extensive number of questions and responses. This approach avoids overwhelming the reader and offers in-depth insights into where they matter most and aligns with best practices in qualitative research that advocate for focused analysis when dealing with extensive datasets [254], [255]. The questions were selected based on their centrality to the research questions, the volume and depth of stakeholder feedback received, and their potential to yield rich insights into the key themes and tensions surrounding the digital pound.

9.2.2 Data Collection

The primary data sources are the publicly available BoE's response papers to the Consultation and Technology Papers:

- Response to the Bank of England and HM Treasury Consultation Paper − The digital pound: A new form of money for households and businesses? [13].
- Response to the digital pound Technology Working Paper [14].

The BoE received 51,529 responses from individuals and organisations across various sectors. The majority of these responses (40,330) were from individuals, while a smaller proportion (555) were from organisations, including large firms, small and medium-sized enterprises (SMEs), and sole traders [13]. For the Technology Working Paper, the bank received 391 responses, reflecting a broad spectrum of perspectives on the proposed technology design considerations and the conceptual model [14]..

9.2.3 Data Preparation and Coding

An Excel spreadsheet (codebook) was created to systematically organise the data, containing the following columns for each question:

- **Question No.:** The specific question number from the Consultation or Technology Paper.
- **Code:** Short labels representing key concepts or ideas identified in the responses.
- **Code definition:** Detailed explanations of each code.
- **Theme:** Broader categories that group related codes.
- **Evidence:** Direct quotes or summaries from the responses supporting each code and theme.
- **BoE response analysis:** Analysis of how the BoE addressed each theme, noting alignment, omissions, or framing in their official responses.

The link to the full codebook can be found in Appendix13. This structured coding framework was chosen to facilitate a systematic and comprehensive analysis of the BoE's responses, allowing for the tracking of specific questions, the identification of recurring themes, and the comparison of public feedback with official responses [256], [257].

9.2.3.1 Integration of Python for Data Handling

The identified themes were imported into Python using the pandas library; Python scripts were used to remove inconsistencies, such as extra whitespace and duplicate entries, ensuring data integrity. Preliminary codes were assigned using keyword matching and frequency analysis. This automated process provided an initial structure that was refined through manual review, combining the strengths of computational tools with human interpretive skills [258].

9.2.4 Data Analysis Strategy

As mentioned in Section 9.2.1, key questions for detailed analysis were selected based on their relevance to RQ4 and the substantial stakeholder feedback they elicited. Analysing every single question in detail would be impractical and could dilute the analysis with repetitive or less informative data.

Consultation Paper:

- Question 1: Future payments landscape
- Question 2: Platform model and public-private partnership
- Question 3: Privacy and data protection

Technology Paper:

- Question 1: Foundational technology considerations
- Question 2: Privacy-enhancing technologies (PETs)
- Question 4: Design considerations

For the remaining questions in both papers, themes were identified and presented in a synthesised manner. This approach maintains the chapter's manageability and highlights significant insights without diluting the focus. In qualitative research, it is common practice to focus in-depth on key areas and provide an overview of the rest [254].

In addition to the qualitative analysis, Python was used to perform quantitative analyses, complementing the qualitative thematic analysis:

- **Frequency analysis**: The Counter class from Python's collections module was used to calculate the frequency of each theme across responses.
- **Co-occurrence analysis**: The itertools library facilitated the examination of theme co-occurrences within responses, identifying patterns of themes that frequently appear together [248].
- **Cluster analysis**: Themes were grouped into clusters based on content similarity using text similarity measures from the difflib library [253].

These quantitative analyses add a layer of quantitative evidence to support the qualitative findings and helps identify potential relationships between themes that might not be immediately apparent through manual review.

9.2.5 Thematic Analysis Process

The thematic analysis followed a six-phase framework by Braun and Clarke [253], a widely recognised and flexible approach well-suited for analysing diverse textual data and generating rich, detailed descriptions of themes. According to the authors, there is no standardised method for thematic analysis. Table 9.1 explains the thematic analysis process for the study under consideration.

Stages	Description
Familiarisation with the Data	All responses were thoroughly reviewed to gain a deep understanding of the content

Generating Initial Codes	Significant statements were identified based on their relevance to the RQ, and descriptive codes were assigned
Searching for Themes	Codes were collated into potential themes, capturing the essence of stakeholder perspectives
Reviewing Themes	Themes were refined (through an iterative process) to ensure they accurately represented the data
Naming and Defining Themes	Each theme was clearly defined and labelled
Producing the chapter	Writing up the analysis

Table 9.1: Thematic analysis process.

An initial set of codes was developed based on the consultation questions and the content of the BoE responses. Manual coding was performed to identify significant statements and assign labels that captured the essence of the content[256]. Codes were examined for patterns and grouped into broader themes that reflected underlying ideas across the responses. Themes were reviewed and refined to ensure they accurately represented the coded data and were distinct [257]. Each theme was clearly defined and named to encapsulate its core message and relevance to the consultation questions.

To ensure reliability and validity:

- **Codebook development**: A comprehensive codebook was developed, providing definitions and examples for each code [259].
- **Reflexivity**: The analysis process involved continuous reflection on assumptions and decisions to mitigate potential biases [260].
- **Audit trail**: Detailed documentation of the analysis process was maintained to ensure transparency and reproducibility [260].

9.3 Thematic Findings: Response to the Consultation Paper

9.3.1 Overview

The Consultation Paper posed twelve specific questions covering various aspects of the digital pound, including its potential impact on the payments landscape, financial inclusion, privacy and data protection, and technological considerations. These diverse themes echo the multifaceted sentiment and topic trends observed in Chapters 7 and 8, where public discourse on X revealed a complex interplay of hopes, concerns, and expectations surrounding the digital pound.

A total of 87 themes were identified across all 12 questions, of which 64 were unique. The total number of themes identified per-question are shown in Table 9.2. References to quotations from 2024 Consultation Paper [13] are cited with specific page numbers throughout this chapter.

Question No.	No. of themes per question
1	6
2	8
3	5
4	4
5	5
6	10
7	7
8	6
9	4
10	11
11	10
12	11

Table 9.2: No. of themes per question in Consulting Paper.

9.3.2 Question1: Future Payments Landscape

Themes identified: Cash Usage Trends; Public Sentiment and Adoption; Security Concerns; Implementation Clarity; Economic and Competitive Advantage; Technological Advancements.

Figure 9.1 illustrates the relationship between themes and codes identified for Question1.

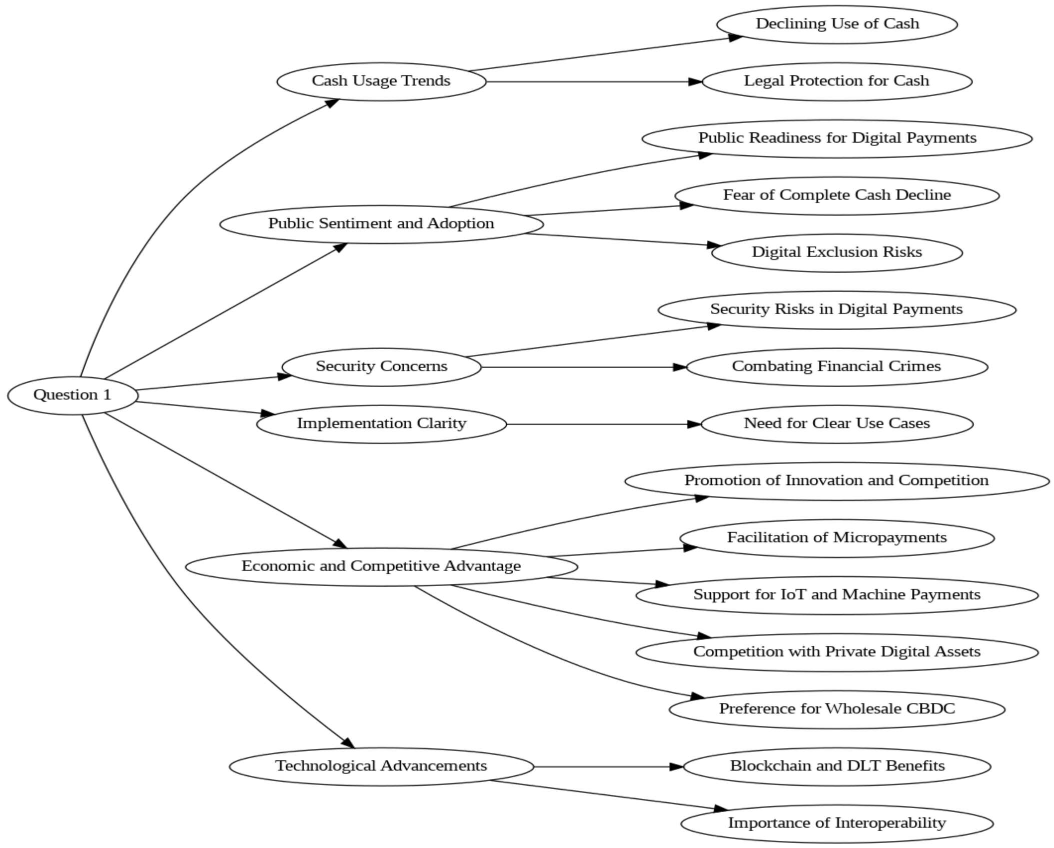


Figure 9.1: Mind map of Question1.

Cash Usage Trends:

Respondents expressed concerns about cash's declining usage and its implications for society and the economy. Many individuals were worried that the reduced use of cash could lead to the exclusion of cash-dependent people from the financial system. The concerns regarding cash usage trends echo financial inclusion literature findings [261], which emphasise the importance of cash for vulnerable populations who may lack access to digital payment methods and can exacerbate the digital divide. This finding is consistent with the early exploratory sentiment observed in Section 7.3 and the temporal trend of cautious optimism in Sections 8.3 and 8.4.

Evidence: "Some respondents stressed the risk of digital exclusion if cash use continued to decline (p.35)."

Public Sentiment and Adoption:

The public called for legal measures to ensure continued access to cash despite declining usage and to legally protect access to cash to address public concerns about its future availability. Ensuring legal protection for cash aligns with efforts to maintain inclusivity in the financial system, as highlighted by Lupo-Pasini [262], who discuss the importance of cash for social groups that rely primarily on cash. Also, these sentiments parallel the evolving public attitudes tracked in Chapters 7 and 8, where concerns about financial exclusion and the role of traditional cash emerged as significant themes.

Evidence: "In that context, some saw a digital pound as maintaining access to risk-free public money and improving payment options where cash is not readily accepted (p.35)."

Security Concerns:

The potential risks associated with digital payments, such as exposure to online fraud, cyberattacks, and the misuse of personal data, were significant concerns among stakeholders. The Technology Acceptance Model (TAM) suggests that "trust" in the security and privacy of digital transactions is fundamental for user acceptance [263], [264]. Moreover, key risks, such as cybersecurity vulnerabilities across infrastructure providers, CBDC operators, and users, must be carefully addressed before implementing CBDCs [265]. This concern aligns directly with the heightened anxieties about privacy and security identified through sentiment and topic analysis in Sections 7.5, 7.7, 7.8 and 7.11, and the increasing focus on these issues over time, as documented in Chapter 8.

Evidence: "Others pointed out the potential risk to digital-payment users from exposure to online fraud, scams and cyberattacks (p.35)."

Implementation Clarity:

Respondents' feedback indicates that the rationale and practical applications of the digital pound need to be better communicated to the public. This means that clear communication is essential for public understanding and acceptance of new financial technologies, as suggested by Balaskas et al. [266], who note that transparent information helps build trust and reduce uncertainty.

Evidence: "Other respondents questioned the need for a digital pound, given that retail payments today are generally fast, digital, and efficient (p.35)."

Economic and Competitive Advantage:

A few fintech respondents saw a strong use case for micropayments, enabling new business models such as paying small amounts for individual services, such as paying to read a single article. Similarly, a small number of respondents from civil society and the technology sector identified use cases where the digital pound could facilitate payments between connected devices, highlighting the potential for machine-to-machine payments. The feedback also emphasised the preference for a wholesale CBDC over a retail digital pound [29], emphasising different use cases focused on high-value transactions and institutional payments.

Evidence: "Agreeing with one of the primary motivations for a digital pound set out in the Consultation Paper, they noted that a digital pound could lead to improvements in payments, including greater choice, convenience, speed, and lower cost for users (p.36)."

Technological Advancements:

Many respondents highlighted the importance of ensuring that the digital pound works seamlessly with other forms of money and payment systems to provide flexibility and user convenience.

Several technology companies and consultancies emphasised the advantages of using blockchain and DLT for transparency, security, and efficiency in payments.

Evidence: "Fintech respondents were largely supportive of a mixed payment ecosystem, where cash, a digital pound and private digital means of payment, co-existed and were used in a complementary way (p.36)."

As central banks explore the adoption of CBDCs, the design process involves balancing consumer needs with technical trade-offs, including the choice between DLT and conventional infrastructures [233], [267]. Ongoing experiments with various prototypes across jurisdictions will help clarify which technological choices are best suited for CBDCs.

BoE & HM Treasury's Response Analysis: The BoE acknowledged the above-mentioned concerns and emphasised that the digital pound would complement, not replace, cash. The BoE also committed to leveraging technological advancements while ensuring interoperability with existing systems. However, it does not explicitly address the call for legal protection for cash beyond general commitment [47].

9.3.3 Question2: Platform Model and Public-Private Partnership

Themes identified: Support for Platform Model; Business Model Viability; Cost and Efficiency Concerns; Economic and Competitive Advantage; Regulatory Clarity and Fairness; Public-Private Partnership; Need for Interoperability; Strategic and Policy Framework. These themes mirror the n-gram and clustering findings in Chapter 7 (Sections 7.6 and 7.8) and are reinforced by the temporal shifts observed in Chapter 8, particularly the increasing focus on practical implementation challenges and regulatory considerations.

Support for Platform Model:

The public endorsed the proposed public-private partnership platform model for the digital pound yet requested further details on implementation, aligning with collaborative ecosystems in finance and fostering innovation through partnership [268].

Evidence: *''There was broad support for the platform model, with an emphasis on the need for further detail being provided in the future, for example regarding the allocation of accountability between the PIPs and the Bank in the case of PIP failure (p.39)."*

Business Model Viability:

Concerns were raised that private wallet providers (PIPs) might struggle to develop viable business models due to compliance costs and lack of revenue sources.

Evidence: *''That concern generally reflected the cost of compliance with AML/KYC regulations. To several respondents it was not clear how PIPs would raise revenue without charging consumers and/or merchants (p.40)."*

Figure 9.2 illustrates the relationship between themes and codes identified for Question2.

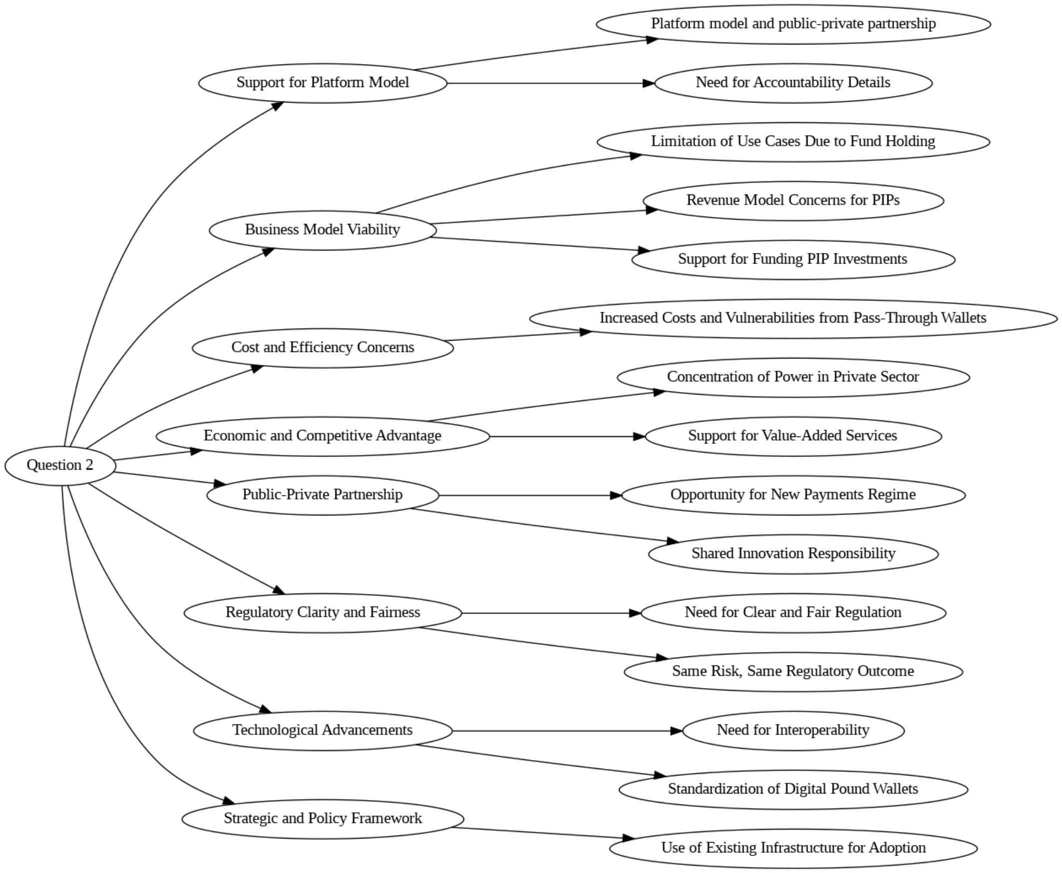


Figure 9.2: Mind map of Question2.

Cost and Efficiency Concerns:

Respondents point to potential inefficiencies and security risks inherent in the platform model. The key concerns include that pass-through wallets hosted by intermediaries might add unnecessary steps, increase end-user costs, and make the system more vulnerable to attacks.

Evidence: *''Others felt that 'pass-through' wallets hosted by intermediaries might add a potentially unnecessary step that could increase costs, including for end-users, and make the system more vulnerable to attacks (p.39).''*

Economic and Competitive Advantage:

A small number of respondents from civil society groups were concerned that the private provision of wallets could concentrate power among dominant fintech and card network providers, weakening competition in payment services and disadvantaging consumers.

Evidence: *''Concerns were raised that such concentration in the hands of financial incumbents would weaken competition in payment services to the detriment of consumers (p.39).''*

Regulatory Clarity and Fairness:

Respondents emphasised the need for clear and fair regulatory frameworks for PIPs to ensure equal competition and high operational standards. To ensure financial system stability, regulatory clarity is essential to prevent regulatory arbitrage [269]

Evidence: *''The majority suggested taking a 'same risk, same regulatory outcome' approach, to ensure that private sector intermediaries compete on an equal footing and are held to rigorous standards for operational resilience, risk management, and compliance (pp.39-40)."*

Public-Private Partnership:

There was a strong emphasis that innovation in the digital pound ecosystem should be a shared responsibility between the Bank and private sector partners. Respondents advocated for collaborative innovation efforts to ensure the platform evolves to meet diverse user requirements and technological advancements.

Evidence: *''Some saw the public-private partnership as an opportunity to design a new payments regime (p.39)."*

Need for Interoperability:

Commercial banks and fintechs emphasised the necessity for digital pound wallets to be interoperable with other payment systems.

Evidence: *''Several stressed the need for a seamless transfer of digital pounds across wallets from different providers (p.40)"*

Strategic and Policy Framework:

Some respondents recommended leveraging existing payments infrastructure and established customer bases to promote merchant adoption and support the digital pound's launch strategy.

Evidence: *''Some respondents suggested using the existing payments infrastructure to promote merchant adoption, as well as the existing customer bases of established payment ecosystems to support the digital pound's launch strategy (p.40)."*

BoE and HM Treasury's Response: The BoE and HM Treasury endorsed the platform model, emphasising its alignment with fostering innovation, competition, and efficiency. They acknowledged the need for further detail in the platform model's implementation, particularly regarding accountability mechanisms and regulatory clarity. While recognising the potential of value-added services and shared innovation, the BoE did not specifically address concerns about micropayments or IoT payments, indicating a potential gap in their current focus, raising questions about how these use cases will be addressed within the proposed platform model.

9.3.4 Question3: Privacy and Data Protection

Themes identified: Privacy and Surveillance; Tiered Access and Inclusion; Privacy Enhancements; Regulatory Clarity; User Autonomy and Control.

Figure 9.3 illustrates the relationship between themes and codes identified for Question3.

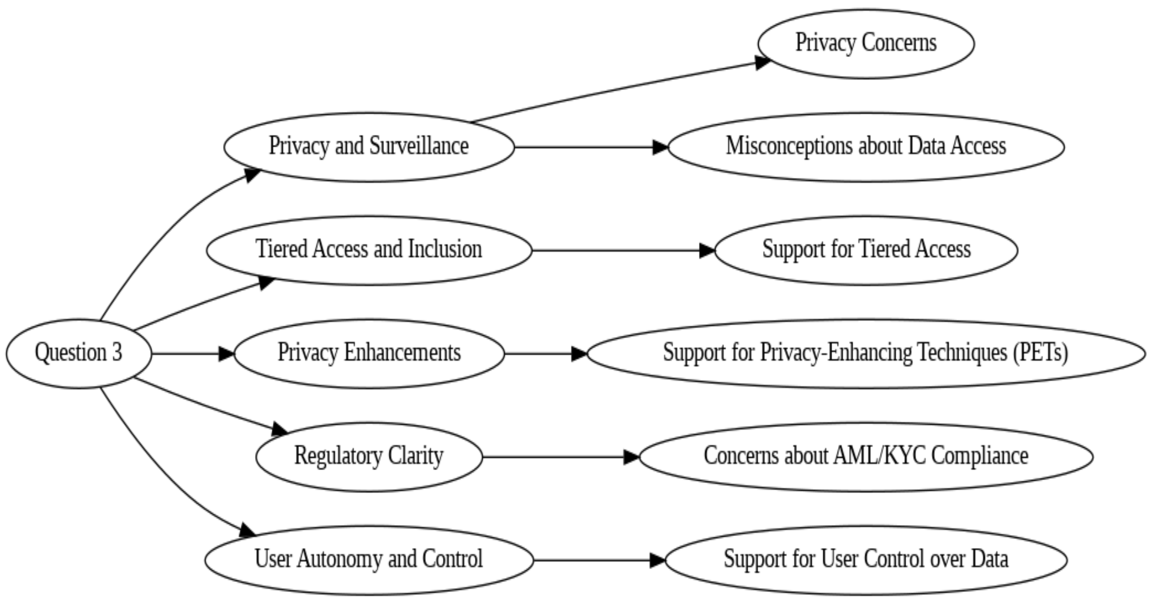


Figure 9.3: Mindmap of Question3.

Privacy and Surveillance Concerns:

Respondents expressed significant concerns about privacy and the potential for government surveillance through the digital pound's infrastructure. This concern is directly in line with the increasing negative sentiment and emphasis on privacy observed in Sections 7.5-7.11 and the evolving topic dynamics in Sections 8.3 and 8.4, particularly the spikes in negative sentiment and discussions related to "privacy concerns" following key policy announcements.

Evidence: *''The main concern was that the Bank and the Government would use the technology and processes of the platform model to breach users' privacy actively for surveillance purposes, for example to track individuals' spending habits (p.42)."*

Tiered Access and Inclusion:

The concept of tiered access was widely supported to enhance financial inclusion by allowing PIPs to offer different levels of service based on user identification.

Evidence: *''Tiered wallets would allow PIPs to offer less stringent ID requirements for low value digital pound holdings and transactions, supporting consumer choice and the inclusion of those who are only able or willing to provide more limited forms of ID (p.43)."*

Privacy Enhancements:

There was strong advocacy for integrating Privacy-Enhancing Techniques (PETs) to give users greater control over their personal data.

Evidence: *''Respondents supported user control of their data, with a range of views on privacy-enhancing functionality. There was support for the exploration of Privacy-Enhancing Techniques (PETs) (p.43).''*

Regulatory Clarity:

Concerns were raised about how Anti-Money Laundering (AML) and Know Your Customer (KYC) checks would apply.

Evidence: *''Some respondents reported it was unclear how AML/KYC checks would apply to tiered wallets, and that without a uniform approach to identity verification, people may be treated differently by different PIPs (p.43).''*

User Autonomy and Control:
Respondents supported the exploration of PETs to ensure data control lies in users' hands, such as functionalities to opt out of sharing data with third parties by default.

Evidence: *''A majority from industry and civil society groups thought that end-users should have a proactive say in how their data is used (p.43)''*

BoE and HM Treasury's Response: The BoE and HM Treasury acknowledged these concerns and committed to addressing them by exploring tiered access models and integrating PETs. They clarified that PIPs would handle identity verification, with the BoE only accessing anonymized transaction data. They emphasised the necessity of balancing privacy enhancements with regulatory compliance, ensuring that AML/CFT requirements are met without compromising user data protections.

9.3.5 Synthesised Analysis Across Remaining Questions (Q4 to Q12)

The analysis of stakeholder responses to Questions 4 through 12 reveals several themes pivotal to the design and implementation of the digital pound. These themes, when considered alongside the public sentiment analysis presented in Chapter 8, provide a comprehensive understanding of public priorities and concerns, which can then be compared against the BoE's official responses to assess the degree of alignment and identify potential gaps.

9.3.5.1 Summary of the Thematic Analysis of Public's Feedback

Question 4: Provision and Utility of Tiered Access

Themes identified: Tiered Access and Inclusion; Technological Advancements; Economic and Competitive Advantage; Regulatory Clarity.

- Broad support for tiered access to the digital pound, allowing services based on varying levels of user identification to include individuals with limited identification capabilities.
- Emphasis on the need for clear guidelines on AML and KYC requirements to ensure consistent compliance among PIPs.

- Advocacy for users to have autonomy over their privacy settings and control over personal data sharing, enhancing trust and aligning with data protection principles.

Question 5: Embedding Privacy-Enhancing Techniques

Themes identified: User Autonomy and Control; Privacy and Regulatory Balance; Privacy Enhancements; Regulatory Clarity; Economic and Competitive Advantage.

- **Privacy Enhancement:** Strong support for integrating PETs into the digital pound to protect user privacy without hindering transaction functionality.
- **Balancing Privacy with Regulation:** Recognition of the need to ensure privacy measures do not impede AML and Counter Financing of Terrorism (CFT) obligations, requiring a careful balance between user privacy and regulatory compliance.

Question 6: Priority Payment Use Cases

Themes identified: Priority Payment Use Cases; Expansion of Use Cases; Innovative Payment Functionalities; Programmable Payments; Privacy and Surveillance; Experimental and Flexible Development; Offline Payment Capabilities; Regulatory Clarity; Business Model Viability; Economic and Competitive Advantage.

- Consensus on prioritising in-store, online, and person-to-person (P2P) payments to encourage immediate adoption.
- Support for extending the digital pound to business-to-business (B2B) and business-to-consumer (B2C) payments to enhance versatility.
- Emphasis on clear guidelines for programmable features to foster innovation while preventing misuse and ensuring privacy.

Question 7: Holding Limits on Individuals

Themes identified: Holding Limits Justification, Implementation Strategy; Risk Management Mechanism; Regulatory Challenges; Corporate Holding Limits; Global Accessibility; Regulatory Framework.

- Majority supported limits to prevent financial instability due to potential disintermediation of bank deposits.
- Agreement on starting with lower limits and adjusting them based on ongoing assessments to manage risks effectively.

Question 8: Corporate Use of Digital Pounds

Themes identified: Corporate Access Justification; Corporate Holding Limits; Risk Management Mechanism; Wholesale Activity Risks; Institutional Impact; Implementation Strategy.

- Highlighted the role of corporates in promoting adoption and economic activity by using the digital pound.

- Support for higher holding limits for corporates due to operational needs, coupled with risk management mechanisms to mitigate potential financial risks.

Question 9: Non-Resident Access to Digital Pounds

Themes identified: Global Accessibility; Global Accessibility Risks; Regulatory Framework; Corporate Access Strategy.

- Majority favoured allowing non-residents to access the digital pound to enhance its international utility and facilitate cross-border transactions.
- Recognized potential impacts on monetary sovereignty, stressing the need for robust regulatory frameworks to mitigate these risks.

Question 10: Primary Design Objectives

Themes identified: Iterative and Flexible Design; Design Features and Principles; Interoperability vs Extensibility; Cost-Benefit Justification; Stakeholder Engagement; Consumer Education; Privacy Enhancements; Remuneration Structure; Operational Challenges; Transparency and Engagement; Future Use Cases Exploration.

- Advocated for a flexible design process incorporating testing and feedback to refine the digital pound’s functionalities.
- Emphasised the need for inclusive design principles to accommodate diverse user needs, enhancing accessibility and user acceptance.
- Highlighted the importance of ensuring compatibility with existing systems while remaining adaptable to future technological advancements.

Question 11: Supporting Financial Inclusion

Themes identified: Offline Payment Capabilities; Tiered Access and Inclusion; User Assistance and Support; Third-Sector Involvement; Public Digital Wallets; Financial Education and Literacy; Inclusive Design Principles; Cash Access and Financial Inclusion; Micro and Split Payments; Financial Inclusion Initiatives.

- Strong support for the digital pound's potential to include vulnerable and financially excluded groups.
- Importance of educational initiatives to improve digital skills and ensure effective use of the digital pound.
- Need for offline payment options to serve users without reliable digital access, thereby bridging the digital divide.

Question 12: Equality Considerations

Themes identified: Financial Inclusion Priority; Digital Literacy and Education; Offline Payment Capabilities; Third-Sector Collaboration; Public Wallet Options; Accessible Digital Tools; Third-Sector Support; Financial Education; Trust Building; Inclusive Technology Design.

- Concerns about disproportionate effects on vulnerable groups, such as the elderly and individuals with disabilities, highlighting the need for equality impact assessments.
- Advocacy for designing the digital pound to be accessible and user-friendly for all, incorporating features that meet diverse needs.

9.3.5.2 BoE and HM Treasury Response Analysis

The BoE and HM Treasury's responses to consultation Questions 4 through 12 demonstrate a comprehensive alignment with public priorities, particularly in enhancing financial inclusion, ensuring privacy, and fostering collaborative innovation. For instance, an analysis of Q4 highlights their commitment to exploring tiered access based on identification levels, aligning with public support for inclusive financial services. The Bank of England's investigation into tiered access based on varying identification levels mirrors the Bahamas' Sand Dollar, which was created to improve financial inclusion for a widely spread population [9], [44]. The Sand Dollar utilises a wallet system with different tiers, enabling users with minimal identification to access essential financial services while adhering to AML/CFT regulations [270]. Nevertheless, in spite of its innovative framework, the Sand Dollar has faced challenges in gaining traction. Studies suggest that there is a lack of public awareness and trust, with many residents of the Bahamas opting for cash or conventional banking services [9], [270]. This brings to light the necessity for robust public education campaigns and stakeholder engagement to ensure adoption. The absence of defined strategies by the BoE for implementing tiered access and insufficient details regarding the resolution of AML/KYC barriers (Q4) could exacerbate similar challenges in the UK. Similarly, in Q8's analysis, BoE supports broad corporate access and higher holding limits to promote adoption and economic activity (similar to Jamaica's Jam-Dex approach), reflecting respondents' emphasis on inclusivity and diverse use cases.

Across multiple questions, BoE and HM Treasury endorse the integration of PETs (Q5) and emphasise the importance of interoperability in digital wallets (Q6), addressing key public concerns regarding data protection and seamless user experiences. This approach aligns with global best practices but lacks the practical implementation details seen in the Sand Dollar and Jam-Dex, which have prioritised interoperability with existing payment systems [34]. However, the Sand Dollar's limited adoption suggests that interoperability alone is insufficient without strong public trust and awareness [44]

In Q9, their support for non-resident access aligns with international commitments to enhance the digital pound's global utility; however, BoE's specific plans for cross-border integration are unclear, a gap that could limit the digital pound's global appeal. Similarly, in Q11, the BoE underscores initiatives to improve digital literacy and collaborate with third-sector organisations to support financial inclusion, aligns with Bank of Jamaica's approach to partner with NGOs and educational institutions to promote digital literacy, particularly among underserved populations.

Nonetheless, several other omissions are evident. For example, Q10 does not explicitly mention the 'start small' strategy or specific design features, and Q12 lacks detailed measures to ensure equitable access beyond conducting equality impact assessments. This omission is notable given the Sand Dollar's challenges in reaching underserved populations [44], which have been exacerbated by a lack of targeted outreach and education.

The framing consistently presents the digital pound as a flexible and inclusive component of a mixed payment ecosystem, emphasising ongoing technological experimentation and stakeholder collaboration. This approach underscores the authorities' intent to balance innovation with regulatory compliance, as seen in Q10 and Q6. Nonetheless, the responses often fall short of providing detailed implementation plans and specific safeguards, indicating a lack of preparedness to address the complex practicalities of implementing such a system [47]. This lack of specificity could undermine public trust and hinder the successful adoption of the digital pound. A more proactive and transparent approach, outlining concrete steps and specific safeguards, is crucial for building public confidence and ensuring that the digital pound truly serves the needs of all members of society.

These themes, when considered alongside the public sentiment analysis presented in Chapter 8, provide a comprehensive understanding of public priorities and concerns, which can then be compared against the BoE's official responses to assess the degree of alignment and identify potential gaps. For example, the consistent emphasis on privacy across both the public feedback and the sentiment analysis on X (Chapters 7 and 8) highlights the critical importance of this issue for public acceptance of the digital pound.

9.3.6 Quantitative Analysis of Consulting Paper Themes

To enhance the rigour and reproducibility of the research, quantitative analysis was integrated using Python with qualitative thematic analysis. This integration allows for triangulation of findings, strengthening the validity of the results [271].

9.3.6.1 Theme Frequencies

A frequency analysis was conducted based on the coding documented in the codebook to assess the relative prominence of individual themes within the dataset. This involved counting the times each theme appeared across all responses and calculating its relative frequency as a percentage of the total number of themes. This quantitative approach provides a clear overview of the most salient themes in the data, highlighting areas of particular concern or interest among stakeholders.

Figure 9.4 illustrates theme frequencies across all questions, emphasising the dominance of "Economic and Competitive Advantage,'' "Privacy Enhancements,'' and ''Technological Advancements." These themes align with critical aspects of CBDC discussions [13], [47]. Lower frequencies for others indicate diverse, specialised subtopics, reflecting a fragmented thematic landscape with occasional overlaps in key policy-relevant areas.

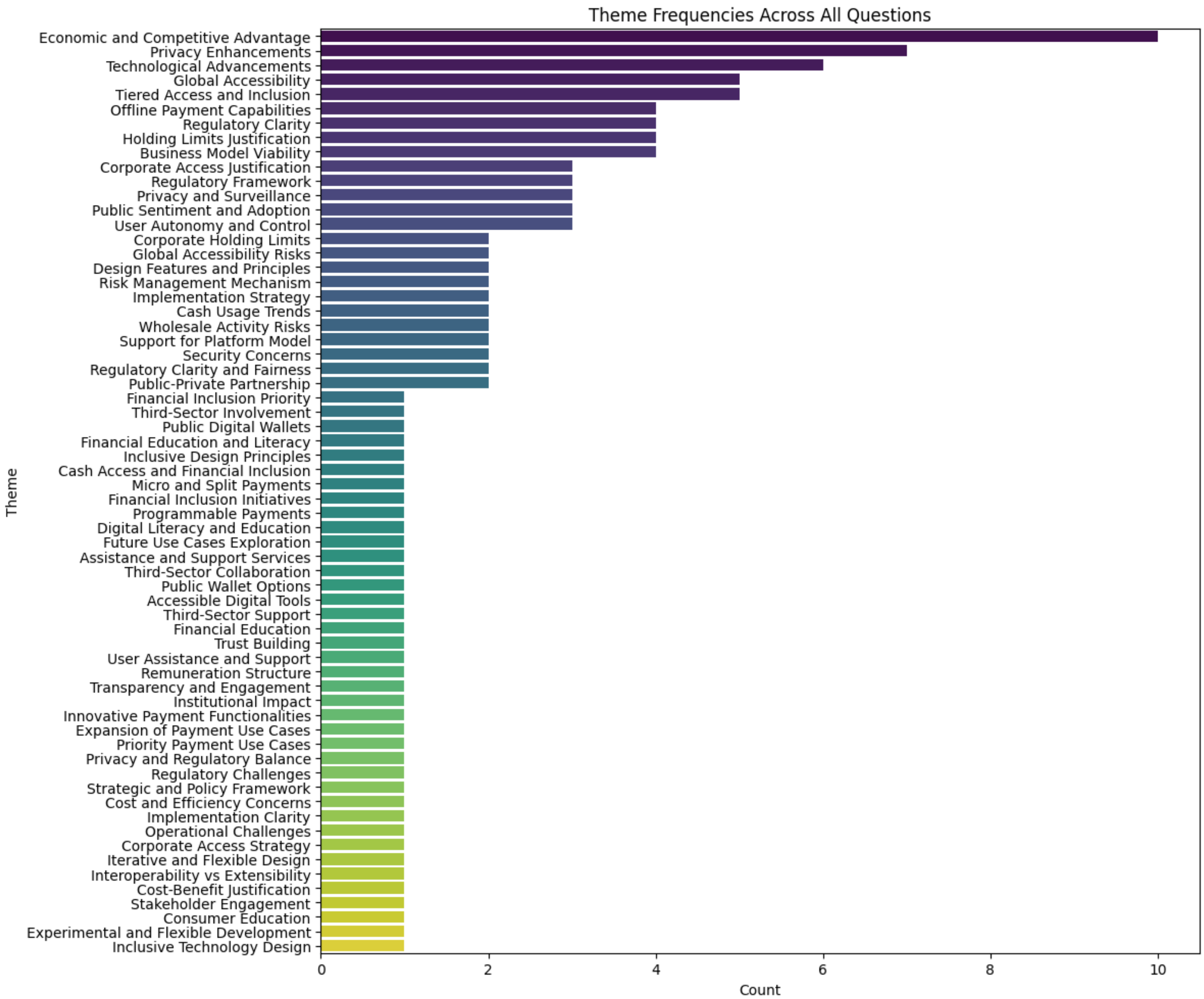


Figure 9.4: Theme frequency across all questions of Consulting Paper.

The prominence of ‘’Economic and Competitive Advantage’’ corroborates earlier qualitative findings (Sections 9.3.2 and 9.3.5), where stakeholders emphasised the potential for the digital pound to enhance payment efficiency, foster innovation, and maintain the UK’s competitive edge in the global financial landscape.

It is important to acknowledge that frequency analysis offers a straightforward metric for assessing theme prominence. While it highlights the most frequently occurring themes, it does not necessarily capture the nuances of meaning or the strength of feeling associated with those themes. Additionally, the process of identifying and labelling themes is inherently subjective, and different researchers may have made different choices, which could affect the results of the frequency analysis [272]. Despite these limitations, frequency analysis can still provide valuable insights into the relative importance of different themes in the data, especially when used in conjunction with qualitative analysis.

9.3.6.2 Theme Co-occurrence Analysis

A co-occurrence analysis was performed to explore the interrelationships between identified themes. This involved examining which themes frequently appear together within the same responses, providing insights into how stakeholders connect different concepts and concerns. The resulting co-occurrence frequencies were visualised using a heatmap (Figure 9.5), allowing for the identification of prominent theme pairings and clusters.

The heatmap shows 297 unique theme pairs; the highest co-occurrence frequency is 3, observed in only two specific pairs:

- **Economic and Competitive Advantage ↔ Technological Advancements:** These reflect the frequent coupling of economic discussions with technological innovation, highlighting their mutual relevance in discourse.
- **Economic and Competitive Advantage ↔ Regulatory Clarity:** This suggests that discussions on economic benefits are closely tied to the need for a well-defined regulatory framework, a point that was also emphasised in the qualitative analysis (Sections 9.3.2 to 9.3.5).

Approximately 96% of theme pairs (285/297) exhibit a single co-occurrence, indicating that most themes are sparsely interlinked or context-specific. These include co-occurrence pairs like Stakeholder Engagement ↔ Inclusive Design Principles, Financial Education and Literacy ↔ Public Sentiment and Adoption. This reflects the complexity and breadth of the domain, where stakeholders focus on individual areas without always integrating them into larger narratives. However, it also suggests a potential need for more integrated discussions that connect these seemingly disparate themes, as they are likely interconnected in practice. For example, effective stakeholder engagement may be crucial for developing inclusive design principles, and public sentiment and adoption are likely influenced by financial education and literacy initiatives.

Moreover, the average co-occurrence count is 1.05, which reflects the general independence of themes in the data. The standard deviation is 0.24 - the low variation further confirms that most co-occurrence counts are clustered around the average, with very few outliers. The heatmap also reveals clusters of themes that co-occur more frequently, such as Regulatory Clarity ↔ Risk Management Mechanisms - this pairing (co-occurrence of 2) indicates a cluster of discussions centred on balancing regulatory frameworks with risk mitigation. This finding reinforces the importance of regulatory considerations, as highlighted in the qualitative analysis, and suggests that stakeholders recognise the complex relationship between regulation and risk in the context of CBDCs. Further analysis of these thematic clusters is presented in Section 9.3.6.3.

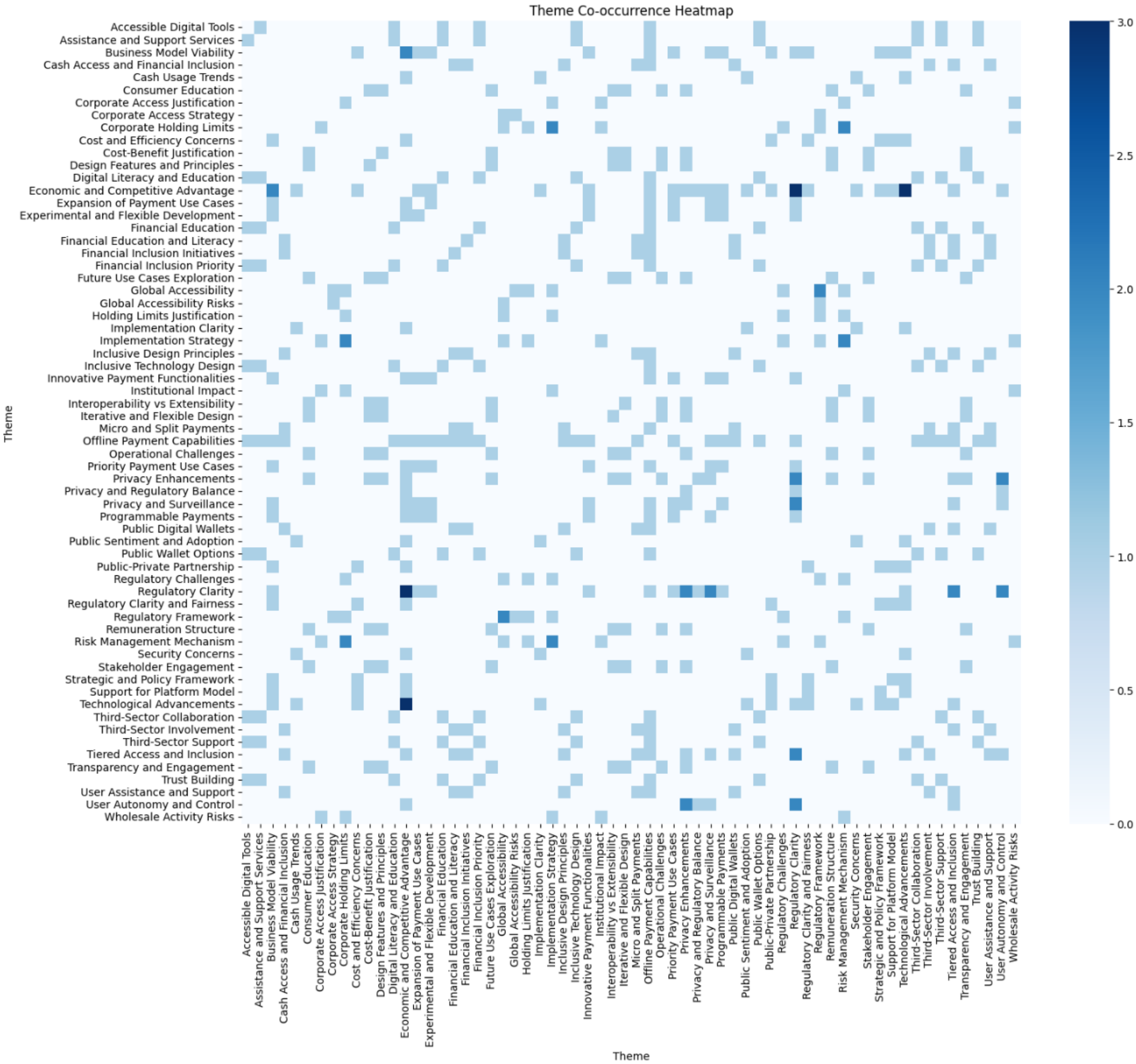


Figure 9.5: Theme co-occurrence heatmap of Consulting Paper questions.

The co-occurrence analysis offers a significant overview of thematic interconnections; nonetheless, it is crucial to recognise that it solely reflects the frequency of topics appearing together, not the nature of their relationship. Two themes may co-occur frequently, but they may be discussed in a positive, negative, or neutral light. Furthermore, the co-occurrence analysis is based on the identified themes, which, as mentioned earlier, are subject to researcher interpretation. Therefore, the results of the co-occurrence analysis should be interpreted in conjunction with the qualitative analysis to gain a more complete understanding of the relationships between themes.

9.3.6.3 Thematic Cluster Trends

To understand how stakeholder concerns evolved across the consultation questions, thematic clusters were manually identified based on semantic similarity. In addition, each thematic cluster's impact, total occurrences, and percentage contributions were quantified (Table 9.3) followed by an analysis of their trends. Nevertheless, this analysis is constrained by the limitations intrinsic to manual clustering, including possible researcher bias and the presumption that all thematic occurrences possess equal significance. Furthermore, it does not capture the nuances of meaning or sentiment associated with the themes.

The identified clusters include:

- Economic and Regulatory
- Accessibility and Support
- Privacy and Security
- Technological Considerations
- Financial Inclusion

Cluster	Total occurrences	Percentage contribution (%)
Economic and Regulatory	51	40.8
Accessibility and Support	24	19.2
Technological Considerations	20	16.0
Privacy and Security	16	12.8
Financial Inclusion	14	11.2

Table 9.3: Cluster frequencies with percentage contribution.

Analysing cluster trends across the consultation questions (Figure 9.6) revealed shifts in stakeholder focus:

- **Economic and Regulatory**: This cluster maintained consistent prominence throughout the consultation, peaking notably in Questions 2, 7, and 10. This reflects the foundational importance of economic and regulatory considerations in the context of the digital pound, aligning with their critical role in policy development. Stakeholders consistently emphasised the need for a robust economic framework and clear regulatory guidelines to support the digital pound's implementation. This finding aligns with the broader literature on CBDC implementation, which highlights the centrality of economic and regulatory factors [35], [83].
- **Technological Considerations**: Themes in this cluster showed intermittent prominence, with peaks in early questions (Questions 1 and 6) and resuming near the end (Questions 11

and 12). This pattern suggests that technology-centric topics were emphasised in response to specific queries about technological infrastructure and future innovations. It indicates stakeholders' interest in how technological choices impact the functionality and success of the digital pound. This focus on technology is consistent with global trends in CBDC development, where technological feasibility and security are paramount concerns [20], [37], [77].

- **Accessibility and Support**: The cluster exhibited steady engagement in the latter half of the consultation, indicating a shift toward inclusion and user-centric design in later questions. This aligns with the qualitative findings in Sections 9.3.5.1 and 9.3.5.2, where stakeholders stressed the importance of designing the digital pound to be accessible to all, particularly vulnerable and financially excluded groups. This emphasis on accessibility reflects a broader societal concern with digital inclusion and the potential for technology to exacerbate existing inequalities [33], [34].
- **Privacy and Security**: Peaks in this cluster occurred in Questions 3, 5, and 8, reflecting focused discussions during targeted queries about data protection and user privacy. Stakeholders expressed significant concerns about surveillance, data misuse, and the need for privacy-enhancing technologies, as detailed in Section 9.3.4. These concerns are consistent with public anxieties about data privacy in the digital age and the potential for surveillance in CBDC systems [36], [37].
- **Financial Inclusion**: This cluster exhibited peaks in Questions 4, 10, and 11, corresponding to moments when social equity and financial democratisation were focal points. Stakeholders advocated for measures to ensure the digital pound promotes financial inclusion, bridging the digital divide and providing equal access to financial services.

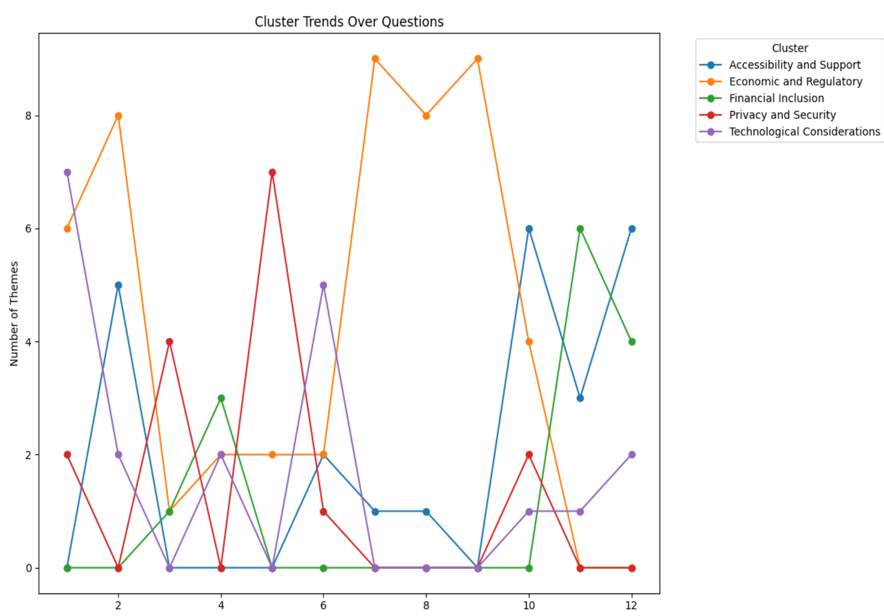


Figure 9.6: Cluster trends over Consulting Paper questions.

These quantitative results presented in this section support the thematic trends identified in Sections 7.8 and 7.9.

9.4 Thematic Findings: Response to the Technology Working Paper

9.4.1 Overview

The Technology Working Paper posed 13 specific questions focusing on the technical considerations essential for designing and implementing the digital pound. These questions covered foundational technology considerations, PETs, performance metrics, ledger design, architectural models, data analysis, APIs, alias services, payment devices, interoperability, programmability, and offline functionality. The emerging technical themes here are consistent with the evolving public discussions on technological and privacy issues observed in Chapters 7 and 8. For instance, the anxieties about data security and the need for robust technological infrastructure, as identified through sentiment analysis (Section 7.2), emotion analysis (Section 7.5), and topic modelling (Section 7.7), are mirrored in the themes emerging from the Technology Working Paper responses.

A total of 70 themes were identified across all 13 questions, of which 63 were unique. The distribution of themes per question is presented in Table 9.4. References to quotations from 2024 Technology Paper [14]are cited with specific page numbers throughout this chapter.

Question No.	No. of themes per question
1	2
2	3
3	3
4	5
5	5
6	5
7	5
8	5
9	5

10	7
11	8
12	8
13	9

Table 9.4: No. of themes per question in Technology Paper.

9.4.2 Question1: Foundational Technology Considerations

Themes identified: Foundational Technology Support; Expanded Technology Considerations (Figure 9.7).

Foundational Technology Support:

Respondents broadly endorsed the six foundational technology considerations outlined in the paper — privacy, security, resilience, performance, extensibility, and energy usage — as critical for ensuring the system's robustness, reliability, and sustainability.

Evidence: *"Most respondents agreed that the six considerations set out in the Technology Working Paper should be the highest priority technology considerations for a digital pound (p.11)."*

This consensus aligns with the literature emphasising the importance of these core principles in designing secure and efficient digital payment systems [273], [274].

Expanded Technology Considerations:

Respondents highlighted additional factors beyond the six foundational aspects: interoperability, usability, accessibility, and scalability. These considerations emphasise the need for the digital pound to integrate seamlessly with existing financial systems and to be user-friendly for all demographics.

Evidence:

- "*Interoperability, usability, accessibility and scalability of a digital pound were cited as additional considerations (p.11)."*
- *"Scalability was cited as important for accommodating increasingly large transaction volumes as future use cases develop and payments needs evolve (p.11)."*

These additional considerations reflect stakeholders' desire for a digital pound that is technically sound, practical, and adaptable to future technological advancements.

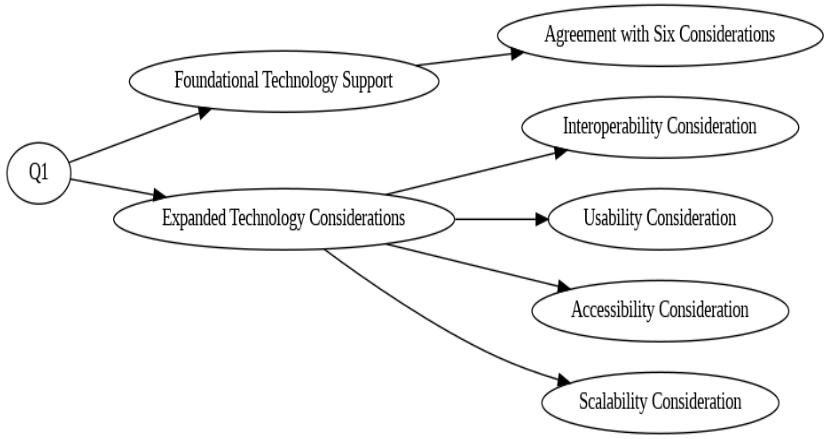


Figure 9.7: Mindmap of Question1.

BoE and HM Treasury's Response Analysis: The authorities acknowledged these themes by incorporating additional considerations into their design principles. The BoE agreed on design principles that include interoperability, usability, accessibility, and scalability, thereby addressing the expanded technology considerations highlighted by respondents. However, while the BoE recognised these additional considerations, specific implementation strategies for achieving scalability and interoperability were not detailed, indicating areas for further development [47]. This omission suggests the BoE needs to provide more concrete plans to address stakeholder concerns fully.

9.4.3 Question 2: Privacy-Enhancing Technologies (PETs)

Themes identified: Privacy Mechanisms; Advanced Privacy Solutions; Expanded Privacy Solutions (Figure 9.8).

Privacy Mechanisms:

There was widespread agreement on the utility of PETs in safeguarding user privacy. Respondents emphasised that PETs are crucial for protecting personal data while enabling necessary data analysis.

Evidence: *"Almost all respondents agreed that PETs could be useful in supporting user privacy (p.11)."*

This aligns with the TAM, which highlights trust and perceived security as fundamental factors influencing user acceptance of new technologies [265], [266].

Advanced Privacy Solutions:

Respondents favoured emerging PETs like zero-knowledge proofs (ZKPs), homomorphic encryption, and blind proofs for their potential to offer robust privacy protections. These advanced technologies can enable transaction verification without revealing sensitive user information.

Evidence: *“Several respondents cited the benefits of emerging types of PETs such as zero-knowledge proofs (ZKPs), homomorphic encryption techniques and blind proofs (p.11).”*

Expanded Privacy Solutions:

Some stakeholders preferred additional PETs like federated learning and secure multi-party computation, among others.

Evidence: *“Additional types of PETs suggested by respondents included confidential computing, secure multi-party computation, federated learning, decentralised identities and group signatures (p.12)."*

This extension highlights a balance between innovation and practicality, with some stakeholders favouring novel methods to ensure protection against unique attack vectors.

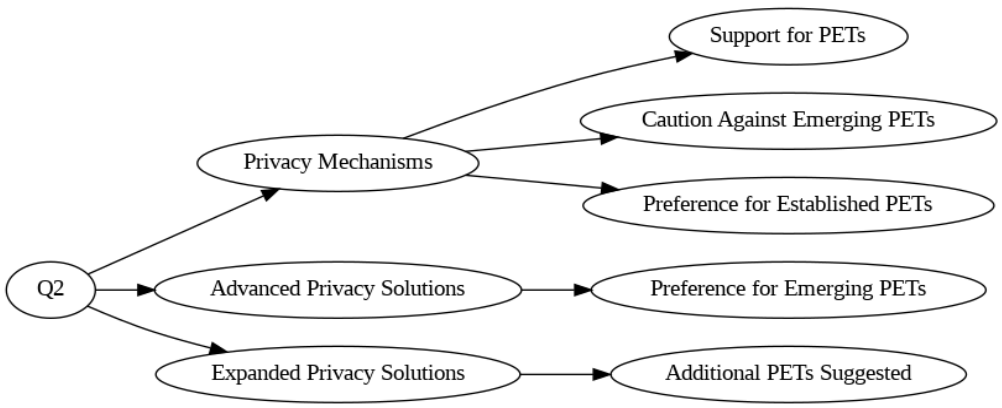


Figure 9.8: Mindmap of Question2.

BoE and HM Treasury’s response analysis: The BoE committed to exploring well-established and emerging PETs to balance innovation and practical implementation. By providing programmability primitives through the API layer, the BoE aims to empower PIPs and External Service Interface Providers (ESIPs) to develop innovative services while maintaining robust privacy protections. However, the BoE’s response lacks detailed strategies for integrating these PETs into the system. There is also an omission regarding the specific regulatory frameworks that will govern the implementation of PETs, indicating further room for more comprehensive planning to fully address stakeholder concerns.

9.4.4 Question4: Architectural Models

Themes Identified: Preferred Architectural Model; Design Considerations; Alternative Architectural Models; Exclusion of Incompatible Models; Security and Inclusivity (Figure 9.9).

Preferred Architectural Model:

Most respondents supported the platform model as the most appropriate architecture for the digital pound, emphasising its effectiveness in meeting policy objectives through a public-private partnership framework.

Evidence: *"The majority of respondents agreed that the platform model was the most appropriate architecture to meet the policy objectives set out in the Consultation Paper (p.14)."*

This preference aligns with collaborative ecosystems in finance that foster innovation through partnership [14].

Design Considerations:

Respondents provided detailed feedback on designing and delivering the platform model's components and activities, advocating for secure, scalable, and interoperable infrastructure.

Evidence: *"Most respondents agreed with the components and activities making up the platform model but had detailed comments on how we might go about designing and delivering them (p.14)."*

Key considerations included ensuring system resilience, adopting open standards, and facilitating seamless integration with existing financial systems.

Alternative Architectural Models:

While the platform model was predominantly favoured, some respondents proposed alternative architectural models that they believed could better address specific requirements, such as decentralised or hybrid models. These alternatives were suggested to enhance aspects like security, scalability, or user control.

Evidence: *"Some respondents proposed alternative models they considered better able to deliver on the objectives set out in the Consultation Paper and the Technology Working Paper (p.14)."*

Exclusion of Incompatible Models:

A few respondents suggested models that the BoE deemed incompatible with the stated policy objectives, such as those based on anonymous bearer instruments. These models were criticised for increasing security risks and excluding less technologically sophisticated users.

Evidence: *"A few respondents suggested models that the Bank judges to not be compatible with the stated policy objectives or design principles, for example models based on anonymous bearer instruments (p.14)."*

Security and Inclusivity:

Emphasis was placed on the need for security measures and inclusivity features to protect users and ensure broad accessibility. Models that compromised on these aspects were largely dismissed.

Evidence: *"Models based on anonymous bearer instruments... increase security risks, would make it difficult for users to recover funds if their devices were lost or stolen, and are not inclusive of the needs of less technologically sophisticated users (p.14)."*

This aligns with the principle that financial technologies should enhance inclusivity and not exacerbate existing inequalities [275].

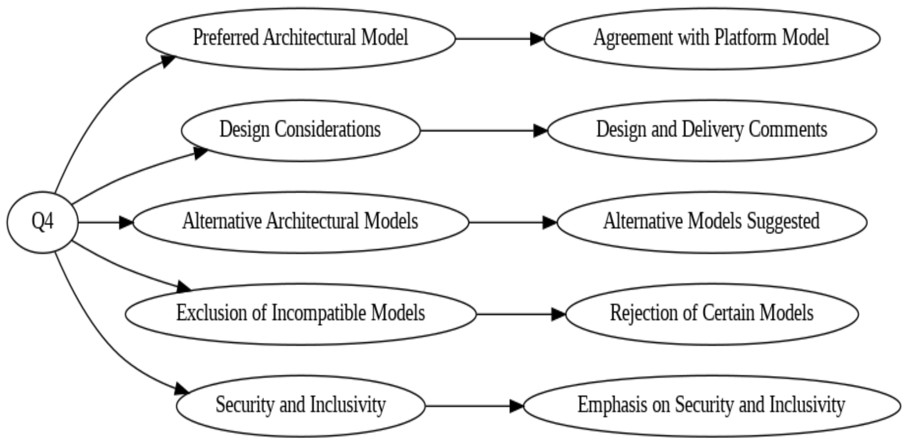


Figure 9.9: Mindmap of Question4.

BoE and HM Treasury's response analysis: The BoE reaffirmed its commitment to the platform model, integrating stakeholder feedback to enhance its design. While prioritising the platform model, the BoE remains open to alternative models that may better address specific requirements, ensuring flexibility in the design phase. However, the BoE's response does not provide detailed consideration of the proposed alternative models or how they might be integrated into the platform model. Additionally, specific strategies for enhancing security and inclusivity within the platform model are not fully elaborated, indicating areas where further development is necessary to address stakeholder concerns comprehensively.

9.4.5 Synthesised Analysis Across Remaining Questions (Q3 and Q5 to Q13)

The analysis of stakeholder responses to the remaining questions reveals several pivotal themes essential to the technical design and implementation of the digital pound.

Question 3: Performance Metrics

Themes identified: Performance Feasibility; Resilience Feasibility; Performance Assurance.

- Approximately half of the respondents deemed the transaction throughput, and uptime targets realistic and appropriate.
- Some respondents expressed that even the higher throughput target of 100,000 transactions per second might not suffice for future needs.
- Calls for detailed plans on testing and measuring performance metrics during the design phase.

Question 5: Alternative Models

Themes identified: Consideration of Alternatives; Preference for Platform Model; Future Consideration of Alternatives; Inclusive Design Process.

- Some respondents believed alternative models could better meet requirements.
- Majority still favoured the platform model.
- Emphasis on engaging stakeholders to evaluate and refine models.

Question 6: Ledger Design

Themes identified: Performance and Scalability; Security and Data Integrity; Reliability and Resilience; Future-Proofing; Architectural Design Choices

- Respondents emphasised the importance of designing the core ledger to handle increasing transaction volumes and maintain high transaction speeds as future use cases develop.
- Ensuring the ledger's security from unauthorised access and maintaining immutable transaction records were identified as vital for maintaining the
- The need for fault-tolerance and redundancy mechanisms was highlighted to ensure continuous operation and resilience against system failures or attacks.
- There was a debate among respondents regarding whether the core ledger should utilise centrally governed technologies, distributed ledger technologies (DLTs), or a hybrid approach.

Question 7: Data Analysis

Themes identified: Data Collection Technologies; Data Analysis Technologies; System Architecture; Data Privacy and Security; Data Integrity.

- Some respondents advocated for using stream processing platforms and tools to facilitate real-time collection and analysis of operational data, enhancing responsiveness.
- Many respondents recommended employing SQL-based systems for data storage, warehousing, and supporting business intelligence and data science operations.
- There was general agreement that analytics should be conducted on a separate platform from the core digital pound system to prevent performance impacts.
- Some respondents emphasised the importance of using management tools to maintain data quality for accurate and reliable analysis.

Question 8: Alias Services

Themes identified: Privacy Enhancement; Interoperability and Usability; Decentralised Hosting; Centralised Hosting; System Simplicity and Efficiency.

- Respondents agreed that disposable aliases could significantly enhance user privacy by preventing the tracking of transactions and protecting user identities.
- The majority favoured hosting the alias service within the wider CBDC ecosystem rather than as part of the Bank-managed infrastructure to distribute management burdens and maintain privacy.
- A minority suggested that the alias service should be hosted centrally to facilitate data recovery and account mobility.
- Some respondents advocated for limiting the number and types of aliases to reduce management overhead and system complexity.

Question 9: API Functionality

Themes identified: API Design Principles; Innovative Payment Functionalities; Transparency and Trust; Developer Support and Innovation; Enhanced Authentication.

- Respondents emphasised that APIs should be secure, standardised, simple to connect to, and easily interoperable with other systems.
- Some respondents suggested that the API layer should support advanced functionalities such as escrow services, push and pull payments, and offline payments.
- A few respondents advocated for APIs that allow users to verify their digital pound holdings directly and obtain verifiable proofs of balance.
- The provision of a sandbox environment was highlighted as critical to enable technologists and stakeholders to test APIs and develop innovative use cases.

Question 10: Payment Devices

Themes identified: Acceptance of Proposed Devices; Seamless Integration; Expanded Payment Modalities; Anticipation of Technological Growth; Standardization and Interoperability; Ongoing Innovation and Assessment; Preference for Established Standards

- Almost all respondents agreed with the proposed list of devices, including smartphones, smart cards, and point-of-sale (POS) systems.
- Respondents emphasised the importance of integrating the digital pound with existing payment infrastructures to ensure easy adoption.
- A few respondents suggested including additional devices and technologies, such as SMS and QR codes, to broaden payment options.
- Several respondents anticipated that the list of supported devices would expand over time to include emerging technologies like those used in autonomous vehicles.
- Respondents encouraged continuous experimentation and assessment of new device form factors to keep the digital pound ecosystem innovative.

Question 11: Interoperability

Themes identified: Leveraging Existing Systems; Targeted Integration Points; Technical Barriers; Regulatory and Legal Considerations; Infrastructure Development; Collaborative Approach; Formalisation of Interoperability; Implementation Strategy.

- Almost all respondents supported using existing payment systems and standards to achieve interoperability between the digital pound and other forms of money.
- Challenges related to aligning messaging standards and protocols were noted as significant obstacles to achieving interoperability.
- Implementing appropriate legislative and regulatory frameworks was emphasised as crucial for supporting interoperability.
- Respondents encouraged the Bank to engage widely with stakeholders and prioritise interoperability during the design phase.
- A small number of respondents suggested that new infrastructure should be developed to better support interoperability.

Question 12: Programmability

Themes identified: Programmability Support; Governance and Control; Infrastructure Provision; Regulatory Oversight; Smart Contract Integration; Infrastructure Concerns; Ethical and Societal Concerns; Innovation and Testing.

- There was broad support for user-initiated programmable payments to foster innovation, simplify transactions, and increase transparency.
- Almost all respondents agreed with the Bank's position that central bank-initiated programmable money should not be implemented.
- Most respondents viewed smart contracts as essential for delivering programmable functionalities within a CBDC system, though some expressed concerns about complexities.
- A few respondents opposed any form of programmability, citing concerns about potential misuse to restrict user payments.
- Some respondents suggested that scheme rules and regulations would be needed to maintain user trust and confidence if programmable payments were offered.

Question 13: Offline Functionality

Themes identified: Offline Functionality Significance; Programmability Support; Data Collection Technologies; Data Analysis Technologies; System Architecture; Data Privacy and Security; Data Integrity; Offline Functionality Support; Offline Implementation Strategies.

- Offline functionality was considered important for supporting financial and digital inclusion and enhancing resilience during system outages.

- Offline capabilities ensure that all population segments, including those with limited access to digital infrastructure, can participate in the digital economy.
- A key consideration was to enhance system resilience by providing a fallback mechanism during technical failures or cyberattacks.
- Respondents suggested methods such as using trusted hardware, ring fencing funds for offline use, and employing communication technologies like NFC, Bluetooth, and QR codes.
- A few respondents raised concerns about potential security risks, fraud, and financial crime associated with offline transactions.

9.4.5.1 BoE and HM Treasury Response Analysis

Analysing the BoE and HM Treasury's responses to Questions 3 and 5-13 reveals a pattern of *acknowledging* stakeholder priorities without always providing *concrete plans* for addressing them. While the responses demonstrate a general alignment with public concerns, a closer examination reveals crucial gaps and ambiguities that require further scrutiny. For example, although the BoE/HM Treasury recognise concerns regarding scalability and performance (Q3), their response does not provide specific strategies for adapting scalability as the digital pound ecosystem develops. Merely acknowledging the issue is not enough; a comprehensive plan detailing how scalability will be proactively managed is crucial for ensuring long-term system resilience.

Regarding the design (Q5), their dedication to an inclusive process and a preference for the platform model is commendable. However, the lack of detailed consideration for alternative models raises concerns about the BoE's willingness to embrace diverse viewpoints and the potential shortcomings of the preferred platform approach. An authentically inclusive design process should examine multiple options and offer clear explanations for the final decision made.

In terms of ledger design (Q6), it is essential to prioritise scalability, security, and extensibility. However, the response does not provide specific details on how to effectively balance these possibly conflicting priorities in practice. What trade-offs are expected, and what measures will be taken to mitigate them? Additionally, while the focus on data privacy and effective analytics (Q7) is commendable, further elaboration on the specific technologies and governance frameworks that will be used to ensure both privacy and optimal data use is necessary.

The consideration of decentralised hosting for alias services (Q8) offers a promising solution for addressing privacy and interoperability issues. Nonetheless, the response lacks information about the specific decentralised technologies being evaluated and their integration within the larger digital pound framework. Likewise, although supporting innovation through APIs and sandboxes (Q9) is vital, the BoE/HM Treasury should offer clearer direction on the types of innovation they seek to promote and how they will align these innovations with overarching policy goals.

The dedication to smoothly integrating with current infrastructure (Q10) and ensuring interoperability (Q11) is essential for gaining user acceptance. Nevertheless, the response does not specify concrete strategies for accomplishing these objectives. Which industry standards will be adopted, and what measures will be taken to guarantee interoperability within a varied and continually changing payment environment? Regarding programmability (Q12), the difference between user-driven and central bank-driven programmability raises significant governance issues. While the rejection of central bank programmability is a commendable measure, the response ought to clarify the regulatory frameworks that will oversee user-driven programmability and how possible risks linked to smart contracts and decentralised finance (DeFi) applications will be addressed.

Finally, although the acknowledgment of offline functionality for promoting financial inclusion (Q13) is a positive indication, the absence of detailed strategies to tackle security issues in offline transactions is a considerable shortcoming. Offline transactions present specific security challenges, and the BoE/HM Treasury must outline a definitive plan to mitigate these risks to uphold the safety and integrity of the system.

9.4.6 Quantitative Analysis of Technology Paper Themes

The same quantitative methods used for the Consulting Paper were applied to the Technology Paper and results are discussed in below sections.

9.4.6.1 Theme Frequencies

Figure 9.10 shows the frequency of themes across all questions, with “Expanded Technology Considerations” and “Privacy Mechanisms” emerging as the most discussed topics. These themes highlight priorities in technological advancements and security concerns within CBDC discussions. Other significant themes include “Performance Feasibility," “Data Collection Technologies,” and "System Architecture,” reflecting a focus on infrastructure and efficiency. Lower-frequency themes like “Innovative Payment Functionalities” and "Offline Implementation Strategies” indicate niche concerns, showcasing a comprehensive but uneven distribution of priorities in CBDC discourse.

The findings from the thematic frequency analysis align strongly with Sections 9.4.2 and 9.4.3 and resonate with broader trends observed in chapters 7 and 8. The bar chart’s dominance of themes like ‘’Expanded Technology Considerations,’’ ‘’Privacy Mechanisms,’’ and ‘’Performance Feasibility’’ reflects stakeholders’ emphasis on foundational technology principles (privacy, security, performance, scalability) and advanced PETs. As detailed in Chapter 7, a range of interconnected themes emerged, including concerns about privacy, security, technological feasibility, and economic implications. Similarly, Sections 9.4.4 and 9.4.5 highlight themes such as ‘’System Architecture,’’ ‘’Interoperability, and Infrastructure Development,’’ which are also prominent in the chart. For instance, ‘’Expanded Technology Considerations’’ maps directly to stakeholders’ emphasis on scalability and integration, while ‘’Privacy Mechanisms’’ and ‘’Data Privacy and Security’’ correspond to concerns raised about robust privacy safeguards, and shows evolving priorities and shifting anxieties, as explored during temporal analysis in Chapter 8. These correlations reinforce the need for scalable, secure, and inclusive CBDC infrastructure, a central theme across the responses.

The limitations of this frequency analysis are the same as those discussed in Section 9.3.6.1. Thus, the quantitative data should be considered as complementary to, not a replacement for, the rich insights provided by the qualitative analysis.

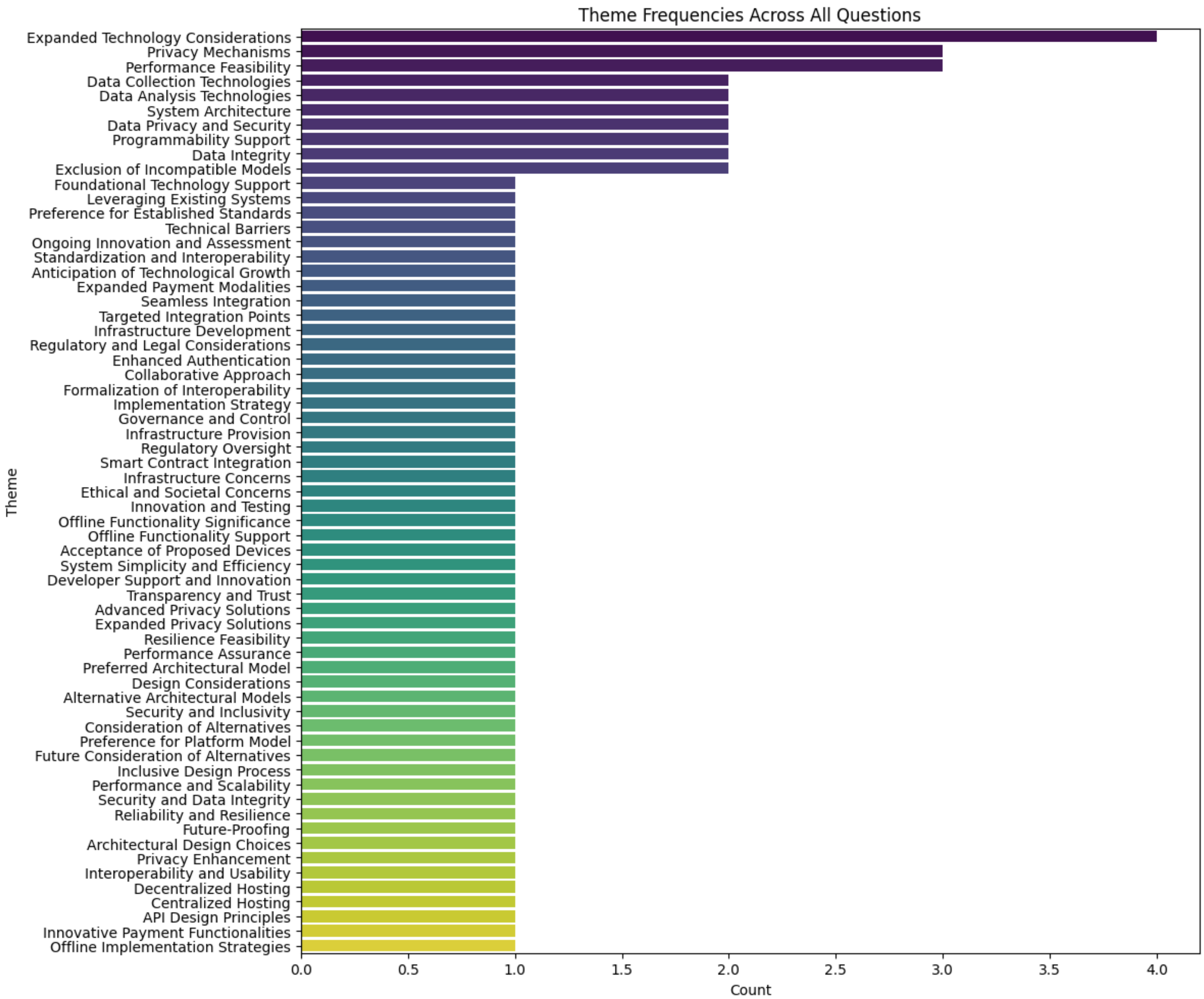


Figure 9.10: Theme frequency across all questions in Technology Working Paper.

9.4.6.2 Theme Co-occurrence Analysis

The heatmap visualises the co-occurrence of 170 thematic pairs in stakeholder discussions (Figure 9.11), with darker cells representing higher co-occurrence frequencies. It provides insights into how often themes are discussed together, offering a quantitative view of interrelationships between concepts critical to designing and implementing a digital pound.

The maximum co-occurrence count is 2, observed in pairs predominantly involving “Data Privacy and Security,” such as:

- Data Privacy and Security ↔ Data Collection Technologies
- Data Privacy and Security ↔ Data Integrity
- Data Privacy and Security ↔ System Architecture

These highlight that privacy concerns are closely tied to data handling (collection, integrity, analysis) and architectural design. Furthermore, themes like "Data Collection Technologies," "Data Analysis Technologies," and "System Architecture" co-occur multiple times with "Data Privacy and Security." An average co-occurrence count of 1.06 indicates that most theme pairs have limited co-occurrence, suggesting specialised and narrowly focused discussions. In addition, a standard deviation of 0.24 refers to low variability and shows that co-occurrence counts are consistent across pairs, with few outliers. 160 out of 170 (approx. 94%) pairs with single co-occurrence reinforces the thematic diversity, as most themes are discussed independently with minimal overlap.

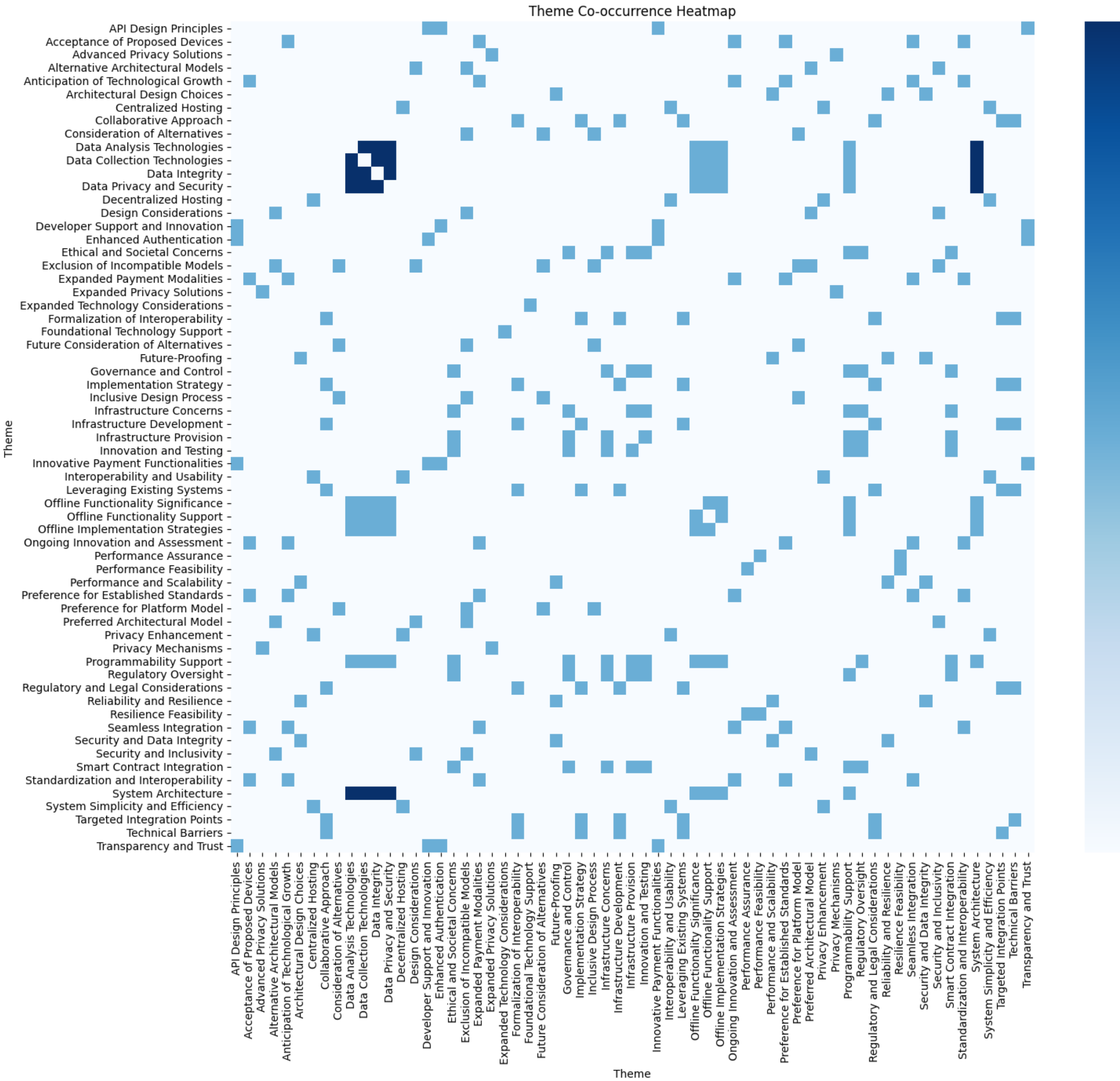


Figure 9.11: Theme co-occurrence heatmap of questions in 2024 Technology Paper.

The frequent mention of ‘’Data Privacy and Security” across multiple themes demonstrates its pivotal role in addressing stakeholder concerns about trust and compliance. The prevalence of single co-occurrence pairs highlights the need for more integrated discussions, as stakeholders address themes in isolation rather than holistically. The identified interconnections can guide design principles, ensuring privacy mechanisms are robustly integrated with data systems and architecture.

9.4.6.3 Thematic Cluster Trends

To trace the evolution of stakeholder concerns across the Technology working paper's questions, thematic clusters were manually identified and analysed. Each cluster's impact, total occurrences, and percentage contributions were quantified to examine trends over time (Figure 9.12).

The identified clusters include:

- Innovation and Integration
- Technical Architecture
- Privacy and Security
- Performance and Reliability
- Regulatory and Legal

These clusters highlight shifting priorities among stakeholders regarding the digital pound:

- **Innovation and Integration:** Gains prominence in Q9–Q12, reflecting stakeholder interest in API functionality (Section 9.4.5), payment device integration, and programmable payments. These later peaks signify a shift toward enabling innovation and adoption through seamless integration. Themes such as developer support, sandbox environments, and advanced payment functionalities align with these peaks.
- **Technological Architecture:** This cluster dominates Q1, reflecting the initial focus on foundational technology considerations (Section 9.4.2). Themes such as interoperability, scalability, and usability align with stakeholder calls for a robust and adaptable architecture. Peaks again in Q5, showing alignment with the emphasis on ledger design (Section 9.4.5), addressing performance and scalability and architectural design choices.
- **Privacy and Security**: Peaks in Q2, linking directly to Section 9.4.3, where Privacy Mechanisms and Advanced Privacy Solutions like ZKPs and homomorphic encryption are central. This underscores privacy as a foundational trust-building element. Another peak in Q8 corresponds to alias services, where privacy enhancement and secure data management were emphasised (Section 9.4.5).
- **Performance and Reliability:** Prominent in Q3 and Q7, tying into discussions in Section 9.4.5 on performance metrics and ledger design. The focus on throughput, uptime targets, and resilience reflects stakeholders' concerns about system robustness and scalability for future use cases.
- **Regulatory and Legal:** Peaks in Q11 and Q13, linking to themes in Section 9.4.5 about interoperability and offline functionality. Regulatory considerations, such as compliance frameworks and ensuring resilience through offline capabilities, dominate these discussions.

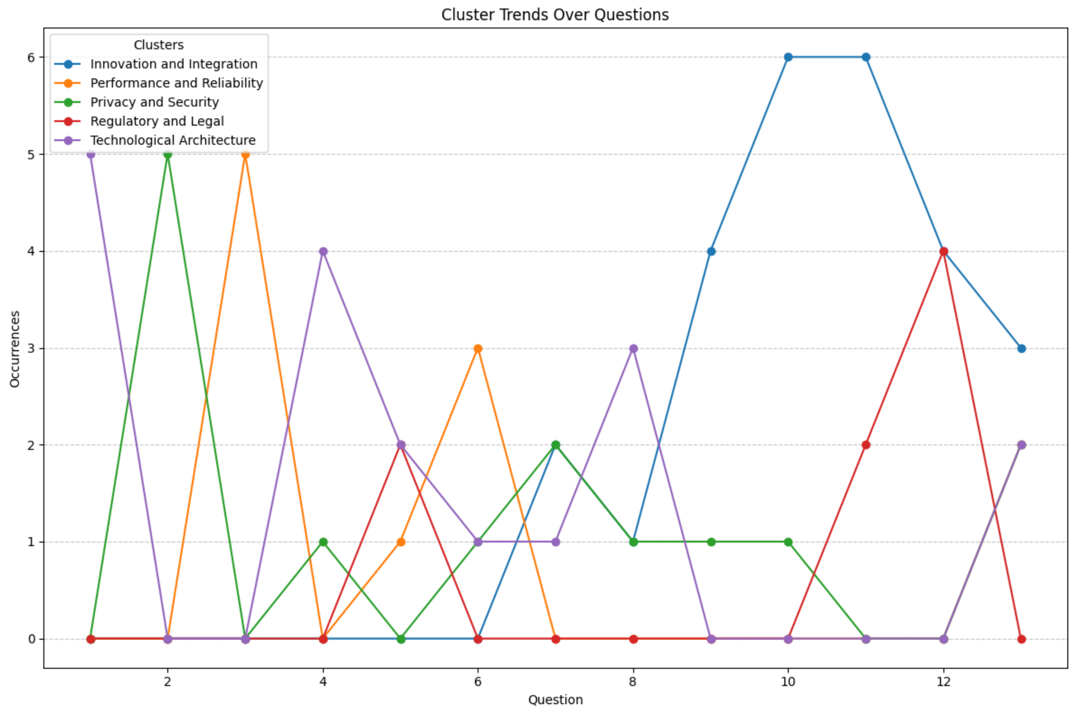


Figure 9.12: Cluster trends over questions in 2024 Technology Paper.

Cluster	Total occurrences	Percentage contribution (%)
Innovation and Integration	26	33.77
Technical Architecture	28	23.38
Privacy and Security	14	18.18
Performance and Reliability	11	14.29
Regulatory and Legal	8	10.39

Table 9.5: Cluster frequencies with percentage contribution in 2024 Technology Paper.

Early questions focus on "Technological Architecture" and "Privacy and Security," reflecting stakeholders' emphasis on ensuring the system's foundations are robust and trustworthy (Sections 9.4.2 and 9.4.3). As the questions progress, clusters like "Innovation and Integration" and "Regulatory and Legal" gain traction, reflecting stakeholders' interest in practical implementation and compliance (Section 9.4.5). This evolution shows a progression from conceptual design to operational readiness and regulatory considerations, mirroring the 'Exploration-Polarisation-Adaptation' sequence observed in Section 8.4.

9.5 Conclusion

The thematic analysis of the Bank of England's responses to the Consultation and Technology Papers reveals a complex interplay between stakeholder concerns and official policy positions. The analysis shows that the authorities are committed to responding to any concerns linked to *financial inclusion, privacy, technological innovation, and regulatory clarity*. For instance, the bank commits to integrating PETs to ensure user autonomy over personal data, emphasising offline functionalities to ensure all segments of society have access to the digital pound. Despite these alignments, several discrepancies and omissions emerge from the lack of detailed plans for implementing key features, such as tiered access models, scalability solutions, and the integration of PETs. Similarly, the authorities do not explicitly address calls for legal measures to protect cash usage; they do not provide concrete solutions or revenue models for PIPs; the debate around enhancing privacy yet abiding by AML and CFT regulations remains inadequately resolved. Stakeholders seek clearer guidelines on how these competing priorities will be balanced. This lack of clarity undermines confidence in the proposed system and raises questions about its long-term viability.

The BoE's overarching narrative portrays the digital pound as a *flexible and inclusive* component of a mixed payment ecosystem, complementing rather than replacing cash. Their communication emphasises *ongoing technological experimentation* and *stakeholder collaboration*, suggesting an iterative and adaptive approach to development. They frame the digital pound as a tool for *innovation* and *modernisation* of the UK's financial infrastructure.

The BoE's communication, while acknowledging key stakeholder concerns, falls short of providing the level of strategic planning and concrete steps necessary to ensure successful implementation and public trust. This gap highlights the importance of providing detailed plans and timelines, having an ongoing dialogue with diverse stakeholders to refine the digital pound's design, and ensuring it genuinely meets the needs of all users. Moreover, striking the right balance between fostering technological innovation and ensuring robust regulatory compliance is essential for navigating the complex landscape of digital finance.

This chapter has highlighted the main themes and narratives communicated by the BoE regarding the digital pound, directly responding to RQ4. This analysis lays the groundwork for the subsequent chapter (addressing RQ5), which will juxtapose these official views with public discussions on X (formerly Twitter) to examine the degree to which public anxieties on social media reflect those noted in the response papers and to pinpoint any additional worries or differing priorities that may need the BoE's attention. By analysing how the BoE presents the digital pound and how this portrayal interacts with public conversations on X, valuable insights into the dynamics of policy communication in today's digital landscape can be obtained to identify alignments and discrepancies in public vs. central bank narratives.

Chapter 10 - Comparative analysis of public discourse on X and the Bank of England’s 2024 response papers

10.1 Introduction

This chapter addresses RQ5 by comparing the key themes and narratives related to the digital pound present in public discourse on X with those identified in the BoE's 2024 response papers (Chapter 9). The analysis aims to identify areas of alignment *(where public concerns and BoE narratives converge)* and divergence *(where they differ)* between public opinion and policy priorities. Identifying these areas of convergence and divergence is crucial for understanding the public's perception of the digital pound and for informing more effective policy communication strategies. Bridging the gap between public concerns and official narratives is essential for building trust and ensuring the successful adoption of the digital currency.

This analysis builds upon the findings presented in Chapters 7 and 8, which provided a comprehensive understanding of public discourse on X. Chapter 7 revealed a predominantly negative sentiment towards the digital pound, particularly in 2023 and 2024, driven by concerns about privacy, government control, and economic stability. These concerns were echoed in the emotion analysis, which showed a rise in fear and anger in response to policy announcements. Chapter 8 further highlighted the dynamic nature of public sentiment, with the temporal analysis demonstrating a distinct *'Exploration-Polarisation-Adaptation'* sequence in response to key policy events.

The comparative approach taken in this study contributes to existing scholarship on participatory policy design that calls for iterative feedback loops and greater stakeholder inclusion in the policymaking process, especially in the highly technical and emerging domains such as digital currencies [276], [277]. Additionally, the chapter advances scholarly debates on participatory policy design, central bank transparency, and the democratic legitimacy of financial innovation as cornerstones of sustainable monetary reforms [273], [277]. This analysis also emphasises the relevance of multistakeholder engagement [278], and communication models that facilitate trust-building and mutual understanding.

Drawing on communication theories, such as framing theory [278] and Grunig and Hunt's [142] models of public relations, this chapter examines the potential for two-way symmetrical communication to enhance understanding and alignment of stakeholder values in the CBDC context. By analysing how the BoE frames the digital pound and how this framing interacts with public discourse on X, valuable insights into the dynamics of policy communication in the digital age can be gained.

10.2 Comparison Framework

A structured comparative framework (Figure 10.1) was developed to systematically compare the public discourse on X with the BoE's 2024 response papers. This framework focuses on mapping the topics generated from the public discourse to the official themes identified in the BoE's documents (as presented in Chapter 9) to determine how both differ in terms of priorities. Topics were assigned to BoE themes based on shared concepts and underlying concerns. This process was primarily manual in nature.

For the purpose of this comparison, X and the BoE response papers *(i.e. both Consulting and Technology Papers are one source)* are treated as two distinct sources, regardless of the number of items (tweets or documents) available for analysis. This approach allows for a direct comparison between the collective public discourse and the official policy narratives, focusing on the thematic content rather than the volume of data from each source.

As the comparative approach integrates both institutional and public narratives, this chapter resonates with emerging best practices in fintech governance research, which advocate multi-stakeholder engagement and collaboration across all actors in the ecosystem [142]. These frameworks reiterate the significance of multidirectional information flows in influencing perceptions and building trust and resemble the two-way symmetrical communication models described in public relations literature [142].

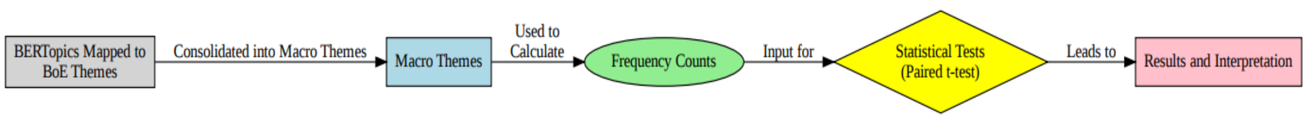


Figure 10.1: Comparative framework.

The key components of the framework are explained as follows:

10.2.1 Mapping BERTopic Topics to BoE Themes

The first step involves mapping the BERTopic topics (detailed in Chapter 8) to the BoE's themes (identified in Chapter 9) to identify possible matches and non-matches. Even though the comparative evaluations in chapter 7 indicated that NMF outperformed LDA in terms of topic coherence and interpretability, BERTopic was selected for this comparative analysis. This decision rests upon its advanced capabilities in handling temporal dynamics and capturing nuanced semantic relationships, which are *essential* for addressing RQ5. LDA assumes a fixed number of topics and struggles with capturing semantic relationships beyond word co-occurrence. NMF, while offering improved interpretability over LDA, also lacks the ability to model topic evolution over time [247]. These methodological considerations align with the significance of sophisticated NLP tools highlighted in the computational social science research to capture temporally sensitive policy debates [279].

The objective of this analysis is to identify alignments and discrepancies in a dynamic public discourse. BERTopic's transformer-based embeddings and hierarchical clustering offer several advantages:

- Given the temporal span of the X data (2020, 2023, and 2024), BERTopic seems crucial as it is adept at modeling topics over time. Neither LDA nor NMF offer this crucial temporal dimension.

- BERTopic is capable of capturing deeper semantic meanings, which NMF may not have identified. This enhanced semantic understanding is crucial for accurately mapping public concerns, often expressed through nuanced language and evolving terminology, to the BoE's policy themes.
- The hierarchical capabilities of BERTopic enhance the interpretability of topics, which is essential for accurate mapping to the BoE's themes.

The following steps were taken for the mapping process:

Step 1: Manual mapping with clear coding protocols

BERTopics were manually mapped to corresponding themes identified in the BoE's response papers and stored in an Excel sheet ('mapping_themes'). The theme mapping process followed definitions of themes provided in codebooks supporting the analysis discussed in Chapter 9. Cross-referencing codebooks and explicit coding protocols are the core tenets of robust qualitative analysis, which helps reduce interpretative bias, enhance replicability and ensure methodological rigor [146], [280].

This helps outline the scope and boundaries of each theme. For example:

- **Regulatory clarity**: Pertains to discussions on the need for clear regulatory frameworks, legal compliance, and oversight mechanisms related to CBDCs.
- **Technological advancements**: Encompasses topics related to technological innovations, system architecture, and infrastructure development for the digital pound.

In addition, inclusion criteria involve including a BoE theme if its primary focus aligns with the defined scope of that BoE theme, including the presence of keywords, phrases, or concepts central to the BoE theme in the BERTopic topic. In contrast, exclusion criteria include excluding BERTopic if it primarily falls outside the defined scope. Moreover, ambiguous topics were carefully reviewed and assigned based on the dominant context i.e., the most frequent or salient keywords and the overall semantic meaning).

Following the above guidelines, each BERTopic topic, characterised by its defining keywords, was manually reviewed and mapped to corresponding BoE themes. For example, BERTopic 28 with keywords: "anonymity, cbdc, degree, mean, transaction" was mapped to BoE Consultation Paper Themes: 'Privacy Enhancements,' 'Privacy and Surveillance' and BoE Technology Working Paper Themes: 'Privacy Mechanisms,' 'Advanced Privacy Solutions.' Similarly, BERTopic topic 91 with keywords: "cbdc, sTable, regulatory, coexistence" was mapped to BoE Consultation Paper Themes: 'Interoperability vs Extensibility,' and 'Regulatory Framework' and BoE Technology Working Paper Theme: 'Standardisation and Interoperability.' A complete 'mapping_themes' table can be accessed via the link provided in Appendix 14.

Step 2: Filtering unmatched topics

Topics from the BERTopic model that did not correspond to any themes in the BoE papers were filtered out. This was achieved by identifying topics where both the consultation and technology working paper themes were marked as '*No Match*.' These topics were excluded from further analysis to focus on relevant thematic overlaps. While these unmatched topics may be of interest in understanding the broader public discourse, they fall outside the scope of RQ5, which focuses on the alignment and discrepancies between public concerns and BoE narratives. This step left 142 BERTopic topics for further analysis.

Step 3: Standardisation of themes

To ensure comparison consistency, all themes and topics were standardised by converting text to lowercase and removing extraneous whitespace.

10.2.2 Defining Macro Themes

Specific themes were grouped into broader categories, known as macro themes, to facilitate meaningful comparisons and statistical analysis. Twelve macro themes were established to balance the need for capturing the intricate thematic landscape while keeping a manageable number of categories for comparison and statistical analysis. This number was arrived at through a repetitive process of examining the initial group of specific themes and categorising them based on their conceptual similarities. Having fewer macro themes could have led to an oversimplification of the analysis, whereas significantly increasing their number would complicate both comparisons and statistical evaluations. These macro themes encompass the wide range of subjects addressed in both the public discourse on X and the BoE's policy responses, thus ensuring that all significant areas are represented in the comparative analysis.

The designated macro themes include:

1. Privacy and Security
2. Regulatory and Legal Considerations
3. Technological Infrastructure and Architecture
4. Implementation Strategies and Challenges
5. User Experience and Financial Inclusion
6. Stakeholder Engagement and Public Sentiment
7. Innovation and Future Use Cases
8. Business Model and Economic Impact
9. Interoperability and Integration
10. Risk Management and Resilience
11. Data Management and Integrity
12. Governance and Control

Each macro theme was precisely defined, outlining its scope and boundaries. A theme (from BERTopic or BoE documents) was mapped to a macro theme if its core focus directly aligned with the macro theme's definition. The presence of keywords related to the macro theme was a strong indicator, but the overall semantic context and the underlying concepts discussed were also considered. In cases of potential overlap between macro themes, the primary focus of the theme was used to determine the most appropriate mapping. The systematic aggregation into macro themes enables the detection of high-level patterns and is consistent with best practices for qualitative analysis while preserving conceptual integrity [146], [281].

For instance:

- **Privacy and Security** combines themes related to data protection, user privacy, and transaction security, which are central concerns in both sources.
- **Technological Infrastructure and Architecture** encompasses themes about system design, technological advancements, and infrastructure development, highlighting the technical underpinnings of a digital currency.
- **Regulatory and Legal Considerations** groups themes about regulatory frameworks, legal compliance, and oversight, reflecting the importance of governance in the implementation of a CBDC.

The next step involved assigning macro themes to both BERTopic topics and BoE themes (Figure 10.2), which was performed manually. Manual mapping was chosen over automated methods to capture subtle distinctions and overlap between themes, which requires expert judgement for accurate categorisation of themes and validity of the comparative analysis. This process involved meticulously reviewing each topic and theme to determine the most appropriate macro theme category based on content and context.

- **For BERTopic Topics**: Each BERTopic topic, characterised by its associated BoE themes (from the mapping process) and defining keywords, was manually assigned to one predefined macro theme based on the primary focus.

Example Topic 10:

    - **Keywords**: “concern, pound, privacy, uk, advance”
    - **Mapped BoE themes**: “Privacy Enhancements,” “Privacy and Surveillance”
    - **Assigned macro theme**: "Privacy and Security”

- **For BoE themes**: Similarly, each theme identified in the BoE Consultation Paper and Technology Working Paper was manually mapped to one macro theme through an examination of the themes to ensure accurate categorisation.

Example:

- **BoE theme**: “Financial Inclusion Priority”
- **Assigned macro theme**: “User Experience and Financial Inclusion”

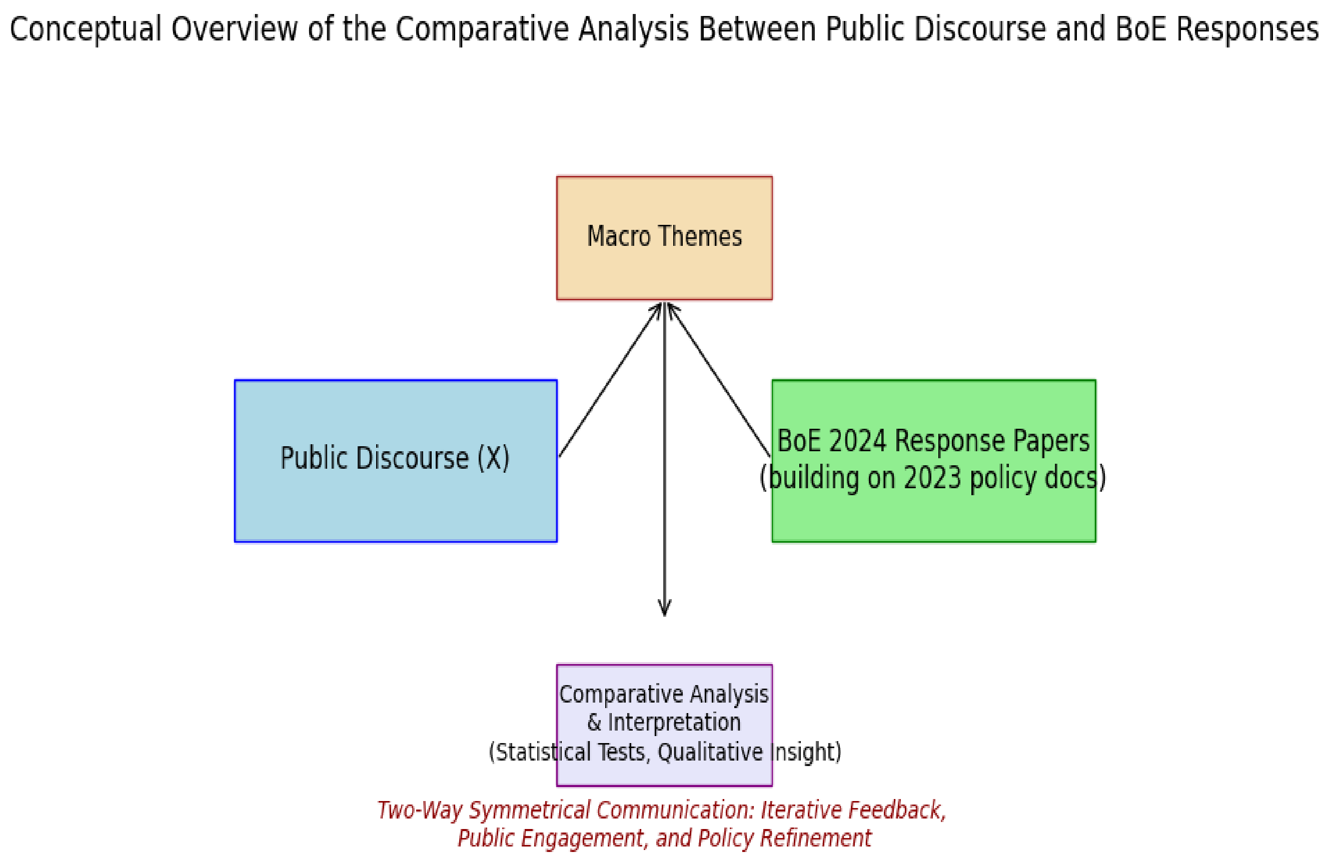


Figure 10.2: Conceptual overview of the comparative analysis.

10.2.3 Frequency Analysis

- **Calculating frequencies**: Calculating the frequency of each macro theme in both the X data and the BoE responses allows us to quantitatively compare the relative prominence of different concerns and priorities. This provides a basis for identifying areas of alignment and divergence between public discourse and official narratives. The frequency of each macro theme was calculated for both sources, which involved counting the number of occurrences of each macro theme in the respective datasets. An "occurrence" of a macro theme in the X data is defined as each instance where a BERTopic topic is assigned to that macro theme. For instance, "Privacy and Security" theme was assigned to topics 10, 27, 28, 40, 58, 59, 61, 62, 77, 86, 96, 97, 98, 103, 113, 116, 125, 133, 134, 135, and 137, resulting in 21 occurrences in the X data. Similarly, in the Consulting Paper and Technology Working Paper, "Privacy and Security" was assigned 5 and 8 times, respectively, leading for a total of 13 occurrences in the BoE responses.
- **Summing up BoE themes' frequencies:** The frequencies of macro theme currencies in Consulting and Technology Working Papers were summed up to derive an aggregate BoE perspective or count. Figure 10.3 shows the distribution of frequencies per paper.

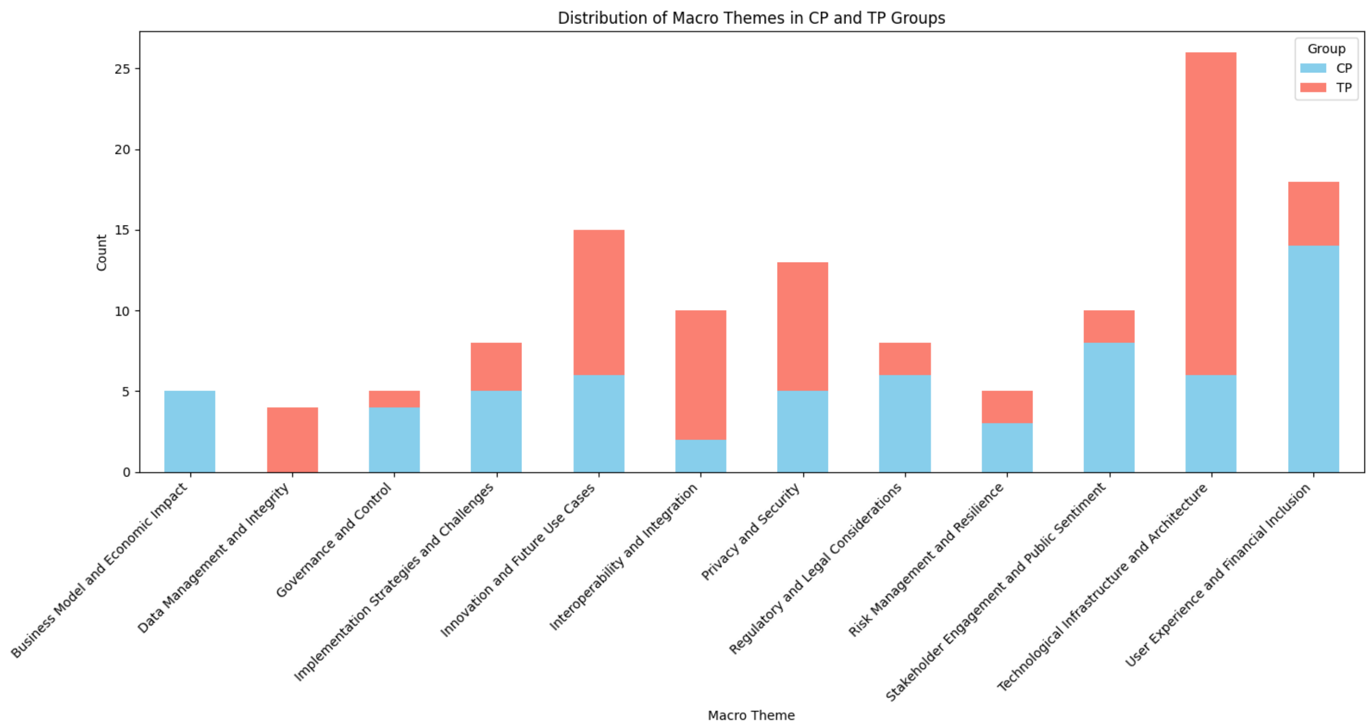


Figure 10.3: Distribution of macro themes in consulting (CP) and Technology Working (TP) groups.

10.2.3.1 Assumptions

The following assumptions were made in conducting the frequency analysis:

- **Independence of themes**: It is assumed that each theme occurrence is independent of others, allowing for straightforward frequency counts.
- **Consistency in assignment**: It is assumed that the manual assignment of macro themes was consistent based on a thorough understanding of the content.
- **Representativeness**: It is assumed that the selected topics and themes are representative of the broader discourse in both data sources (X and policy documents).

While the above assumptions are standard, intercoder reliability checks could be incorporated in future research [282].

10.3 Results: Observed Frequencies and Interpretation

Table 10.1 presents the observed frequencies of each macro theme in both datasets (public discourse and BoE documents). It also includes a "Divergence" column, which quantifies the difference in emphasis on each macro theme by the public and the BoE. Divergence is calculated as the frequency difference for each macro theme between the public discourse and the BoE's aggregated frequencies. Formally:

$$\text{Divergence} = \text{Frequency}_{\text{Public}} - \text{Frequency}_{\text{BoE}}$$

A positive divergence value indicates that the public emphasizes that macro theme more than the BoE does. Conversely, a negative divergence suggests that the BoE prioritises that macro theme more than the public discourse does. A divergence value of zero reflects equal emphasis.

Macro Themes	BERTopic Count (Public)	BoE Count	Divergence (Public - BoE)	Assessment
Business Model and Economic Impact	7	5	2	Aligned: Similar emphasis indicates mutual recognition of economic considerations.
Data Management and Integrity	0	4	-4	Divergent: Absent in public discourse, indicating lack of awareness on data integrity issues.
Governance and Control	12	5	7	Divergent: Public emphasizes governance more, reflecting concerns over control mechanisms.
Implementation Strategies and Challenges	16	8	8	Divergent: Public shows more concern over implementation details than the BoE documents reflect.
Innovation and Future Use Cases	6	15	-9	Divergent: BoE focuses on innovation, public discourse less so, possibly due to immediate concerns.
Interoperability and Integration	10	10	0	Aligned: Equal emphasis suggests shared priority on seamless integration.
Privacy and Security	21	13	8	Aligned: High emphasis in both datasets indicates

				mutual concern over privacy issues.
Regulatory and Legal Considerations	21	8	13	Divergent: More emphasized in public discourse, reflecting public demand for regulatory clarity.
Risk Management and Resilience	3	5	-2	Aligned: Low emphasis in both, but BoE slightly more focused possibly due to technical nature.
Stakeholder Engagement and Public Sentiment	25	10	15	Divergent: High public emphasis indicates desire for engagement, less reflected in BoE documents.
Technological Infrastructure and Architecture	10	26	-16	Divergent: BoE places greater emphasis, indicating a policy focus on technical aspects not mirrored by public interest.
User Experience and Financial Inclusion	9	18	-9	Divergent: BoE emphasizes inclusivity, suggesting the public may be unaware of these efforts.

Table 10.1: Frequency of Macro Themes in Public Discourse and BoE Response Papers.

The observed divergences in macro theme frequencies raise important questions about agenda-setting. While the BoE may prioritise certain technical or policy aspects, the public's emphasis on other areas, like regulatory clarity and stakeholder engagement, suggests a potential disconnect in what each considers to be most important. This discrepancy in agenda-setting could lead to miscommunication and a perception that the BoE is not adequately addressing public concerns [140].

The divergence heatmap (Figure 10.4) visually represents the difference in emphasis between public discourse and BoE documents. Positive divergence (red tones) indicates themes more emphasised in public discourse than BoE documents, such as "Stakeholder Engagement and Public Sentiment" (+15) and "Regulatory and Legal Considerations" (+13), with darker shades of red representing larger differences. Negative divergence (blue tones) highlights themes prioritised by

BoE but less discussed publicly, such as "Technological Infrastructure and Architecture" (-16) and “User Experience and Financial Inclusion” (-9), with darker shades of blue representing larger differences. Neutral values (gray) like “Interoperability and Integration” indicate near-equal emphasis (zero divergence).

The heatmap identifies mismatches and suggests areas where BoE needs to strengthen communication and responsiveness, including engagement and legal considerations. Conversely, BoE’s technical focus indicates institutional priorities that may not resonate with the public, risking a lack of public support or understanding. Balancing these disparities is crucial for effective CBDC adoption and acceptance. This frequency-based evidence aligns with the studies and discussions on CBDCs, where the public wants to know why CBDCs are being introduced and how their privacy rights will be protected, while the institutional narrative revolves around resilience, infrastructure, and design. For instance, Frieman [283] disclosed that public scepticism persists regarding the necessity of a CBDC even though the Bank of Canada has engaged financial sector stakeholders, civil society, and the public through consultations. On the contrary, CBDCs have become a topic of political debate in the United States, which misses broader issues like public participation, governance, and regulatory frameworks, remain underexplored in the discourse [283], [284].

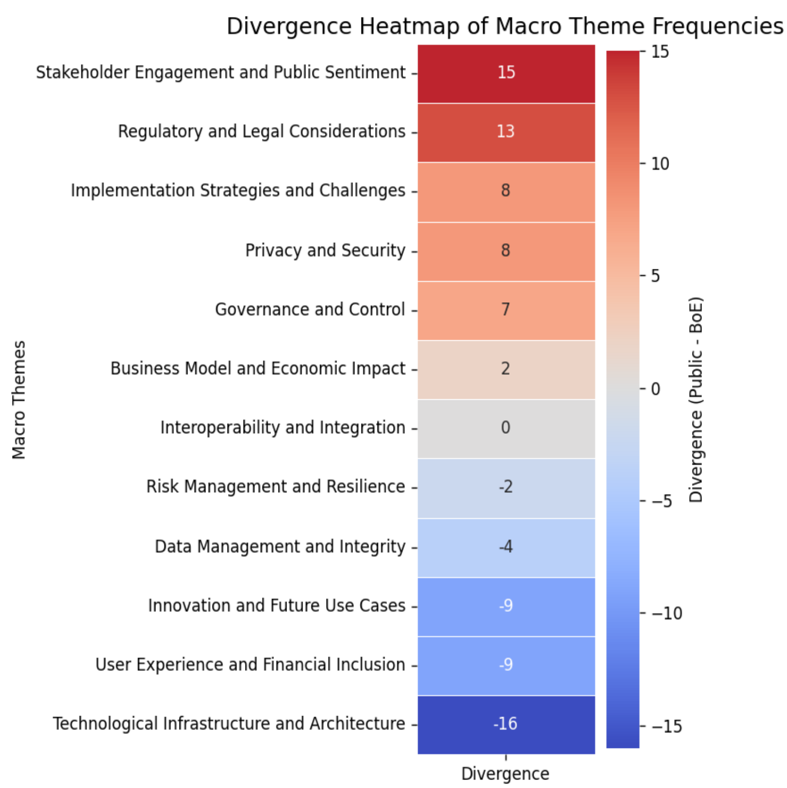


Figure 10.4: Divergence heat map.

The public concerns observed from this analysis mirror concerns highlighted by Pirgmann [285], who noted widespread public resistance to replacing physical currency due to similar anxieties, and the ECB’s public consultation, which identified privacy as the top priority for individuals regarding the digital euro [285]. This convergence suggests that privacy and security are consistent and crucial factors influencing public acceptance of CBDCs across different contexts. The challenges faced by the Bahamian Sand dollar, as highlighted by Ozili [286], provide further context for the findings of this study. The low adoption rate of the Sand dollar due to public concerns about privacy [32] underscores the importance of addressing these issues proactively in the design and implementation of the digital pound. Moreover, the lack of public awareness about CBDCs and reliance solely on pilot results to make policy-related decisions may not prove productive, offering insights regarding CBDC implementation measures for other countries.

10.4 Statistical Analysis and Interpretation

In line with the growing use of mixed methods approaches in finance and regulatory research [287], [288], statistical tests were employed to assess the significance of observed differences in macro theme emphasis between public discourse and BoE documents.

10.4.1 Dependent and Independent Variables and Hypotheses Formulation

Dependent and independent variables for

- **Independent variable**: The source of discourse (*public discussion on X vs. BoE policy documents*), which influences how macro themes are discussed.
- **Dependent variable**: The frequency of macro theme mentions, reflecting the emphasis placed on different topics in public sentiment vs. institutional discourse.

The following null and alternative hypotheses were formulated to guide the statistical analysis, directly addressing RQ5:

- **Null Hypothesis (H_0):** There is no statistically significant difference in the *emphasis* (frequency of discussion) of macro themes between public discourse on X and the BoE's documents, indicating no substantial discrepancy in priorities.
- **Alternative Hypothesis (H_1):** There is a statistically significant difference in the *emphasis* (frequency of discussion) of at least one macro theme between public discourse on X and the BoE's documents, suggesting discrepancies in priorities.

10.4.2 Normality and Homogeneity Assessment

To assess the normality of the frequency distributions for both datasets, the Shapiro-Wilk test was used (as discussed in Chapter 5).

Shapiro-Wilk test results are as follows:

- **Test statistic**: 0.9666
- **P-value**: 0.8720

Interpretation: *A p-value greater than 0.05 suggests that the data does not significantly deviate from a normal distribution. Therefore, the differences between the paired frequencies are normally* *distributed and parametric tests like the paired t-test are Suitable for further analysis. This test is appropriate because it compares the means of two related samples (the public and BoE frequencies for the same macro themes) and accounts for the correlation between them.*

The assumption of homogeneity of variances is inherently met because each difference score is considered independent of the others in the paired t-test, where the differences in frequencies of each macro theme between the two datasets (i.e., differences between paired observations) are analysed. Additionally, given the relatively small sample size (12 macro themes) and the normality test results, the further analysis hinges upon the assumption that variance homogeneity is not violated.

10.4.3 Comparative Testing

Given the results of the normality test, paired t-test was used to compare the same set of macro themes across two datasets.

Paired t-test results:

- **t-statistic**: 0.3938
- **Degrees of Freedom**: 11 (number of pairs minus one)
- **P-value**: 0.7012

Interpretation*: A p-value greater than 0.05 suggests no statistically significant difference exists (at the macro level) between the frequencies observed in the public discourse and the BoE documents.*

In addition, measures like Cohen's d (effect size) and Pearson's correlation coefficient (Figure 10.5) were calculated to understand the magnitude of any observed differences and the relationship between the datasets.

Cohen's d for effect size (magnitude of differences between the datasets):

- **Cohen's d**: 0.1137

Interpretation*: A value of 0.1137 indicates a very small effect size, supporting the conclusion that the difference between the datasets is negligible.*

Pearson Correlation Results:

- **Correlation coefficient**: 0.1076
- **P-value**: 0.7393

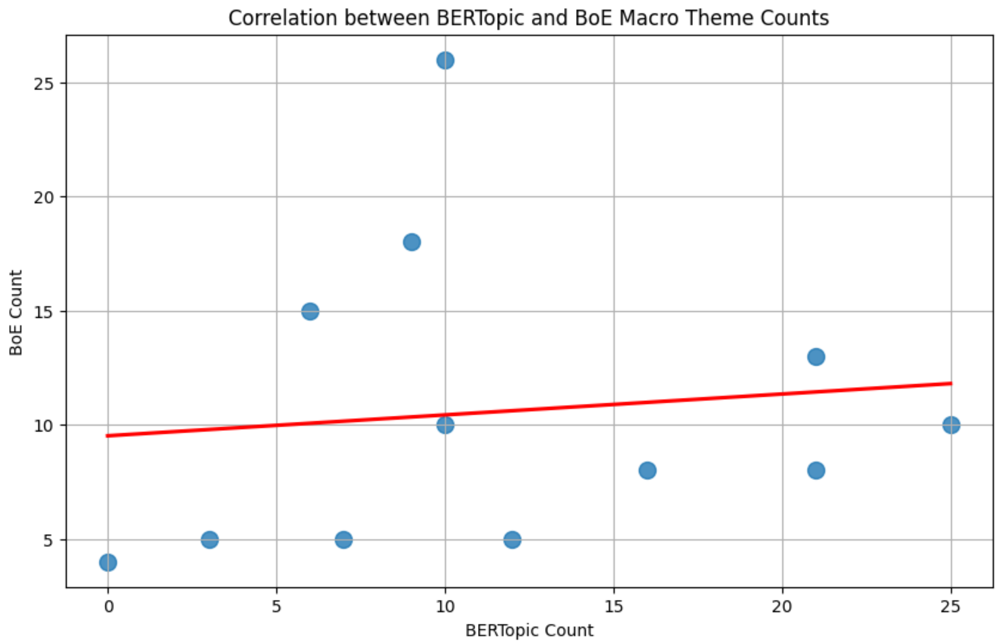


Figure 10.5: Correlation between BERTopic and BoE macro theme counts.

Interpretation*: A correlation coefficient close to 0 indicates a very weak linear relationship between the two sources or datasets. The high p-value suggests that the correlation is not statistically significant, highlighting gaps in the coverage of institutional focus in BERTopic-generated themes. It simply means the two variables tend to move together in a linear fashion. There could be other factors influencing both BERTopic and BoE theme counts. This phenomenon is noted in various regulatory studies, where broad agreement between stakeholders exists alongside profound differences, showing that consensus on certain aspects can mask disagreements about others [286], [289]. The weak and non-significant correlation suggests that the BoE's communication may not be fully aligned with the public's focus. This aligns with existing research [20], [33], [44], which has shown that even when there is broad agreement on some issues, significant differences can exist on others.*

10.5 A Communication Lens on Digital Pound Narratives

This section analyses the alignments and discrepancies between public opinion and policy priorities regarding the digital pound through the lens of established communication theories, directly addressing RQ5. The findings provide insightful observations on both the alignment and divergence between public opinion and policy priorities regarding the digital pound.

The following approach was adopted to evaluate alignment and divergence:

- Themes with similar frequencies across public discourse and BoE documents were initially flagged as potential areas of alignment, indicating a shared emphasis. In contrast, significant discrepancies in frequencies suggested divergence, pointing to differing priorities or concerns.

- The qualitative analysis of sample tweets and policy statements enabled a richer understanding beyond mere frequency counts, shedding light on the substantive nature of these thematic convergences and divergences.
- The insights gained from the frequency and qualitative analyses were then interpreted through the lens of relevant communication theories, including framing theory, agenda-setting theory, and two-way symmetrical communication, to explain the observed patterns of alignment and divergence.

10.5.1 Alignment and Divergence of Themes

10.5.1.1 Alignment of Themes

Through this analysis, *'Privacy and Security'* emerged as the substantial emphasis by both the BoE and the public (Table 1), suggesting the mutual recognition and understanding of data protection and security and the potential erosion of privacy rights in a digital financial landscape. Such alignment resonates with existing scholarship on fintech adoption, where data protection, trust, and security are consistently identified as critical factors influencing their success [290], [291]. Also, this shared concern is consistent with the findings from Chapter 7, where sentiment analysis (Section 7.2) and emotion analysis (Section 7.5) revealed a strong public emphasis on privacy, with negative sentiment and heightened emotions like fear and anger observed in response to perceived threats to data protection. The temporal analysis in Chapter 8 further highlighted the prominence of privacy-related topics, particularly following policy announcements that triggered spikes in public discussion and volatility (Section 8.4).

Furthermore, studies on emerging digital currencies, such as the digital euro and China's digital yuan pilots [292], [293], have similarly highlighted the centrality of privacy concerns in shaping stakeholder perceptions [294]. Even the commonly debated 'privacy paradox' - where people act opposite to their belief in privacy protection - was absent in the context of CBDCs, which suggests that, unlike some other digital technologies, individuals are acutely aware of the potential privacy implications and are not willing to trade privacy for perceived convenience or other benefits [292]. This contrasts with the findings of Jabbar et al. [20], who, using the lens of privacy calculus theory, suggest that individuals may be willing to compromise on privacy if the perceived benefits of CBDCs are substantial. The significant focus on privacy in this study indicates that the public does not easily accept this trade-off, underscoring the importance of adequately resolving privacy concerns.

The public's anxieties about privacy are clearly articulated in social media posts on X, which reveal fears of data breaches, unauthorised surveillance, and the potential for government overreach. For example,

Example 1: *"our currency is already digital. we have notes, but the majority of transactions are done digitally. cbdc's will remove anonymity, allow for data mining and eventually limit transactions based on personal algorithms according to social credit scores in line with SDGs"*

Example 2: *"central bank digital currency (cbdc) will end human freedom. Don't fall for the assurances of safeguards, the promises of anonymity and of data protection. they are all deceptions and diversions."*

The example 1 shows concerns about the removal of anonymity, data mining, and the potential for transaction limitations based on social credit scores. This tweet reveals a deep distrust of centralised control and a fear that a digital pound could be used as a tool for social engineering. Similarly, example 2 expresses a more generalised anxiety about the potential for government surveillance and the erosion of individual freedoms. The use of terms like "human freedom," "deceptions," and "diversions" underscores the level of suspicion and distrust that exists regarding the assurances of safeguards and data protection.

The bank acknowledges these concerns in Technology Working Paper, emphasising the implementation of robust security measures:

''The Bank will explore technological design options that would prevent access to personal data. This will include practical experimentation to assess the benefits and trade-offs of both well-established and emerging PETs."

The Consulting Paper further reinforces this commitment, suggesting that private-sector firms could develop PETs for digital pound wallets:

''Private-sector firms could also explore building Privacy-Enhancing Techniques (PETs) into digital pound wallets to provide users with control of personal data generated by transactions."

These statements indicate that the BoE is aware of the public's anxieties about privacy and is attempting to address them through technical solutions. Nonetheless, although both the public and the Bank of England prioritise privacy, their *framing* of the issue differs (as noted in Section 9.4.5.1). The public discourse focuses on the *sociopolitical* dimensions of privacy, emphasising concerns about surveillance, control, and individual liberties. Section 8.3's sentiment analysis also showed that public trust hinges on explicit data protection commitments. The BoE's framing, on the other hand, tends to be more *technocratic (as seen in the above example)*, focusing on the technical mechanisms for data protection. This gap illustrates a potential 'privacy framing gap' and aligns with framing theory, which suggests that the BoE's technical framing, while perhaps reassuring to experts, may not fully resonate with a public concerned about surveillance and control. This disconnect highlights a limitation in the BoE's communication strategy.

Moreover, this technocratic tendency, reminiscent of the traditional, expert-driven approach to central banking [78], is also observed in the context of government technology initiatives like the UK Identity Cards Scheme [295], which often prioritises technical solutions while downplaying broader concerns about surveillance and civil liberties. Whitley's [295]analysis underscores that technocratic approaches may engender public suspicion and distrust, especially in perceived deficiencies in openness and accountability. The public, as Whitley's work suggests, may view these technical assurances as insufficient, fearing that they do not adequately address the potential for government overreach and the erosion of privacy rights.

This pessimism is reflected in other studies, such as Ghafur et al.'s [296]investigation on public attitudes toward AI and big data in healthcare, which revealed that despite the potential advantages, individuals continue to be apprehensive about data breaches and the misuse of personal information. Similarly, a UK survey on crime-fighting surveillance [297] revealed a strong lack of trust in personal data protection and widespread unease about surveillance practices, particularly when conducted without public knowledge. These findings underscore the recurring tension between the promise of technological solutions and the public's deep-seated concerns about privacy and control. This difference in framing, while perhaps subtle, can lead to miscommunication and a perception of misalignment, even when there is a shared underlying concern.

The pronounced focus on privacy in public discourse, reflected by the quantity and fervour of tweets regarding this subject, indicates that the public is proactively shaping the agenda on this matter. This aligns with agenda-setting theory, which posits that the media (and, increasingly, social media) can influence the public's perception of what issues are important [140]. The BoE's responsiveness in its policy documents indicates that it is aware of and reacting to this public agenda-setting. However, the BoE's response could be interpreted as reactive rather than proactive, suggesting that it is not fully controlling the narrative around privacy.

In addition to the above theme, both the public and the bank emphasised a shared understanding of the necessity for the digital pound to integrate seamlessly with existing financial systems. *'Interoperability and Integration'* (10 occurrences each) reflect a consensus on minimising disruption and enhancing user convenience. For example,

- **Tweet:** ''just in: uk lobby groups and crypto companies support digital pound but call for interoperability with crypto for future-proofing.''
- **BoE (user feedback):** The platform model also considers devices that could allow users to access and use a digital pound, as well as functionality for offline payments, and interoperability with other forms of money.
- **BoE and HM Treasury response (Technology Working Paper):** The platform model also considers devices that could allow users to access and use a digital pound, as well as functionality for offline payments, and interoperability with other forms of money.

However, while the interoperability principle is widely agreed upon, there may still be differences in how the problem is framed. As the tweet demonstrates, the public conversation centres on the usefulness for consumers and highlights the necessity of smooth integration with both conventional finance and the expanding cryptocurrency industry. Despite recognising the value of interoperability, the BoE's statements frequently concentrate on the technical details of attaining it (as discussed in Sections 9.4.5 and 9.4.5.1). If the BoE's technocratic approach fails to sufficiently meet the public's expectations for integration with the wider digital asset sector, this mismatch in framing may cause a divide. Moreover, this potential mismatch in framing highlights the importance of the BoE engaging in more two-way symmetrical communication to understand and address the public's specific expectations regarding interoperability.

Beyond the themes discussed in detail, "Business Model and Economic Impact" and "Risk Management and Resilience" also exhibited alignment. As already explored in Chapter 9, the shared emphasis on economic considerations and systemic resilience suggests a mutual, if less prominent, understanding of these factors' importance in digital pound development. This broader alignment reinforces the potential for constructive dialogue between the public and the BoE.

Overall, alignment on privacy and interoperability shows that certain aspects of two-way symmetrical communication may already be emerging [142]. While these areas of convergence suggest some degree of shared understanding, the nuances in framing and the potential for miscommunication highlight the need for more proactive and symmetrical communication from the BoE. Simply acknowledging public concerns is not enough; the BoE must actively engage with the public to understand and address their specific needs and expectations.

10.5.1.2 Divergence of Themes

Despite some alignments, the analysis reveals significant divergences in thematic emphasis, indicating areas where public concerns and BoE priorities diverge.

A) Public emphasis on regulatory and legal considerations:

The public’s strong focus on 'Regulatory and Legal Considerations' (21 occurrences) contrasts with the BoE's less developed approach to regulatory communication. This divergence is consistent with the specific concerns identified in Chapter 7, where topic modelling (Section 7.7) highlighted public anxieties about regulatory clarity and the potential impact of the digital pound on existing legal frameworks. The volatility analysis in Section 8.4 further demonstrated the sensitivity of public sentiment to regulatory announcements, with spikes in volatility observed following key policy publications.

As discussed in Chapter 9, many respondents requested further clarity on expanding existing legal and regulatory regimes to accommodate a digital pound. Public concerns expressed on X include:

Example 1: *"legal experts foresee a legal overhaul in the #uk as plans for a digital pound gain momentum. introducing a #cbdc requires new legislation and regulatory amendments."*

Example 2: *"best new thing in what sense? with what has been built, is just another digital asset issued by the government… aside the regulatory body issuing it and how it being issued (another discussion) how is it applicable in utility in reference to usd, euro, pounds etc…"*

These instances show that the public wants to understand the *entire* legal and regulatory landscape surrounding the digital pound, not just specific technicalities. Example 1 emphasises the need to comprehend the extent of impending major legislative changes. A more profound concern regarding the digital pound's role in the current financial system is seen in Example 2, which raises concerns about both its usefulness and compatibility with other currencies. This implies that the public is interested in the practical application of the digital pound as well as its relationship to other forms of currency, in addition to the regulations.

Analysis of BoE and HM Treasury responses *(detailed in Sections 9.3.5.2 and 9.4.5.1 and codebooks)* reveals a less developed approach to regulatory *communication.* The BoE has not clarified how AML/KYC requirements will be uniformly applied across tiered wallets, instead framing regulatory development as an ongoing effort in the design phase without detailing specific measures. This suggests a potential disconnect between the BoE's communication strategy and the public's desire for participatory governance. The lack of genuine two-way symmetrical communication could lead to feelings of disenfranchisement and a perception that the BoE is not responsive to public concerns.

Similarly, the lower occurrences of 'Stakeholder Engagement,' in BoE documents highlight that public engagement may not be prioritised to the extent that the public desires, suggesting the need for a two-way symmetrical communication approach [142], wherein feedback loops ensure that public concerns are not passively received and rather *meaningfully* influence policy outcomes. The public's demand for clearer regulatory frameworks and more transparent governance structures aligns with legitimacy theory in financial governance, which stresses that institutional credibility and public trust rely upon robust oversight mechanisms and transparent policymaking [298]. A lack of transparency can undermine public trust, even if the BoE is actively working on regulatory solutions.

The BoE's engagement approach appears more aligned with a two-way asymmetric model [142]; however, regulatory details are not presented as a collaboratively shaped agenda but rather as an ongoing development process, lacking the openness and reciprocity characteristic of *genuine* symmetrical communication. This can create the perception that the BoE is simply informing the public of its decisions rather than engaging in a true dialogue.

B) Institutional prioritisation of technological aspects:

The BoE places significantly more emphasis on 'Technological Infrastructure and Architecture' (26 occurrences) compared to the public discourse (10 occurrences), reflecting the BoE's focus on the technical underpinnings essential for the secure and efficient operation of the digital pound. The BoE delves into technical details about the digital pound, highlighting the importance of system architecture, scalability, resilience, and technological innovation, as seen in the thematic analysis in Chapter 9. On the other hand, public discussion is more centred on the practical implications of the UK CBDC rather than technical implementation. This divergence may be linked to a lack of knowledge and digital illiteracy [286], creating a gap between the BoE's technical focus and the public's more practical concerns.

This aligns with the analysis of 3 tweets extracted from the dataset by using terms like "technological infrastructure," "architecture design," "technology stack," "digital pound infrastructure," "payment architecture," "system design," and "digital currency infrastructure." For example, the phrase '*hope this post helps to clear the cbdc' in a* tweet such as *"thanks dwight as always great examination of the digital pound infrastructure development hope this post helps to clear the cbdc,"* suggests that the post intends to simplify CBDC infrastructure development for the audience, likely addressing common misconceptions or questions. Furthermore, this highlights that user acceptance depends not solely on technical soundness but also on clear frameworks, procedural fairness, and governance norms [299] i.e., the public wants to know *how* the technology will affect them, not just *what* the technology is.

From the lens of two-way symmetrical communication, providing technical details in *an accessible way*, considering public interpretation, and addressing misconceptions should be the priority of institutional actors. Instead, the observed emphasis on technological aspects could reflect an expert-driven, top-down narrative, which does not resonate well with the public information model [142]. While the regulatory authorities are not dismissive of public input, there is a lack of knowledge translation into public trust, practical value, or accessibility. The challenge lies in bridging the gap between technical expertise and public understanding.

C) A Potential convergence space:

‘User Experience and Financial Inclusion' theme receives more attention from the BoE documents (18 occurrences) than public (9 occurrences), indicating that the institution recognises the importance of making the digital pound more accessible. The BoE specifically mentioned in the response papers that the design of the digital pound will consider financial inclusion. This suggests the BoE is aware of the potential for the digital pound to address financial exclusion, but the lower public emphasis may indicate a lack of awareness of these efforts or perhaps scepticism about their effectiveness.

To understand public discourse, 11 tweets discussing “User Experience and Financial Inclusion,” as well as related terms like “underbanked,” “unbanked,” “inclusive ux,” “digital inclusion,” and “financial accessibility,” were extracted and analysed. Examples include:

- **Positive sentiment:** “Interesting news: A digital pound could be a great way to increase financial inclusion and help spur economic growth,” reflecting optimism about the role of digital currencies in fostering inclusivity.
- **Neutral sentiment:** “The digital pound aims to support financial inclusion; Bitcoin is borderless and open access by design,” highlighting the comparison of approaches without strong opinion.
- **Negative sentiment:** “Why would I need to use a CBDC when I can pay with my debit card? I agree that more people who are underbanked and unbanked will,” expressing scepticism about the relevance of certain financial innovations.

The BoE's stated intentions regarding financial inclusion are positive, but these examples express scepticism and comparative assessments with other digital currencies, showcasing a lack of public awareness or engagement and suggesting that its communication strategy in this area needs improvement. A more proactive and transparent approach, involving genuine dialogue with the public *(such as how the digital pound will improve financial inclusion in reality than just stating its aims)*, is essential to build trust and ensure that the digital pound genuinely serves the needs of all citizens.

The statistical analysis, as discussed in the Section 10.4.2., provides quantitative support for these observations but also highlights limitations. No statistical difference in overall theme frequencies suggests a general alignment at a macro level, i.e., aggregated theme frequencies appear similar. However, low correlation implies that specific themes may vary independently, aligning with the qualitative findings of divergence in certain areas and point to an underlying communication model that is not fully symmetrical. These findings underscore the need for the BoE to move beyond a top-down, information-dissemination approach and embrace a more two-way symmetrical model of communication, characterised by genuine dialogue, active listening, and responsiveness to public concerns.

10.5.1.3 Factors Contributing to Misalignment

The divergences identified can be attributed to several interconnected factors:

- **Public priorities**: The public prioritises tangible and immediate concerns that directly impact their daily lives and individual rights. Themes like privacy, security, governance, and regulatory clarity resonate strongly because they are perceived as having direct consequences for individuals. This focus on "lived experience" explains the prominence of these themes in public discussions and highlights the public's active role in agenda-setting. This suggests that the public is not simply a passive recipient of information but actively shapes the discourse around the digital pound, influencing which issues are considered important. The BoE's communication strategy must acknowledge and respond to this public agenda.
- **Information asymmetry**: Less emphasis on themes like technological infrastructure and data management could be due to the lack of public knowledge about technical concepts; this asymmetry highlights the need for the BoE to communicate and emphasise complex topics in an accessible manner. It also underscores the need for the BoE to adopt a "knowledge translation" approach, communicating complex technical topics in an accessible and engaging manner.
- **Communication gaps**: The policy documents are formal in delivery, so they may not effectively convey specific priorities to the public. These documents, while valuable for experts, may not effectively convey key priorities to a broader audience. This suggests a need for more engaging, transparent, and *user-centred* communication strategies that go beyond traditional policy publications because a lack of two-way communication can exacerbate misalignments and erode trust.
- **Framing differences:** As mentioned previously in Section 10.5.1.1, although the public and the Bank of England use similar terminology (such as "privacy"), their perspectives on these matters can vary greatly. These framing differences can lead to misinterpretations and a perception of misalignment, even when there is shared concern. The Bank of England should pay closer attention to how the public perceives these topics and adjust its communication to reflect that understanding. This aligns with framing theory, which posits that how information is presented can significantly influence its interpretation and acceptance [77]

10.5.1.4 Implications for Policy Communication and Recommendations

Above-identified divergences have significant implications for the successful implementation and public acceptance of the digital pound. A critical examination of these implications, through the lens of communication theory, reveals several key areas requiring attention:

A) Policy legitimacy and democratic governance:

BoE's policy formulation may not fully reflect public concerns, as observed from the misalignment in themes such as 'Regulatory and Legal Considerations' and 'Governance and Control.' This disconnect undermines the principles of participatory governance and two-way symmetrical communication. According to Ngo et al. [299], to enhance the acceptance of CBDC projects, governments must carefully address the public's issues and concerns. The heightened emphasis on regulatory clarity and governance in the public feedback indicates public demand for transparent and participatory-policy-making processes. If such concerns remain under-addressed, the public may resist adopting CBDCs, which could undermine the legitimacy of the digital pound initiative. This resonates with the findings from Chapter 7 and 8, where negative sentiment and volatility surrounding policy announcements highlighted the public's desire for greater transparency and involvement in the policymaking process.

Additionally, the public's high emphasis on 'Stakeholder Engagement and Public Sentiment' shows that BoE could benefit from adopting an inclusive strategy where the public and policymakers consult on crucial matters around the digital pound. This may include setting up public discussions, online feedback sessions, or other methods to collect and integrate public opinions. By showing a dedication to community involvement, the BoE can build trust and ensure that the digital pound aligns with the needs and values of the citizens it aims to serve. An approach that encourages participation can also result in more effective policy outcomes, as it allows decision-makers to draw on the varied viewpoints and expertise of the public. Disregarding public opinion risks fostering a perception of top-down governance, which can undermine trust and fuel dissent.

B) Communication strategies, information asymmetry and digital literacy:

An information asymmetry between the BoE and the public exists, as highlighted by the divergence in emphasis on technical themes like 'Technological Infrastructure and Architecture.' As Grunig and Hunt's two-way symmetrical communication model emphasises the importance of mutual understanding and trust-building, the BoE must adopt communication strategies that go beyond simply disseminating information [142]. They need to actively engage with the public, listen to their concerns, and tailor their communication to address those concerns in a clear and accessible manner.

Furthermore, the absence of themes like 'Data Management and Integrity' in public discourse suggests the need to enhance digital literacy. This means that BoE should create educational initiatives and resources that simplify these complex subjects. This will enable citizens to make educated choices regarding the digital pound and engage more actively in community conversations. Closing this knowledge gap is essential for promoting informed agreement and ensuring the public embraces the digital pound.

C) Public trust and acceptance:

The public's greater emphasis on 'Privacy and Security' and 'Regulatory and Legal Considerations' highlights 'Trust' as a critical factor in adopting new financial technologies [266]. Such issues, if unaddressed, could erode trust in the digital pound (Figure 10.6). This requires BoE to demonstrate a clear use for the digital pound and *communicate* that value proposition effectively, its commitment to protecting user privacy and providing clear regulatory frameworks to foster trust and encourage adoption [292]. Furthermore, the BoE should proactively address potential risks and concerns *before* they become major issues, demonstrating a commitment to responsible innovation.

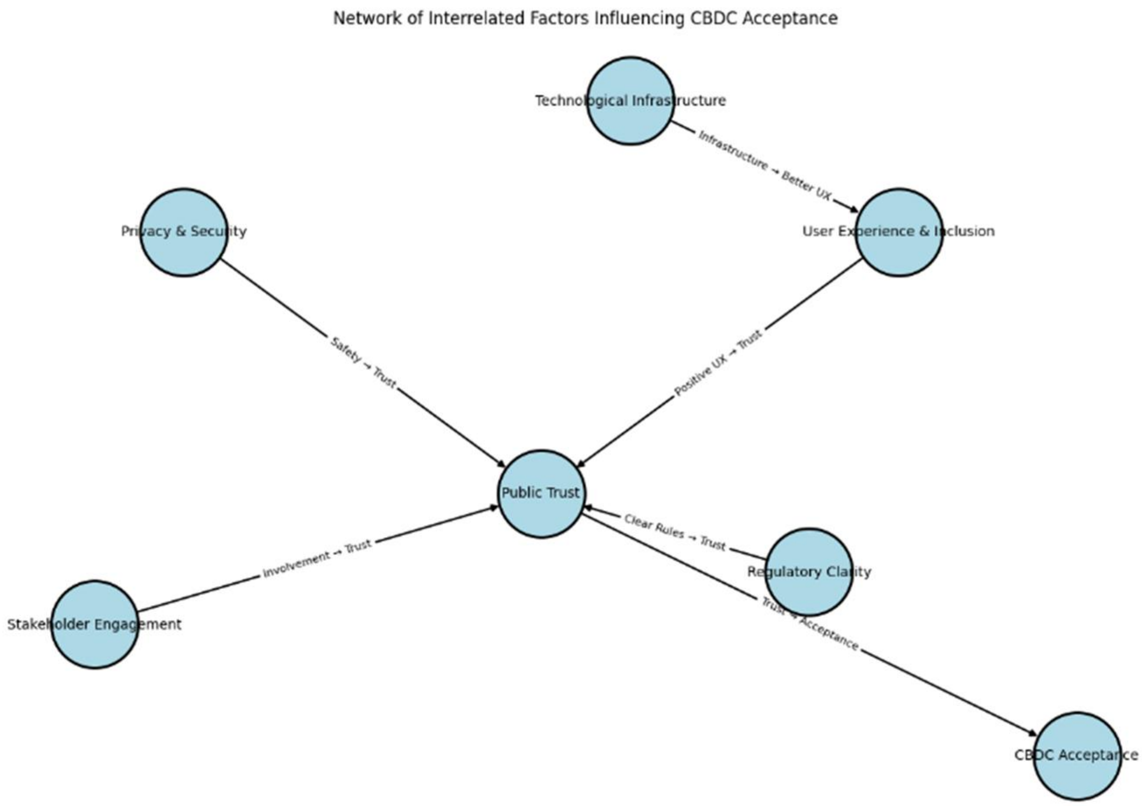


Figure 10.6: Factors influencing CBDC acceptance.

D) Addressing the digital divide and financial inclusion:

The BoE's efforts to enhance 'User Experience and Financial Inclusion' are not recognised by the public, meaning that the BoE must ensure to not exacerbate existing inequalities by introducing digital pound without addressing accessibility issues and promoting inclusive design. In fact, it should be leveraged as an opportunity to *reduce* them.

The digital divide remains a significant barrier to financial inclusion [261]. Without proactively addressing this divide, the digital pound risks being inaccessible to, or underutilised by, digitally vulnerable populations, including the elderly, low-income individuals, and those in rural areas. This could lead to increased resistance against the digital pound, undermining its potential and increasing the risk of policy failure. To foster inclusivity, the BoE should take initiative in tackling the digital divide by enhancing accessibility on various devices and platforms, advancing digital literacy via focused initiatives, collaborating with communities to provide training and support to digitally vulnerable populations, and consistently assess and analyse the impact of the digital pound on financial inclusion and make adjustments as needed.

10.6 Limitations

The comparative analysis offers meaningful insights, yet its evidential power is constrained by several methodological and contextual factors that must temper any generalisation of the results. First, the study drew exclusively on publicly available English-language posts from the platform X. As a result, viewpoints expressed on alternative social media such as Reddit, Facebook, or TikTok, or in offline settings such as news articles and parliamentary hearings, remain outside the scope of observation. The dataset, therefore, reflects the priorities of an engaged, technologically oriented subset of users rather than the attitudes of the wider UK population. Second, the collection windows were centred on three policy milestones in 2020, 2023, and 2024. Public sentiment surrounding central-bank digital currency initiatives can shift in response to economic shocks, electoral cycles, or new pilot results. Hence, the findings reflect public discourse during selected policy windows rather than offering continuous longitudinal coverage, and the thematic balance may evolve as the digital pound debate progresses.

Also, the mapping of 142 BERTopic clusters and numerous Bank of England sub-themes into twelve macro-themes involved interpretive judgment despite the use of explicit coding schemes. Boundaries between concepts such as privacy, data governance, and surveillance are porous, and intercoder reliability was not formally calculated. Researchers and practitioners should therefore treat the thematic allocations as indicative rather than definitive. Moreover, aggregation into macro-themes simplifies the discourse but also masks variation within each category. Fine-grained issues, for instance, distinctions between algorithmic discrimination and general surveillance fears, may be subsumed under a single privacy label. This compression can obscure divergent sub-concerns that possess different policy implications.

The frequency analysis rests on assumptions that each occurrence is independent, that all references carry equal rhetorical weight, and that policy documents faithfully represent institutional priorities. In practice, tweets vary in reach and influence, themes often co-occur, and official publications may reflect negotiated positions rather than precise measures of salience. Also, the institutional and regulatory environment of the Bank of England differs from those of other central banks. Consequently, extrapolating the observed divergence patterns to jurisdictions such as the United States or the Eurozone would require careful local validation.

Finally, the analytical framing draws primarily on communication theories that emphasise stakeholder symmetry, framing, and agenda setting. Alternative theoretical perspectives, for example, behavioural finance or network diffusion, could yield different readings of the same empirical patterns. For all these reasons, the chapter should be read as a demonstration of how structured comparisons can illuminate communication gaps rather than as a definitive map of all public and institutional priorities surrounding a prospective digital pound.

10.7 Conclusion

This chapter addresses RQ5 by comparatively analysing public discourse on X and the BoE 2024 response papers, revealing both alignment (e.g., privacy, interoperability) and critical divergences (e.g., regulatory clarity, technological infrastructure). These divergences underscore the necessity of enhanced two-way communication and engagement for successful digital pound implementation.

Crucially, these findings underscore that the uptake of a CBDC can't be ensured with technical robustness alone. Public acceptance is contingent upon how the regulators address core societal concerns around privacy and security, provide regulatory clarity, establish meaningful stakeholder engagement, and set a clear use case for the digital pound. Failure to respond to these issues effectively could erode public trust in broader monetary policy intervention and, ultimately, the success of a digital pound. The observed disconnect suggests that the messaging efforts of the regulatory authorities should resonate more strongly with the public experiences and anxieties. A communicated approach - rooted in two-way symmetrical engagement is essential rather than only focusing on the legal and engineering aspects of CBDCs.

A top-down communication style, even with consultations, falls short of genuine two-way symmetrical engagement, which requires ongoing dialogue throughout the policy cycle. This difference emphasises that genuine participatory engagement necessitates ongoing, responsive communication with the public and goes beyond sporadic consultations.

This study contributes to the literature by interpreting findings through the lens of communication theories — specifically, Grunig and Hunt's two-way symmetrical model, agenda-setting theory, and framing theory — to the emerging CBDC domain, reaffirming prior findings about the significance of trust, privacy, and regulatory clarity in fintech adoption. Previous research has primarily focused on digital currencies' technological, legal, and financial facets. This study fills a significant gap by empirically investigating the communication dynamics between the public and policymakers, shedding light on the co-evolution of institutional narratives, stakeholder participation, and public perceptions. In doing so, it draws attention to how increased policy transparency, digital literacy, and democratic legitimacy can all work together to influence the adoption and user acceptance of new forms of money like CBDCs.

These observations also result in practical recommendations. First, the BoE and other policy makers should proactively educate the public on intricate technical ideas and solicit critical opinion through easily accessible consultations. Second, by incorporating digital literacy programs into CBDC rollouts, citizens can be better equipped to interact with design and governance issues, lessening information asymmetry and boosting public confidence. Third, the democratic legitimacy of CBDC initiatives can be strengthened by incorporating public sentiment research into policy cycles to help detect new issues early and direct prompt, responsive changes.

In addition, this research lays the groundwork for further empirical inquiry into the evolving relationships between public sentiment and institutional communications as the digital pound efforts step into the next stage. Future research may include comparative international instances, longitudinal evaluations of shifting public narratives, and larger data sources to refine communication strategies and mitigate risks. Ultimately, the true litmus test of a CBDC's viability is a combination of technical sophistication and the extent to which it harmonises with the values, needs, and aspirations of the society it aims to serve.

Chapter 11 - Concluding remarks and future research directions

11.1 Summary

This study aimed to determine how public sentiment on X evolves in response to key policy milestones and whether it aligns or diverges from official Bank of England narratives regarding a prospective digital pound. Preceding chapters addressed the five RQs individually, this chapter translates those insights into overarching conclusions, policy implications, and directions for future research. By intersecting the fields of monetary policy, advanced sentiment analysis, and crucially, communication theory, this study revealed where the public and policymakers agree or disagree on the prospects of a CBDC in the UK across three distinct timelines/periods: the BOE's 2020 discussion paper (Jan-Jun), the 2023 consulting and technology papers (Feb-Jun), and the 2024 responses to public feedback on the 2023 papers (Jan-Mar). This longitudinal approach allowed for a nuanced understanding of how public sentiment evolves in response to key policy milestones and how the BoE's communication strategies influence public perception.

Contributions of this thesis span two primary domains, including data science methodology and the financial/policy domain:

- **Data science (NLP methodology) contributions:** A central methodological contribution of this thesis lies in the creation and release of a policy-specific gold-standard corpus comprising 778 manually annotated tweets that capture the distinctive vocabulary and rhetorical devices employed in discussions of the Bank of England's digital-pound proposal on X. By filling a clear gap in the literature, where extant sentiment resources are either generic (e.g., product reviews) or oriented toward cryptocurrencies, this dataset enables sentiment modelling attuned to the privacy, governance, and sovereignty frames unique to a central-bank digital-currency debate [19], [25], [134]. Leveraging this corpus, the study undertook a systematic comparison of three transformer architectures, including DistilBERT, RoBERTa, and XLM-RoBERTa, fine-tuned under identical conditions; RoBERTa trained for precisely three epochs delivered the best balance of accuracy and generalisation (Chapter 5), reinforcing evidence that robust pre-training coupled with targeted domain adaptation yields superior performance in specialised policy text. Beyond benchmarking, the research subjected the chosen model to adversarial word substitutions and punctuation noise, and interpreted its decisions with LIME, thereby uncovering failure modes around sarcasm, ambiguous valence cues, and orthographic clutter. These robustness and explainability tests together constitute a replicable validation blueprint for high-stakes, policy-oriented NLP, demonstrating how sentiment insights can be audited for reliability before they inform public-sector decision-making. Critically, the temporal analysis of 5,702 tweets spanning three key policy milestones (2020, 2023, and 2024) enabled the identification of a distinctive *"Exploration → Polarisation → Adaptation"* trajectory in social media discourse on X regarding the digital pound. This empirically grounded sequence extends existing research in agenda-setting and framing theory, while also contributing to social media-focused NLP by demonstrating how public sentiment toward a novel monetary technology evolves over time in response to shifting institutional narratives. Beyond tracing sentiment change, the study provides a structured framework for comparing quantitative sentiment trends (derived from transformer-based classification and temporal modelling) with qualitative themes found in official policy documents and public consultation responses. In doing so, it enables a more holistic evaluation of alignment or divergence between public opinion and institutional communication, offering a replicable framework for future policy discourse analysis at the intersection of computational and interpretive methods.

- **Financial/policy domain contributions:** Substantively, the thesis provided a longitudinal mapping of public sentiment on X from 2020 through 2024, revealing an evolution from early cautious interest to heightened negativity in 2023 (after detailed policy proposals), followed by a mildly less negative tone in early 2024. This timeline analysis identified key recurring themes, including privacy, surveillance, financial freedom, etc. (Chapter 9) and showed how these surged or waned in tandem with major BoE announcements. Such insights extend existing understanding of public opinion formation around CBDCs beyond what prior cross-sectional studies offered. Crucially, the findings demonstrate that public sentiment must be closely monitored throughout the policy design lifecycle, particularly before assuming 'product–policy fit'; failing to anticipate or address sentiment-driven resistance risks undermining legitimacy, regardless of technical soundness. Furthermore, by comparatively analysing the public discourse against official BoE communications, the study pinpointed alignments (e.g., a shared emphasis on privacy and security) and clear divergences (e.g., public fears of government control vs. the BoE's focus on innovation and feasibility). These findings highlight specific "framing gaps" between public demands and official narratives, offering empirical evidence of where central bank messaging is or isn't resonating. Lastly, the research applied communication theories, including framing theory, agenda-setting, and Grunig's two-way symmetrical model, to interpret the interplay between the BoE's messaging and public reactions [77], [140], [142]. This theoretical lens provided a deeper explanation for the observed sentiment trends: for instance, how institutional framing of issues like privacy or technological readiness shaped (or clashed with) public views. In doing so, the thesis bridges data science and finance, demonstrating that technical robustness alone cannot guarantee CBDC uptake; *how* the policy is communicated and how public concerns are addressed emerge as indispensable factors for building trust and legitimacy in monetary innovation.

11.2 Revisiting the Research Questions

11.2.1 RQ1

RQ1 (Model Selection and Justification): *Which transformer-based model (DistilBERT, RoBERTa, or XLM-RoBERTa) performs optimally for sentiment analysis on Twitter/X data related to the digital pound when fine-tuned on a domain-specific gold standard dataset, and what are the theoretical and empirical justifications for using fine-tuning in this context?*

Key findings:

In line with domain-adaptation literature [124], [125], [126], the study demonstrated that fine-tuning is crucial for optimising sentiment models on specialised financial discourse like CBDCs. After comparing DistilBERT, RoBERTa, and XLM-RoBERTa (each trained on the new gold-standard dataset of the digital pound–related tweets), RoBERTa emerged as the best performer (Chapter 5), particularly the version trained for three epochs (RoBERTa_3), as discussed in Section 5.9. These results confirm broader evidence that robust pre-training and careful domain-specific fine-tuning capture nuanced language related to CBDC discussions more effectively (Section 5.9.2) than lighter or general-purpose models [125].

The superior performance of RoBERTa_3 aligns with Liu et al. [115], who highlighted RoBERTa's effectiveness in a range of NLP tasks and this study extends those findings to the specialised context of CBDC discourse on social media. In contrast, DistilBERT underperformed slightly, consistent with the trade-off documented by Sanh et al. [116], wherein increased computational efficiency can come at a marginal cost in accuracy. XLM-RoBERTa's multilingual capacity also provided no clear advantage, supporting the idea that a powerful monolingual model often outperforms multilingual counterparts for domain-specific, English-only data.

Statistical analyses further underscore these conclusions. The Kruskal-Wallis H Test yielded an H statistic of 147.539 ($df = 5$, $p < 0.0000$), indicating that there are statistically significant differences in accuracy among the six groups (three models at two epoch levels). Dunn's Post-Hoc Test revealed, for example, that DistilBERT_5 outperformed DistilBERT_3 by 19.90% ($Z = -9.7025$, $p = 0.0000$), while RoBERTa_5 showed a 4.5% improvement over RoBERTa_3 ($Z = -5.0445$, $p = 0.0000$). However, these benefits come with trade-offs. These quantitative gains illustrate that while extended training can boost performance metrics, it may also increase the risk of overfitting, thereby compromising generalisation. Balancing these considerations, the ultimate selection of RoBERTa_3 reflects a deliberate trade-off: its performance is robust and less susceptible to overfitting compared to the 5-epoch variants, which, despite slightly higher raw metrics, exhibit signs of reduced stability on unseen data.

By creating and fine-tuning on a bespoke gold-standard dataset of digital pound–related tweets, this study underscores the importance of domain-specific training. Such an approach addresses limitations observed in broad financial models like FinBERT [129], which often fail to capture the distinctive emerging slang, rhetorical styles, and policy-specific terminology present in social media discourse [74], [100]. The findings thus both confirm earlier suggestions that financial text requires special treatment [56], [57] and challenge any assumption that generic finance embeddings suffice for nuanced policy-related analysis.

In summary, RoBERTa_3 was theoretically justified (due to its robust pre-training and adaptability) and empirically validated as the optimal model for this context, demonstrating how fine-tuning on domain data can significantly enhance sentiment classification performance in niche domains.

11.2.2 RQ2

RQ2 (Model Capabilities and Limitations): *What are the capabilities and limitations of the selected transformer model (identified in RQ1) in accurately predicting sentiments in the digital pound discourse, and how can its robustness and explainability be evaluated using techniques like LIME and robustness testing?*

Key findings:

- **General performance:** This research identified RoBERTa_3 as the most suitable model (as identified in RQ1) for sentiment analysis of UK CBDC tweets due to its balance of performance and generalisation ability, as demonstrated through rigorous statistical analysis and overfitting assessment (Chapter 5). While RoBERTa_5 initially showed higher training performance, it was found to be more susceptible to overfitting, making RoBERTa_3 a more robust choice for real-world applications.

- **Robustness tests:**
  - **Adversarial word substitutions**: Chapter 6 described robustness tests on RoBERTa_3, highlighting that it is relatively resilient to adversarial word substitutions. Accuracy rose from 70% to 72% under specific perturbations (Table 6.1), showing that RoBERTa_3 can handle minor lexical changes. Confusion matrices (Figures 6.5 and 6.6) illustrate that neutral vs. positive misclassifications persist, but negative and positive classes remain relatively stable.
  - **Noise injections (random punctuation):** When random punctuation was injected, accuracy dropped to 68.42% from 70.04%, suggesting that typographic 'noise' can disrupt tokenisation, consistent with prior NLP fragility findings [210], [211]. Specifically, the classification report (Table 6.2) shows that the F1-score for the neutral class dropped from 0.70 on the original data to 0.66 on the noisy data, highlighting the difficulty in distinguishing neutral sentiments in noisy text.
- **Error patterns and analysis**: A local explainability analysis using LIME [136], [137] revealed recurring misclassifications when negative sentiments were superficially cloaked in seemingly positive wording. Privacy concerns (e.g., "digital IDs," "anonymity") expressed via rhetorical or sarcastic constructs often confuse the model. *Section 6.4.3* analysed misclassified tweets. *Tables 6.3* summarise misclassification results, revealing that RoBERTa_3 frequently misclassified negative tweets as neutral or positive, particularly when the negative sentiment is implicit or expressed through complex sentence structures. It also struggles to distinguish between neutral and positive tweets, especially in the presence of noise. Analysis of specific misclassified examples (e.g., tweets about the "great reset" or those containing quotations about anonymity) exposed terms like "freedom," "privacy," "trust" frequently appearing in misclassifications, indicating that certain high-frequency words can lead to overemphasis of positivity (even if context is negative or neutral).
- **Model explainability:** *Section 6.4.4* details the use of LIME to understand which words most influenced RoBERTa_3's predictions. This analysis revealed that while the model generally performs well on clearly expressed sentiment, it struggles with more nuanced language. For example, the confusion matrix (*Figure 6.3*) shows that on the validation data, 8 out of 46 neutral tweets (containing positive-sounding words like "anonymity" or "freedom") were misclassified as positive. It also misclassified negative tweets that use positive words to frame a negative point (e.g., "freedom of money" used to argue against digital IDs). LIME plots (Figures 6.11–6.13) confirm that certain keywords heavily sway the model's output. Consequently, even a fine-tuned model could struggle to distinguish between factual reporting and actual positive/negative sentiment. Similar challenges in sentiment detection have been noted in other finance or policy contexts where implicit negativity is disguised by formal or "positive-sounding" language [56], [74].
- **Implications for policy research**: While RoBERTa_3 is robust for much domain-specific text, caution is necessary when handling sarcasm, irony, or subtle rhetorical inversions — phenomena previously noted in political discourse analysis [89], [90]. Hence, multi-

method sentiment checks (e.g., comparing results with multiple lexicon-based approaches) remain important for critical policy analysis to ensure that implicit negativity is detected accurately. In future, this analysis can be extended by retraining or augmenting the model with more examples of sarcastic and ironic tweets that might further boost performance. Similarly, robustness improvements could involve specialised tokenisation or noise-handling modules that better handle random punctuation or slang common in social media. Overall, RQ2 highlighted that RoBERTa_3 is a powerful tool for understanding public sentiment in this domain, but its predictions must be interpreted with awareness of its limitations. This echoes a broader lesson in applied NLP: model accuracy on paper does not guarantee complete reliability in practice, especially in socially and politically nuanced contexts.

11.2.3 RQ3

RQ3 (Sentiment Trends, Topics, and Evolution Over Time): *What key themes, topics, and sentiment patterns emerge from multifaceted analysis of Twitter/X data concerning the digital pound across three major timelines linked to BoE policy events, and how do sentiments, emotions, and semantic relationships shift in response to these events? What patterns of change appear from the temporal analysis?*

Key findings:

This study employed a multi-faceted approach, comparing RoBERTa_3 against VADER (as discussed in Section 7.2.2) for sentiment classification and utilising LDA, NMF, and BERTopic for topic modelling (see Sections 7.7 and 8.5). VADER frequently misclassifies negative and neutral tweets as positive, highlighting the importance of fine-tuned, domain-focused NLP models in capturing subtle sentiments around the digital pound. While LDA captured broad conceptual themes, NMF offered higher coherence, and BERTopic provided temporal granularity, unearthing evolving subtopics linked to privacy, regulation, and public trust. This longitudinal analysis revealed a clear evolution of discourse from the initial period (2020), exhibiting tentative optimism or "cautious curiosity," which deteriorated significantly in 2023 following the release of the BoE's Consultation and Technology Papers and then stabilising at a mildly less negative state in 2024. However, it is important to note that the following findings are drawn from a limited X dataset, comprising 250 tweets from early 2020, 4,271 tweets from early 2023, and 1,181 tweets from early 2024. These tweets represent a small, self-selected subset of the public. Consequently, the identified patterns reflect the engaged online discourse on X and should not be over-generalised to the entire population.

- **Emerging themes and topics across timelines**:
    - **2020 (exploration)**: Themes like general CBDC design, central bank functions, and blockchain technology were dominant, reflecting a period of cautious curiosity and exploration. Bigrams and trigrams (Section 7.6) often referenced "central bank digital" and "bank digital currency," underscoring foundational interest in CBDC's feasibility. This phase aligned with an early-stage knowledge gap about CBDCs.

- **2023 (polarisation)**: A surge in privacy-related anxieties, strong apprehension about government overreach, and an intensified focus on regulatory implications emerged following the BoE's Consultation and Technology Papers. This suggests that the release of specific policy details triggered a significant shift in public concerns, moving from general exploration to concrete anxieties. Confusion matrices (Section 7.2.2) and topic modelling using BERTopic (Section 8.6) confirmed *heightened negativity* around data control, digital IDs, and perceived erosion of personal freedoms. As noted in Section 4.3.4, this shift may also be reflected in annotator disagreements, potentially indicating a rise in polarised opinions as the public grapples with the implications of specific policies.
- **2024 (adaptation)**: Continuing privacy anxieties, plus talk of "control" and "surveillance," offset by discussions of "business opportunities" and "financial freedom" surfaced, reflecting a maturing debate. In addition, response papers from the BoE and HM Treasury slightly alleviated negativity, as shown by an increase in positive sentiment from 14.6% in 2023 to 17.7% in 2024 (Sections 8.3 and 8.4). Nonetheless, overall negativity remained high at 50.7%, highlighting a persistent undercurrent of scepticism towards the digital pound despite some improvement in public optimism. This indicates a process of adaptation and a diversification of public perspectives as the debate matured.

- **Trajectory of sentiment trends:**
  - **Downward shift in sentiment:** A clear trajectory toward negative sentiment emerged over the study period. Initially, 2020 discussions displayed cautious optimism (mean sentiment score of 0.94). However, following the 2023 release of the BoE's Consultation and Technology Papers, sentiment declined sharply (mean score of 0.69), showing rising scepticism. Although sentiment improved marginally in 2024 (mean score of 0.67) after the release of BoE and HM Treasury's response papers, overall negativity (50.7%) remained dominant. Polarity analysis based on RoBERTa *(Section 7.10.2)* further confirmed this downward trend, shifting from −0.11 in 2020 to −0.30 by 2024, reinforcing the impact of policy details on public attitudes. This quantitative evidence underscores the significant impact of policy announcements on public sentiment and highlights the persistent negativity surrounding the digital pound. This study extends earlier Twitter-based CBDC sentiment research by scholars like Prodan et al. [100] by incorporating domain-specific annotated datasets, robust noise handling, and, crucially, linking sentiment shifts directly to official BoE announcements.
  - **Event-driven changes:** Temporal analysis in Chapter 8 shows that public sentiment dropped markedly with each key policy milestone (e.g., the 2023 Consultation Paper), reflecting an uptick in volatility as debates intensified. Although sentiment rebounded slightly in 2024, suggesting partial reassurance from the BoE and HM Treasury's clarifications, it did not fully dispel entrenched concerns. In particular, weekly and monthly volatility metrics *(Section 8.4)* confirm that heightened discussion often coincided with spikes in negative

sentiment, indicating persistent scepticism even amidst incremental improvements in public optimism.

- **Semantic network shifts**: Semantic network analysis revealed a clear evolution in the language used to discuss the digital pound. Early references (2020) to "bank-central-digital-currency" gave way to 2023 discourse centring on "digital-pound," *(Section 7.9)* reflecting heightened public familiarity [29], [30]. By 2024, nodes like "digital-privacy" co-occurred frequently with "control," revealing a pronounced orientation towards surveillance fears [37]. These semantic shifts demonstrate how public understanding and concerns evolve as the digital pound proposal progresses. Complementing these semantic shifts, the analysis of sentiment association with predefined aspects *(Section 7.11)* reveals a parallel trend. The radar chart (Figure 7.13) shows a marked increase in negative sentiment association with privacy, anonymity, surveillance, and regulation in 2023, mirroring the emergence of 'digital-pound' as a central term and the growing focus on control. These negative associations, particularly with privacy (rising from below 500 in 2020 to approximately 3000-3500 in 2023) and surveillance (increasing to around 2500-3000), become significantly more pronounced. While these negative associations moderated somewhat in 2024 (with privacy, for example, decreasing to the 2000-2500 range), similar to the overall sentiment trend, they remained elevated compared to 2020, suggesting that underlying concerns persisted even as the discourse adapted.
- **Emotional patterning and change points**: The analysis revealed a strong connection between key policy milestones and emotional responses, supporting the "issue-attention cycle" [236] and echoing event-driven sentiment shifts observed in earlier policy research [81], [82]. Applying change-point detection with the PELT algorithm *(Section 8.7),* this study identified notable sentiment shifts in 2023, specifically around February 4th, February 7th, and February 27th, coinciding with the release and discussion of the BoE's Consultation and Technology Papers. Another shift on June 18th, 2023, pointed to evolving views on practical implementation and financial impacts. Further change points appeared on February 7th and March 15th, 2024, following the BoE and HM Treasury's feedback responses or policy announcements. Statistical tests, Kruskal-Wallis and Dunn's post-hoc comparisons, confirmed significant differences ($p < 0.05$) between 2020 vs. 2023/2024, but no statistically significant difference between 2023 and 2024 ($p = 0.3469$), suggesting persisting negativity despite partial sentiment improvement. NRCLex-based emotion analysis (*Section 8.5)* showed a notable spike in fear, sadness, and anger in 2023, presumably triggered by the release of Consultation and Technology Papers, while trust and anticipation remained comparatively steady across all periods. Although negative emotions declined slightly in 2024, they persisted overall, indicating a less polarised yet still cautious public discourse. These findings highlight the importance of considering emotional responses in policy communication and the need for proactive strategies to address public anxieties.

Overall, these findings reveal an "Exploration-Polarisation-Adaptation" sequence in public sentiment. This framework provides a valuable lens for understanding the dynamics of public opinion formation in response to complex policy initiatives. The approach adopted for this research

— via a domain-specific gold-standard dataset, advanced robustness checks, and multi-method topic modelling — aligns with the focus of some Twitter-based sentiment analysis studies [74], [104], [105] yet diverges by illustrating direct links between official BoE announcements and temporal sentiment shifts. As such, the study underscores the importance of timely, transparent communication around newly released CBDC details, extending prior works that have documented similar phenomena in broader fintech discussions.

11.2.4 RQ4

RQ4 (Analysis of Official Communications): *What are the key themes and narratives presented in the official Bank of England policy documents and responses related to the digital pound?*

Key findings:

Chapter 9's qualitative thematic analysis of the questions and BoE's response to the public feedback discussed in the 2024 response papers highlighted several central narratives:

- **Innovation vs. risk**: The BoE frames the digital pound as an innovative means to safeguard monetary sovereignty and maintain the UK's position in the global CBDC landscape [83], [84], paralleling findings in Chapter 7 *(Sections 7.3, 7.7),* where public discourse similarly connected the digital pound's adoption to broader economic and geopolitical considerations. This framing emphasises the potential benefits of the digital pound (e.g., speed, efficiency, innovation) while acknowledging the inherent risks associated with introducing a new form of digital currency (e.g., financial instability, technological vulnerabilities). As analysed in Chapter 8's temporal trends (Sections 8.4 and 8.7), these risks correlate with volatility in public sentiment, where negativity spikes often emerge following policy milestones clarifying risk factors. In addition, official documents devote pronounced focus to infrastructure (e.g., ledger design, scalability), reflecting a technocratic lens on policy goals. Although user experience and social dimensions appear in the BoE's narrative, public awareness of these issues seems limited, suggesting that technical emphasis has overshadowed direct engagement on societal implications.
- **Privacy acknowledgement but vague mechanisms**: Recurrent mentions of privacy and data protection appear, but specific implementation details remain ambiguous, paralleling criticisms voiced about privacy provisions in other CBDC projects (e.g., digital euro) [36], [292]. This gap between acknowledging the importance of privacy and an ambiguous roadmap to address those concerns shows that public/respondents demanded more concrete security guarantees than the BoE currently provides.
- **Financial inclusion**: The BoE stresses a public–private partnership, emphasising that a well-designed digital pound could bring unbanked communities into the financial mainstream [33], [34]. However, the lack of detailed strategies (e.g., offline usage, bridging digital literacy gaps) and measurable targets for achieving financial inclusion weakens this narrative and raises questions about the BoE's approach to addressing the digital divide.
- **Coexistence with cash**: Mirroring other countries' approaches [9], [34], the Bank assures stakeholders that digital currency would complement, not replace, physical notes and coins. This narrative aims to reassure the public that the introduction of a digital pound will not

disrupt existing payment habits or undermine the role of cash in society. Chapter 8's analysis of monthly sentiment distributions (Section 8.4.5) indicates that such assurances do mitigate some negativity spikes related to "cash displacement" concerns. Although "cash usage" remains a priority, no firm legal mandate or "protections" for cash are outlined, echoing Section 7.6 (n-gram analysis) findings that the public remains unsure if "cash is king" will truly endure.

Taken together, thematic analysis shows that while the BoE recognises the significance of privacy, financial inclusion, and maintaining cash, its response to public feedback fell short of offering robust, transparent plans to implement these goals. By comparing these findings with public discourse (Chapters 7 and 8), RQ5 ascertained where official narratives effectively address or fail to address the persistent negativity and privacy anxieties shaped by real-world user sentiments.

11.2.5 RQ5

RQ5 (Comparative Analysis and Communication Theories' Lens): *What alignments and discrepancies exist between public concerns expressed on X and the narratives in BoE policy documents, and how do established communication theories (e.g., framing theory, agenda-setting theory) explain these alignments and discrepancies? What are the implications of these findings for effective policy communication?*

Key findings:

Social media data (via BERTopic in Section 8.6) was systematically compared against the themes found in BoE's 2024 response papers (Chapter 9) using a comparative framework (Section 10.2) to evaluate thematic alignments and divergences. "Public concerns" here specifically refer to those voiced by the engaged X users in this study's sample; these may not represent the full public sentiment but rather the vocal subset actively discussing the digital pound.

- **Areas of alignment:**
    - **Privacy and security (a shared concern, but divergent frames)**: Both the public and BoE recognise privacy as crucial. However, the BoE frames it through a technical lens, focusing on cryptography and system resilience, whereas the public underscores potential surveillance and overreach [77], [141]. This "privacy framing gap" highlights the limitations of a purely technocratic approach to communication. As Whitley [295] demonstrated in the context of the UK Identity Cards Scheme, a heavy focus on technical data protection measures, without accompanying transparency and accountability, can foster public distrust. This study extends Whitley's findings to the CBDC context, revealing that despite technological and policy advances, public concerns about government overreach and privacy erosion persist. This indicates that while the BoE's emphasis on technical safeguards is necessary, it is not sufficient to allay public fears, underscoring the need for the BoE to move beyond technical assurances and address the broader civil liberties concerns driving public anxieties. This aligns

with framing theory, which posits that how an issue is presented can significantly influence its interpretation.
    - **Interoperability and integration:** S*ection 9.4.5*, indicates the BoE seeks seamless integration with existing payment rails, while social media discourse desires compatibility with broader digital asset ecosystems. This mutual concern supports the potential for two-way symmetrical communication if the BoE actively engages public input on how interoperability should unfold.
- **Key divergences**
    - **Regulatory clarity and public scepticism**: The public is highly sceptical of "government control," "stakeholder engagement," and "regulatory overreach," (section 10.5.1), whereas BoE documents emphasise operational feasibility and public–private synergy. The result is a weak correlation between the frequency of these themes in official papers and X discourse, consistent with studies showing that broad agreement (e.g., on "privacy") can mask deeper disagreements [33], [34]. While statistical tests (paired t-test, Cohen's d, Pearson correlation) did not reveal statistically significant differences at the macro-theme level, the qualitative analysis highlighted important practical differences in emphasis.
    - **Stakeholder engagement and governance**: The public strongly calls for deeper, two-way engagement around the digital pound, whereas official communications lean more top-down, prompting negative sentiment spikes when policy announcements lack inclusive or co-creative strategies (as observed in Sections 7.3, 7.5, 7.10, 8.4, 8.7, 9.3 and 9.4). Simultaneously, the BoE devotes considerable attention to technological infrastructure, overshadowing the public's focus on practical usage, privacy, and potential government overreach.
- **Two-way symmetrical communication (aspirational vs. actual)**: While feedback is solicited by the BoE, the overall communication structure remains closer to a two-way asymmetrical model [142] as the Bank largely aims to persuade the public of the digital pound's benefits rather than engaging in a reciprocal exchange of ideas (Section 10.5). Existing research insufficiently explores the role of narrative and communication in central banking. While scholars like Blinder et al. [79]acknowledged that truly reciprocal communication is challenging for central banks, they also emphasise that it is essential for building lasting public trust. Grunig and Hunt's [142] the framework further underscores this point, arguing that true symmetrical communication is absent if the institution primarily aims to persuade or merely inform. By applying this theory to assess BoE narratives, this study extends existing literature by demonstrating the impact of specific narratives, particularly those reflecting asymmetric communication, on public perception. The strong public emphasis on "stakeholder engagement" vs. moderate official references underscores a shortfall in symmetrical exchange and iterative feedback loops. This finding emphasises the need for the BoE to move beyond consultation and embrace genuine two-way symmetrical communication, characterised by active listening, responsiveness to public concerns, and a willingness to adapt policy in light of public feedback.
- **Implications for policy communication:** Bridging the "privacy framing gap" stands out as the most pressing challenge. The public's uptake of a digital pound hinges on trust,

which in turn is closely tied to how convincingly the BoE addresses privacy and security vulnerabilities [36], [37]. In addition, recognising the information asymmetry in "Technological Infrastructure" is vital. Without concrete transparency and technical guarantees, scepticism could undercut adoption. This finding extends Koziuk et al.'s argument [143] that trust is *sine qua non* for CBDC acceptance. While Koziuk et al. [143] establish the fundamental importance of trust, this research demonstrates that, in the context of the digital pound, trust is specifically contingent on the BoE's ability to credibly address public privacy concerns through clear narrative or communication, transparency, and robust technical safeguards. Also, to regain control of the narrative, authorities might need to put *their own* concern about public trust on the agenda and take actions to bolster that trust (transparency, independent oversight committees, pilot programs with citizen observers, etc.).

Given the nascent nature of CBDCs, these insights serve as an empirical foundation for policymakers and researchers seeking to understand how best to communicate about, and ultimately implement, a digital national currency. By understanding the dynamics of public opinion, the influence of framing and agenda-setting, and the importance of two-way symmetrical communication, the study underscores that privacy, regulatory clarity, and meaningful stakeholder engagement are pivotal determinants of trust. It is important to note, however, that the evidence rests on a modest, X-sourced corpus (~5,700 tweets and a 778-item gold standard), which captures the views of highly engaged social-media users rather than the wider UK public; the findings should therefore be read as directional signals, to be triangulated with survey and qualitative data before firm policy action. Interpreted in this cautious light, the results nonetheless underscore that, as the digital pound moves from concept toward possible deployment, transparent and genuinely interactive communication channels will be indispensable for securing broad societal acceptance and safeguarding the democratic legitimacy of this financial innovation.

11.3 Practical Recommendations

This study reveals a crucial and consistent finding that should inform the BoE's approach to the digital pound: *public concerns about privacy, anonymity, and surveillance have been a dominant and persistent theme throughout the observed discourse, intensifying significantly in response to specific policy announcements*.

This concern, evident in Chapters 7 and 8, sentiment analysis indicates that privacy and government control feature prominently in negative spikes post-policy milestones; similarly, the thematic analysis in Chapter 9 echoes these anxieties about data protection and oversight.

Based on these insights, the following recommendations are proposed to policymakers to navigate the challenges and opportunities associated with the digital pound:

- **Targeted public engagement on privacy:** The BoE should initiate dedicated communication campaigns detailing precisely *how* user data will be protected and by whom, akin to transparency measures in the Bahamas' Sand Dollar project [27], [43]. Such campaigns must clarify technical safeguards (e.g., privacy-enhancing technologies) and highlight accountability mechanisms to mitigate fears of unrestrained government surveillance.

- **Transparent implementation roadmaps:** Publishing phased pilot timelines, technical milestones, and open risk assessments can foster trust among stakeholders, as advocated by UK Finance [46]and academic critiques [47], [295].
- **Two-way symmetrical dialogue:** Incorporate ongoing "listening sessions" to crowdsource user concerns, fostering iterative feedback loops across the pilot phase [79], [81]. These sessions should not be mere formalities but genuine opportunities for dialogue, where public input is actively sought, acknowledged, and incorporated into policy decisions.
- **Refinement of sentiment modelling and topic discovery:** Given RoBERTa_3's susceptibility to sarcasm and typographic noise, practitioners should adopt a multi-model approach (e.g., corroborating with lexicon-based or sarcasm-detection methods [215], [218]) to ensure robust classification.
- **Policy coexistence with stablecoins:** As adoption of stablecoins (e.g., USDT, USDC) grows, the digital pound may not likely not operate in a vacuum. Policymakers must anticipate potential overlap in payment use-cases, ensuring interoperability and regulatory clarity so that the digital pound is not rendered obsolete by more user-friendly or better-known cryptocurrencies. This requires proactive engagement with the stablecoin industry to understand their operations and develop policies that promote innovation while mitigating risks.
- **Adaptive policy communication:** Integrate real-time data analytics into policymaking by establishing dashboards that track sentiment metrics alongside policy milestones, enabling adaptive communication strategies that respond to emerging public concerns. In addition, organise workshops and training sessions in collaboration with academic institutions and data scientists to ensure policymakers are proficient in interpreting and acting upon real-time sentiment data.
- **Adopt a regulatory sandbox approach:** Create a regulatory sandbox for the digital pound that allows controlled, iterative testing alongside existing payment systems. Also, the bank could include selected businesses and consumer groups in pilot tests, then use the resulting data to iteratively improve both technology and communication strategies.

In addition, here are technical guidelines for incorporating sentiment analysis into digital currency policy decisions:

- **Prioritise generalisation:** Avoid relying on training performance alone; validate with real-world data.
- **Balance model complexity:** Bigger isn't always better. Carefully consider the trade-off between model size, performance, and computational cost. Evaluate whether a simpler model, like RoBERTa_3, offers a better balance than larger, more complex alternatives.
- **Rigorous overfitting checks:** Avoid relying on training performance alone; validate with real-world data.

- **Training epoch optimisation:** More epochs can cause overfitting. Determine an optimal epoch count through empirical experimentation and close monitoring of validation metrics. Track validation loss or other robust measures to identify performance peaks.

11.4 Limitations

Despite its contributions, this study is subject to several limitations:

- The analysis relied primarily on data sourced from X, which skews toward particular demographics and patterns of social media engagement. Consequently, groups less active on X, such as older adults or low-tech communities, may be underrepresented, potentially distorting the full range of public sentiment toward the digital pound.
- The quantity of data analysed, while sufficient for exploratory analysis, is relatively modest. Only 778 tweets were manually annotated to create the sentiment classification model's gold-standard training set. The longitudinal sentiment analysis then examined approximately 250 tweets from early 2020, 4,271 tweets from Feb–Jun 2023, and 1,181 tweets from Jan–Mar 2024. These sample sizes are small in comparison to the potentially vast conversation around the digital pound and were gathered via specific keyword queries, which may not capture all relevant discussions. Consequently, the dataset may not reflect the full diversity of opinions, particularly moderate or indifferent positions might be under-sampled, and likely overrepresents particularly vocal users (who often are the ones with strong negative or positive views). Therefore, any observed sentiment patterns and topical emphases in this study should be generalised to the broader UK public with caution. The findings indicate how a segment of engaged users reacted, but not necessarily how *all* citizens feel.
- While major policy announcements (2020, 2023, 2024) were captured, micro-level sentiment shifts (e.g., hourly or real-time reactions) remained beyond the scope of the present research.
- Resource considerations limited the exploration of hyperparameter settings beyond epoch counts. Although RoBERTa_3 proved superior among tested configurations, performance might be further optimised with systematic tuning or larger-scale experiments.
- The study concentrated on textual tweets, omitting other modalities (e.g., memes, images, audio-video content), which can heavily influence public sentiment in modern online discourse. Thus, the findings should be interpreted considering the chosen narrower lens.
- Although the analysis identified notable correlations between policy events and shifts in sentiment, it does not establish direct causality. External factors, ranging from macroeconomic developments to concurrent social or political dynamics, could also influence public attitudes. Therefore, the associations should be interpreted as indicative rather than conclusive.

In light of these limitations, the findings of this thesis should be viewed as indicative rather than definitive. They highlight trends and gaps worthy of attention, but they are not a perfect mirror of the entire UK population's stance on a digital pound. Recognising these constraints is important for anyone looking to apply these insights to policy or further research.

11.5 Discussion on Generalisability

This study offers an empirically grounded look at how public sentiment toward a prospective central-bank digital currency (digital pound) unfolds over multiple policy cycles, using a rigorously validated, domain-specific NLP pipeline. By linking temporal shifts in social-media discourse to concrete Bank of England milestones, it demonstrates both the diagnostic power of sentiment analytics for real-time policy feedback and the pivotal role of communication strategy in shaping public trust. However, it is also important to consider the contexts in which this study's conclusions might hold true versus where they may not. The insights were derived from a specific use case (UK digital pound) and a specific medium (X) within a certain timeframe. Generalising these findings to other settings should be done only under certain assumptions. For example, one might expect similar sentiment trajectories and communication gaps in studies of other countries' CBDC initiatives *if* those countries have a comparable social media landscape and similar levels of public trust in institutions. Under assumptions of a similar socio-technical context (i.e., an open online discourse, engaged citizens, and a central bank making public communications), the "exploration–polarisation–adaptation" pattern observed here might emerge elsewhere.

Indeed, applying this study's methodological framework to, say, the discourse around a potential digital euro (in Europe) or digital rupee (in India) could test whether public concerns like privacy and control consistently dominate once policy specifics are revealed. However, without those conditions, one should be careful not to generalise this study's results. In contexts where social media usage is low or tightly controlled, or where cultural attitudes towards authority and privacy differ greatly, the public reaction to a CBDC could diverge significantly from the UK case. Likewise, within the UK, X-based findings might not generalise to the wider public without corroboration from surveys or other data sources that capture more diverse demographics.

Beyond the CBDC domain, the transferability of both method and findings diminishes as data genre and communicative setting diverge. For example, financial-statement sentiment analysis relies on highly structured, compliance-driven prose written by corporations, not spontaneous public commentary, rendering social-media–tuned models less effective. Similarly, fraud-detection tasks in accounting or insurance focus on identifying anomalous numerical or textual patterns rather than tracking collective sentiment and therefore demand different feature sets and validation standards. Finally, in low-connectivity or heavily censored information environments, the open agenda-setting dynamics that underlie the present study may be absent altogether. Researchers should therefore recalibrate both their data strategy and interpretive lens when migrating this framework to such settings, ensuring that domain-specific linguistic norms, incentive structures, and communication channels are appropriately accounted for before drawing substantive conclusions.

Accordingly, the methodological template provided by this study, i.e., domain-specific annotation, transformer fine-tuning, robustness auditing, and narrative comparison, remains portable, but its *substantive* insights (e.g., the primacy of privacy fears) should not be presumed universal. Researchers venturing into other financial-text arenas or different socio-political settings must recalibrate both the data strategy and the interpretive frame to match the linguistic norms, incentive structures, and communication channels of their chosen domain.

11.6 Future Research Directions

Although this thesis provides novel insights, multiple avenues remain open for further inquiry:

- Incorporating user demographics, e.g., age, location, socioeconomic background, could illuminate whether certain groups are systematically more sceptical or supportive of CBDCs.
- Expanding beyond Twitter/X to Reddit, LinkedIn, or Discord would capture broader, possibly more niche communities of interest.
- Conduct automated hyperparameter searches (e.g., grid search, Bayesian optimisation) to systematically refine model accuracy and robustness.
- XLM-RoBERTa could prove effective if the digital pound conversation extends to international audiences or devolved UK nations, testing cross-lingual sentiment. Moreover, replicating the study in another country's context (for example, studying the public discourse around a prospective digital rupee in India) can test which findings are context-specific and which are global. It would be interesting to see if, under different political cultures, the same issues (privacy, trust, control) dominate or if other concerns emerge (perhaps more about efficacy or economics in some places).
- Linking sentiment developments to actual policy decisions (e.g., legislative revisions, pilot expansions) would clarify how real-time feedback shapes or fails to shape central bank strategy.
- Additional interpretability methods, such as SHAP or integrated gradients, would deepen trust in automated analytics, particularly for mission-critical policy decision-making.
- Deploy an ensemble that blends the domain-tuned RoBERTa _3 with a sarcasm-aware GPT classifier and a lightweight lexicon ruleset. A voting or stacking scheme would dampen each model's idiosyncratic errors, yielding more robust sentiment estimates for policy dashboards.
- Extend the corpus to include memes, GIFs, and infographics and apply vision-language models (e.g., CLIP) to detect sentiment and framing cues in both text and imagery. This would reveal how visual rhetoric amplifies or attenuates privacy anxieties and trust signals around the digital pound. Correlating multimodal sentiment peaks with policy events could offer an even richer early-warning system for communicative misalignment.

In closing, this thesis underscores that the prospective digital pound's ultimate success depends on bridging communication gaps, a move from top-down policy pronouncements to two-way symmetrical dialogue. By ensuring transparency, addressing fears about surveillance, and demonstrating tangible benefits, policymakers can foster the trust essential for a successful CBDC. The methodological framework presented in this thesis, integrating domain-adapted NLP, temporal sentiment analysis, and communication theory, provides a blueprint for future research on digital monetary innovations. This framework can inform how data-driven insights and reciprocal stakeholder engagement can coalesce to shape the future of money and ensure that financial innovation serves the public interest.

References

[1] S. Bibi, ‘Money in the time of crypto’, *Res Int Bus Finance*, vol. 65, 2023, doi: 10.1016/j.ribaf.2023.101964.

[2] G. Y. Wang, ‘Book Review: Another Economy Is Possible: Culture and Economy in a Time of Crisis’, *Global Media and Communication*, vol. 15, no. 3, 2019, doi: 10.1177/1742766519838186.

[3] CoinGecko, ‘Crypto Market Cap Charts’, CoinGecko.

[4] W. J. Lee, S. T. Hong, and T. Min, ‘Bitcoin distribution in the age of digital transformation: Dual-path approach’, *Journal of Distribution Science*, vol. 16, no. 12, 2019, doi: 10.15722/jds.16.12.201812.47.

[5] P. R. Cunha, P. Melo, and H. Sebastião, ‘From bitcoin to central bank digital currencies: Making sense of the digital money revolution’, *Future Internet*, vol. 13, no. 7, 2021, doi: 10.3390/fi13070165.

[6] Rupert Jones, ‘Cashless society draws closer with only one in six payments now in cash’, The Guardian.

[7] NetWest Group, ‘From Cash to Contactless: Understanding the future of payment habits across the UK ’, NatWest.

[8] Credit Connect, ‘Cash payments decline as consumers embrace mobile contactless payments’, Credit Connect.

[9] K. Wenker, ‘Retail Central Bank Digital Currencies (CBDC), Disintermediation and Financial Privacy: The Case of the Bahamian Sand Dollar’, *FinTech*, vol. 1, no. 4, 2022, doi: 10.3390/fintech1040026.

[10] Bank of England, ‘Central Bank Digital Currency Opportunities, challenges and design’, 2020. Accessed: Jan. 23, 2023. [Online]. Available: https://www.bankofengland.co.uk/-/media/boe/files/paper/2020/central-bank-digital-currency-opportunities-challenges-and-design.pdf

[11] Bank of England, ‘The digital pound: a new form of money for households and businesses?’, Feb. 2023. Accessed: Feb. 10, 2023. [Online]. Available: https://www.bankofengland.co.uk/-/media/boe/files/paper/2023/the-digital-pound-consultation-working-paper.pdf

[12] Bank of England, ‘The digital pound: Technology Working Paper’, Feb. 2023. Accessed: Feb. 10, 2023. [Online]. Available: https://www.bankofengland.co.uk/-/media/boe/files/paper/2023/the-digital-pound-technology-working-paper.pdf

[13] Bank of England, ‘Response to the Bank of England and HM Treasury Consultation Paper: The digital pound: a new form of money for households and businesses?’, 2024. Accessed: Jan. 25, 2024. [Online]. Available: https://www.bankofengland.co.uk/-/media/boe/files/paper/2024/responses-to-the-digital-pound-consultation-paper.pdf

[14] Bank of England, ‘Response to the digital pound Technology Working Paper’, Jan. 2024. Accessed: Jan. 25, 2024. [Online]. Available: https://www.bankofengland.co.uk/-/media/boe/files/paper/2024/response-to-the-digital-pound-technology-working-paper.pdf

[15] E. Owen and D. P. Quinn, ‘Does Economic Globalization Influence the US Policy Mood?: A Study of US Public Sentiment, 1956-2011’, 2014. doi: 10.1017/S000712341400009X.

[16] P. Burstein, ‘The impact of public opinion on public policy: A review and an agenda’, 2003. doi: 10.1177/106591290305600103.

[17] L. Mosley and D. A. Singer, ‘Taking stock seriously: Equity-market performance, government policy, and financial globalization’, *International Studies Quarterly*, vol. 52, no. 2, 2008, doi: 10.1111/j.1468-2478.2008.00507.x.

[18] S. Rouhani and E. Abedin, ‘Crypto-currencies narrated on tweets: a sentiment analysis approach’, *International Journal of Ethics and Systems*, vol. 36, no. 1, 2020, doi: 10.1108/IJOES-12-2018-0185.

[19] N. Walia, P. Bandha, and N. Goyal, ‘Unmasking the Twitterverse: Analyzing Sentiments towards DeFi’, *MUDRA: Journal of Finance and Accounting*, vol. 11, no. 1, pp. 36–57, Jun. 2024, doi: 10.17492/jpi.mudra.v11i1.1112403.

[20] A. Jabbar, A. Geebren, Z. Hussain, S. Dani, and S. Ul-Durar, ‘Investigating individual privacy within CBDC: A privacy calculus perspective’, *Res Int Bus Finance*, vol. 64, 2023, doi: 10.1016/j.ribaf.2022.101826.

[21] M. K. Brunnermeier and D. Niepelt, ‘On the equivalence of private and public money’, *J Monet Econ*, vol. 106, pp. 27–41, Oct. 2019, doi: 10.1016/j.jmoneco.2019.07.004.

[22] S. Tabinda Kokab, S. Asghar, and S. Naz, ‘Transformer-based deep learning models for the sentiment analysis of social media data’, *Array*, vol. 14, 2022, doi: 10.1016/j.array.2022.100157.

[23] L. Ramos and O. Chang, ‘Sentiment Analysis of Russia-Ukraine Conflict Tweets Using RoBERTa’, *Uniciencia*, vol. 37, no. 1, 2023, doi: 10.15359/ru.37-1.23.

[24] M. Hasan, L. Islam, I. Jahan, S. M. Meem, and R. M. Rahman, ‘Natural Language Processing and Sentiment Analysis on Bangla Social Media Comments on Russia-Ukraine War Using Transformers’, *Vietnam Journal of Computer Science*, vol. 10, no. 3, 2023, doi: 10.1142/S2196888823500021.

[25] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. Ali Almazroi, 'A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification', *J Healthc Eng*, vol. 2022, 2022, doi: 10.1155/2022/3498123.

[26] ANDREW STANLEY, 'The Ascent of CBDCs', International Monetary Fund.

[27] S. Branch, L. Ward, and A. Wright, 'The Evolution of SandDollar', *Intereconomics*, vol. 58, no. 4, 2023, doi: 10.2478/ie-2023-0038.

[28] M. Committee, *Committee on Payments and Market Infrastructures Markets Committee Central bank digital currencies*. 2018. [Online]. Available: www.bis.org

[29] C. Borio *et al.*, 'Central bank cryptocurrencies 1', 2017. [Online]. Available: www.blockchain.info;

[30] C. M. Kahn and W. Roberds, 'Why pay? An introduction to payments economics', *Journal of Financial Intermediation*, vol. 18, no. 1, 2009, doi: 10.1016/j.jfi.2008.09.001.

[31] Atlantic Council, 'CBDC Tracker', Atlantic Council .

[32] H. Sisodia, 'Case Study on the Sand Dollar CBDC of the Bahamas', 2024. doi: 10.4018/979-8-3693-1882-9.ch018.

[33] Taiwo O. Soetan and Emmanuel Mogaji, *Financial Services in Nigeria. The Path Towards Financial Inclusion, Economic Development and Sustainable Growth*. Palgrave Macmillan, 2024.

[34] C. Baker, 'Examining the Potential of the JAM-DEX® CBDC to Improve Financial Inclusion in Jamaica', 2024, pp. 170–193. doi: 10.4018/979-8-3693-5588-6.ch012.

[35] K. Ahiabenu and S. A. Olaleye, 'The Views of Regulators Around the World', IGI Global, 2024, pp. 32–53. doi: 10.4018/979-8-3693-5588-6.ch003.

[36] S. Choi, Y. S. Kim, B. Kim, and O. Kwon, 'Central Bank Digital Currency and Privacy: A Randomized Survey Experiment', *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4204110.

[37] N. Kshetri and E. Loukoianova, 'Data Privacy Considerations for Central Bank Digital Currencies in Asia-Pacific Countries', *Computer (Long Beach Calif)*, vol. 55, no. 3, 2022, doi: 10.1109/MC.2022.3141228.

[38] H. Xia, Y. Gao, and J. Z. Zhang, 'Understanding the adoption context of China's digital currency electronic payment', *Financial Innovation*, vol. 9, no. 1, 2023, doi: 10.1186/s40854-023-00467-5.

[39] D. Broby, 'Central Bank Digital Currencies − Lessons from China', *Journal of Information Science and Engineering*, vol. 40, no. 4, pp. 699–715, 2024, doi: 10.6688/JISE.202407_40(4).0002.

[40] E. Monye, 'Why Nigeria's Controversial Naira Redesign Policy Hasn't Met Its Objectives', 2024. [Online]. Available: https://www.worldbank.org/en/publication/globalfindex/Data.

[41] G. Tolfsen, 'Privacy Risks Associated With a Central Bank Digital Currency', 2023. Accessed: Feb. 11, 2024. [Online]. Available: https://www.duo.uio.no/bitstream/handle/10852/103278/5/680.pdf

[42] S. Guo, J. Kreitem, and T. Moser, 'DLT Options for CBDC', *Journal of Central Banking Theory and Practice*, vol. 13, no. 1, 2024, doi: 10.2478/jcbtp-2024-0004.

[43] S. L. N. Alonso, J. Jorge-Vazquez, and R. F. R. Forradellas, 'Central banks digital currency: Detection of optimal countries for the implementation of a CBDC and the implication for payment industry open innovation', *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 1, 2021, doi: 10.3390/joitmc7010072.

[44] D. Syrrakos, S. Kuehnlenz, and R. Shand, 'Currency, Identity and trust: Cryptocurrencies, central bank digital currencies and the case for the Bahamian sand dollar', in *Intercultural Issues in the Workplace: Leadership, Communication and Trust*, 2023. doi: 10.1007/978-3-031-42320-8_10.

[45] G. Kaur, 'Privacy implications of central bank digital currencies (CBDCs): a systematic review of literature', *EDPACS*, vol. 69, no. 9, pp. 87–123, Sep. 2024, doi: 10.1080/07366981.2024.2376794.

[46] UK Finance, 'UK Finance response to BoE HMT Digital Pound consultation', 2024, Accessed: Oct. 02, 2024. [Online]. Available: https://www.ukfinance.org.uk/system/files/2023-06/UK%20Finance%20response%20to%20BoE%20HMT%20Digital%20Pound%20consultation.pdf

[47] G. Goodell, 'Response to "The digital pound: a new form of money for households and businesses"', May 2023, [Online]. Available: http://arxiv.org/abs/2305.09059

[48] V. Bonta, N. Kumaresh, and N. Janardhan, 'A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis', *Asian Journal of Computer Science and Technology*, vol. 8, no. S2, 2019, doi: 10.51983/ajcst-2019.8.s2.2037.

[49] A. Sadia, F. Khan, and F. Bashir, 'An Overview of Lexicon-Based Approach For Sentiment Analysis', *International Electrical Engineering Conference*, vol. 1, no. IEEC, 2018.

[50] S. A. S. Neshan and R. Akbari, 'A Combination of Machine Learning and Lexicon Based Techniques for Sentiment Analysis', in *2020 6th International Conference on Web Research, ICWR 2020*, 2020. doi: 10.1109/ICWR49608.2020.9122298.

[51] C. S. G. Khoo and S. B. Johnkhan, ‘Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons’, *J Inf Sci*, vol. 44, no. 4, 2018, doi: 10.1177/0165551517703514.

[52] W. N. S. W. Min and N. Z. Zulkarnain, ‘Comparative Evaluation of Lexicons in Performing Sentiment Analysis’, 2020.

[53] M. Wankhade, A. C. S. Rao, and C. Kulkarni, ‘A survey on sentiment analysis methods, applications, and challenges’, *Artif Intell Rev*, vol. 55, no. 7, 2022, doi: 10.1007/s10462-022-10144-1.

[54] V. Moshkin, N. Yarushkina, and I. Andreev, ‘Approaches to sentiment analysis of the social network text data’, in *CEUR Workshop Proceedings*, 2020.

[55] D. A. K. Khotimah and R. Sarno, ‘Sentiment detection of comment titles in booking.com using probabilistic latent semantic analysis’, in *2018 6th International Conference on Information and Communication Technology, ICoICT 2018*, 2018. doi: 10.1109/ICoICT.2018.8528784.

[56] A. Rutkowska and M. Szyszko, ‘Dictionary-based sentiment analysis of monetary policy communication: on the applicability of lexicons’, *Qual Quant*, Dec. 2024, doi: 10.1007/s11135-024-01896-9.

[57] T. Loughran and B. Mcdonald, ‘When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks’, *Journal of Finance*, vol. 66, no. 1, 2011, doi: 10.1111/j.1540-6261.2010.01625.x.

[58] H. Bennani and M. Neuenkirch, ‘The (home) bias of European central bankers: new evidence based on speeches’, *Appl Econ*, vol. 49, no. 11, 2017, doi: 10.1080/00036846.2016.1210782.

[59] M. Apel and M. B. Grimaldi, ‘How Informative Are Central Bank Minutes?’, *Review of Economics*, vol. 65, no. 1, 2014, doi: 10.1515/roe-2014-0104.

[60] M. Hu and B. Liu, ‘Mining and summarizing customer reviews’, in *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004. doi: 10.1145/1014052.1014073.

[61] K. Barik and S. Misra, ‘Analysis of customer reviews with an improved VADER lexicon classifier’, *J Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00861-x.

[62] B. Hayo and J. Zahner, ‘What is that noise? Analysing sentiment-based variation in central bank communication’, *Econ Lett*, vol. 222, 2023, doi: 10.1016/j.econlet.2022.110962.

[63] A. M. Rahat, A. Kahir, and A. K. M. Masum, ‘Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset’, in *Proceedings of the*

2019 8th International Conference on System Modeling and Advancement in Research Trends, SMART 2019, 2020. doi: 10.1109/SMART46866.2019.9117512.

[64] M. Guia, R. R. Silva, and J. Bernardino, ‘Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis’, in *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2019. doi: 10.5220/0008364105250531.

[65] N. Banik and M. Hasan Hafizur Rahman, ‘Evaluation of Naïve Bayes and Support Vector Machines on Bangla Textual Movie Reviews’, in *2018 International Conference on Bangla Speech and Language Processing, ICBSLP 2018*, 2018. doi: 10.1109/ICBSLP.2018.8554497.

[66] I. Suasnawa, I. Caturbawa, I. Widharma, A. Sapteka, I. Indrayana, and I. Sunaya, ‘Twitter Sentiment Analysis on the Implementation of Online Learning during the Pandemic using Naive Bayes and Support Vector Machine’, 2023. doi: 10.5220/0010945500003260.

[67] K. S. Yogi, V. Dankan Gowda, D. Sindhu, H. Soni, S. Mukherjee, and G. C. Madhu, ‘Enhancing Accuracy in Social Media Sentiment Analysis through Comparative Studies using Machine Learning Techniques’, in *2024 International Conference on Knowledge Engineering and Communication Systems, ICKECS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICKECS61492.2024.10616441.

[68] M. U. Salur and I. Aydin, ‘A Novel Hybrid Deep Learning Model for Sentiment Classification’, *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2982538.

[69] M. S. Islam *et al.*, ‘“Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach”’, *Artif Intell Rev*, vol. 57, no. 3, 2024, doi: 10.1007/s10462-023-10651-9.

[70] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, ‘A review of sentiment analysis: tasks, applications, and deep learning techniques’, 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s41060-024-00594-x.

[71] F. S. gizi Akbarova, ‘The Characteristics of Discourse in Economic Texts’, *Path of Science*, vol. 10, no. 2, pp. 1015–1020, Feb. 2024, doi: 10.22178/pos.101-15.

[72] T. Loughran and B. McDonald, ‘Textual Analysis in Finance’, 2020. doi: 10.1146/annurev-financial-012820-032249.

[73] M. Maia, A. Freitas, and S. Handschuh, ‘FinSSLx: A Sentiment Analysis Model for the Financial Domain Using Text Simplification’, in *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018*, 2018. doi: 10.1109/ICSC.2018.00065.

[74] I. Cioroianu, S. Corbet, C. Larkin, and L. Oxley, ‘The Effects of Central Bank Digital Currency Communication and Associated Social Media Sentiment on Cryptocurrency Markets’, Jul. 2022.

[75] I. Siregar, F. Rahmadiyah, and A. F. Q. Siregar, ‘Linguistic Intervention in Making Fiscal and Monetary Policy’, *International Journal of Arts and Humanities Studies*, vol. 1, no. 1, 2021, doi: 10.32996/ijahs.2021.1.1.8.

[76] D. Masciandaro, O. Peia, and D. Romelli, ‘Central bank communication and social media: From silence to Twitter’, *J Econ Surv*, vol. 38, no. 2, 2024, doi: 10.1111/joes.12550.

[77] J. Choi, K. K. Kim, and Y. Kim, ‘Emotional analysis with news using text mining for framing theory’, in *Studies in Computational Intelligence*, vol. 849, 2020. doi: 10.1007/978-3-030-25213-7_7.

[78] P. Conti-Brown and D. A. Wishnick, ‘Technocratic pragmatism, bureaucratic expertise, and the federal reserve’, *Yale Law Journal*, vol. 130, no. 3, 2021, doi: 10.2139/ssrn.3754523.

[79] A. S. Blinder, M. Ehrmann, J. de Haan, and D.-J. Jansen, ‘Central Bank Communication with the General Public: Promise or False Hope?’, *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4180010.

[80] *The Oxford Handbook of the Economics of Central Banking*. 2019. doi: 10.1093/oxfordhb/9780190626198.001.0001.

[81] I. Korhonen and E. Newby, ‘Mastering Central Bank Communication Challenges via Twitter’, *Bank of Finland Economics Review*, vol. 7, 2019.

[82] A. Haldane and M. McMahon, ‘Central Bank Communications and the General Public’, *AEA Papers and Proceedings*, vol. 108, 2018, doi: 10.1257/pandp.20181082.

[83] R. Auer, G. Cornelli, and J. Frost, ‘Rise of the Central Bank Digital Currencies: Drivers, Approaches and Technologies’, *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3724070.

[84] F. Panetta, ‘A digital euro to meet the expectations of Europeans’, 2021.

[85] M. Janssen and N. Helbig, ‘Innovating and changing the policy-cycle: Policy-makers be prepared!’, *Gov Inf Q*, vol. 35, no. 4, 2018, doi: 10.1016/j.giq.2015.11.009.

[86] D. Kahneman, ‘Daniel Kahneman - thinking, fast and slow’, *Int J Advert*, vol. 31, no. 3, 2012.

[87] R. Shahani and S. Fayaz Ahmed, ‘Psychological and Social Factors Determining Investment Decisions in Cryptocurrency: Exploring the Mediating Role of Cognitive

Biases', *Journal of Organisational Studies and Innovation*, vol. 9, no. 4, 2023, doi: 10.51659/josi.22.159.

[88] J. Bollen, H. Mao, and X. Zeng, 'Twitter mood predicts the stock market', *J Comput Sci*, vol. 2, no. 1, 2011, doi: 10.1016/j.jocs.2010.12.007.

[89] M. M. Chandio and M. Sah, 'Brexit Twitter Sentiment Analysis: Changing Opinions About Brexit and UK Politicians', in *Learning and Analytics in Intelligent Systems*, 2020. doi: 10.1007/978-3-030-38501-9_1.

[90] M. Mora-Cantallops, S. Sánchez-Alonso, and A. Visvizi, 'The influence of external political events on social networks: the case of the Brexit Twitter Network', *J Ambient Intell Humaniz Comput*, vol. 12, no. 4, 2021, doi: 10.1007/s12652-019-01273-7.

[91] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, 'COVID-19 public sentiment insights and machine learning for tweets classification', *Information (Switzerland)*, vol. 11, no. 6, 2020, doi: 10.3390/info11060314.

[92] K. Mohamed Ridhwan and C. A. Hargreaves, 'Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore', *International Journal of Information Management Data Insights*, vol. 1, no. 2, 2021, doi: 10.1016/j.jjimei.2021.100021.

[93] M. Andreotta *et al.*, 'Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis', *Behav Res Methods*, vol. 51, no. 4, 2019, doi: 10.3758/s13428-019-01202-8.

[94] E. Ainley, C. Witwicki, A. Tallett, and C. Graham, 'Using twitter comments to understand people's experiences of uk health care during the COVID-19 Pandemic: Thematic and sentiment analysis', *J Med Internet Res*, vol. 23, no. 10, 2021, doi: 10.2196/31101.

[95] Akif Mustafa, Imaduddin Ansari, Subham Kumar Mohanta, and Shalem Balla, 'PUBLIC REACTION TO COVID-19 ON TWITTER: A THEMATIC ANALYSIS', *EPRA International Journal of Multidisciplinary Research (IJMR)*, 2020, doi: 10.36713/epra4518.

[96] S. Noor, Y. Guo, S. H. H. Shah, M. S. Nawaz, and A. S. Butt, 'Research synthesis and thematic analysis of twitter through bibliometric analysis', *Int J Semant Web Inf Syst*, vol. 16, no. 3, 2020, doi: 10.4018/IJSWIS.2020070106.

[97] P. Varma, S. Nijjer, K. Sood, S. Grima, and R. Rupeika-Apoga, 'Thematic Analysis of Financial Technology (Fintech) Influence on the Banking Industry', 2022. doi: 10.3390/risks10100186.

[98] A. Kosari, M. Khajeh, H. Jahangirni, and R. Sanavifard, 'Identifying Influential Components on the Self-Regulation of Real Traders in Competitive Financial Markets Using a Thematic Analysis Approach', 2023.

[99] P. Sahoo, P. K. Saraf, and R. Uchil, 'Moving toward Bank 4.0 in India: a qualitative thematic analysis and future research agenda', *Journal of Science and Technology Policy Management*, 2024, doi: 10.1108/JSTPM-08-2023-0128.

[100] S. Prodan, D.-C. Dabija, and L. Marincean, 'Exploring Consumer Sentiment on Central Bank Digital Currencies: A Twitter Analysis from 2021 to 2023', *Proceedings of the International Conference on Business Excellence*, vol. 17, no. 1, 2023, doi: 10.2478/picbe-2023-0098.

[101] T. Pano and R. Kashef, 'A complete vader-based sentiment analysis of bitcoin (BTC) tweets during the ERA of COVID-19', *Big Data and Cognitive Computing*, vol. 4, no. 4, pp. 1–17, 2020, doi: 10.3390/bdcc4040033.

[102] I. Gurrib and F. Kamalov, 'Predicting bitcoin price movements using sentiment analysis: a machine learning approach', *Studies in Economics and Finance*, vol. 39, no. 3, pp. 347–364, Apr. 2022, doi: 10.1108/SEF-07-2021-0293.

[103] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, 'Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model', *IEEE Access*, vol. 10, pp. 39313–39324, 2022, doi: 10.1109/ACCESS.2022.3165621.

[104] D. Mukesh, S. Sangeetha, and A. Latha, 'Sentimental Analysis of CBDC Tweets Using Machine Learning and Deep Learning Techniques', in *2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2023*, 2023. doi: 10.1109/ICAECA56562.2023.10199561.

[105] R. F. Astuti and A. Alamsyah, 'Measuring CBDC and DeFi Public Discourse Using BERT and RoBERTa', in *2024 International Conference on Smart Computing, IoT and Machine Learning, SIML 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 150–155. doi: 10.1109/SIML61815.2024.10578226.

[106] M. Kulakowski, F. Frasincar, and E. Cambria, 'Sentiment Classification of Cryptocurrency-Related Social Media Posts', *IEEE Intell Syst*, vol. 38, no. 4, pp. 5–9, Jul. 2023, doi: 10.1109/MIS.2023.3283170.

[107] X. Huang *et al.*, 'LSTM Based Sentiment Analysis for Cryptocurrency Prediction', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021. doi: 10.1007/978-3-030-73200-4_47.

[108] D. Guégan and T. Renault, 'Does investor sentiment on social media provide robust information for Bitcoin returns predictability?', *Financ Res Lett*, vol. 38, 2021, doi: 10.1016/j.frl.2020.101494.

[109] L. White, R. Togneri, W. Liu, and M. Bennamoun, ‘Generating bags of words from the sums of their word embeddings’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. doi: 10.1007/978-3-319-75477-2_5.

[110] S. Qaiser and R. Ali, ‘Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents’, *Int J Comput Appl*, vol. 181, no. 1, 2018, doi: 10.5120/ijca2018917395.

[111] A. H. Ribeiro, K. Tiels, L. A. Aguirre, and T. B. Schön, ‘Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness’, in *Proceedings of Machine Learning Research*, 2020.

[112] A. Vaswani *et al.*, ‘Attention is all you need’, in *Advances in Neural Information Processing Systems*, 2017.

[113] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, ‘BERT: Pre-training of deep bidirectional transformers for language understanding’, in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019.

[114] U. B. Mahadevaswamy and P. Swathi, ‘Sentiment Analysis using Bidirectional LSTM Network’, in *Procedia Computer Science*, 2022. doi: 10.1016/j.procs.2022.12.400.

[115] Y. Liu *et al.*, ‘RoBERTa: A Robustly Optimized BERT Pretraining Approach’, Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.11692

[116] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, ‘DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter’, Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

[117] A. Conneau *et al.*, ‘Unsupervised Cross-lingual Representation Learning at Scale’, Nov. 2019, [Online]. Available: http://arxiv.org/abs/1911.02116

[118] G. Hinton, O. Vinyals, and J. Dean, ‘Distilling the Knowledge in a Neural Network’, Mar. 2015, [Online]. Available: http://arxiv.org/abs/1503.02531

[119] J. Mozafari, A. Fatemi, and P. Moradi, ‘A Method for Answer Selection Using DistilBERT and Important Words’, in *2020 6th International Conference on Web Research, ICWR 2020*, 2020. doi: 10.1109/ICWR49608.2020.9122302.

[120] A. Bris *et al.*, ‘KNIGHTS, RAIDERS, AND TARGETS - THE IMPACT OF THE HOSTILE TAKEOVER - COFFEE,JC, LOWENSTEIN,L, ROSEACKERMAN,S’, *J Bank Financ*, vol. 37, no. 1, 2021.

[121] A. F. Adoma, N. M. Henry, and W. Chen, ‘Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition’, in *2020 17th International*

Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020, 2020. doi: 10.1109/ICCWAMTIP51612.2020.9317379.

[122] X. Jiang, Y. Liang, W. Chen, and N. Duan, 'XLM-K: Improving Cross-Lingual Language Model Pre-training with Multilingual Knowledge', in *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, 2022. doi: 10.1609/aaai.v36i10.21330.

[123] Y. Lu, L. Chen, X. Tong, Y. Peng, and H. Zhu, 'Research on cross-lingual multi-label patent classification based on pre-trained model', *Scientometrics*, vol. 129, no. 6, pp. 3067–3087, Jun. 2024, doi: 10.1007/s11192-024-05024-0.

[124] Y. Gu *et al.*, 'Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing', *ACM Trans Comput Healthc*, vol. 3, no. 1, 2022, doi: 10.1145/3458754.

[125] Y. Li, S. Wang, H. Ding, and H. Chen, 'Large Language Models in Finance: A Survey', in *ICAIF 2023 - 4th ACM International Conference on AI in Finance*, 2023. doi: 10.1145/3604237.3626869.

[126] J. Zhou, J. Tian, R. Wang, Y. Wu, W. Xiao, and L. He, 'SENTIX: A Sentiment-Aware Pre-Trained Model for Cross-Domain Sentiment Analysis', in *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 2020. doi: 10.18653/v1/2020.coling-main.49.

[127] G. Zhang, J. Wu, M. Tan, Z. Yang, Q. Cheng, and H. Han, 'Learning to Predict U.S. Policy Change Using New York Times Corpus with Pre-Trained Language Model', *Multimed Tools Appl*, vol. 79, no. 45–46, 2020, doi: 10.1007/s11042-020-08946-y.

[128] Z. Liu, W. Lin, Y. Shi, and J. Zhao, 'A Robustly Optimized BERT Pre-training Approach with Post-training', in *CCL 2021 - Proceedings of the 20th Chinese National Conference on Computational Linguistics*, 2021.

[129] D. Araci, 'FinBERT: Financial Sentiment Analysis with Pre-trained Language Models', Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.10063

[130] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, 'FinBERT: A pre-trained financial language representation model for financial text mining', in *IJCAI International Joint Conference on Artificial Intelligence*, 2020. doi: 10.24963/ijcai.2020/622.

[131] D. Licari and G. Comandè, 'ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain', *Computer Law and Security Review*, vol. 52, 2024, doi: 10.1016/j.clsr.2023.105908.

[132] P. Lobanova, P. Bakhtin, and Y. Sergienko, 'Identifying and Visualizing Trends in Science, Technology, and Innovation Using SciBERT', *IEEE Trans Eng Manag*, vol. 71, 2024, doi: 10.1109/TEM.2023.3306569.

[133] J. Grimmer, M. E. Roberts, and B. M. Stewart, *Text as Data: A New Framework for Machine Learning and the Social Sciences*. 2022. doi: 10.1177/00943061231214609o.

[134] S. Gössi, Z. Chen, W. Kim, B. Bermeitinger, and S. Handschuh, 'FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment Focus Method for Enhancing Sentiment Analysis of FOMC Minutes', in *ICAIF 2023 - 4th ACM International Conference on AI in Finance*, 2023. doi: 10.1145/3604237.3626843.

[135] S. Kumar and R. Chaturvedi, 'Evaluating The Efficacy Of Distilled Transformer Models For Sentiment Analysis In Financial Texts: A Comparative Study', 2024. [Online]. Available: www.ijcspub.org

[136] D. Garreau and U. von Luxburg, 'Explaining the Explainer: A First Theoretical Analysis of LIME', in *Proceedings of Machine Learning Research*, 2020.

[137] Aditya Bhattacharya, *Applied Machine Learning Explainability Techniques*. 2022.

[138] I. Goodfellow, P. McDaniel, and N. Papernot, 'Making machine learning robust against adversarial inputs', *Commun ACM*, vol. 61, no. 7, 2018, doi: 10.1145/3134599.

[139] X. Wang, J. Li, X. Kuang, Y. an Tan, and J. Li, 'The security of machine learning in an adversarial setting: A survey', *J Parallel Distrib Comput*, vol. 130, 2019, doi: 10.1016/j.jpdc.2019.03.003.

[140] M. E. McCombs, D. L. Shaw, and D. H. Weaver, 'New Directions in Agenda-Setting Theory and Research', *Mass Commun Soc*, vol. 17, no. 6, 2014, doi: 10.1080/15205436.2014.964871.

[141] R. M. Entman, 'Framing: Toward Clarification of a Fractured Paradigm', *Journal of Communication*, vol. 43, no. 4, 1993, doi: 10.1111/j.1460-2466.1993.tb01304.x.

[142] J. E. Grunig and L. A. Grunig, 'Models of public relations and communication', in *Excellence in Public Relations and Communication Management*, 2013. doi: 10.4324/9780203812303-20.

[143] V. Koziuk, Y. Ivashuk, Y. Hayda, and O. Dluhopolskyi, 'CBDC and Trust in A Central Bank: Transitivity Of Preferences Vs. The Privacy Paradox', *Financial Internet Quarterly*, vol. 20, no. 4, pp. 32–47, Dec. 2024, doi: 10.2478/fiqf-2024-0025.

[144] S. Rouhani and E. Abedin, 'Crypto-currencies narrated on tweets: a sentiment analysis approach', *International Journal of Ethics and Systems*, vol. 36, no. 1, pp. 58–72, Jan. 2020, doi: 10.1108/IJOES-12-2018-0185.

[145] B. Simpson, 'Pragmatism: A Philosophy of Practice', in *The SAGE Handbook of Qualitative Business and Management Research Methods: History and Traditions*, 2018. doi: 10.4135/9781526430212.n4.

[146] U. Kuckartz, 'Qualitative Text Analysis: A Systematic Approach', 2019. doi: 10.1007/978-3-030-15636-7_8.

[147] J. Lukito and M. L. Pruden, 'Critical computation: mixed-methods approaches to big language data analysis', *Review of Communication*, vol. 23, no. 1, 2023, doi: 10.1080/15358593.2022.2125821.

[148] N. Medagoda and S. Shanmuganathan, 'Keywords based temporal sentiment analysis', in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, 2016. doi: 10.1109/FSKD.2015.7382152.

[149] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, 'Finding high-quality content in social media', in *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008. doi: 10.1145/1341531.1341557.

[150] M. F. A. Gadi and M. Á. Sicilia, 'A sentiment corpus for the cryptocurrency financial domain: the CryptoLin corpus', *Lang Resour Eval*, Jun. 2024, doi: 10.1007/s10579-024-09743-x.

[151] Divesh, G. Prasad, G. Sharma, and D. K. Vishwakarma, 'Sentiment Analysis on Cryptocurrency using Youtube Comments', in *Proceedings - 6th International Conference on Computing Methodologies and Communication, ICCMC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 730–733. doi: 10.1109/ICCMC53470.2022.9753723.

[152] M. González, 'An analysis of twitter corpora and the differences between formal and colloquial tweets', in *CEUR Workshop Proceedings*, 2015.

[153] L. Borin, 'All that glitters...', *Nordlyd*, vol. 46, no. 1, Aug. 2022, doi: 10.7557/12.6348.

[154] A. Zaenen, 'Last Words Mark-up Barking Up the Wrong Tree', 2006.

[155] J. Medero, K. Maeda, S. Strassel, and C. Walker, 'An efficient approach to gold-standard annotation: Decision points for complex tasks', in *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*, 2006.

[156] L. Lavitas, O. Redfield, A. Lee, D. Fletcher, M. Eck, and S. Janardhanan, 'Annotation Quality Framework-Accuracy, Credibility, and Consistency', *NEURIPS 2021 Workshop for Data Centric AI*, 2021.

[157] R. Artstein and M. Poesio, 'Inter-coder agreement for computational linguistics', 2008. doi: 10.1162/coli.07-034-R2.

[158] J. Pustejovsky and a Stubbs, *Natural language annotation for machine learning*. 2013.

[159] J. R. Landis and G. G. Koch, 'The Measurement of Observer Agreement for Categorical Data', *Biometrics*, vol. 33, no. 1, 1977, doi: 10.2307/2529310.

[160] J. L. Fleiss, ‘Measuring nominal scale agreement among many raters’, *Psychol Bull*, vol. 76, no. 5, 1971, doi: 10.1037/h0031619.

[161] K. Krippendorff, ‘Reliability in content analysis: Some common misconceptions and recommendations’, *Hum Commun Res*, vol. 30, no. 3, 2004, doi: 10.1093/hcr/30.3.411.

[162] C. O’Connor and H. Joffe, ‘Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines’, *Int J Qual Methods*, vol. 19, 2020, doi: 10.1177/1609406919899220.

[163] G. Hripcsak and A. S. Rothschild, ‘Agreement, the F-measure, and reliability in information retrieval’, *Journal of the American Medical Informatics Association*, vol. 12, no. 3, 2005, doi: 10.1197/jamia.M1733.

[164] Y. Mathet, A. Widlöcher, and J. P. Métivier, ‘The unified and holistic method gamma (¡) for inter-annotator agreement measure and alignment’, *Computational Linguistics*, vol. 41, no. 3, 2015, doi: 10.1162/COLI_a_00227.

[165] I. Mozetič, M. Grčar, and J. Smailović, ‘Multilingual twitter sentiment classification: The role of human annotators’, *PLoS One*, vol. 11, no. 5, May 2016, doi: 10.1371/journal.pone.0155036.

[166] V. Bobicev and M. Sokolova, ‘Inter-annotator agreement in sentiment analysis: Machine learning perspective’, in *International Conference Recent Advances in Natural Language Processing, RANLP*, Incoma Ltd, 2017, pp. 97–102. doi: 10.26615/978-954-452-049-6_015.

[167] A. Go, R. Bhayani, and L. Huang, ‘Twitter Sentiment Classification using Distant Supervision’, *Processing*, vol., 2009.

[168] R. Goel *et al.*, ‘The social dynamics of language change in online networks’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016. doi: 10.1007/978-3-319-47880-7_3.

[169] S. García-Méndez, F. de Arriba-Pérez, A. Barros-Vila, and F. J. González-Castaño, ‘Detection of temporality at discourse level on financial news by combining Natural Language Processing and Machine Learning’, *Expert Syst Appl*, vol. 197, 2022, doi: 10.1016/j.eswa.2022.116648.

[170] A. McCallum and K. Nigam, ‘A Comparison of Event Models for Naive Bayes Text Classification’, *AAAI/ICML-98 Workshop on Learning for Text Categorization*, 1998, doi: 10.1.1.46.1529.

[171] D. Learning, ‘Deep Learning - Goodfellow’, *Nature*, vol. 26, no. 7553, 2016.

[172] M. G. Qabasiyu, M. A. Zayyad, and S. Abdullahi, ‘An Ensembled Based Machine Learning Technique of Sentiment Analysis’, *Journal of Telecommunication*.

[173] Shai. Shalev-Shwartz and Shai. Ben-David, *Understanding machine learning : from theory to algorithms*. Cambridge University Press, 2014.

[174] T. Wolf *et al.*, ‘Transformers: State-of-the-Art Natural Language Processing’, in *EMNLP 2020 - Conference on Empirical Methods in Natural Language Processing, Proceedings of Systems Demonstrations*, 2020. doi: 10.18653/v1/2020.emnlp-demos.6.

[175] C. Raffel *et al.*, ‘Exploring the limits of transfer learning with a unified text-to-text transformer’, *Journal of Machine Learning Research*, vol. 21, 2020.

[176] X. Bouthillier *et al.*, ‘ACCOUNTING FOR VARIANCE IN MACHINE LEARNING BENCHMARKS’, 2021.

[177] A. Komatsuzaki, ‘One Epoch Is All You Need’, Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.06669

[178] L. Aitchison, ‘Why you don’t overfit, and don’t need Bayes if you only train for one epoch’, Nov. 2024, [Online]. Available: http://arxiv.org/abs/2411.14478

[179] R. Kohavi, ‘A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection’, in *IJCAI International Joint Conference on Artificial Intelligence*, 1995.

[180] T. D1Etterich, ‘Overfitting and Undercomputing in Machine Learning’, *ACM Computing Surveys (CSUR)*, vol. 27, no. 3, 1995, doi: 10.1145/212094.212114.

[181] F. Hutter, L. Kotthof, and J. Vanschoren, *Automated machine learning: Methods, Systems, Challenges*. 2020.

[182] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ‘Dropout: A simple way to prevent neural networks from overfitting’, *Journal of Machine Learning Research*, vol. 15, 2014.

[183] M. Mosbach, M. Andriushchenko, and D. Klakow, ‘ON THE STABILITY OF FINE-TUNING BERT: MISCONCEPTIONS, EXPLANATIONS, AND STRONG BASELINES’, in *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[184] Z. Fu, A. M.-C. So, and N. Collier, ‘A Stability Analysis of Fine-Tuning a Pre-Trained Model’, Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.09820

[185] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, and N. Smith, ‘Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping’, Feb. 2020, [Online]. Available: http://arxiv.org/abs/2002.06305

[186] J. Howard and S. Ruder, ‘Universal language model fine-tuning for text classification’, in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018. doi: 10.18653/v1/p18-1031.

[187] N. S. Keskar, J. Nocedal, P. T. P. Tang, D. Mudigere, and M. Smelyanskiy, 'On large-batch training for deep learning: Generalization gap and sharp minima', in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.

[188] M. Popel and O. Bojar, 'Training Tips for the Transformer Model', Mar. 2018, doi: 10.2478/pralin-2018-0002.

[189] P. Goyal *et al.*, 'Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour', Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.02677

[190] I. Loshchilov and F. Hutter, 'Decoupled weight decay regularization', in *7th International Conference on Learning Representations, ICLR 2019*, 2019.

[191] D. P. Kingma and J. L. Ba, 'Adam: A method for stochastic optimization', in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[192] H. He and E. A. Garcia, 'Learning from imbalanced data', *IEEE Trans Knowl Data Eng*, vol. 21, no. 9, 2009, doi: 10.1109/TKDE.2008.239.

[193] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. 2018. doi: 10.1007/978-3-319-98074-4.

[194] S. Singla and N. Ramachandra, 'Comparative analysis of transformer based pre-trained NLP Models', *International Journal of Computer Sciences and Engineering*, vol. 8, no. 11, 2020.

[195] M. Sokolova and G. Lapalme, 'A systematic analysis of performance measures for classification tasks', *Inf Process Manag*, vol. 45, no. 4, 2009, doi: 10.1016/j.ipm.2009.03.002.

[196] D. M. W. POWERS, 'Estimation of high affinity estradiol binding sites in human breast cancerEVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION', *Journal of Machine Learning Technologies*, vol. 2, no. 1, 2011.

[197] X. Glorot and Y. Bengio, 'Understanding the difficulty of training deep feedforward neural networks', in *Journal of Machine Learning Research*, 2010.

[198] L. Bottou, 'Large-scale machine learning with stochastic gradient descent', in *Proceedings of COMPSTAT 2010 - 19th International Conference on Computational Statistics, Keynote, Invited and Contributed Papers*, 2010. doi: 10.1007/978-3-7908-2604-3_16.

[199] Y. You *et al.*, 'The Limit of the Batch Size', 2020. Accessed: Sep. 25, 2024. [Online]. Available: https://arxiv.org/pdf/2006.08517

[200] W. H. Kruskal and W. A. Wallis, 'Use of Ranks in One-Criterion Variance Analysis', *J Am Stat Assoc*, vol. 47, no. 260, 1952, doi: 10.1080/01621459.1952.10483441.

[201] D. M. Hawkins, 'The Problem of Overfitting', 2004. doi: 10.1021/ci0342472.

[202] S. R. Sain and V. N. Vapnik, 'The Nature of Statistical Learning Theory', *Technometrics*, vol. 38, no. 4, 1996, doi: 10.2307/1271324.

[203] T. Hastie, R. Tibshirani, and J. Friedman, *Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, vol. 2nd. 2009.

[204] L. Prechelt, 'Early stopping - But when?', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7700 LECTURE NO, 2012, doi: 10.1007/978-3-642-35289-8_5.

[205] H. Taherdoost, H. Business, S. Sdn, C. Group, and K. Lumpur, 'Sampling Methods in Research Methodology ; How to Choose a Sampling Technique for', *International Journal of Academic Research in Management (IJARM)*, vol. 5, no. 2, 2016.

[206] R. Jia and P. Liang, 'Adversarial examples for evaluating reading comprehension systems', in *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2017. doi: 10.18653/v1/d17-1215.

[207] M. T. Ribeiro, S. Singh, and C. Guestrin, 'Semantically equivalent adversarial rules for debugging NLP models', in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018. doi: 10.18653/v1/p18-1079.

[208] J. Ebrahimi, A. Rao, D. Lowd, and D. Dou, 'Hotflip: White-box adversarial examples for text classification', in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018. doi: 10.18653/v1/p18-2006.

[209] E. Pavlick and J. Tetreault, 'Erratum: "An Empirical Analysis of Formality in Online Communication"', *Trans Assoc Comput Linguist*, vol. 4, 2016, doi: 10.1162/tacl_a_00243.

[210] D. Pruthi, B. Dhingra, and Z. C. Lipton, 'Combating adversarial misspellings with robust word recognition', in *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020. doi: 10.18653/v1/p19-1561.

[211] Y. Belinkov and Y. Bisk, 'Synthetic and natural noise both break neural machine translation', in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[212] M. Moradi and M. Samwald, ‘Evaluating the Robustness of Neural Language Models to Input Perturbations’, in *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2021. doi: 10.18653/v1/2021.emnlp-main.117.

[213] E. Wallace, S. Feng, N. Kandpal, M. Gardner, and S. Singh, ‘Universal Adversarial Triggers for Attacking and Analyzing NLP’, 2019. doi: 10.18653/v1/d19-1221.

[214] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, ‘New avenues in opinion mining and sentiment analysis’, *IEEE Intell Syst*, vol. 28, no. 2, 2013, doi: 10.1109/MIS.2013.30.

[215] A. Rajadesingan, R. Zafarani, and H. Liu, ‘Sarcasm detection on twitter:A behavioral modeling approach’, in *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 2015. doi: 10.1145/2684822.2685316.

[216] B. Saberi and S. Saad, ‘Sentiment analysis or opinion mining: A review’, 2017. doi: 10.18517/ijaseit.7.5.2137.

[217] O. D. Okey, E. U. Udo, R. L. Rosa, D. Z. Rodríguez, and J. H. Kleinschmidt, ‘Investigating ChatGPT and cybersecurity: A perspective on topic modeling and sentiment analysis’, *Comput Secur*, vol. 135, 2023, doi: 10.1016/j.cose.2023.103476.

[218] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, ‘Sarcasm as contrast between a positive sentiment and negative situation’, in *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2013.

[219] C. J. Hutto and E. Gilbert, ‘VADER: A parsimonious rule-based model for sentiment analysis of social media text’, in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 2014. doi: 10.1609/icwsm.v8i1.14550.

[220] S. M. Mohammad, S. Kiritchenko, and X. Zhu, ‘NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets’, in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

[221] E. Kouloumpis, T. Wilson, and J. Moore, ‘Twitter Sentiment Analysis: The Good the Bad and the OMG!’, in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media, ICWSM 2011*, 2011. doi: 10.1609/icwsm.v5i1.14185.

[222] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, ‘Sentiment in short strength detection informal text’, *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, 2010, doi: 10.1002/asi.21416.

[223] W. Liao, B. Zeng, X. Yin, and P. Wei, ‘An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa’, *Applied Intelligence*, vol. 51, no. 6, 2021, doi: 10.1007/s10489-020-01964-1.

[224] M. Abdullah, J. Khrais, and S. Swedat, ‘Transformer-Based Deep Learning for Sarcasm Detection with Imbalanced Dataset: Resampling Techniques with Downsampling and Augmentation’, in *2022 13th International Conference on Information and Communication Systems, ICICS 2022*, 2022. doi: 10.1109/ICICS55353.2022.9811196.

[225] Y. Younes and B. Mathiak, ‘Handling Class Imbalance when Detecting Dataset Mentions with Pre-trained Language Models’, in *ICNLSP 2022 - Proceedings of the 5th International Conference on Natural Language and Speech Processing*, 2022.

[226] S. M. Mohammad and P. D. Turney, ‘Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon’, *CAAGET ’10 Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, no. June, 2010.

[227] T. A. Nurdin, M. B. Alexandri, W. Sumadinata, and R. Arifianti, ‘SENTIMENT ANALYSIS OF USER PREFERENCE FOR OLD VS NEW FINTECH TECHNOLOGY USING SVM AND NB ALGORITHMS’, *Management Systems in Production Engineering*, vol. 31, no. 4, 2023, doi: 10.2478/mspe-2023-0041.

[228] D. Pattnaik, S. Ray, and R. Raman, ‘Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review’, *Heliyon*, vol. 10, no. 1, 2024, doi: 10.1016/j.heliyon.2023.e23492.

[229] A. M. Priyatno and F. I. Firmananda, ‘N-Gram Feature for Comparison of Machine Learning Methods on Sentiment in Financial News Headlines’, *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 1, no. 1, 2022, doi: 10.31004/riggs.v1i1.4.

[230] M. M. Haider, M. A. Hossin, H. R. Mahi, and H. Arif, ‘Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm’, in *2020 IEEE Region 10 Symposium, TENSYMP 2020*, 2020. doi: 10.1109/TENSYMP50017.2020.9230670.

[231] M. Hoang, O. A. Bihorac, and J. Rouces, ‘Aspect-Based Sentiment Analysis Using BERT’, in *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 2019.

[232] E. Hoque and M. S. Islam, ‘Natural Language Generation for Visualizations: State of the Art, Challenges and Future Directions’, *Computer Graphics Forum*, Oct. 2024, doi: 10.1111/cgf.15266.

[233] R. Auer and R. Böhme, ‘The Technology of Retail Central Bank Digital Currency’, *SSRN*, no. March, 2020.

[234] M. Brodo D and A. Levin T, ‘Central bank digital currency and the future of monetary policy - Global Business Outlook’, *NATIONAL BUREAU OF ECONOMIC RESEARCH*, vol. 93, no. I, 2017.

[235] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. Zhao, ‘Analyzing patterns of user content generation in online social networks’, in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009. doi: 10.1145/1557019.1557064.

[236] A. Downs, ‘Up and down with ecology: The “issue-attention cycle”’, in *Agenda Setting: Readings on Media, Public Opinion, and Policymaking*, 2016. doi: 10.4324/9781315538389.

[237] R. E. Kasperson *et al.*, ‘The Social Amplification of Risk: A Conceptual Framework’, *Risk Analysis*, vol. 8, no. 2, 1988, doi: 10.1111/j.1539-6924.1988.tb01168.x.

[238] J. W. Dearing, ‘Diffusion of Innovations’, in *The Oxford Handbook of Organizational Change and Innovation*, 2021. doi: 10.1093/oxfordhb/9780198845973.013.23.

[239] O. Dedehayir and M. Steinert, ‘The hype cycle model: A review and future directions’, *Technol Forecast Soc Change*, vol. 108, 2016, doi: 10.1016/j.techfore.2016.04.005.

[240] L. Xie, D. Wang, and F. Ma, ‘Analysis of individual characteristics influencing user polarization in COVID-19 vaccine hesitancy.’, *Comput Human Behav*, vol. 143, 2023, doi: 10.1016/j.chb.2022.107649.

[241] C. S. G. Khoo, A. Nourbakhsh, and J. C. Na, ‘Sentiment analysis of online news text: A case study of appraisal theory’, 2012. doi: 10.1108/14684521211287936.

[242] A. Wildavsky and K. Dake, ‘Theories of Risk Perception: Who Fears What and Why?’, in *The Institutional Dynamics of Culture, Volume I and II: The New Durkheimians*, vol. 1–2, 2018. doi: 10.4324/9781315238975-10.

[243] R. Egger and J. Yu, ‘A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts’, *Frontiers in Sociology*, vol. 7, 2022, doi: 10.3389/fsoc.2022.886498.

[244] A. C. Nanayakkara and G. A. D. M. Thennakoon, ‘Enhancing Social Media Content Analysis with Advanced Topic Modeling Techniques: A Comparative Study’, *International Journal on Advances in ICT for Emerging Regions (ICTer)*, vol. 17, no. 1, pp. 40–47, May 2024, doi: 10.4038/icter.v17i1.7276.

[245] Y. Yang, Y. Wei, M. Gao, Z. Ran, and Q. Wang, ‘Sentiment Analysis of Social Network Text Based on HDBSCAN and SO-PMI’, in *Journal of Physics: Conference Series*, 2023. doi: 10.1088/1742-6596/2504/1/012055.

[246] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, ‘Hierarchical density estimates for data clustering, visualization, and outlier detection’, *ACM Trans Knowl Discov Data*, vol. 10, no. 1, 2015, doi: 10.1145/2733381.

[247] H. Axelborn and J. Berggren, 'Topic Modeling for Customer Insights A Comparative Analysis of LDA and BERTopic in Categorizing Customer Calls', 2023.

[248] A. D. Susanto, S. Andrian Pradita, C. Stryadhi, K. E. Setiawan, and M. Fikri Hasani, 'Text Vectorization Techniques for Trending Topic Clustering on Twitter: A Comparative Evaluation of TF-IDF, Doc2Vec, and Sentence-BERT', in *2023 5th International Conference on Cybernetics and Intelligent Systems, ICORIS 2023*, 2023. doi: 10.1109/ICORIS60118.2023.10352228.

[249] I. Mozetič, L. Torgo, V. Cerqueira, and J. Smailović, 'How to evaluate sentiment classifiers for Twitter time-ordered data?', *PLoS One*, vol. 13, no. 3, 2018, doi: 10.1371/journal.pone.0194317.

[250] M. S. Asyaky and R. Mandala, 'Improving the Performance of HDBSCAN on Short Text Clustering by Using Word Embedding and UMAP', in *Proceedings - 2021 8th International Conference on Advanced Informatics: Concepts, Theory, and Application, ICAICTA 2021*, 2021. doi: 10.1109/ICAICTA53211.2021.9640285.

[251] G. Dorcas Wambui, 'The Power of the Pruned Exact Linear Time(PELT) Test in Multiple Changepoint Detection', *American Journal of Theoretical and Applied Statistics*, vol. 4, no. 6, 2015, doi: 10.11648/j.ajtas.20150406.30.

[252] K. Okoye and S. Hosseini, 'Mann–Whitney U Test and Kruskal–Wallis H Test Statistics in R', in *R Programming*, Singapore: Springer Nature Singapore, 2024, pp. 225–246. doi: 10.1007/978-981-97-3385-9_11.

[253] V. Braun and V. Clarke, 'Using thematic analysis in psychology', *Qual Res Psychol*, vol. 3, no. 2, 2006, doi: 10.1191/1478088706qp063oa.

[254] E. Hashimov, 'Qualitative Data Analysis: A Methods Sourcebook and The Coding Manual for Qualitative Researchers', *Technical Communication Quarterly*, vol. 24, no. 1, 2015, doi: 10.1080/10572252.2015.975966.

[255] E. Fitkov-Norris and N. Kocheva, 'Are we There yet? Thematic Analysis, NLP, and Machine Learning for Research', in *Proceedings of the European Conference on Research Methods in Business and Management Studies*, 2023. doi: 10.34190/ecrm.22.1.1616.

[256] J. Corbin and A. Strauss, *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory*. 2012. doi: 10.4135/9781452230153.

[257] C. Marshall and G. B. Rossman, 'Designing qualitative research. 3rd edition', 1999.

[258] J. Pääkkönen and P. Ylikoski, 'Humanistic interpretation and machine learning', *Synthese*, vol. 199, no. 1–2, 2021, doi: 10.1007/s11229-020-02806-w.

[259] A. N. Soloshenko, Y. A. Orlova, V. L. Rozaliev, and A. V. Zaboleeva-Zotova, 'Establishing semantic similarity of the cluster documents and extracting key entities in the problem of

the semantic analysis of news texts', *Mod Appl Sci*, vol. 9, no. 5, 2015, doi: 10.5539/mas.v9n5p246.

[260] V. A. Anfara, K. M. Brown, and T. L. Mangione, 'Qualitative Analysis on Stage: Making the Research Process More Public', *Educational Researcher*, vol. 31, no. 7, 2002, doi: 10.3102/0013189X031007028.

[261] R. Cullen, 'Addressing the digital divide', *Online Information Review*, vol. 25, no. 5, 2001, doi: 10.1108/14684520110410517.

[262] F. Lupo-Pasini, 'Financial Inclusion And The "War For Cash"', *Law Contemp Probl*, vol. 84, no. 1, 2021.

[263] F. D. Davis, 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS Q*, vol. 13, no. 3, 1989, doi: 10.2307/249008.

[264] F. Bélanger and R. E. Crossler, 'Privacy in the digital age: A review of information privacy research in information systems', 2011. doi: 10.2307/41409971.

[265] A. Zarifis and X. Cheng, 'The six ways to build trust and reduce privacy concern in a central bank digital currency (CBDC)', in *Business Digital Transformation: Selected Cases from Industry Leaders*, 2023. doi: 10.1007/978-3-031-33665-2_6.

[266] S. Balaskas, M. Koutroumani, K. Komis, and M. Rigou, 'FinTech Services Adoption in Greece: The Roles of Trust, Government Support, and Technology Acceptance Factors', *FinTech*, vol. 3, no. 1, 2024, doi: 10.3390/fintech3010006.

[267] A. Iwańczuk-Kaliska, 'Potential Implications of Retail Central Bank Digital Currency for Banking Systems Identified in the Literature and by Central Banks', *Accounting, Economics and Law: A Convivium*, vol. 14, no. 2, 2024, doi: 10.1515/ael-2022-0005.

[268] Michael A. Cusumano, Annabelle Gawer, and David B. Yoffie, 'The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power HarperCollins, 2019, 320p.', *HarperCollins*, 2019.

[269] H.-W. Teng *et al.*, 'Mitigating Digital Asset Risks', *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4594467.

[270] Cordier, '"The central banks" evil plan : What if digital banks were a gateway to Central Bank Digital Currencies?"""', Belgium, 2024. [Online]. Available: http://hdl.handle.net/2078.1/thesis:47209

[271] R. S. Clark and V. L. P. Clark, 'The use of mixed methods to advance positive psychology: A methodological review', *International Journal of Wellbeing*, vol. 12, no. 3, 2022, doi: 10.5502/ijw.v12i3.2017.

[272] A. George, 'Quantitative and qualitative approaches to content analysis', in *Trends in content analysis*, 1959.

[273] S. Allen *et al.*, 'Design Choices for Central Bank Digital Currency: Policy and Technical Considerations', *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3670469.

[274] R. S. Samudrala and S. K. Yerchuru, 'Central bank digital currency: risks, challenges and design considerations for India', *CSI Transactions on ICT*, vol. 9, no. 4, pp. 245–249, Dec. 2021, doi: 10.1007/s40012-021-00344-5.

[275] A. Demirguc-Kunt, L. Klapper, D. Singer, S. Ansar, and J. Hess, *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. 2018. doi: 10.1596/978-1-4648-1259-0.

[276] A. Carstens, 'Digital Currencies and the Future Monetary System', *Bank for International Settlements Hoover Institution policy seminar*, vol. 89, no. 1, 2021.

[277] Á. Kerényi and J. Müller, 'Brave New Digital World? – Financial Technology and the Power of Information', *Financial and Economic Review*, vol. 18, no. 1, 2019, doi: 10.33893/fer.18.1.532.

[278] E. Brown and D. Piroska, 'Governing Fintech and Fintech as Governance: The Regulatory Sandbox, Riskwashing, and Disruptive Social Classification', *New Political Economy*, vol. 27, no. 1, 2022, doi: 10.1080/13563467.2021.1910645.

[279] A. Karabuk, T. Hauptberichter, S. Padó, and B. Plank, 'Challenges of Computational Social Science Analysis with NLP Methods Erenay Dayanik', 2022.

[280] K. Roberts, A. Dowell, and J. B. Nie, 'Attempting rigour and replicability in thematic analysis of qualitative research data; A case study of codebook development', *BMC Med Res Methodol*, vol. 19, no. 1, 2019, doi: 10.1186/s12874-019-0707-y.

[281] S. Mishra and A. K. Dey, 'Understanding and Identifying "Themes" in Qualitative Case Study Research', 2022. doi: 10.1177/22779779221134659.

[282] C. MacPhail, N. Khoza, L. Abler, and M. Ranganathan, 'Process guidelines for establishing Intercoder Reliability in qualitative studies', *Qualitative Research*, vol. 16, no. 2, 2016, doi: 10.1177/1468794115577012.

[283] D. A. Kochergin, 'Central banks digital currencies: World experience', *World Economy and International Relations*, vol. 65, no. 5, 2021, doi: 10.20542/0131-2227-2021-65-5-68-77.

[284] Y. Huang and M. Mayer, 'Digital currencies, monetary sovereignty, and U.S.–China power competition', *Policy Internet*, vol. 14, no. 2, 2022, doi: 10.1002/poi3.302.

[285] F. Tronnier, P. Biker, E. Baur, and S. Löbner, ‘An Evaluation of Information Flows in Digital Euro Transactions Using Contextual Integrity Theory’, in *ACM International Conference Proceeding Series*, 2023. doi: 10.1145/3590777.3590779.

[286] P. K. Ozili and S. L. N. Alonso, ‘Central Bank Digital Currency Adoption Challenges, Solutions, and a Sentiment Analysis’, *Journal of Central Banking Theory and Practice*, vol. 13, no. 1, 2024, doi: 10.2478/jcbtp-2024-0007.

[287] N. J. Dewasiri, Y. K. B. Weerakoon, and A. A. Azeez, ‘Mixed Methods in Finance Research: The Rationale and Research Designs’, *Int J Qual Methods*, vol. 17, no. 1, 2018, doi: 10.1177/1609406918801730.

[288] M. B. Legowo, S. Subanidja, and F. A. Sorongan, ‘Model of Sustainable Development Based on FinTech in Financial and Banking Industry: A Mixed-Method Research’, in *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, 2020. doi: 10.1109/IC2IE50715.2020.9274605.

[289] A. As-Salafiyah, A. S. Rusydiana, and Ihsanul Ikhwan, ‘CENTRAL BANK DIGITAL CURRENCY (CBDC): A SENTIMENT ANALYSIS AND LEGAL PERSPECTIVE’, *Journal of Central Banking Law and Institutions*, vol. 2, no. 2, 2023, doi: 10.21098/jcli.v2i2.177.

[290] U. Bindseil, ‘Central Bank Digital Currency: Financial System Implications and Control’, *International Journal of Political Economy*, vol. 48, no. 4, 2019, doi: 10.1080/08911916.2019.1693160.

[291] O. Freiman, ‘From Central Banks to Governments and Standards: CBDC Governance in Canada About the Hub’, 2024. [Online]. Available: www.cigionline.org

[292] F. Tronnier and W. Qiu, ‘How do privacy concerns impact actual adoption of central bank digital currency? An investigation using the e-cny in China’, *Quantitative Finance and Economics*, vol. 8, no. 1, 2024, doi: 10.3934/QFE.2024006.

[293] M. Pirgmann, ‘THE SENTIMENT FOR THE DIGITAL EURO: A EUROPEAN SURVEY NÁZORY NA DIGITÁLNE EURO: EURÓPSKY PRIESKUM’, 2024. [Online]. Available: http://www.mladaveda.sk

[294] M. Cayford, W. Pieters, and P. H. A. J. M. van Gelder, ‘Wanting it all – public perceptions of the effectiveness, cost, and privacy of surveillance technology’, *Journal of Information, Communication and Ethics in Society*, vol. 18, no. 1, 2020, doi: 10.1108/JICES-11-2018-0087.

[295] E. A. Whitley, ‘Perceptions of government technology, surveillance and privacy: the UK Identity Cards Scheme’, in *New Directions in Surveillance and Privacy*, 2013. doi: 10.4324/9781843927266-9.

[296] S. Ghafur, J. Van Dael, M. Leis, A. Darzi, and A. Sheikh, 'Public perceptions on data sharing: key insights from the UK and the USA', 2020. doi: 10.1016/S2589-7500(20)30161-8.

[297] N. Brockdorff, S. Appleby-Arnold, M. Krlic, D. Trottier, and E. Beaman, 'The citizens' perspective: Awareness, feelings and acceptance of surveillance and surveillance systems for fighting crime in the UK. A quantitative study', in *Rules, Expectations & Security through privacy-enhanced convenient technologies*, 2015.

[298] L. Fornaciari and C. Pesci, 'Global financial crisis and relevance of GRI disclosure in Italy. Insights from the stakeholder theory and the legitimacy theory', *FINANCIAL REPORTING*, no. 1, 2018, doi: 10.3280/fr2018-001003.

[299] V. M. Ngo, P. Van Nguyen, H. H. Nguyen, H. X. Thi Tram, and L. C. Hoang, 'Governance and monetary policy impacts on public acceptance of CBDC adoption', *Res Int Bus Finance*, vol. 64, 2023, doi: 10.1016/j.ribaf.2022.101865.

Appendices

Appendix 1- Link for code and related files: https://github.com/Code-Fintech-AI/CBDCs-Project.git

Appendix 2 - Model Experimentation results, as discussed in Chapter 5-

Run	Model	Epochs	Training Loss	Validation Loss	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0	DistilBERT	3	0.734	0.4183	87.18%	88.17%	87.18%	86.52%
1	DistilBERT	3	0.6847	0.4296	84.62%	86.05%	84.62%	84.17%
2	DistilBERT	3	0.708	0.4246	82.05%	83.54%	82.05%	79.33%
3	DistilBERT	3	0.717	0.4434	82.05%	84.39%	82.05%	79.90%
4	DistilBERT	3	0.7405	0.4842	84.62%	85.63%	84.62%	83.74%
5	DistilBERT	3	0.7108	0.4342	82.69%	84.68%	82.69%	80.80%
6	DistilBERT	3	0.6804	0.3944	87.18%	88.43%	87.18%	86.13%
7	DistilBERT	3	0.7439	0.5596	78.21%	85.02%	78.21%	73.51%
8	DistilBERT	3	0.7416	0.5008	79.49%	82.83%	79.49%	76.32%
9	DistilBERT	3	0.7161	0.4798	80.13%	80.57%	80.13%	79.85%
10	DistilBERT	3	0.7262	0.4819	83.97%	85.60%	83.97%	82.60%
11	DistilBERT	3	0.6989	0.4368	82.69%	86.26%	82.69%	80.89%
12	DistilBERT	3	0.7129	0.4258	80.13%	80.43%	80.13%	77.79%
13	DistilBERT	3	0.6997	0.511	80.13%	83.24%	80.13%	79.26%
14	DistilBERT	3	0.6931	0.434	81.41%	82.52%	81.41%	79.95%
15	DistilBERT	3	0.705	0.5847	77.56%	80.68%	77.56%	77.31%

16	DistilBERT	3	0.7328	0.4152	83.97%	84.45%	83.97%	83.17%
17	DistilBERT	3	0.7469	0.4623	80.13%	83.61%	80.13%	76.47%
18	DistilBERT	3	0.6423	0.3726	87.18%	86.92%	87.18%	87.00%
19	DistilBERT	3	0.7334	0.4626	86.97%	83.91%	83.97%	83.91%
20	DistilBERT	3	0.7175	0.5058	78.21%	81.32%	78.21%	76.87%
21	DistilBERT	3	0.7175	0.5058	78.21%	81.32%	78.21%	76.87%
22	DistilBERT	3	0.7086	0.3806	84.62%	85.36%	84.62%	83.69%
23	DistilBERT	3	0.7448	0.4304	86.54%	86.58%	86.54%	86.20%
24	DistilBERT	3	0.7448	0.4304	86.54%	86.58%	86.54%	86.20%
25	DistilBERT	3	0.6876	0.4471	83.33%	84.80%	83.33%	82.74%
26	DistilBERT	3	0.7329	0.4858	79.49%	82.47%	79.49%	77.19%
27	DistilBERT	3	0.7329	0.4858	79.49%	82.47%	79.49%	77.19%
28	DistilBERT	3	0.6924	0.4323	82.05%	83.89%	82.05%	81.79%
29	DistilBERT	3	0.717	0.438	82.69%	84.22%	82.69%	80.70%
0	DistilBERT	5	0.2375	0.0548	99.36%	99.37%	99.36%	99.36%
1	DistilBERT	5	0.284	0.0701	100.00%	100.00%	100.00%	100.00%
2	DistilBERT	5	0.2564	0.06	99.36%	99.37%	99.36%	99.36%
3	DistilBERT	5	0.248	0.0628	100.00%	100.00%	100.00%	100.00%
4	DistilBERT	5	0.2031	0.0584	99.36%	99.37%	99.36%	99.36%
5	DistilBERT	5	0.2757	0.0774	98.72%	98.76%	98.72%	98.72%
6	DistilBERT	5	0.2553	0.06	100.00%	100.00%	100.00%	100.00%

7	DistilBERT	5	0.2508	0.1643	94.87%	95.75%	94.87%	94.92%
8	DistilBERT	5	0.2731	0.0838	98.08%	98.17%	98.08%	98.05%
9	DistilBERT	5	0.2698	0.0572	100.00%	100.00%	100.00%	100.00%
10	DistilBERT	5	0.3135	0.0796	98.08%	98.12%	98.08%	98.04%
11	DistilBERT	5	0.2539	0.0752	99.36%	99.37%	99.36%	99.36%
12	DistilBERT	5	0.3016	0.0665	99.36%	99.37%	99.36%	99.36%
13	DistilBERT	5	0.2478	0.0733	99.36%	99.38%	99.36%	99.36%
14	DistilBERT	5	0.2404	0.0738	100.00%	100.00%	100.00%	100.00%
15	DistilBERT	5	0.2611	0.0686	98.72%	98.73%	98.72%	98.72%
16	DistilBERT	5	0.2662	0.0705	98.72%	98.74%	98.72%	98.70%
17	DistilBERT	5	0.2329	0.067	100.00%	100.00%	100.00%	100.00%
18	DistilBERT	5	0.2861	0.0726	98.72%	98.74%	98.72%	98.70%
19	DistilBERT	5	0.2861	0.0726	98.72%	98.74%	98.72%	98.70%
20	DistilBERT	5	0.2374	0.068	98.72%	98.74%	98.72%	98.70%
21	DistilBERT	5	0.2063	0.0643	99.36%	99.37%	99.36%	99.36%
22	DistilBERT	5	0.2986	0.098	96.15%	96.33%	96.15%	95.98%
23	DistilBERT	5	0.2319	0.067	98.08%	98.10%	98.08%	98.07%
24	DistilBERT	5	0.2321	0.0716	100.00%	100.00%	100.00%	100.00%
25	DistilBERT	5	0.3048	0.0966	97.44%	97.75%	97.44%	97.47%
26	DistilBERT	5	0.2555	0.0776	98.08%	98.12%	98.08%	98.04%
27	DistilBERT	5	0.2568	0.0718	98.72%	98.76%	98.72%	98.71%

28	DistilBERT	5	0.2568	0.0718	98.72%	98.76%	98.72%	98.71%
29	DistilBERT	5	0.2765	0.0885	99.36%	99.37%	99.36%	99.36%
0	RoBERTa	3	0.3717	0.1923	93.59%	93.55%	93.59%	93.56%
1	RoBERTa	3	0.3551	0.1652	95.51%	95.49%	95.51%	95.48%
2	RoBERTa	3	0.3573	0.1884	94.23%	94.35%	94.23%	94.14%
3	RoBERTa	3	0.361	0.1811	95.51%	95.48%	95.51%	95.48%
4	RoBERTa	3	0.3556	0.1602	95.51%	95.49%	95.51%	95.48%
5	RoBERTa	3	0.3489	0.1474	94.87%	94.84%	94.87%	94.83%
6	RoBERTa	3	0.3385	0.1391	96.15%	96.16%	96.15%	96.13%
7	RoBERTa	3	0.3569	0.1679	95.51%	95.59%	95.51%	95.51%
8	RoBERTa	3	0.3677	0.1405	96.79%	96.83%	96.79%	96.80%
9	RoBERTa	3	0.3874	0.1978	92.31%	92.59%	92.31%	92.28%
10	RoBERTa	3	0.3763	0.2006	96.15%	96.20%	96.15%	96.16%
11	RoBERTa	3	0.3692	0.2309	91.67%	92.15%	91.67%	91.49%
12	RoBERTa	3	0.359	0.1908	92.95%	93.75%	92.95%	93.04%
13	RoBERTa	3	0.3665	0.2018	92.31%	92.75%	92.31%	92.23%
14	RoBERTa	3	0.381	0.1893	94.87%	94.87%	94.87%	94.87%
15	RoBERTa	3	0.3524	0.1652	95.51%	95.53%	95.51%	95.47%
16	RoBERTa	3	0.3682	0.1535	93.59%	93.70%	93.59%	93.56%
17	RoBERTa	3	0.368	0.1537	94.87%	95.17%	94.87%	94.84%
18	RoBERTa	3	0.3639	0.1858	92.95%	93.44%	92.95%	92.90%

19	RoBERTa	3	0.3561	0.1575	94.87%	94.87%	94.87%	94.82%
20	RoBERTa	3	0.358	0.1689	94.87%	95.06%	94.87%	94.92%
21	RoBERTa	3	0.3877	0.1876	94.23%	94.29%	94.23%	94.25%
22	RoBERTa	3	0.3711	0.155	94.87%	94.87%	94.87%	94.87%
23	RoBERTa	3	0.3684	0.2359	90.38%	90.70%	90.38%	90.19%
24	RoBERTa	3	0.361	0.2282	92.31%	92.56%	92.31%	92.03%
25	RoBERTa	3	0.365	0.1933	94.23%	94.29%	94.23%	94.15%
26	RoBERTa	3	0.3405	0.1437	96.15%	96.14%	96.15%	96.13%
27	RoBERTa	3	0.3907	0.2415	92.31%	92.83%	92.31%	92.12%
28	RoBERTa	3	0.38	0.2239	91.67%	91.64%	91.67%	91.53%
29	RoBERTa	3	0.3731	0.1789	93.59%	94.11%	93.59%	93.38%
0	RoBERTa	5	0.17	0.0722	97.44%	97.71%	97.44%	97.46%
1	RoBERTa	5	0.1585	0.0479	99.36%	99.37%	99.36%	99.36%
2	RoBERTa	5	0.1833	0.035	100.00%	100.00%	100.00%	100.00%
3	RoBERTa	5	0.1206	0.0458	98.08%	98.07%	98.08%	98.07%
4	RoBERTa	5	0.1888	0.1403	91.67%	92.94%	91.67%	91.22%
5	RoBERTa	5	0.1588	0.0753	97.44%	97.61%	97.44%	97.40%
6	RoBERTa	5	0.1485	0.0519	98.72%	98.79%	98.72%	98.72%
7	RoBERTa	5	0.1435	0.0303	99.36%	99.37%	99.36%	99.36%
8	RoBERTa	5	0.1377	0.0336	99.36%	99.37%	99.36%	99.36%
9	RoBERTa	5	0.1607	0.0364	98.72%	98.73%	98.72%	98.72%

10	RoBERTa	5	0.1409	0.0631	97.44%	97.45%	97.44%	97.42%
11	RoBERTa	5	0.1548	0.1074	97.44%	97.54%	97.44%	97.41%
12	RoBERTa	5	0.1616	0.0427	98.72%	98.76%	98.72%	98.71%
13	RoBERTa	5	0.1867	0.0641	98.08%	98.16%	98.08%	98.07%
14	RoBERTa	5	0.1506	0.0302	99.36%	99.37%	99.36%	99.36%
15	RoBERTa	5	0.178	0.0515	98.08%	98.11%	98.08%	98.07%
16	RoBERTa	5	0.1524	0.0477	98.72%	98.76%	98.72%	98.71%
17	RoBERTa	5	0.1453	0.0253	100.00%	100.00%	100.00%	100.00%
18	RoBERTa	5	0.1638	0.0368	100.00%	100.00%	100.00%	100.00%
19	RoBERTa	5	0.1668	0.0906	98.08%	98.10%	98.08%	98.07%
20	RoBERTa	5	0.1976	0.0864	96.79%	96.97%	96.79%	96.82%
21	RoBERTa	5	0.1284	0.0328	100.00%	100.00%	100.00%	100.00%
22	RoBERTa	5	0.1585	0.055	98.08%	98.07%	98.08%	98.07%
23	RoBERTa	5	0.1525	0.0373	98.72%	98.76%	98.72%	98.71%
24	RoBERTa	5	0.1449	0.0423	99.36%	99.37%	99.36%	99.36%
25	RoBERTa	5	0.1725	0.0663	98.72%	98.76%	98.72%	98.71%
26	RoBERTa	5	0.1284	0.0227	100.00%	100.00%	100.00%	100.00%
27	RoBERTa	5	0.1524	0.036	98.72%	98.76%	98.72%	98.70%
28	RoBERTa	5	0.2051	0.0853	96.79%	96.89%	96.79%	96.79%
29	RoBERTa	5	0.1548	0.0514	99.36%	99.37%	99.36%	99.36%
0	XLM-RoBERTa	3	0.4208	0.208	94.23%	94.41%	94.23%	94.28%

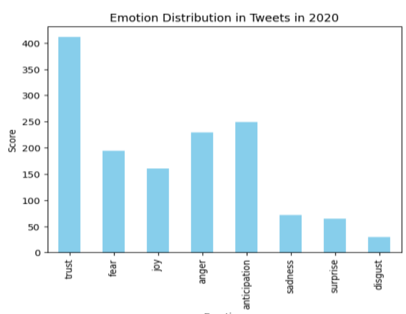
1	XLM-RoBERTa	3	0.4386	0.206	92.95%	93.19%	92.95%	92.97%
2	XLM-RoBERTa	3	0.3932	0.2147	92.95%	93.21%	92.95%	92.88%
3	XLM-RoBERTa	3	0.4472	0.2804	93.59%	93.86%	93.59%	93.64%
4	XLM-RoBERTa	3	0.4024	0.2613	91.67%	93.43%	91.67%	91.87%
5	XLM-RoBERTa	3	0.401	0.2262	94.23%	94.54%	94.23%	94.28%
6	XLM-RoBERTa	3	0.4389	0.2946	88.46%	90.84%	88.46%	88.69%
7	XLM-RoBERTa	3	0.4432	0.2489	92.95%	93.33%	92.95%	92.98%
8	XLM-RoBERTa	3	0.4177	0.1776	96.15%	96.20%	96.15%	96.17%
9	XLM-RoBERTa	3	0.3876	0.1872	95.51%	95.65%	95.51%	95.55%
10	XLM-RoBERTa	3	0.4331	0.2207	93.59%	93.52%	93.59%	93.53%
11	XLM-RoBERTa	3	0.405	0.2205	94.23%	94.47%	94.23%	94.25%
12	XLM-RoBERTa	3	0.4298	0.1874	95.51%	95.61%	95.51%	95.55%
13	XLM-RoBERTa	3	0.3932	0.1895	94.87%	95.22%	94.87%	94.90%
14	XLM-RoBERTa	3	0.4591	0.4507	77.56%	86.34%	77.56%	77.65%
15	XLM-RoBERTa	3	0.3674	0.1937	93.59%	94.27%	93.59%	93.60%
16	XLM-RoBERTa	3	0.3933	0.2139	93.59%	93.63%	93.59%	93.60%
17	XLM-RoBERTa	3	0.4087	0.1879	94.23%	94.23%	94.23%	94.23%

18	XLM-RoBERTa	3	0.4193	0.2669	91.03%	91.71%	90.13%	91.05%
19	XLM-RoBERTa	3	0.6924	0.4664	82.69%	85.52%	82.69%	82.06%
20	XLM-RoBERTa	3	0.7494	0.533	77.56%	78.02%	77.56%	76.56%
21	XLM-RoBERTa	3	0.6817	0.4004	86.54%	87.09%	86.54%	85.63%
22	XLM-RoBERTa	3	0.7276	0.48	85.26%	85.03%	85.26%	84.52%
23	XLM-RoBERTa	3	0.6548	0.4022	87.82%	89.78%	87.82%	87.83%
24	XLM-RoBERTa	3	0.7347	0.4855	87.82%	87.59%	87.82%	87.57%
25	XLM-RoBERTa	3	0.6395	0.353	89.74%	89.99%	89.74%	89.40%
26	XLM-RoBERTa	3	0.6973	0.4302	86.54%	89.06%	86.54%	84.91%
27	XLM-RoBERTa	3	0.7037	0.3928	90.38%	90.49%	90.38%	90.14%
28	XLM-RoBERTa	3	0.7073	0.4195	83.33%	85.03%	83.33%	82.42%
29	XLM-RoBERTa	3	0.7405	0.5275	80.77%	84.76%	80.77%	77.89%
0	XLM-RoBERTa	5	0.2558	0.1049	98.08%	98.13%	98.08%	98.07%
1	XLM-RoBERTa	5	0.1938	0.0646	99.36%	99.38%	99.36%	99.36%
2	XLM-RoBERTa	5	0.1873	0.0636	99.36%	99.37%	99.36%	99.36%
3	XLM-RoBERTa	5	0.1707	0.0719	97.44%	97.57%	97.44%	97.45%
4	XLM-RoBERTa	5	0.196	0.1567	93.59%	94.22%	93.59%	93.65%

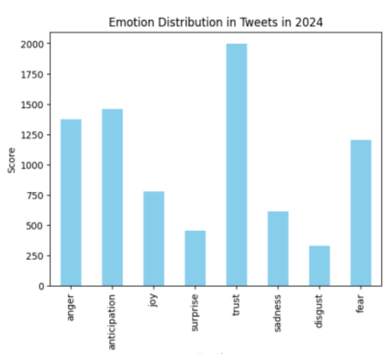
5	XLM-RoBERTa	5	0.1699	0.1645	94.23%	95.07%	94.23%	94.33%
6	XLM-RoBERTa	5	0.181	0.0512	98.08%	98.08%	98.08%	98.07%
7	XLM-RoBERTa	5	0.1689	0.1473	95.51%	95.70%	95.51%	95.54%
8	XLM-RoBERTa	5	0.1735	0.1333	96.25%	96.21%	96.15%	96.17%
9	XLM-RoBERTa	5	0.1992	0.1268	96.15%	96.18%	96.15%	96.16%
10	XLM-RoBERTa	5	0.2116	0.1405	94.87%	94.90%	94.87%	94.86%
11	XLM-RoBERTa	5	0.173	0.1485	94.87%	95.18%	94.87%	95.90%
12	XLM-RoBERTa	5	0.1645	0.0756	97.44%	97.60%	97.44%	97.45%
13	XLM-RoBERTa	5	0.1645	0.0756	97.44%	97.60%	97.44%	97.45%
14	XLM-RoBERTa	5	0.159	0.0765	97.44%	97.48%	97.44%	97.44%
15	XLM-RoBERTa	5	0.1959	0.1054	96.15%	96.21%	96.15%	96.17%
16	XLM-RoBERTa	5	0.186	0.1022	96.15%	96.52%	96.15%	96.17%
17	XLM-RoBERTa	5	0.2092	0.13	94.87%	95.26%	94.87%	94.92%
18	XLM-RoBERTa	5	0.1741	0.1797	92.95%	93.60%	92.95%	93.01%
19	XLM-RoBERTa	5	0.1922	0.1156	96.15%	96.27%	96.15%	96.18%
20	XLM-RoBERTa	5	0.171	0.1212	96.15%	96.15%	96.15%	96.15%
21	XLM-RoBERTa	5	0.1714	0.0975	97.44%	97.46%	97.44%	97.44%

22	XLM-RoBERTa	5	0.1725	0.0733	96.79%	96.95%	96.79%	96.79%
23	XLM-RoBERTa	5	0.1866	0.1029	96.15%	96.52%	96.15%	96.17%
24	XLM-RoBERTa	5	0.1456	0.1179	95.51%	95.85%	95.51%	95.56%
25	XLM-RoBERTa	5	0.2171	0.0947	96.15%	96.38%	96.15%	96.16%
26	XLM-RoBERTa	5	0.1698	0.0865	96.15%	96.21%	96.15%	96.15%
27	XLM-RoBERTa	5	0.1892	0.1307	94.87%	94.87%	94.87%	94.86%
28	XLM-RoBERTa	5	0.2154	0.1005	96.15%	96.18%	96.15%	96.16%
29	XLM-RoBERTa	5	0.1917	0.0975	96.79%	96.79%	96.79%	96.79%

Appendix 3- Positive emotions dominated in 2020



Appendix 4- Emotion distribution in 2024



Appendix 5- LDA hyperparameter tuning results across all years-

Year	Alpha	Eta	Coherence score
2020	symmetric	auto	0.4213
2020	symmetric	0.01	0.4111
2020	symmetric	0.1	0.4192
2020	**asymmetric**	**auto**	**0.4728**
2020	asymmetric	0.01	0.4663
2020	asymmetric	0.1	0.4722
2020	0.01	auto	0.4287
2020	0.01	0.01	0.4069
2020	0.01	0.1	0.4177
2020	0.1	auto	0.4353
2020	0.1	0.01	0.4153
2020	0.1	0.1	0.4093
2023	symmetric	auto	0.3724
2023	symmetric	0.01	0.4105
2023	symmetric	0.1	0.4234
2023	asymmetric	auto	0.3951
2023	asymmetric	0.01	0.4412
2023	**asymmetric**	**0.1**	**0.4556**
2023	0.01	auto	0.3802

2023	0.01	0.01	0.4181
2023	0.01	0.1	0.4211
2023	0.1	auto	0.3824
2023	0.1	0.01	0.4339
2023	0.1	0.1	0.4144
2024	symmetric	auto	0.4068
2024	symmetric	0.01	0.4087
2024	symmetric	0.1	0.429
2024	asymmetric	auto	0.4002
2024	asymmetric	0.01	0.3871
2024	asymmetric	0.1	0.4346
2024	0.01	auto	0.4074
2024	0.01	0.01	0.4019
2024	**0.01**	**0.1**	**0.4381**
2024	0.1	auto	0.4179
2024	0.1	0.01	0.4211
2024	0.1	0.1	0.4322

Appendix 6- LDA topics for 2020

Topic no.	Topic label	Top words
1	European News and Cryptocurrency	health, get, market, world, cbdc, news, boe, issue, dont, offer

2	Future Digital Currencies and Economic Considerations	digital, say, cbdc, mark, carney, risk, currency, monetary, planned, banking
3	CBDC Design and Cash Discussions	cbdc, anonymity, bank, system, time, account, would, deposit, rather, design
4	International Crypto Developments	like, france, many, cbdc, sector, crypto, work, even, way, look
5	CBDCs in Global Central Banks	bank, england, central, potential, digital, paper, currency, cbdc, discussion, government
6	Technological Advancements in Finance	digital, pound, new, billion, technology, service, year, money, financial, thats
7	Economic and Settlement Discussions	china, cbdc, going, consider, without, could, law, cryptocurrency, work, way
8	Bank Risks and Anonymity Concerns	central, bank, currency, digital, sweden, canada, switzerland, japan, settlement, european
9	Monetary Systems and Financial Benefits	want, via, england, future, model, benefit, financial, system, cbdcs, cryptocurrency
10	European Banking and Token-Based Systems	retail, bank, europe, based, similar, towards, central, would, token, much
11	Banking Business and Trading	trading, business, cbdc, year, used, project, explore, money, form, bank
12	Retail Banking Models and Fintech Projects	cash, king, economy, already, point, year, working, even, way, take
13	Cash vs Anonymity in Markets	multimillion, people, see, feature, risk, sterling, day, government, gold, world
14	Mark Carney and Digital Pound Challenges	could, challenge, pound, present, bitcoin, digital, read, crypto, governance, cryptocurrency
15	Government Projects and Digital Pound	pound, digital, million, left, dollar, project, could, company, currency, bank
16	China’s Blockchain and CBDC Initiatives	payment, cbdc, blockchain, need, digital, currency, anonymity, way, crypto, cryptocurrency

17	International Financial Players and CBDCs	multimillion, people, see, feature, risk, sterling, day, government, gold, world
18	Payment Systems and Anonymity	multimillion, people, see, feature, risk, sterling, day, government, gold, world
19	Global Financial Systems and Work Practices	multimillion, people, see, feature, risk, sterling, day, government, gold, world
20	Fintech Sector and Digital Money	fintech, firm, british, money, major, help, crypto, digital, cryptocurrency, law

Appendix 7- LDA topics for 2023

Topic no.	Topic label	Top words
1	UK Bank and Digital Pound Developments	pound, new, plan, launch, cbdc, via, news, cap, call, briton
2	Britcoin and Leadership in Digital Currency	pound, could, britcoin, sTablecoins, would, access, cbdc, foundation, part, cunliffe
3	Policy Concerns and Societal Impact	bitcoin, way, work, find, policy, even, industry, every, live, general
4	Tokens and Crypto Challenges	want, use, one, launched, transaction, tax, many, council, cbdc, bring
5	Blockchain Adoption and Private Sector Involvement	private, know, blockchain, open, here, coexist, recommends, adopting, egbp, mixed
6	Bitcoin and Global Crypto Regulation	government, control, restrict, get, step, week, cbdc, card, state, still
7	Regulatory Roadmap and Future Implementation	money, cash, already, freedom, feb, everyone, illegal, spending, explaining, danger
8	Financial System and Individual Credit	would, could, limit, decade, payment, financial, potential, system, wallet, launching
9	Project Payments and Ripple Support	pound, say, project, likely, need, cbdc, end, using, help, support

10	Privacy, Risk, and Cryptocurrency Usage	crypto, privacy, risk, first, cbdc, user, interest, data, question, called
11	CBDC and Future Treasury Developments	future, read, paper, needed, boe, take, used, published, cbdc, soon
12	British Monetary Design and International Collaboration	time, people, good, looking, developing, dollar, cbdc, another, btc, period
13	Freedom, Security, and Global Threats	cbdcs, working, qnt, two, cant, yet, sec, xrp, etn, avax
14	Challenges in Digital Currency Adoption	dont, wont, make, british, ahead, thats, impact, rather, push, forget
15	Government Control and CBDC Implementation	ceo, spend, big, value, come, benefit, june, would, fiat, much
16	Cryptocurrency Launching and Authority Responses	preserved, iso, preserve, consider, extent, method, everyones, compliant, steele, involving
17	Cash Usage and Payment News	pound, bank, england, treasury, pay, likely, announced, later, hunt, trusted
18	Economy and Finance Global Trends	economy, finance, next, going, ripple, given, thing, innovation, cbdc, month
19	Consumer Decisions and WEF Influence	think, like, within, coming, decision, year, rosalind, done, developed, safe
20	Retail, Company Strategies, and Tracking	case, retail, company, probably, purchase, isnt, cbdc, track, saying, use
21	Central Bank and Global Monetary System	bank, central, currency, reserve, instead, federal, coexist, holding, adopting, recommends
22	STablecoins and Asset Management	form, business, money, said, payment, right, limited, see, idea, happen
23	Consultations and Public Crypto Commentary	consultation, public, also, may, thought, might, important, running, attention, cbdc
24	CEO Discussions and Carbon-Related Projects	preserved, iso, preserve, consider, extent, method, everyones, compliant, steele, involving

25	Tax Reforms and Political Campaigns in Digital Currency	preserved, iso, preserve, consider, extent, method, everyones, compliant, steele, involving

Appendix 8- LDA topics for 2024

Topic no.	Topic label	Top words
1	Government Adoption and Economic Considerations	cbdc, pound, need, people, time, get, way, coming, know, one
2	BoE, Legislation, and Parliamentary Control	cbdc, plan, government, control, boe, legislation, week, proposed, whether, security
3	CBDC Design, Privacy, and Banking Concerns	pound, privacy, concern, design, future, labour, approach, party, issue, banking
4	Freedom, Economic Control, and Public Opinion	use, still, yet, undecided, freedom, solution, via, economic, used, create
5	UK Digital Pound and Financial Governance	pound, bank, england, treasury, currency, central, consultation, cbdc, privacy, potential
6	Government Decision-Making and Public Voting Trends	see, want, thats, already, government, world, interesting, today, building, sign
7	Legal Frameworks for CBDCs and Bitcoin	would, say, present, cbdc, move, law, economy, bitcoin, well, part
8	Innovation, Finance, and Data Privacy in CBDCs	also, innovation, pound, data, news, finance, remains, push, balance, including
9	Global Systems and CBDC Implementation Challenges	system, cbdc, form, wef, zero, stop, implement, usa, right, bill
10	Life Changes, Global Conflict, and CBDC Adaptations	china, problem, cbdc, life, war, pushing, live, follow, without, lead
11	Surveillance and City-Based Digital Identification Systems	cbdcs, dont, etc, working, city, surveillance, must, passport, idea, actually
12	Global STablecoins and Retail Measures	control, good, thing, every, keep, terrorist, understand, certainly, scandal, listen

13	Technology Advancements in the Digital World	new, technology, service, question, far, experience, rapidly, looking, detail, led
14	Citizen Control, Tax, and Surveillance Concerns	whats, citizen, track, fight, currently, free, one, purchase, best, everywhere
15	Monetary, Potential, and Impact on Adoption	monetary, potential, interesting, read, excited, impact, adoption, cbdc, learn, article
16	Business Opportunities, Interest, and Sunak's Priorities	opportunity, big, business, official, pound, global, say, major, interest, provides
17	Cash, Access Rights, and Privacy Commitments	cash, pound, money, payment, access, crypto, bank, britcoin, towards, introduced
18	Economic Power, Fintech, and Future Planning	financial, could, next, year, take, privacy, find, freedom, account, fraud
19	Cryptocurrency Development and Britcoin Trends	promise, policing, robust, responded, respect, protected, serious, legal, mass, spent
20	Coming Contract Scandals and Digital Initiatives	coming, contract, scandal, launch, australia, passed, fujitsu, platform, swift, awarded

Appendix 9- NMF hyperparameter tuning results across all years-

Year	n_components	l1_ratio	Coherence score
2020	**15**	**0**	**0.7616**
2020	15	0.5	0.7616
2020	15	1	0.7616
2020	20	0	0.6556
2020	20	0.5	0.6556
2020	20	1	0.6556
2020	25	0	0.6228
2020	25	0.5	0.6228

2020	25	1	0.6228
2023	**15**	**0**	**0.5739**
2023	15	0.5	0.5739
2023	15	1	0.5739
2023	20	0	0.5411
2023	20	0.5	0.5411
2023	20	1	0.5411
2023	25	0	0.5332
2023	25	0.5	0.5332
2023	25	1	0.5332
2024	15	0	0.5016
2024	15	0.5	0.5016
2024	15	1	0.5016
2024	**20**	**0**	**0.5046**
2024	20	0.5	0.5046
2024	20	1	0.5046
2024	25	0	0.4876
2024	25	0.5	0.4876
2024	25	1	0.4876

Appendix 10- NMF topics for 2020

Topic no.	Topic label	Top words
1	CBDC-Based Retail Systems	based, retail, central, allow, similar, token, manner, ecb, circulate, anonymity
2	Cash and Economic Flow	cash, king, market, profit, economy, pandemic, flow, care, world, going
3	Challenges and Leadership in Digital Currency	present, challenge, mark, say, carney, pound, digital, bitcoin, crypto, coindesk
4	Digital Pound Services	pound, digital, service, year, billion, new, multimillion, british, health, money
5	CBDC Anonymity and Payments	cbdc, anonymity, payment, account, need, like, deposit, idea, anonymous, look
6	International CBDC Initiatives	china, planned, little, common, appear, technology, currency, unitedkingdom, cryptocurrencies, digital
7	Central Bank Papers and Designs	bank, central, england, paper, currency, discussion, digital, boe, design, exploring
8	Central Bank Collaborations	central, bank, japan, sweden, switzerland, canada, currency, ass, group, explore
9	Crypto Market Movements	coinbase, ceo, zeeshan, yuan, coin, usdc, pump, listing, battle, cbdcs
10	Financial Feedback and Retail Banking	model, england, feedback, invite, platform, retail, fintech, linklaters, blog, cbdc
11	Governance and Monetary Policy	monetary, governance, potential, risk, carney, mark, say, highlighted, governor, outgoing
12	Project Transformation and Economy	transformation, project, company, left, million, manufacturing, missing, unclaimed, exist, scheme
13	Legal Frameworks and Blockchain	china, law, cryptography, firm, encryption, private, way, blockchain, standard, new
14	Future Financial Benefits	future, financial, benefit, cbdcs, sterling, looking, banknote, point, pandemic, word

15	Digital Pound Characteristics	mar, gbp, size, zip, pdf, pound, oven, shape, tonight, gold

Appendix 11- NMF topics for 2023.

Topic no.	Topic label	Top words
1	Bank of England's Plans	england, bank, plan, briton, cap, face, treasury, governor, deputy, create
2	Support and Project Needs	need, likely, treasury, support, say, project, england, bank, pound, cbdc
3	Digital Pound Foundations	pound, digital, foundation, limit, britcoin, consumer, case, pay, mean, work
4	Programmable CBDCs and Privacy	cbdc, programmable, petition, coming, anonymity, introduction, year, implement, prevent, soon
5	Everyday Digital Payments	new, form, household, money, business, payment, digital, everyday, pound, news
6	Central Bank Issuance	currency, central, bank, digital, country, issued, potential, fiat, cbdcs, end
7	Government Control and Cryptocurrency	government, bitcoin, pursues, restrict, ceo, access, bank, zerohedge, gold, lunarcrush
8	STablecoins and Coexistence	coexist, private, sTablecoins, central, bank, think, pound, belief, egbp, mixed
9	Public Concerns and Development	decade, developed, read, age, according, safe, priority, likely, think, public
10	Privacy and Technology	privacy, boe, focus, pseudonymous, chief, technology, cbdc, blockchain, anonymity, user
11	Consultation Papers and Deadlines	consultation, paper, june, response, respond, public, deadline, published, close, released
12	Project Launch and Research	launch, project, rosalind, closer, study, step, launching, england, britcoin, bank

13	Business Opportunities and Growth	opportunity, big, say, present, official, business, provides, england, bloomberg, bank
14	Cash, Control, and Public Freedom	cash, control, want, people, dont, cbdcs, money, freedom, king, use
15	Corporate Strategy in Crypto	crypto, interoperable, lobbyist, say, britcoin, future, like, group, company, needed

Appendix 12- NMF topics for 2024

Topic no.	**Topic label**	**Top words**
1	Future of the Digital Pound	pound, digital, foundation, future, decision, money, working, launch, legislation, finance
2	Opportunities and Official Statements	opportunity, big, present, england, bank, cbdc, pound, digital, say, breaking
3	Advancements and Stability	feasibility, advancing, exploring, cbdc, news, crypto, january, cryptocurrency, britcoin, stability
4	Public Opinion and Freedom	wef, people, net, zero, vote, cbdc, reform, party, freedom, cbdcs
5	Privacy and Progress	persist, worry, progress, plan, privacy, cbdc, pound, digital, learn, feed
6	Government and Privacy Concerns	concern, privacy, approach, expert, help, manage, pound, digital, remains, undecided
7	Currency Design and Implementation	currency, central, bank, digital, cbdc, affect, designing, potential, country, launch
8	Treasury Decisions and Public Response	treasury, england, bank, decision, undecided, reach, remain, boe, introduced, middle
9	Security and Policy Initiatives	labour, tokenization, hub, work, security, advance, party, want, aim, make
10	Digital Pound Design Feedback	design, amid, criticism, banking, accelerates, privacy, concern, working, start, pound

11	Managing Public Expectations	say, boe, expert, manage, help, coexist, cbdcs, sign, loud, quiet
12	Cash Access and Privacy Preservation	cash, access, tread, carefully, prioritizing, privacy, government, code, commitment, reiterate
13	Political Commitments and Privacy	financial, sunak, rishi, plan, freedom, govt, equality, introduce, year, privacy
14	Future Control and Freedom	want, future, start, acquiring, money, head, consider, thought, domain, control
15	Monetary Impact and Adoption	monetary, potential, interesting, read, excited, impact, adoption, cbdc, learn, article
16	Business Growth and Opportunities	new, business, official, according, lucrative, boe, opportunity, bring, provides, broadbent
17	Public Control and Cryptography	cbdc, like, need, way, know, dont, going, look, control, crypto
18	Consultation and Legislation Responses	consultation, government, response, published, legislation, release, result, thursday, received, responded
19	Social Credit and Surveillance	credit, social, score, carbon, city, minute, china, deducted, cbdc, buy
20	Government and Business Collaboration	coming, contract, scandal, launch, australia, passed, fujitsu, platform, swift, awarded

Appendix 13- Codebook link: https://github.com/Code-Fintech-AI/CBDCs-Project/commit/1e3f88c876c86afcc95b10b24119e477079842bb

Appendix 14- Mapping Table link: https://github.com/Code-Fintech-AI/CBDCs-Project/commit/a7bbd6fa4282f6ba3d86f1f793bcd15e0157cb12#diff-9e00700add058b1aa9cdd78e33182d3020fea9be1612b9138a85879982849606