A HUMAN-CENTRED APPROACH TO SOCIAL MEDIA ANALYSIS FOR DECISION-MAKING

XIN YU LIEW



Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

September 2025

Supervisors:
Prof Joel E. Fischer
Dr Jeremie Clos
Dr Nazia Hameed

Xin Yu Liew: *A Human-Centred Approach to Social Media Analysis for Decision-Making*, Thesis submitted to the University of Nottingham for the degree of , Doctor of Philosophy, © September 2025

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof Joel Fischer, Dr Jeremie Clos, and Dr Nazia Hameed. This thesis and my entire PhD journey would not have been possible without the trust, guidance, and faith that they placed in me. Their advice and encouragement carried me through both the challenging and rewarding stages of this research, and the opportunities they introduced have shaped my growth as a researcher. Thank you for providing me with the mentorship that has shaped how I think and approach problems. I am forever grateful to have built a supportive relationship with them as my supervisor during my time as a PhD student - one that I will always cherish.

Special thanks to the external advisors and examiners who took an interest in my research and generously provided me with their time and expertise. I am especially grateful to my examiners, Prof Gina Neff and Dr Kai Xu, for their support and encouragement during my viva, making the experience far more rewarding than I had anticipated. Their constructive feedback helped improve the final refinement of this thesis. I am also deeply thankful to Prof Neff for kindly connecting me with Dr Stefanie Felsberger and Dr Hugo Leal from the AI4TRUST project. Their insights, knowledge, and perspectives helped inspire the early stages of this research. I would also like to thank Prof Colin Strong from Ipsos for having conversations about my early ideas and providing valuable input at various stages of this research.

I am also grateful to the European Conference on Information Retrieval (ECIR) 2023 Doctoral Consortium for the opportunity to present my early research ideas and receive valuable feedback from peers and mentors. I am especially grateful to Prof Allan Hanbury for provid-

ing constructive feedback at an early stage of my work, which helped refine the direction of my research.

This research was supported by the Engineering and Physical Sciences Research Council [Grant number: EP/V00784X/1]. I am grateful to the University of Nottingham for funding my PhD and for providing the resources and opportunities that facilitated my academic journey. I also wish to thank my peers in the Mixed Reality Lab for their warm welcome, encouragement, and continued support throughout my PhD.

On a more personal note, I would like to thank the people who have been my source of strength and comfort outside of academia.

To my beloved family - Daddy, Mummy, my sister, and my baby brother - thank you for your unwavering love, patience, and support throughout my life, and especially during the years I spent working on this PhD.

To my partner Cheehong, who shared every up and down of this challenging yet exciting journey. Thank you for being my ever-patient listener, for your enthusiasm, and for spending countless hours exploring the concepts and ideas I devised over the years. I could not have done this without you by my side.

To my best friend, Kheeyen, who understood this journey better than anyone and always held my hand with confidence at every step. To all my closest friends at home, thank you for being my biggest encouragers since the beginning. Finally, to all the friends I made in Nottingham who made it feel like home, I am forever grateful for the much-needed Malaysian spirit you brought into my life in Notts.

ABSTRACT

Social media has become indispensable in the digital age, frequently used for engaging with contemporary social agendas. These platforms not only shape public discourses but also influence individuals' perceptions and behaviours. This influence was especially evident during critical moments such as the COVID-19 pandemic, where social media played a central role in disseminating information and shaping public responses. As a result, key stakeholders, including policymakers, journalists, and communication professionals, increasingly rely on interpreting these dynamic social interactions to inform strategic decisions. This is due to the persuasive power embedded in social media, making it instrumental in guiding organisational objectives. However, the complexity and volume of social media discourses pose significant challenges for existing analytical tools in providing useful insights required for effective decision-making. Therefore, there is a growing need to enhance these analytical tools to support informed decision-making.

Human-centred approaches to designing tools that integrate human expertise with the qualitative analyses of social media discourses remain underexplored. Addressing this gap requires understanding the limitations of both machine and human perspectives and how they can effectively complement each other. This thesis approaches the problem from two disciplinary perspectives. First, it applies Natural Language Processing (NLP) methods to establish the technical feasibility of analysing social media discourse and to identify where computational learning's limitations can be enhanced with human input. Second, it adopts Human-Computer Interaction (HCI) methods to design an interactive system informed by these technical findings, allowing design choices to be grounded in empirical evidence that highlights where and how human input can complement computational

analysis. Then, the resulting interactive design is evaluated through user studies to examine how grounding the design in both computational limitations and human strengths influences human experiences during interaction with these automated analyses.

For this purpose, the first part of this thesis experimented with a computational learning task, *stance detection*. This is a commonly chosen NLP method in social media analysis for misinformation because it involves identifying the attitude or opinions expressed on social media texts towards a target topic. Such positions reveal how narratives are spread and how discussions become divided or amplified, which are all common patterns in misinformation. The key challenge identified early in the study is the limited availability of datasets that capture the surrounding discourse to model contextual knowledge. To address this, the study adopted data augmentation, a widely used technique for mitigating data limitations in producing more relevant training samples. The study concluded by developing augmentation methods using advances in prompt engineering with Large Language Model (LLM) to generate synthetic conversations. The results demonstrated improved performance over baseline methods. While exploratory, the findings established the technical feasibility of modelling context learning through social media discourses and revealed limitations where the design of interactions can adopt the complementary human judgment.

To bridge these computational insights with human expertise, the thesis introduces *DiscourseDash*, an interactive dashboard designed to support the exploration and interpretation of social media discourse. This design bridges the technical capabilities and human strengths as a combined effort to make more informed judgments by combining computational outputs with their interpretive skills. Guided by human-centred design principles, a functional prototype of *DiscourseDash* was developed to facilitate its evaluation through two phases of user studies. The first study explores how general users interact with *DiscourseDash* across a set of tasks designed to reflect real-world workflow scenarios. To test the hypothesis that *DiscourseDash*

offers a more effective way of understanding social media discourse compared to navigating unstructured platforms, we examined participants' performance and experiences across two conditions. A mixed-methods approach was employed, integrating quantitative and qualitative data to assess the overall effectiveness of *DiscourseDash* and the usefulness of its proposed analytical features, stance, sentiment, popularity, and discourse-level insights in facilitating decision-making processes. The findings demonstrate both strong user preference for *DiscourseDash* and the perceived value of these analysis measures. The study also identified design limitations that outline opportunities for future improvements.

To further assess its practical relevance, *DiscourseDash* was evaluated through a semi-structured interview study with industry experts. This aimed to explore how the tool aligns with established professional workflows and to capture perspectives more than those of general users. Experts recognised the tool's potential within its capacity to support discourse-level analysis and offer more qualitative insights. However, the study also revealed key misalignments between the design and expectations embedded in expert practices, which may affect adoption. Experts consistently emphasised the need for tools that move above surface-level statistical outputs to enhance support for the complexity and contextual sensitivity of social media discourses, strengthening the core motivation behind *DiscourseDash*'s design. These two studies provided complementary perspectives. The first captures how non-expert users engage with the tool in tasks that simulate real-world decision-making scenarios. The second study contributes expert applicability by examining how *DiscourseDash* aligns with established professional practices. Finally, we discuss the broader implications of these findings for designing tools that focus on providing discourse-level insights.

This thesis makes several key contributions by demonstrating how integrating discourse-level qualitative analysis with an interactive tool can enhance social media analysis for more informed decision-making. It introduces an interactive design approach through *DiscourseDash*,

a dashboard that combines the strengths of computational methods with human judgment. Grounded in both technical findings and human-centred design principles, *DiscourseDash* emphasises contextual understanding to support decision-making workflows. The contributions include: (1) methodological reflections on the practical trade-offs and limitations of NLP methods addressing data limitation through augmentation and context modelling in a computational learning task; (2) the design and implementation of a functional prototype, *DiscourseDash*, informed by these reflections on computational learning capabilities and limitations, and (3) evidence from two user studies, one with general users and one with industry experts, that demonstrate the tool's effectiveness by showcasing its applicability across different expertise levels, and identify considerations for real-world adoption. Overall, this research emphasises the importance of ongoing investigation in discourse-level insights into analytical tools for social media analysis by demonstrating their relevance and applicability in real-world contexts. *DiscourseDash* represents a foundational step towards developing advanced interactive tools for social media analysis, built upon an empirical understanding of the practical workflows in which these tools are embedded.

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis titled "A Human-Centred Approach to Social Media Analysis for Decision-Making" is my own work from the time of my PhD at the University of Nottingham under the supervision of Prof Joel E. Fischer, Dr Jeremie Clos and Dr Nazia Hameed, and that it has never been submitted for the award of any degree, diploma, or any similar type of recognition.

September 2025

Xin Yu Liew

CONTENTS

1 INTRODUCTION 1

1.1 Problem definition 4

1.2 Research questions 6

1.3 Research areas 7

1.4 Contributions 8

1.5 Publications 9

1.6 Structure of the thesis 11

I BACKGROUND AND APPROACH 13

2 LITERATURE REVIEW 15

2.1 Establishing the research contexts 16

2.1.1 Social media and discourses 16

2.1.2 Decision-making in practice 20

2.1.3 Section summary 22

2.2 Computational approaches for social media analysis 23

2.2.1 Supportive indicators for interpretation 23

2.2.2 Improved learning through modelling contexts from social media discourse 31

2.2.3 Section summary 33

2.3 Human-centred design for decision-making support 34

2.3.1 Conversational interfaces 34

2.3.2 Explainability interfaces 35

2.3.3 Simulation and nudging 36

2.3.4 Visual analytics tools 37

2.3.5 Section summary 38

2.4 Chapter summary 39

3 APPROACH 43

3.1 Methodological approach 44

3.2 Development of the design concept 47

3.2.1 Conducting computational learning 47

3.2.2 Connecting computational insights with design 58
3.2.3 Prototyping 58
3.3 Data collection for evaluation 60
3.3.1 Mixed-method user study 60
3.3.2 Consideration of expert practices 62
3.4 Data analysis 64
3.4.1 Statistical analysis 64
3.4.2 Thematic Analysis 65
3.5 Chapter summary 66

II EMPIRICAL WORK 69

4 STANCE DETECTION: AN EXPLORATORY STUDY 71
4.1 Introduction 72
4.2 Problem Definition 74
4.3 Methodology 75
4.3.1 Contextualise 76
4.3.2 Generation 78
4.3.3 Learning 81
4.4 Experimental Setup 83
4.4.1 Datasets 83
4.4.2 Baseline justification 84
4.4.3 Evaluation details 85
4.4.4 Implementation details 86
4.5 Findings 86
4.5.1 Macro-averaged F1 scores 86
4.5.2 Prompt-based augmentation 88
4.5.3 Impact based on the depth of discourses 89
4.5.4 Limitations and future work 91
4.5.5 Relating findings to the interactive design 92
4.6 Chapter summary 93

5 CREATING *discoursedash*: AN INTERACTIVE PROTOTYPE DESIGN 95
5.1 Introduction 96
5.2 Design rationale 97

5.2.1 Providing a visual overview 98
5.2.2 Further detail of the analysis 99
5.3 Interface design 100
5.3.1 *Overview* Page 100
5.3.2 *Detailed Analysis* Page 104
5.4 System architecture 107
5.4.1 Basic requirements 107
5.4.2 Database 108
5.4.3 Visualisation 109
5.4.4 Interaction 109
5.5 User flow diagram 110
5.6 System overview 110
5.6.1 Client 111
5.6.2 Server 112
5.7 Prototype implementation 112
5.8 Chapter Summary 113
6 A MIXED-METHOD STUDY WITH GENERAL USERS 115
6.1 Introduction 116
6.2 Study design 117
6.2.1 Baseline 118
6.2.2 Recruitment and participants 119
6.2.3 Procedure 119
6.2.4 Data collection 121
6.2.5 Analysis 123
6.3 Findings 124
6.3.1 Quantitative results 124
6.3.2 Qualitative results 126
6.3.3 Thematic analysis 142
6.4 Chapter summary 152
7 AN INTERVIEW STUDY WITH INDUSTRY EXPERTS 155
7.1 Introduction 155
7.2 Study design 157
7.2.1 Recruitment and participants 157
7.2.2 Procedure 158

7.2.3 Data collection and analysis 160
7.3 Findings 161
7.3.1 Understanding expert's existing workflows 161
7.3.2 Perceptions during expert's interactive experience with *DiscourseDash* 180
7.3.3 Expert's reflection regarding to *DiscourseDash*'s real-world applicability 190
7.4 Chapter summary 201

III SYNOPSIS 203

8 DISCUSSION 205
8.1 Designing for social media analysis 205
8.1.1 Grounding design concepts 206
8.1.2 The values within the design 208
8.1.3 Connecting the value to future directions 210
8.2 Understanding the real-world applicability 211
8.2.1 Shaping values with experts 212
8.2.2 The evolution of perception 213
8.2.3 Key takeaways from experts' perspectives 215
8.3 Bridging human judgment with analytical tools 216
8.4 Implications for design 217
8.4.1 Improving analysis outputs 218
8.4.2 Handling variability in data sources 219
8.4.3 Incorporating demographics 220
8.4.4 Balancing the needs of different users 221
8.4.5 Inclusivity and representation 223
8.4.6 Layering multi-levels of trust 224
8.5 Chapter Summary 225

9 CONCLUSION 227
9.1 Summary 227
9.2 Main contributions 229
9.2.1 Bridging NLP computational methods with HCI design principles 230
9.2.2 An interactive design prototype - *DiscourseDash* 231

9.2.3 Evidence-based validation through human experiences for general applicability 233
9.2.4 Real-world applicability implications for discourse-level analytical tools 235
9.3 Critical reflections 237
9.4 Future work 238

IV APPENDIX 241

A APPENDIX 243
A.1 Preliminary Data Augmentation Experiments 243
A.1.1 Motivation 243
A.1.2 Data Augmentation Methods 243
A.1.3 Datasets 244
A.1.4 Results 244
A.1.5 Conclusion 246
A.2 Questionnaires 246
A.3 Consent procedures 250
A.4 Recruitment procedures 255
A.4.1 General users 255
A.4.2 Industry experts 255
A.4.3 Conditions to filter experts 256
A.4.4 Recruitment materials 257
A.4.5 Interest forms 258
A.4.6 Scheduling and follow-up procedures 263

BIBLIOGRAPHY 267

LIST OF FIGURES

Figure 1 Overview of the thesis 3

Figure 2 An example of a (short) discourse with additional context from corresponding textual interactions (e. g., replies, comments). 5

Figure 3 An example of a 3-layer neural network. 53

Figure 4 The transformer architecture obtained from the original paper by Vaswani et al. (2023). 54

Figure 5 An example of k-fold cross-validation where $k = 5$. 55

Figure 6 Problem definition of stance detection. 75

Figure 7 Overview of the proposed framework for stance detection. 76

Figure 8 Violin plot of macro-averaged F1 scores for different approaches corresponding to the standard deviation across evaluation folds. 88

Figure 9 Boxplot illustrating the macro-averaged F1 scores across approaches for each discourse depth. The dashed line indicates the trend of the median. 91

Figure 10 *Overview* page showing (A) the *User Inputs* component located at the top of the page. 101

Figure 11 *Overview* page showing (B) the *Trending Topics* table directly below the *User Inputs* component. 102

Figure 12 *Overview* page showing (C) *Stance Distribution by Topic* charts directly below the *Trending Topics* table. 102

Figure 13 *Overview* page showing (D) *Sentiment Distribution by Topic* charts directly below the *Stance Distribution by Topic* chart. 103

Figure 14 *Overview* page showing (E) the *Popularity Over Time* graph directly at the bottom of the page as the final component. 104

Figure 15 *Detailed Analysis* Page showing (A) the *User Inputs* component located at the top of the page. 105

Figure 16 *Detailed Analysis* Page showing (B) the *Summary* combining *Top Keywords*, *Stance Distribution* and *Sentiment Distribution* for the selected discourse directly below the *User Inputs* component. 106

Figure 17 *Detailed Analysis* Page showing (C) the *Discourses* in collapsible tables to present each comment with stance and sentiment annotations and filter functionalities directly at the bottom of the page as the final component. 107

Figure 18 A user flow diagram for *DiscourseDash*. 110

Figure 19 A system overview of *DiscourseDash* presented based on a Model-View-Controller (MVC) architecture. 111

Figure 20 Overview of the study procedure involving general users. 121

Figure 21 A stacked bar chart shows the proportion of responses from participants evaluating each feature. 125

Figure 22 Overview of the study procedure with industry experts. The two key interactive pages are denoted as (i): when $i = 1$, this refers to the *Overview* Page; when $i = 2$, this refers to the *Detailed Analysis* Page. 160

Figure 23 Bar graph comparing the impact of augmentation methods when applied versus when none are applied. 245
Figure 24 Box plots comparing the performance improvements of different augmentation methods across training set sizes. 245
Figure 25 Questionnaire for *DiscourseDash* condition (page 1). 247
Figure 26 Questionnaire for *DiscourseDash* condition (page 2). 248
Figure 27 Questionnaire for *Reddit* condition (page 1). 249
Figure 28 Questionnaire for *Reddit* condition (page 2). 250
Figure 29 An example of the consent form (page 1). 252
Figure 30 An example of the consent form (page 2). 253
Figure 31 An example of the consent form (page 3). 254
Figure 32 An example of the consent form (page 4). 255
Figure 33 Recruitment post for general users shared via Microsoft Teams. 257
Figure 34 Recruitment post targeting experts shared on LinkedIn. 258
Figure 35 Prolific recruitment advertisement for experts. 258
Figure 36 Interest form for general user participants (page 1). 259
Figure 37 Interest form for general user participants (page 2). 260
Figure 38 Interest form for industry experts participants (page 1). 261
Figure 39 Interest form for industry experts participants (page 2). 262
Figure 40 Interest form for industry experts participants (page 3). 262
Figure 41 Interest form for industry experts participants (page 4). 263

Figure 42 Interest form for industry experts participants (page 5). 263
Figure 43 Example of email invitation sent to qualified LinkedIn experts. 264
Figure 44 An invitation survey posted via Prolific to selected experts (part 1). 264
Figure 45 An invitation survey posted via Prolific to selected experts (part 2). 265

LIST OF TABLES

Table 1 Overview of research approaches. 44
Table 2 Example confusion matrix for a three-class classification task with classes A, B, and C. The arrows → indicate the direction of misclassification. 56
Table 3 Examples of stance labels towards a target from a text. 73
Table 4 Summary of macro-averaged F1 scores (mean ± standard deviation) across varying approaches evaluated on the dataset. Best performance is in bold and marked with †. 88
Table 5 Depth-specific macro-averaged F1 scores (mean ± standard deviation) for each approach. Best performance in each column is in bold and marked with †. 90
Table 6 Translation of key findings from Chapter 4 into specific design implications for *DiscourseDash*. 97
Table 7 Comparison of Task 1 summarised findings across sub-questions and conditions. 128

Table 8 Comparison of Task 2 summarised findings across sub-questions and conditions. 133
Table 9 Comparison of Task 2 summarised results across sub-questions and conditions. 137

ACRONYMS

BERT Bidirectional Encoder Representations from Transformers

CNN Convolutional Neural Network

EDA Easy Data Augmentation

XAI eXplainable AI

HCD Human-Centred Design

HCI Human-Computer Interaction

LDA Latent Dirichlet Allocation

LLM Large Language Model

LSTM Long Short-Term Memory

ML Machine Learning

NLP Natural Language Processing

RAG Retrieval-Augmented Generation

RtD Research-through-Design

SUS System Usability Scale

UK United Kingdom

GPT Generative Pretrained Transformer

1

INTRODUCTION

The Global Digital Report 2024 reported that 82.8% of the total population in the United Kingdom (UK) are social media users (Kemp, 2024). This statistic reflects the extensive reach and influence of these platforms in the modern digitisation (Kizgin et al., 2020; Lupinacci, 2021; Subrahmanyam et al., 2008). This influence can be illustrated through the COVID-19 pandemic, during which individuals relied on social media platforms to express opinions, stay informed, and form connections (Appel et al., 2020; Bouvier and Machin, 2020; Saud, Mashud, and Ida, 2020). As a result of these engagements on social media, organisations have had to adapt their strategies and decision-making processes to these dynamic environments (Persily, Tucker, and Tucker, 2020). As noted by Jim Coleman, CEO of *We Are Social*, a global socially led creative agency, industries must remain responsive to evolving online behaviours by staying "on top of these continuous shifts in the way that people are spending time online and adapt their approach" (Underwood, 2024).

The influence of social media extends across commercial enterprises, government bodies, and public health sectors, shaping how decisions are made, communication strategies are developed, and societal challenges are addressed. While social media offers advantages for these stakeholders in designing effective strategies and public engagement, it also presents a new set of challenges. These include the rapid spread of misinformation during crises such as the pandemic ("infodemic") (Cinelli et al., 2020), reputational harm (Jahng, 2021), and the manipulation of public narratives and ideologies (Reisach, 2021). These challenges are consequential because misinformation and harmful narratives can directly impact vulnerable individuals and communities (Drouin et al., 2020). Not only that, they complicate the work

of policymakers, fact-checkers, and government agencies tasked with navigating and responding to the fast-evolving online narratives (Azzaakiyyah, 2023; Mahoney and Tang, 2024).

This evolving complexity of social media environments presents significant opportunities to develop analytical approaches that more effectively support stakeholders' decision-making processes (Moe and Schweidel, 2017; Rathore, Kar, and Ilavarasan, 2017). Many existing social media analytical tools have focused predominantly on quantitative measures, frequently overlooking the qualitative aspects that are essential for capturing context and meaning within a social media discourse (Chen, Lin, and Yuan, 2017). While these existing approaches remain effective at highlighting large-scale trends, they are insufficient to capture the complexity of social media discourses or interpret the meaning behind quantitative measures for effective decision-making that are dependent on context. In response to these influences of social media, the AI4TRUST project provides a relevant example that aims to promote collaboration between AI systems and human fact-checkers through a responsible platform for monitoring and analysing social media disinformation (Leal et al., 2025).

Despite these existing efforts, there is an underexplored opportunity to design analytical tools that provide discourse-level insights to help interpret qualitative patterns and better understand complex phenomena such as misinformation, polarisation, and controversy (Vasist, Chatterjee, and Krishnan, 2024). This thesis explores this further in Chapter 2. Without such insights, analyses risk oversimplifying the dynamics of social media discourses, limiting their value in countering misinformation or targeted propaganda. An analytical tool is needed to support navigation through these narrative shifts by helping stakeholders identify underlying causes or issues and enabling more informed decisions. At the same time, its design must recognise that users interpret analytical outputs through their prior expectations, experiences, and social context (Neff et al., 2017).

Addressing these challenges requires designing an analytical tool grounded in empirical insights and evaluated for practical relevance

through targeted studies with end-users. This thesis, therefore, investigates how integrating computational methods with human expertise in an interactive design can address their respective limitations and improve decision-making outcomes. This investigation starts with an examination of how such a tool can hypothetically support professional workflows (see Section 2.1.2). With these in mind, this thesis unfolds the extent to which an interactive tool can support the interpretation of social media discourse for informed decision-making. To achieve this, the research is guided by two key objectives: (1) to design an interactive tool based on examining existing approaches in interpreting social media discourse, and (2) to explore how such interactive design can be considered and enhanced to support professional decision-making contexts.

The rest of this chapter presents the problems that motivate this thesis, the research objectives designed to address them, and the research questions that frame the investigation. Finally, the chapter positions this work within the broader discipline of HCI and defines the scope of the research domain explored. An overview of the structure and focus of this thesis is provided in Figure 1.

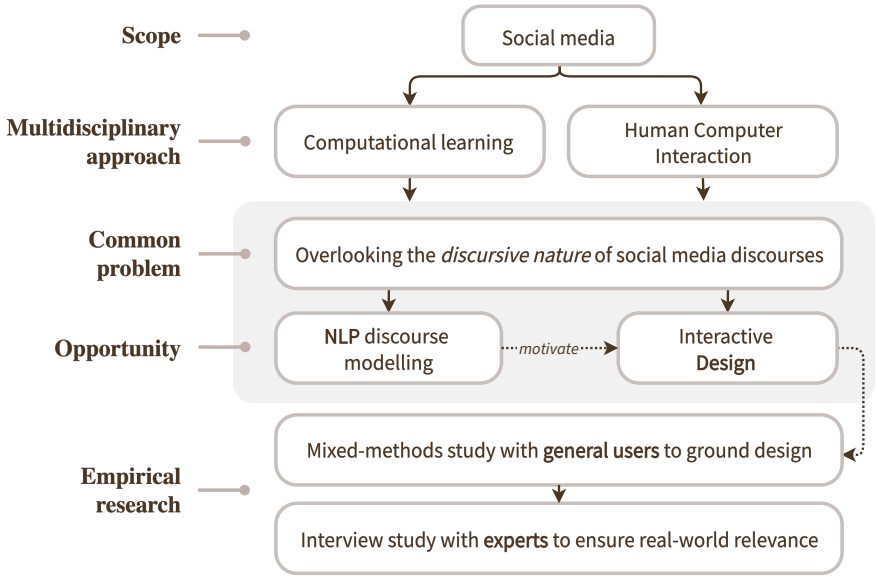


Figure 1: Overview of the thesis

1.1 PROBLEM DEFINITION

Despite the growing reliance on social media as a key indicator of public opinion and behavioural trends, analysing these conversations at scale remains a complex and unresolved challenge. Misinformation is a well-explored topic in social media analysis, providing a foundational example for this research to understand how various analytical approaches have been applied in that context (Aïmeur, Amri, and Brassard, 2023; Zhao et al., 2023). Early interventions have explored strategies such as content labelling (Spradling, Straub, and Strong, 2021), removing misleading narratives (Sharma et al., 2019), and implementing suggestive measures (Vicario et al., 2019). With advances in computational approaches, i. e., machine learning, research efforts have focused widely on analysing stance (Hardalov et al., 2022) and sentiment (Alonso et al., 2021; Bhutani et al., 2019) to capture the scale and spread of narratives. However, these methods predominantly focus on root-level content, such as initial posts or primary threads on social media platforms (Egelhofer and Lecheler, 2019). They frequently overlook the discursive nature of social media, which comprises the root-level content and any subsequent replies and comments.

A recent study highlighted that root-level comments frequently function as 'clickbait' for drawing attention rather than providing meaningful context (Bu et al., 2023; Lama et al., 2020). This can lead to ambiguity as machine learning models struggle to interpret accurate references without a full discourse context. For instance, in Figure 2, where machine learning models struggle to resolve what "this" refers to without additional context, resulting in incomplete or inaccurate insights regarding what the *@rootuser* is "against". This thesis incorporates the full discourse to achieve a more accurate interpretation by broadening the analysis beyond the root-level comment. In the same example, analysing the full discourse reveals that "this" refers to "vaccines", as supported by follow-up textual interactions where *@replyuser1* argues "for" with reasons supporting vaccines and *@re-*

plyuser2 argues "against", illustrating how *@rootuser* potentially influenced the spread.



Figure 2: An example of a (short) discourse with additional context from corresponding textual interactions (e. g., replies, comments).

Addressing this gap requires shifting analytical focus to connecting contextual understanding of discourse-level interactions, rather than treating these connected texts individually. This perspective reflects real-world practices, where analysts must examine not only what is being said but also why it is being said (Hamm, 2024). Their analysis not only observes surface statistical measures but also investigates underlying causes and broader conversational dynamics to support more informed decision-making.

Even so, generating accurate discourse-level analyses poses challenges, e. g., dynamics of linguistic patterns on social media. As a result, human expertise is an integral part of contextualising and validating machine-generated insights for making the analytical process collaborative. The core problem, then, lies in how to advance existing approaches in ways that address the limitations of both computational methods and human judgment, while also providing effective support for interpreting the meaning behind qualitative measures. A reflection on the broader concern arises as many existing tools are designed around technical affordances rather than the practical needs

(Neff et al., 2017). This essentially leads to potential gaps regarding relevance within established real-world workflows. It is therefore not only important to ground the design of an interactive tool in empirical insights but also to involve users throughout the design process to create solutions that accurately address practical needs.

1.2 RESEARCH QUESTIONS

So far, research into social media analysis has primarily focused on delivering quantifiable metrics to support decision-making, with less emphasis on qualitative designs and an understanding of the practical needs within this contextual perspective. To explore the potential of integrating computational learning with an interactive design to support users in better understanding and interpreting social media discourses, this thesis will first address the following research question:

RQ1 How can interactive tools be designed to enhance social media analysis to support more effective decision-making workflows?

An effective design of a solution requires understanding how humans and machines can address each other's limitations in the context of analysing social media discourses. To achieve this, it is first necessary to establish the capabilities and limitations of computational learning methods in handling such complex and unstructured data. A computational exploration study serves two primary purposes. Firstly, implementing a computational learning analysis task can inform and inspire the overall design direction. Secondly, it identifies aspects of analysis where human judgment remains essential. The findings from a computational learning study will provide a foundation for the broader design process presented in this thesis and support the rationale for an interactive approach to social media analysis. In doing so, the study not only guides design decisions but also demonstrates the complementary strengths of human and computational methods.

Guided by an HCI approach, a mixed-methods evaluation will then be conducted following the establishment of the design prototype to ground the design rationale in empirical evidence. This study aims to validate the design concept by assessing its effectiveness with manual navigation of unstructured social media discourses in decision-making scenarios. This evaluation captures general users' interactions across two experimental conditions and generates complementary insights by combining statistical analysis with qualitative feedback. As a fundamental part of the design process, this evaluation validates the design concept and establishes the foundation for the subsequent research question:

RQ2 How do industry experts' workflows and experiences in analysing social media discourse shape the real-world applicability of *DiscourseDash*?

This research question focuses on aligning the tool by reflecting the needs and practices of industry experts. While the general user study provided insights into potential benefits for non-expert users, expert perspectives are important to ensure the tool's practical relevance and appropriateness in professional workflows. Plus, there remains a limited understanding of how experts' knowledge and practices can be effectively integrated into the design of analytical tools for highlighting social media discourse. This study addresses this gap by engaging with industry experts to investigate the workflows, strategies, and challenges of social media analysis within the context of their decision-making. These expert insights will situate our design within the practical industry practices by outlining how it complements existing professional processes and implications for future developments.

1.3 RESEARCH AREAS

This thesis connects several intersecting research areas, with a particular focus on advancing the understanding and support of discourse-

level social media analysis through both computational methods and human-centred approaches.

- **HCI** - This research contributes to the growing body of HCI work focused on designing interactive systems that support social media analysis. This thesis addresses an emerging area of interest that bridges interaction design and computational social media research by gathering insights to ground the design of a discourse-level analytical tool. The thesis contributes to the understanding of how users interact with complex data and how design can support interpretive and decision-making processes.
- **Mixed-methods research** - A mixed-methods approach was adopted to evaluate the interactive tool, integrating both quantitative and qualitative methods. This methodological design enabled the examination of general user engagement, perception, and task performance while interacting with the design concept, reflecting the practicality of the design.
- **Computational methods** - This research focuses on leveraging existing NLP techniques, including data augmentation and the use of LLM to improve the automated interpretation of discourse-level context in social media analysis. The research explores how current computational methods can be integrated into interactive systems to support analysis tasks in discourse environments.

1.4 CONTRIBUTIONS

The main contributions of this thesis are as follows:

1. NLP methods for stance detection: This contribution lies in the methodological framework that informs the interactive design concept for social media analysis of the thesis. It involves implementing data augmentation through LLM prompt-generation

to improve discourse-level context learning. While the augmentation techniques themselves are established, this thesis demonstrates how the limitations observed in computational experiments (Chapter 4) guided the design of a user-centred interactive tool to bridge technical insights with HCI principles.

2. Conceptual design towards actionable impacts: The novelty of *DiscourseDash* lies in its integration of discourse-level analysis as context for high-level visual summaries. This establishes a design that shifts from static data presentation toward a "data for action" framework for supporting industry experts in interpreting social media discourses for decision-making (Chapter 5).

3. Mixed-method evaluation with general users: This study establishes a baseline for how non-experts interact with complex discourse-level features. Evaluating general user interactions provides empirical evidence that connecting contexts from discourse-level analysis with visual summaries facilitates decision-making. To this end, it allows the research to validate the design's conceptual reasoning (Chapter 6) before engaging expert users.

4. Interview study with industry expert to align *DiscourseDash* with real practices: This contribution identifies key insights between the design of an automated tool in supporting ongoing workflows of industry experts. These findings provide a theoretical and practical roadmap for bridging the gap between computational development and professional application (Chapter 7).

1.5 PUBLICATIONS

Part of this thesis has been presented in peer-reviewed venues. Some publications contributed to the refinement of this thesis, while others

correspond to the empirical chapters. The following list highlights these publications:

- The formulated research questions in this chapter were presented during the early stages and refined based on advice from a doctoral consortium:

  Liew, Xin Yu (2023). "Monitoring Online Discussions and Responses to Support the Identification of Misinformation." In: *European Conference on Information Retrieval*. Springer, pp. 450–455.

- The preliminary experimentation plan for the empirical study in Chapter 4 was presented in a poster paper to gather feedback and refine the research directions:

  Liew, Xin Yu, Nazia Hameed, Jeremie Clos, and Joel E Fischer (2023). "Predicting stance to detect misinformation in few-shot learning." In: *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, pp. 1–5.

- The earlier version of the empirical design in Chapter 5 and the planned evaluations were presented in a poster paper:

  Liew, Xin Yu, Nazia Hameed, Jeremie Clos, and Joel E Fischer (2023). "Predicting stance to detect misinformation in few-shot learning." In: *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, pp. 1–5.

- Parts of the empirical work from 5 & 6 & 7 were published in a journal:

  Liew, Xin Yu, Jeremie Clos, Nazia Hameed, and Joel Fischer (2026). "DiscourseDash: designing and evaluating a tool for analysing social media discourses to support decision-making." In: *PeerJ Computer Science* 12, e3632.

1.6 STRUCTURE OF THE THESIS

This thesis is organised into three parts across nine chapters. Below is an overview of the structure and the contributions of each part and chapter.

Part I lays the foundation for the thesis through two chapters. Chapter 2 presents a literature review that explores the relationship between social media, discourses, and decision-making in practice. The chapter then examines existing approaches to social media analysis by reviewing HCI and NLP approaches for a multidisciplinary perspective. This informs and motivates the design choices and methodological approaches used in the empirical studies. Chapter 3 outlines the methodological framework drawn from this literature to support the empirical investigations. The chapter outlines the NLP approach to computational learning for textual analysis and HCI approaches used to inform and facilitate design concepts that address human-centred needs.

Part II organises the empirical work of this thesis across four chapters. Chapter 4 investigates the role of NLP in data augmentation for stance detection. This chapter examines how synthetic data generated by LLMs can be used to improve stance detection. To translate these computational insights into a design, Chapter 5 details the development of the prototype design, grounded in those computational explorations, which will then be evaluated in the following user studies. Chapter 6 details the first evaluation of the prototype through a mixed-method study involving general users without prior analytical experience. This study compares the effectiveness of decision-making when using the *DiscourseDash* versus manually navigating unstructured social media data, helping to assess the relevance and usability of the design concept. Recognising the importance of real-world applicability, Chapter 7 explores how the *DiscourseDash* aligns with the workflows of industry experts. Through semi-structured interviews, this research investigates the decision-making responsibilities of professionals who work with social media. I. e., how they manage the

volume and complexity of discourse, the challenges they encounter, and their reflections on *DiscourseDash*'s design. These insights help position the concept within real-world contexts for bridging the gap between computational methods and practical decision-making in social media analysis.

Part III concludes the thesis with a reflective discussion and final remarks. Chapter 8 provides a synthesis of findings by critically reflecting on the broader implications of combining computational learning techniques to motivate human-centred design for social media analysis. The chapter also outlines the contributions of this work to the research community. The chapter then included a discussion on its limitations and areas for future research. Finally, Chapter 9 summarises the key contributions of the thesis, revisits the research questions, and outlines the significance of the work in supporting decision-making through analysis of social media discourses.

Part I

BACKGROUND AND APPROACH

2

LITERATURE REVIEW

This chapter introduces the basis of the research that informed the empirical work of this thesis and how it connects with the formulation of our research questions outlined in Chapter 1.

To set the research context, section 2.1 first understands the nature of social media. The section defines what this thesis means by discourses by emphasising why they are the fundamental drivers of impact on social media. Then, the chapter outlines the decision-making roles that rely on analysing these discourses for decision-making outcomes. The purpose of this section is to formulate the motivation within the scope of this thesis. The section achieves this by highlighting the impact of social media and how addressing it within the wider field of HCI can contribute to potential impact in professional contexts.

Following this, section 2.2 provides an overview of computational approaches discussed in the literature. The section summarises existing methods designed to support the analysis of social media challenges. This includes reviewing approaches from stance detection, sentiment analysis, harmfulness detection, to popularity analysis, all of which are valued in the literature for understanding social media dynamics. This establishes the computational learning background and serves to inspire reflections on how these analyses can be integrated when designing a tool. From here, the section navigated further by examining studies that structured discourse-level as context to improve model learning.

Next, section 2.3 reviews the current efforts in the literature that have designed human-centred approaches for social media analysis. The section reviews HCI approaches by categorising them into conversational interfaces, explainability interfaces, simulation and nudg-

ing, and visual analytics tools. The purpose of this section is to motivate the methodological approach of this thesis for identifying opportunities that we can draw on to improve.

This literature review situates the contribution of this thesis within its interdisciplinary grounding. Each of the sectioned reviews in this chapter positions the prior works to motivate the considerations that shape the scope of the research. Prior literature outlines (1) how social media analysis plays an integral role in society and (2) how an understanding of both computational approaches and human-centred perspectives can inform the design of an interactive tool. All in all, the purpose is to demonstrate how this thesis approaches the problem by drawing on insights and reviewing prior studies.

2.1 ESTABLISHING THE RESEARCH CONTEXTS

2.1.1 *Social media and discourses*

Social media refers to digital platforms, such as websites and applications that allow individuals and groups within society to interact virtually through activities like commenting, sharing, posting, and messaging (Meikle, 2016). This form of digital communication supports a wide range of social engagements amongst individuals, communities, and organisations (Kaplan and Haenlein, 2010; Lovejoy and Saxton, 2012). One of the fundamental capabilities of these platforms is their ability to facilitate the formulation of public discourses. This means that these platforms enable continuous and active conversations where anyone can exchange ideas, opinions, and arguments (Majchrzak et al., 2013).

With individuals globally engaging with these platforms daily, social media is rooted in the routines of everyday life. On average, individuals dedicate a significant portion of their daily time to using social media. According to recent statistics, the typical internet user spends approximately 2.5 hours per day on these platforms (Kemp,

2025). In the UK, approximately 77% of adults reported using social media regularly. Usage remains high across all age groups and is evidently more common amongst younger people (Ofcom, 2024). This scale and influences are reflected not only in user numbers, but also in their economic and societal impact. The global social media advertising market alone was valued at over $276.72 billion in 2025 and is projected to exceed $406.45 billion by 2029 (Statista, 2025).

More than their economic footprint, social media are instrumental in shaping public discourse, cultural trends, and even political outcomes (Ausat, 2023). In a day, a typical social media user may encounter hundreds of posts, interact with multiple communities, and be exposed to a wide range of discourses from personal, professional, or societal domains. Platforms such as *Facebook*, *X* (formerly *Twitter*), and *Reddit* allow users freedom to share opinions, build communities, or even interact directly with public institutions. These capabilities have transformed how information on social media is disseminated (the way information spreads), consumed (the way people interpret), and debated (the way people discuss). These dynamics overlap and create a complex environment of interaction (how information is understood and acted upon) between experts, public sentiment, and misinformation (Johnson et al., 2020).

As individuals become dependent on social media for information consumption, the content they encounter influences the following actions: how they perceive social issues, form personal opinions, and engage with societal concerns. Public perceptions are frequently shaped by the perceived 'authenticity' of social media discourse compared to traditional news media. This eventually contributes to a so-called 'echo chamber' phenomenon and the reinforcement of existing biases (Flintham et al., 2018; Han et al., 2021). With that said, the dynamics underlying social media discourses illustrate how sociotechnical systems structure and influence the public's collective understanding of a matter.

Though social media contributes beneficial means within a society, these influential capabilities are accompanied by challenges. A promi-

nent example is the widespread circulation of health-related misinformation during the COVID-19 pandemic. Research has shown how malicious actors use social media to amplify misleading health claims that contributed to public confusion and undermined trust in scientific guidance (Muric, Wu, and Ferrara, 2021). Reports estimated that nearly half the UK population was exposed to COVID-19 misinformation (Ofcom, 2020). This exposure has been associated with reduced compliance with public health measures, such as vaccine uptake (K et al., 2021).

These are also evident in political and commercial contexts. For instance, in the UK, online discourse around Brexit and immigration has intensified ideological polarisation and shaped public sentiment through emotionally charged narratives (Moss, Robinson, and Watts, 2020). There is increasing concern that these political discourses fuelled by far-right groups pose a threat to the public's safety as they lead to offline violence (Wahlström and Törnberg, 2021). On the other hand, brand campaigns by Nike and Gillette have provoked both support and backlash (Mirzaei, Wilkie, and Siuki, 2022). Gillette's *The Best Men Can Be* challenged toxic masculinity (Hsu, 2019), while *Nike's For Once, Don't Do It* was interpreted by some as racially insensitive (Cohen, 2020). These campaigns provoked widespread debate on social media as users who engaged with them amplified and reinterpreted the information. These examples showed how corporate communication can rapidly become part of broader cultural debates and disputed discourses on social media. These interactions that fuelled the impact of a brand's reputation suggested that analysing them can provide support to engage better with broader sociocultural narratives.

Taking everything into consideration, these examples highlight the complex and evolving nature of discourse on social media platforms. Misinformation, polarisation, and discourse manipulation are not isolated problems. This is because they reflect deeper structural tensions within digital communication environments that are interconnected (Caled and Silva, 2022). Addressing these complexities requires effective monitoring of discourse dynamics where public perception and

decision-making processes are directly related and impacted. This thesis, therefore, adopts this perspective and defines social media discourse as sequences of textual interactions such as discussions, threads, and conversations.

2.1.1.1 *Defining discourse*

For purpose of this research, "discourse" is technically defined as consisting of two main components: **root-level comments** (C_0), such as those initiated as a post or thread on social media, and **child-level comments** (C_i, where $i \geqslant 1$), which are hierarchically structured replies to the root-level comments, also referred to as "replies." A complete discourse (D), can be represented as: $D = C_0, C_1, C_2, ..., C_n$, where $n \geqslant 1$. This implies that a complete discourse must include both root-level comments (C_0) and one or more child-level comments (C_i).

If a root-level comment (C_0) does not have corresponding child-level comments ($n = 0$), this research does not consider it as a discourse. These child-level comments (C_i) represent a form of engagement on social media, contributing to the evolution of narratives and influence (Epstein et al., 2021; Neubaum and Krämer, 2017). A root-level comment (C_0) that does not generate engagement and therefore lacks a discourse ($D = C_0$ where $n = 0$), reflects a low impact on public narratives. For example, a Reddit post without replies generally has limited influence, whereas a post with multiple layers of interaction indicates a more impactful presence (Choi et al., 2015).

This allows the research to emphasise discourses as emergent and interactive processes instead of isolated textual artefacts. All things considered, prior studies mentioned that these replies reflect the dynamics of narrative from opinion shifts, argumentation, and other factors that influence our understanding of them (Benamara, Inkpen, and Taboada, 2018).

2.1.2 *Decision-making in practice*

Understanding the dynamics of social media discourse has become increasingly important for informing decision-making processes due to the affordances of social media that shape how information is created, shared, and interpreted in organisational contexts (Treem and Leonardi, 2013). While misinformation is a prominent evidence of the risks in social media, the broader challenge concerns how professionals make sense of these constant changes in public conversations to inform timely and appropriate action. This section examines how decision-making across domains is influenced by the challenge of interpreting fast-changing social media discourses.

As mentioned briefly previously about public health, analysing social media discourse has shown its importance during critical moments such as the COVID-19 pandemic (Nihlén Fahlquist, 2021). Health authorities have relied on such analyses to monitor the shifts in public sentiment, misinformation, and emerging health concerns (Banerjee and Rao, 2020; Puri et al., 2020). Observing how these insights are put into use by authorities demonstrates that monitoring such interactions can directly inform communication strategies. These strategies aimed to increase public trust, address vaccine hesitancy, and combat the spread of deceptive treatments (Kennedy, Rorie, and Benson, 2021). It can therefore be inferred that health organisations depend on these discourses to adjust interventions in countering harmful behaviours by monitoring how narratives unfold and identifying influential voices within them (Litman et al., 2020).

On the other hand, policy-making and regulatory organisations engage with social media discourse to monitor risks associated with harms such as hate speech, extremism, and the targeting of vulnerable communities (Dekavalla, 2022). While public health practitioners focus on response strategies for mitigating the impacts of emerging harms (Madhav et al., 2017), policy actors prioritise preventative approaches (Mickoleit, 2014; Yeung, 2018). From this, it can be assumed that societal authorities need to analyse social media discourses to

understand how they may intensify inequalities, drive polarisation, or contribute to harm against marginalised groups. Insights from social media discourse could therefore benefit governments when it comes to developing strategies for harm reduction, e. g., reshaping regulatory frameworks to protect individuals at heightened risk of online abuse or targeted misinformation campaigns (Digital and Office, 2020).

Similarly, the workflow of journalists and fact-checkers relies on understanding the flow of information and the narratives gaining traction within public discourse (Cavaliere, 2020). These experts face increasing pressure during timely events, e. g., to identify which claims gain exponential traction, to verify impactful information, and to understand how narratives are shaped within these evolving social environments (Sehat et al., 2024). Even within the comparatively lower-impact environments of modern journalism, such as podcasts, journalists frequently extract opinions and sentiments from social media discourse to inform reporting and shape narratives (Stephanie Soo, n.d.). These can be perceived as a way to decide what to share within their platform that contributes towards building awareness or educating the public about a matter. From these experts' practices, the formulated understanding is that social media discourse influences decisions about what to report or even how to intervene within them through counter-narratives or verification efforts.

In commercial or brand marketing, social media discourses can have been used to shape organisational reputation and inform strategic decision-making (Di Domenico et al., 2021; Mills and Robson, 2020; Visentin, Pizzi, and Pichierri, 2019). On one end of profitable purposes, these organisations monitor them to identify opportunities for strategic positioning, brand engagement, and market advantage. On the other end of reputational purposes, they are tasked to manage the risks associated with the unpredictable spread of misleading narratives that undermine consumer trust and damage brand reputation (Bakir and McStay, 2018; Cock Buning, 2018; Weiss, 2021). With this in mind, it is evident that analysing social media discourses becomes

central to both proactive strategies for business growth and reactive efforts to safeguard organisational integrity.

Across these domains, a shared concern centres around the ability to make sense of and intervene in the evolving dynamics of social media discourse. These examples from different organisations illustrate that public discourse is a fast-paced and ongoing process. Their impacts can be seen from how they actively shape and are shaped by shifting perceptions and behaviours. Hence, this research refers to decision-making as interpretive and strategic judgements made by professional communicators through social media discourse. E. g., these decisions could include determining whether and how to respond to emerging narratives or misinformation, adapting messaging strategies based on behaviours, prioritising issues that require immediate intervention, or evaluating the impact of communicative actions.

2.1.3 *Section summary*

This section has established the relevance of social media as a site where discourse unfolds and shapes societal responses to societal matters. Through examples drawn from several domains, the section within this chapter so far has articulated how decision-making practices rely on interpreting these constantly evolving social media discourses. This includes how analysing social media discourses can support organisations in fulfilling their responsibilities and achieving their intended impact. The issues discussed here demonstrate a clear need for tools or approaches that provide the ability to support those responsible for analysing and responding to such practical societal concerns. This situates the problem that the thesis is addressing within decision-making by emphasising the importance of designing solutions that align with the real-world relevance and responsibilities of those managing these environments. The next section reviews the

multidisciplinary perspective that contributes to the empirical work of this thesis.

2.2 COMPUTATIONAL APPROACHES FOR SOCIAL MEDIA ANALYSIS

Computational approaches, such as NLP methods, have been widely used in the field of Machine Learning (ML) to extract meaning from textual data (Hirschberg and Manning, 2015). This section first reviews computational approaches that have been proposed as supportive indicators. I. e., stance detection, sentiment analysis, harmfulness detection, and popularity analysis. These approaches are categorised in this section based on their wide adoption in the literature for supporting the interpretation of complex social media data.

The section then proceeds to review approaches that model discourse structures by incorporating them as contextual information during computational learning. This focuses on methods within the problem scope defined in Figure 1 of those that integrate discourse structures into the analysis. The purpose of this section provide an understanding of how discourse as contextual information has been applied in the literature and how it benefits the learning and interpretation. The reviewed computational approaches for social media analysis will inform the computational grounding of this thesis. Nonetheless, these methods should be viewed not as complete automated solutions, but guidance towards the development of an interactive design concept.

2.2.1 *Supportive indicators for interpretation*

2.2.1.1 *Stance detection*

Stance detection refers to the computational task of identifying the position expressed in a text towards a specific target, commonly categorised as *in-favour*, *against*, or *neutral* (Hardalov et al., 2022; Moham-

mad et al., 2016). This computational approach is a target-specific task that reflects how a social media user positions themselves with respect to an entity, claim, or issue (Lillie and Middelboe, 2019). With that said, these positions within social media are what make stance detection valuable for analysing social media discourses that convey a clear position on polarised issues.

Early work in stance detection was influenced by shared tasks such as SemEval 2016 that provided benchmark datasets and evaluation standards in political and health-related contexts (Mohammad et al., 2016). Many of the approaches relied on manually crafted lexicon-based features, sentiment features, and machine learning classifiers (Küçük and Can, 2021). However, these approaches struggled with the complexities of social media data due to the limitation in extracting semantic meaning between words and phrases and did not take into account the interpretation of meaning within broader discourses. When transformer-based models were introduced (Vaswani et al., 2023), their architecture inspired pre-trained encoder models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) for many classification tasks in NLP. Not only does it improve in providing embeddings that consider these semantic meanings for context, but it also significantly improves classification tasks across diverse topics or domain-specific tasks (ALDayel and Magdy, 2021; Hardalov et al., 2022).

Recent advances in LLMs have shifted this methodological approach of stance detection (Zhang et al., 2024). For instance, the Chain-of-Stance framework was proposed to use multi-step reasoning to produce more interpretable stance predictions instead of the direct route of classification (Ma et al., 2024). Approaches within LLM represent an improvement when compared to earlier transformer-based methods that relied heavily on task-specific parameter optimisation. Multi-agent LLM systems have also been proposed to address multi-aspect stance reasoning targeting scalability challenges in traditional approaches (Lan et al., 2024; Yan, Joey, and Ivor, 2024). Their work

demonstrated the efficacy of using a multi-agent architecture for the stance detection task.

Another important trend is the use of synthetic data generated by LLMs to overcome data scarcity and imbalance. Studies such as Wagner et al. (2024, 2025) proposed that augmenting stance detection datasets with LLM-generated examples improves robustness and reduces bias in domains like political discourse. This is because domain-specific annotated data is limited, and traditional supervised machine learning tasks rely on it for accurate prediction. However, the evolution of stance detection approaches has shifted from static and resource-intensive pipelines towards a more adaptive approach in handling data from social media environments. This is because fine-tuning LLM approaches has shown significant performance in dealing with social media's complex linguistic patterns and contexts.

Applications of stance detection span misinformation detection (De Magistris et al., 2022), rumour verification (Alsaif and Aldossari, 2023; Pamungkas, Basile, and Patti, 2019), credibility analysis (Karande et al., 2021), and uptake of misleading narratives (Weinzierl, Hopfer, and Harabagiu, 2021). These applications highlight the broader significance of stance detection as an indicator for understanding public perceptions for informing decision-making around societal issues. Within this thesis, stance detection is therefore conceptualised as a supportive indicator for designing interactive systems that support stakeholders working within those social media analysis environments to interpret discourses.

2.2.1.2 *Sentiment analysis*

Sentiment analysis refers to the computational study of affective states expressed in text by categorising emotional tone into positive, negative, or neutral classes (Liu, 2020). This differs from stance detection, which captures an author's position towards a target; sentiment analysis reflects the emotional valence of the text. This distinction is important in analysing social media discourses because comments

may convey strong emotions without explicitly revealing a position on a given issue.

Early approaches to sentiment analysis relied on lexicon-based methods and supervised classifiers trained on annotated corpora (Bonta, Kumaresh, and Janardhan, 2019). These approaches achieved significant sentiment analysis results in domains such as product reviews and political opinion mining (Medhat, Hassan, and Korashy, 2014). However, the complexity of linguistic patterns like informal languages, multilingual, and multimodal introduced challenges for these computational methods. E. g., these approaches do not perform well in recognising languages commonly used on social media comments that carry sarcasm or ironic meanings. These presented challenges of those traditional methods in analysing sentiments. The advances in deep learning, such as recurrent and convolutional neural networks, were proposed to address challenges of those traditional methods for learning distributed text representations (Zhang, Wang, and Liu, 2018).

There is an observed shift to sentiment analysis, similar to stance detection, where approaches were proposed using transformer-based architectures and LLMs (Krugmann and Hartmann, 2024). The Generative Pretrained Transformer (GPT) model was adopted in sentiment classification to tackle problems like understanding context and detecting sarcasm (Kheiri and Karimi, 2023). Their work reported improvements over baseline methods and a reduction in reliance on large labelled datasets. Studies have also highlighted the potential of LLMs in cross-lingual by presenting the significant results in sentiment analysis across different online communities in low-resource settings, e. g., minority languages (Chen, Shang, and Wang, 2025; Miah et al., 2024).

Sentiment analysis has been widely applied across a range of domains that are relevant to the focus of this thesis. For instance, this perspective is positioned to be valuable for detecting social media comments designed to provoke strong emotional reactions (Liu, 2020). Prior studies have shown that sentiment analysis can be used to

track social media phenomena such as the spread of misinformation, which was shown to be correlated to emotionally charged information (Alonso et al., 2021; Preston et al., 2021). Negative emotions such as fear and anger have also been shown to increase the circulation of social media information regardless of its factual accuracy (Wang et al., 2019; Weeks, 2015).

Sentiment analysis has also been applied to track shifts in public mood during elections, policy debates, or crises in political communication (Ceron et al., 2014). Besides, sentiment can serve as a proxy for reputation management for organisations to detect negative public reactions and implement corrective strategies in brand communication (Guo, Fan, and Zhang, 2020). Meanwhile, it was highlighted that monitoring sentiment in real time can support timely interventions in health, disaster, and humanitarian contexts within crisis informatics research (Reuter, Hughes, and Kaufhold, 2018).

This thesis hence positions sentiment analysis as an important aspect for understanding the emotional dimension of social media discourses during decision-making. Across the reviewed literature so far, their influence has been shown to impact virality and engagement within discourse on social media for decision-making. E. g., this means that identifying a surge of negative reactions around a public health campaign or brand event can potentially inform timely interventions for management of strategies within organisational goals.

2.2.1.3 *Harmfulness detection*

Harmfulness in computational tasks involves concepts such as toxicity (Sheth, Shalin, and Kursuncu, 2022) and hate speech (Cinelli et al., 2021). This perspective has become an important focus in research related to false information on social media (Alam et al., 2022). This focus arises from the observation that malicious actors frequently use inflammatory and harmful language to amplify the reach and impact of misleading content (Sharevski, Jachim, and Florek, 2020). Prior studies have demonstrated a clear link between harmfulness and misinformation spread. E. g., showing that toxic language (Pascual-

Ferrá et al., 2021; Salminen et al., 2020) and hate speech (Evolvi, 2018; Sharma et al., 2020) contribute to increased user engagement and attention. Recent research involving fact-checkers has also highlighted that harm assessment can assist in prioritising intervention efforts (Sehat et al., 2024).

A recurring debate within this area of research is how harmfulness is conceptualised and annotated. Researchers have treated harm categories such as hateful, offensive, toxic, racist, or sexist as interchangeable (D'Sa, Illina, and Fohr, 2020), although these concepts represent distinct phenomena (Waseem et al., 2017). This lack of clarity has influenced shared tasks providing various datasets (e. g., Offensive Language Identification Dataset (OLID) (Zampieri et al., 2019), Hate Speech and Offensive Content Identification (HASOC) (Mandl et al., 2019), Hatebase (Hatebase, 2022)). Across these examples, differences in annotation guidelines make direct comparisons between models and tasks challenging (Alkomah and Ma, 2022; Chhabra and Vishwakarma, 2023). As a result of these broad definitions, it is safe to say that harmfulness detection is not a straightforward computational problem, but is closely related to shifts in social norms and annotator subjectivities.

Early approaches relied on lexicon-based methods or supervised machine learning classifiers such as Support Vector Machines (SVM) and Naïve Bayes (NB). One limitation of traditional classifier models is their inability to capture contextual meanings and the diverse linguistic expressions found in social media (Fortuna and Nunes, 2018). With the rise of deep learning, transformer-based architectures have demonstrated comparable performance and significance in stance detection and sentiment analysis, as discussed earlier. For instance, pseudo-label fine-tuning of transformers was proposed to improve zero-shot hate speech detection in non-English languages (Zia et al., 2022). Their work addresses the common problem within low-resource scenarios in NLP. The benefits of pre-trained deep learning models are that downstream tasks can adapt them using fine-tuning strategies without requiring a large dataset to train from scratch. Sim-

ilarly, a hypothesis engineering with natural language inference models was proposed to improve zero-shot detection without the need for domain-specific fine-tuning (Goldzycher and Schneider, 2022).

The observed effectiveness of LLMs similarly extends to detecting hate speech (Albladi et al., 2025). Some benchmark work was conducted using GPT-3.5 and GPT-4 on Twitter datasets covering hate speech, offensiveness, and emotions (Bauer, Preisig, and Volk, 2024). Their work demonstrated that while LLMs perform well in general, but are sensitive to the design of prompts and selection of examples introduced to them. A structured evaluation framework and a toxicity metric were proposed by outlining the challenges of bias, reliability, and consistency when deploying LLMs for harmfulness detection at scale (Koh et al., 2024).

A promising and emerging direction of research has used LLMs to uncover coded or evolving forms of harmful speech, such as novel antisemitic terms circulating in extremist communities (Kikkisetti et al., 2024). Recent surveys underline that although these models generalise across multiple forms of abuse (hate, harassment, toxicity), persistent issues remain around data imbalance, evolving vocabularies, and false positives (Diaz-Garcia and Carvalho, 2025). The harmfulness detection approaches reviewed in this section suggest that their impact within social media can be regarded as a supportive indicator. This thesis, therefore, hypothesises that social media analysis requires such a perspective as it captures dimensions of social media discourse that can affect the safety of entities exposed to social media harms.

2.2.1.4 *Popularity analysis*

Popularity, commonly discussed interchangeably with the term *virality*, denotes the degree of engagement a piece of content achieves on social media platforms (Güner, Cebeci, and Aydemir, 2025; Ngai, Singh, and Yao, 2022; Solovev and Pröllochs, 2022). Common engagement indicators on social media include likes, shares, comments, and reposts. These are all quantifiable measures of social media users' interaction. This is relevant for analysing social media discourses as

supportive measures because they have been widely used to understand how information circulates within social media networks. Not only that, these have been used to calculate virality metrics. These calculated scores have been used to inform early intervention strategies in managing the spread of misinformation, rumours, or harmful narratives (Bessi et al., 2015; Kim, 2018; Shao et al., 2016; Shu, Bernard, and Liu, 2019).

Research suggests that popularity on social media information can be influenced by factors such as psychological, social, and algorithmic. For instance, emotionally charged and morally framed content tends to propagate more rapidly online (Solovev and Pröllochs, 2022). Meanwhile, coordinated inauthentic behaviour and platform recommendation algorithms can be manipulated to amplify narratives accompanied by agendas (Shao et al., 2016). Monitoring virality patterns has been highlighted as a way to help stakeholders anticipate public reactions to interventions, such as vaccination campaigns, and to evaluate potential risks to public trust (Xu and Guo, 2018). As a result of these emphases on popularity measures, they can offer a complement to text-based analyses for providing a more holistic understanding of the evolution within social media discourses.

Alongside popularity metrics, topic modelling has been commonly adopted to identify dominant themes and recurring discussions within large-scale social media textual data. Approaches such as Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan, 2003) have been applied to cluster discourses into topics. Neural extensions such as neural variational inference (Srivastava and Sutton, 2017) and topic models (Dieng, Ruiz, and Blei, 2020) facilitate the modelling of more complex discourse structures. Recent method like ViralBERT was proposed to combine content embeddings from BERT with user-level metadata (e. g., follower counts, hashtags) to predict retweet[1] virality. Their work considers both textual feature analysis and temporal dynamics on social media, including aspects such as popularity trends.

1 A retweet (on the platform formerly known as *Twitter*, now *X*) is a re-posting of another user's post.

A recent study proposed a multi-layer temporal Graph Neural Network (GNN) framework for popularity prediction in social media (Jin, Liu, and Murata, 2024). Their work primarily focused on understanding how information spreads across social media networks over time. A Popularity-Aligned Language Model (PopALM) was proposed to improve predictions of social media responses by aligning language representations with popularity trends (Yu, Li, and Xu, 2024). Their work demonstrates the potential of integrating textual and engagement signals in predictive models. Meanwhile, Elmas, Stephane, and Houssiaux (2023) examined Twitter's (currently known as *X*) "Viral Tweet" feature and argued that the ratio of retweets to followers offers a more reliable indicator of virality than absolute retweet counts. Their work critically mentioned that conventional thresholds frequently produce false positives. Popularity is therefore motivated within the mentioned literature and formulated as a supportive feature in analysing social media discourses in this thesis.

2.2.2 *Improved learning through modelling contexts from social media discourse*

The previous section reviewed computational indicators that capture various aspects of social media texts. While these methods have proven effective, a common limitation within these approaches is that they analyse social media texts as separate instances (as related to the problem scope outlined in Chapter 1, Section 1.1). In doing so, they overlook the interconnected nature of social media discourse, where meaning is shaped through sequences of interactions across discourse threads. This gap motivates the present research and aligns with a growing body of work that highlights the importance of contextual modelling. Some prior studies have taken on this direction to capture discourse structures by recognising how ideas, arguments, and responses unfold within interactions (Haupt, Li, and Mackey, 2021; Kalantari, Liao, and Motti, 2021; Silva et al., 2020).

The following review focuses on stance detection as a representative task to illustrate how discourse-level context has been integrated into computational learning models. Although this thesis primarily emphasises stance detection (with the rationale detailed in Chapter 3), the challenges and motivations discussed here are relevant to a broader set of supportive indicators. Early research proposed using recurrent neural networks, such as Long Short-Term Memory (LSTM)-based models with attention mechanisms, to model branches of social media discourses for stance detection (Poddar et al., 2018; Zubiaga et al., 2018). Their work improved performance by explicitly representing the connections within discourse interactions. However, the limitations in their work were the assumptions that social media discourse can be reduced to sequential data rather than framing it as a network of responses.

Building on this limitation, hierarchical models were introduced. For instance, a combined hierarchical and attention-based mechanism was proposed to capture both local and global dependencies across discourse (Yu et al., 2020). Their work showed improvements in stance detection when discourse structures were incorporated for automatic rumour verification. The introduction of pre-trained language models further advanced this direction. Branch-BERT was proposed to encode discourse structure with transformer-based architectures (Li et al., 2023a), while target-specific contextual encoders improved stance detection for claims and entities (Li et al., 2023b).

More recent approaches continue to emphasise discourse-level contexts for model learning. A Global–Local Attention Network (GLAN) was proposed on a multi-turn conversational stance detection dataset (Niu et al., 2024). Their work demonstrated how both short-range and long-range dependencies across conversational threads influence stance prediction. Similarly, the Transformer-based Architecture for Stance Detection with Embeddings *TASTE* was proposed to integrate structural embedding of conversations with transformer-based text encoders to jointly capture content and discourse structure (Barel, Tsur, and Vilenchik, 2025). Their work underscores the benefits of

modelling social interaction patterns. Prior studies addressed limitations of stance detection through contextual reasoning. For instance, *Stance Reasoner* was proposed to combine pre-trained models with explicit reasoning steps and background knowledge to improve zero-shot stance detection (Taranukhin, Shwartz, and Milios, 2024). *MultiClimate* was proposed to integrate textual transcripts (similar to a format of social media discourse) from climate change videos with visual content to model multimodal context for stance detection (Wang et al., 2024). This thesis draws on prior work to motivate the integration of social media discourse as a source of context for computational models. Apart from addressing the problem definition (see Section 1.1), the broader goal within this thesis is to illustrate how discourse-level computational methods can be used to innovate human-centred design.

2.2.3 *Section summary*

This section has reviewed computational approaches for obtaining interpretive indicators from social media texts, including stance detection, sentiment analysis, harmfulness detection, and popularity analysis. These indicators have demonstrated their relevance in representing complex data in addressing challenges within social media. However, their limitation were present in how model learning of texts does not consider the context of connecting instances that formulate a discourse. This overlooks the meaning that is constructed through these very interactions that contribute the most impact toward issues within social media platforms.

Recent advances in NLP models for stance detection addressed this using hierarchical architectures, attention mechanisms, and pre-trained language models adapted to incorporate discourse structures. These approaches have shown performance improvements but highlight the need for models to recognise these contexts during learning. This review thus motivates the thesis's focus on developing computa-

tional methods that capture the discourses of social media as context in Chapter 4. The findings from this study contribute to the HCI design process by providing foundational evidence for design considerations that are motivated by the capabilities and limitations of current computational approaches for social media analysis.

2.3 HUMAN-CENTRED DESIGN FOR DECISION-MAKING SUPPORT

Computational approaches have demonstrated clear benefits in addressing challenges related to understanding social media discourses. However, their practical effectiveness depends on their integration with human-centred design principles that connect computational outputs with human interpretation. This section reviews existing human-centred design approaches by categorising them in various ways they have been approached to support decision-making within the scope of social media analysis.

2.3.1 *Conversational interfaces*

Interactive conversational agents have been developed to support or nudge social media users to make better decisions about what content to share or engage with on social media. Piccolo et al. (2021) developed a multi-agent system that first alerts social media users to potential misinformation, then uses a conversational bot to provide evidence from fact-checkers. Their work encourages users to engage in reflective evaluation and informed judgment when deciding what social media content to consume. Similarly, Gunson et al. (2021) developed a conversational AI system for debunking COVID-19 misinformation. Their work supports users' information consumption decisions by exposing circulated myths and providing reliable alternatives from authoritative sources. These approaches can be centred around either general users of social media or expert users for decision-making related to their respective roles. However,

when examining the effectiveness of these conversational interfaces in broader terms, such as generalisability and scalability, they remain constrained. More recent developments, such as ChatGPT (OpenAI, 2025), Copilot (Microsoft, 2025), and Claude (Anthropic, 2025). These recent conversational agents have demonstrated their capability in performance through linking social media discourse output to information sources (Haleem, Javaid, and Singh, 2022)

Although conversational interfaces provide flexibility through user interaction, integrating them into professional workflows can present challenges. For instance, there is a lack of structured support when navigating a complex social media environment. This is because they require users to initiate the interaction actively to obtain relevant insights. In the workflows of social media analysis professionals, they typically work under time constraints that require a more structured process (Cavaliere, 2020). This open-ended mode of interaction can be inefficient, as it places additional operational demands on the users and problems targeted by this research. Hence, conversational systems can be more suitable as supportive components within broader analytical workflows. Though they are constrained if treated as standalone solutions for social media analysis. Nonetheless, their strengths are valued in facilitating related tasks such as summarising a document, clarifying the meaning behind claims, or retrieving context for information. Hence, for this thesis, we did not consider conversational interfaces for the initial phase, as the focus is centred more on providing structured interfaces for making informed decisions around social media.

2.3.2 *Explainability interfaces*

Explainable AI techniques have been applied to fact-checking systems for building transparency and user trust around automated systems. A fact-checking platform that generates knowledge graphs was proposed to help users evaluate the veracity of statements by

providing a clear representation of the underlying evidence (Vedula and Parthasarathy, 2021). Next, a collaborative frameworks that extract relevant explanations was proposed to advance benchmarks for combating misinformation on issues such as COVID-19 (Yang et al., 2019a). In studies explicitly targeting expert users, explainability methods such as feature importance, partial dependence plots, and SHAP (Shapley Additive exPlanations) (Lundberg and Lee, 2017) values have been integrated into fake news detection systems to demonstrate how different analytical techniques presented through interactive user interfaces can enhance user trust (Purificato, Shahania, and {De Luca}, 2022).

However, these works oversimplify the complexities of social media by reducing the task to a binary classification, such as true versus false content (Cho, Li, and Lopez, 2024). Plus, these studies do not provide empirical evidence on how expert users interpret and apply these explanations in real-world decision-making as claimed. These approaches were developed with a focus on promoting transparency in established techniques for explaining the outputs of automated machine analysis. However, the challenges within social media analysis extend further than fact-checking pieces of information (Humayun and Ferrucci, 2022). They can be considered useful within the workflow pipeline and for linking facts with automated predictions. However, they cannot be viewed as a generalised or unified solution for dealing with the broader range of challenges present in social media.

2.3.3 *Simulation and nudging*

Simulation-based tools and nudging techniques primarily aim to influence social media users' decision-making by engaging them through active learning to improve media literacy. A web game *Fakey* was designed to simulate social media environments, providing an interactive way to educate users and train them to detect misinformation (Micallef et al., 2021). A web browser extension, *NudgeCred*,

was designed to label the credibility of *Twitter*(formerly *X*) tweets (Bhuiyan et al., 2021). Their work focuses on nudging social media users to critically evaluate information before deciding whether to take any action or engage with it. These methods emphasise behavioural change and aim to guide decisions that influence social media users who play a key role in the dissemination of information online.

Several limitations would impact their practical relevance. First, evaluating their effectiveness in real-world settings would be challenging, given that social media users are not bound by external validation in their beliefs or expressions. Social media users may be more susceptible to cognitive biases and exhibit reluctance to adopt new interventions or remain unaware of these issues to which they are exposed (Modgil et al., 2024). Second, the societal impact of targeting general social media users is limited because they do not hold decision-making authoritative roles to represent societal challenges. Third, the scalability of such interventions raises concerns about the efficiency and sustainability. This is because they are targeting a vast number of social media users for behavioural strategies.

2.3.4 *Visual analytics tools*

Visual analytics translates computational linguistic analyses into visual formats and has been widely adopted to reduce cognitive load in decision-making (Endert et al., 2017; Salamkar, 2024). While these tools are primarily designed for experts, they can be extended to any user interested in understanding social media dynamics. *exRumourLens* was developed to summarise rumour propagation and credibility (Phan et al., 2022). *TweetCred* provides real-time credibility scoring (Gupta et al., 2014). *MisVis* developed a visualisation of misinformation prevalence on websites (Lee et al., 2022). Despite their advantages in summarising automated outputs into visuals, these design concepts tend to oversimplify the complexity of social media dis-

course and do not consider the nature of social media discourses in design. As a result, exploring social media discourse limits users' ability to connect reasoning within these visual analyses. Several commercial platforms have developed tools centred around human decision making, such as *CrowdTangle* (Garmur et al., 2019) and *PeakMetrics*[2]. These were designed to provide trend monitoring and narrative analysis. Although they are designed for humans, they did not address transparency or document any design knowledge through user-level iterations or evaluations.

Many design approaches in this section are limited in scope as they focus on specific issues such as misinformation or rumours. While these are important concerns, they represent only a limited perspective of the broader challenges associated with social media. For this reason, such approaches cannot generalise well across domains. This is because their limitation present a risk of overlooking different ways in which varying social media users and communities engage with social media. For this thesis, design efforts focus on creating tools that support a broader range of users and inform decisions with societal impact. Visual analytics tools offer a unified approach by integrating multiple computational indicators within a single interface. This approach allows users to understand complex social media environments and make informed decisions based on the design's effectiveness.

2.3.5 *Section summary*

Human-centred design in social media analysis involves developing strategies that make complex online data interpretable in ways to support real-world workflows. While progress is evident, significant challenges remain in improving these computational outputs through interactive design, as key decision-makers' responsibilities lie in acting on these insights. This review examined various interaction meth-

2 https://www.peakmetrics.com/

ods and concluded that visual analytical tools offer a more unified approach. The other approaches reviewed in this section are limited in their impact, but are useful as subcomponents that contribute to or are integrated within visual analytic systems for future enhancement.

Moreover, this research focuses on expert users who are directly responsible for addressing these challenges in their professional roles, to support broader impacts. Designing for this user group holds particular value, not necessarily because targeting expert users is more effective, but because the gap between the design and real-world impact can be bridged more directly. These individuals already recognise the value of analysing social media to inform decisions and drive societal impact. Plus, they hold influential positions within their organisations to translate analytical insights into purposeful interventions. All in all, these factors support our conclusion that the design of more effective human-centred analytic tools is needed to facilitate informed decision-making.

2.4 CHAPTER SUMMARY

This chapter established the foundation for the thesis by reviewing the role and impact of social media and discourse in shaping public opinion, institutional trust and decision-making (see Section 2.1). The chapter then examined real-world settings where social media plays a central role in influencing actions, drawing examples from government, public health, policy-making, and commercial management. These examples highlight the need for supporting professionals to interpret social media discourse better in their decision-making workflows.

To ground the empirical work of this multidisciplinary research, the chapter also presented a review of both computational approaches and human-centred design approaches in relevant categories. This thesis addresses a common limitation across both aspects: the overlooking of the contextual foundations within discourses on

social media. Although discourse plays a critical role in shaping influence within social media environments, limited studies have considered incorporating discourse-level insights within computational modelling to inspire an interactive design of a tool.

The empirical work of this thesis addresses this gap, starting with a study on implementing computational approaches for analysing social media discourses. The review in Section 2.2 guided the study's direction and highlighted the importance of indicators such as stance, sentiment, and harmfulness. Then, Section 2.3 examined human-centred design approaches for social media analysis. Many systems have been proposed to bridge computational outputs with user interpretation, e. g., conversational agents, explainability interfaces, simulation and nudging, and visual analytics tools. However, a consistent limitation remains, which they often fail to scale across domains or adequately support the interpretation of discourses in complex social media environments. Although visual analytics tools offer a more unified approach by integrating multiple computational outputs, many still struggle to represent discourse structures that shape meaning across interactions.

This thesis proposes a computational framework for modelling discourse context as a foundation for interface design that better supports expert workflows (see Chapter 4). Building on insights from the computational learning study, the resulting prototype addresses these limitations by incorporating discourse-level interactions into the design (see Chapter 5). These empirical chapters reinforce the thesis's focus on integrating discourse context not only into computational modelling but also into interface design for expert users.

Designing for professionals already engaged in impactful decision-making provides a practical route to real-world relevance. These users are positioned to act on analytical insights and to guide improved design choices. The design approach is further evaluated through two empirical studies: a mixed-method study with general users to validate the design rationale (see Chapter 6) and an expert interview

study to assess its practical utility in real-world decision-making contexts (see Chapter 7).

3

APPROACH

This chapter outlines the methodological approach that guides the conceptual design and the data collection and analysis methods used in this thesis. The chapter explains the rationale for adopting a multi-phase, mixed-methods strategy and details how each empirical study contributes to answering the research questions set out in this thesis. The main approach integrates computational learning, prototyping, and empirical evaluation to ensure both conceptual and practical relevance. The research process is organised into four interconnected studies: (1) an investigation of computational methods to inform and justify the design rationale; (2) the creation of *DiscourseDash*, a functional dashboard interface to support interactive evaluations; (3) a mixed-methods study, to gather insights from general users participants, and (4) an interview study to to gather insights aligning design with industry experts as participants.

STUDIES	DESCRIPTION
Computational methods (Chapter 4)	Experiment with practical model development to motivate design rationale.
Prototyping (Chapter 5)	Producing a functional design interface to support interactive evaluations.
Mixed-method (Chapter 6)	Combining surveys and interviews of general user insights; analysed using statistical and thematic approaches.
Consideration of experts (Chapter 7)	Interviews with industry experts to examine alignment with professional practice; analysed thematically.

Table 1: Overview of research approaches.

3.1 METHODOLOGICAL APPROACH

This research is situated within a design-oriented paradigm, where designing for social media analysis serves as a method for exploring broader questions about human-AI interaction and user agency in algorithmic systems. Research-through-Design (RtD) provides an overall framework regarding the iterative creation, evaluation, and reflection on artefacts as a means of generating knowledge (Zimmerman, Forlizzi, and Evenson, 2007). Let us draw a parallel to the field of game design, where Zimmerman and Salen (2003) argues that design itself is a form of research and defines design in relation to research as follows:

> Design is a way to ask questions. Design Research, when it occurs through the practice of design itself, is a way to ask larger questions beyond the limited scope of a particular design problem.
>
> — Zimmerman and Salen (2003, p. 184)

In this thesis, a functional prototype, also known as an "artefact" is treated as a research instrument, rather than an end product for social media analysis. Its design processes support the research to investigate how analysing social media discourses can be integrated into decision-making workflows for addressing limitations reflected in Chapter 2. While much RtD work emphasises the final prototype and presents design steps as standard illustrations for research, it is within these iterative design activities, including experimentation, reflection and adaptation, that valuable insights are produced.

> The designing act of creating prototypes is in itself a potential generator of knowledge (if only its insights do not disappear into the prototype, but are fed back into the disciplinary and cross-disciplinary platforms that can fit these insights into the growth of theory).
>
> — Stappers (2007, p. 87)

RtD frames the design process as a generative research practice, but translating these insights into actionable knowledge requires grounding the prototype development in established frameworks from HCI. HCI in broader terms provides the theoretical foundation for understanding human-computer interactions (Dix, 2010; Lazar, Feng, and Hochheiser, 2017). Within HCI, there is an associated design philosophy referred to as Interaction Design or Human-Centred Design (HCD) (these terms will be used interchangeably in this thesis) that focuses on human factors in the heart of design processes. HCD offers an established principle to practically guide the creation of systems that respond to users' goals, tasks, and constraints (Giacomin, 2014; Saffer, 2010). Following these allows the research to bridge conceptual and applied perspectives, informing design decisions that remain grounded in user needs.

The challenge, however, within HCI research is integrating computational innovations with users' workflows. However, grounding design decisions in practical technical realities helps mitigate the risks of over-optimistic or unfeasible design concepts (Baldassarre et al.,

2024). In the broader HCI and AI design literature, making technical boundaries explicit is argued to prevent unrealistic expectations. For instance, Shneiderman, 2022 encourages designs that inspire without exaggerating capabilities. Bender et al., 2021 emphasises that pre-development exercises are a form of evaluation to fit planning into broader goals in research. From a transparency point of view, Mitchell et al., 2019 argues for documenting the model's limitations, performance characteristics and intended uses.

The challenge of designing human-AI interactions is frequently linked to low-quality ideation (Yang et al., 2019b, 2020; Yildirim et al., 2024). These studies suggest a common gap where engineers propose solutions that fail to meet user needs, whereas designers may generate concepts that are technically infeasible. Taking these into account, this research ensures the design concepts remain grounded in the constraints of the underlying computational architecture and the realities of implementation. This step was placed as a core component of the iterative design process, not only to anticipate computational realities but also to motivate our design concept.

The success of design solutions was "judged based on the relevance to the individual who ultimately must use the creation" (Kolko, 2011). With that, users are involved in a two-phase evaluation study to assess both practicality and professional relevance in this thesis (Lazar, Feng, and Hochheiser, 2017). The first phase engaged general users through scenario-based tasks to simulate real-world system use (Carroll, 2003), while the second phase involved industry experts evaluating alignment with professional practices. Incorporating both perspectives allows this research to provide complementary insights. Meaning that general users can perform a detailed exploration of interactive features, whereas experts contribute higher-level insights grounded in real-world professional contexts. This integration bridges design theory with practical application (Zimmerman, Stolterman, and Forlizzi, 2010). Overall, these established approaches provide a structured yet adaptable framework for this thesis to guide

the interactive design of computational innovations while keeping user needs central.

3.2 DEVELOPMENT OF THE DESIGN CONCEPT

In line with the overall methodological approach outlined in the previous section, this section details the development of the design concept through a sequential process from computational learning to connecting computational insights with design and prototyping.

3.2.1 *Conducting computational learning*

> Suppose you have an application that you think machine learning might be good for. The first problem facing you is the bewildering variety of learning algorithms available. Which one to use?
>
> — Domingos (2012, p. 79)

This question highlights the importance of adopting a systematic approach when applying machine learning in both research and practical applications. In practice, this process can be broadly characterised by three key stages: (1) defining the computational task; (2) identifying limitations in current approaches; and (3) selecting and adapting appropriate modelling techniques.

3.2.1.1 *Defining the computational task*

A computational task formalises what a human is trying to obtain from a dataset in a form that can be learned by a machine (Domingos, 2012; Lake et al., 2017). For instance, when a human is working with a large, noisy, and linguistically complex dataset from social media, the process of manual interpretation becomes impractical. For automated analysis to support that process, it is necessary to clarify the intended outcome. In supervised machine learning (Caruana

and Niculescu-Mizil, 2006), this means specifying a target objective that guides model development and ensures alignment with broader aims. This allows a computational task to replicate aspects of human analysis in a way that is automated and suitable for integration into a design pipeline. This definition is fundamental in guiding a targeted implementation of a standard machine learning pipeline to solve a problem.

3.2.1.2 *Identifying the limitations in current approaches*

Despite considerable algorithmic progress, computational learning is constrained as much by the characteristics of available data as by the design of the algorithms themselves (Akter et al., 2022). Data limitations consistently represent the primary limitation preventing effective model performance and broader adoption (Mumuni and Mumuni, 2022). This has resulted in limited training data with relevant labels to specific computational tasks. The machine learning community has responded to these challenges through targeted solutions, such as transfer learning (Weiss, Khoshgoftaar, and Wang, 2016), data augmentation (Mumuni and Mumuni, 2022), and few-shot learning (Yin, 2020).

This concern is relevant to the context of social media discourses in this research. As outlined in Chapter 2, the majority of available datasets and experiments are designed around classification tasks and do not reflect the discursive structures of social media environments. Through these considerations, this thesis addresses computational learning in the context of data limitations by leveraging data augmentation, which allows the creation of additional training examples to reflect the structure of social media discourses. By situating the pipeline within this problem framing, the research adopts a problem-driven approach that ensures computational learning processes reflect the contexts in which they are applied. As prior work has similarly emphasised, the construction of computational learning pipelines begins with identifying and refining a problem formulation,

as informative outcomes depend on translating domain-specific questions into computationally tractable tasks (Domingos, 2012).

3.2.1.3 *Selecting and adapting modelling techniques*

After defining the task and identifying limitations in existing methods, the next stage involves selecting and adapting appropriate modelling techniques. Approaching this through an NLP structured pipeline, prepare a model to achieve the formulated computational task. This typically involves the following sequential steps:

DATA PRE-PROCESSING

Pre-processing of textual data is a fundamental step in preparing a dataset for any computational learning tasks. Decisions made at this stage directly influence model performance, as textual data typically exhibits noise and inconsistencies that can affect how a model learn informative patterns from it. The following are some common techniques applied to preserve relevant information while reducing irrelevant variability.

- **Tokenisation** (Palmer, 2000) breaks text into units, i. e., tokens, which can include words, phrases, or symbols. This process handles sentence segmentation in informal or noisy text, handling unconventional grammar, abbreviations, and emojis, and converting special characters into machine-readable tokens.
- **Noise removal** (Al Sharou, Li, and Specia, 2021) removes content that can interfere with pattern recognition, such as URLs, non-alphanumeric characters, extra whitespace, stop words, repeated punctuation, etc. .
- **Normalisation** (Patro and Sahu, 2015) standardises text by applying methods such as lowercasing, consistent encoding (e. g., UTF-8), and converting emojis to descriptive texts.

- **Lemmatisation and stemming** (Khyani et al., 2021; Singh and Gupta, 2017) reduce morphological variation by converting words to their root forms. Lemmatisation preserves semantic meaning via dictionary lookups (e. g., *"jumping"* → *"jump"*), whereas stemming applies algorithmic rules that produce non-words but are more computationally efficient (e. g., *"studies"* → *"studi"*).

- **Data resampling** addresses class imbalance using techniques such as Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) or data augmentation (Maharana, Mondal, and Nemade, 2022). E. g., synthetically increasing instances of the minority class to balance the dataset with the majority class to reduce potential bias against underrepresented categories.

FEATURE REPRESENTATIONS

Machines do not process words or understand meaning in the same way humans do. During their learning process, they rely on numerical representations of text, which encode patterns, relationships, and contexts. The process of translating raw text into these numerical forms is a foundational step in NLP known as feature representation. Techniques to achieve that can range from simple encodings of word presence or frequency to distributed embeddings that capture semantic relationships in continuous vector spaces. Traditionally, NLP relied on manually engineered features designed using linguistic knowledge. Deep learning, on the other hand, automatically learn latent feature representations from large corpora. In both cases, textual data must be converted into numerical vectors, a process known as *vectorisation*, to be compatible with machine learning algorithms.

- **Traditional representations** focus on explicit word-level patterns rather than underlying meaning. The following are some techniques. Bag-of-Words (BoW) (Zhang, Jin, and Zhou, 2010), N-gram encoding (Brown et al., 1992) (e. g., unigrams, bigrams,

trigrams), Term Frequency–Inverse Document Frequency (TF-IDF) (Qaiser and Ali, 2018) represent words using frequency-based statistics. Although these methods are computationally efficient, they fail to capture semantic relationships in recognising conceptually similar words or phrases that convey related meanings.

- **Word embeddings** overcome this limitation by representing words as dense vectors that capture both semantic and syntactic relationships. Word2Vec (Mikolov et al., 2013), GloVe (Pennington, Socher, and Manning, 2014), FastText (Bojanowski et al., 2017) represent words in a continuous vector space by placing those with similar meanings close together. However, these embeddings remain static and do not adjust to the specific context of a sentence.

- **Contextualised embeddings** through deep learning allow models to capture semantic and syntactic information dynamically. This is different from static word embeddings, as it adapts to surrounding words by allowing the representation of meaning at the sentence level. Transformer-based architectures (Vaswani et al., 2023), built on attention and encoder–decoder mechanisms for producing representations that can be adjusted for specific tasks. BERT (Devlin et al., 2019) uses a bidirectional encoder architecture to learn context-dependent representations. Other approaches include Convolutional Neural Network (CNN) (O'Shea and Nash, 2015), which inspired TextCNN (Guo et al., 2019) by enabling the extraction of local dependencies in text sequences.

Amongst these, contextual representations are evolved approaches that capture both meaning and structure in ways that static embeddings cannot. In turn, this leads to improved performance across various types of NLP tasks. This makes contextual embeddings well-suited to the computational learning task explored in this thesis around social media discourses.

MODEL LEARNING

Model learning refers to the design of algorithms that automatically identify and generalise patterns from data. This is where machines learn to perform predictive tasks in automating aspects of human analysis.

Neural networks are a class of machine learning models that approximate complex functions through layers of interconnected computational units (inspired by biological networks of neurons). Early developments began with the perceptron for supervised learning (Rosenblatt, 1958) and were later extended into multilayer perceptrons. A neural network consists of an input layer, one or more hidden layers, and an output layer. Each unit computes a weighted sum of its inputs, followed by a non-linear activation and training process is typically performed using back-propagation (Rumelhart, Hinton, and Williams, 1986). Figure 3 presents a simple neural network for text classification, where input text is represented as embeddings, transformed through hidden layers, and mapped to output predictions. Although a single perceptron is limited to linear decision boundaries, the concept of stacking multiple layers allows the network to learn non-linear mappings. This progressively transforms the input space into a representation where the data becomes more linearly separable. Then, the final output layer applies a simple linear classifier to these transformed features.

Deep neural networks are also known as deep learning. When computational hardware advanced, the evolution brought forward the capability of neural networks to accommodate greater depth for representational capacity. This advancement eventually led to the development of deep residual network architectures incorporating hundreds of hidden layers (He et al., 2016). To address data where order and temporal relationships are important, such as text, speech, or time-series signals, RNNs were developed to capture these sequential dependencies. However, gradients within RNNs can either diminish towards zero (vanishing gradient) or grow uncontrollably large (ex-

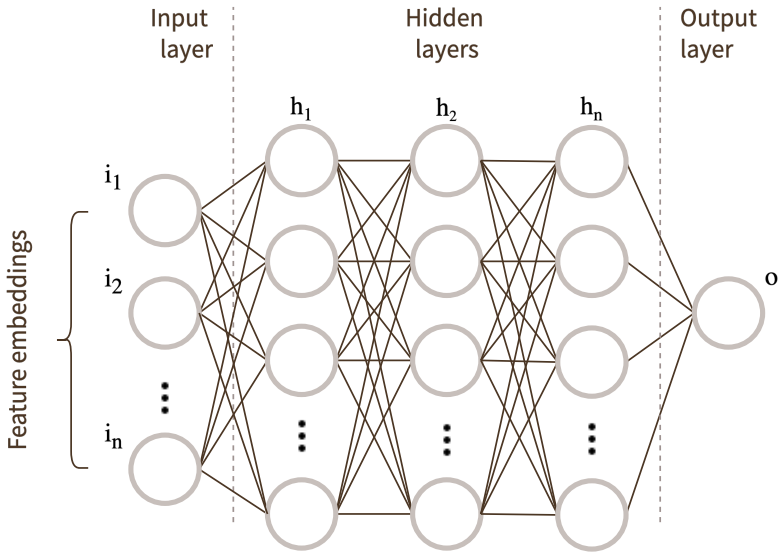


Figure 3: An example of a 3-layer neural network.

ploding gradient). This made training RNNs difficult as information is back-propagated through many time steps. As a result, this limits a model's ability to learn long-range dependencies. With that, popular approaches such as LSTM (Graves, 2012) and Gated Recurrent Units (GRU) (Cho et al., 2014) were designed to mitigate these issues by introducing gating mechanisms that regulate how information is stored, updated, and forgotten over time. Although effective in managing gradient issues, these models remain inefficient due to their sequential processing and their difficulty in capturing global context as the dataset scales.

Transformers overcome the limitations of RNNs by introducing a self-attention mechanism. This mechanism models dependencies simultaneously between all elements of a sequence. The development of this approach addresses the limitations of sequential computation by performing better at capturing long-range dependencies. In NLP, transformer-based architectures have become the state-of-the-art across a wide range of tasks (refer to Chapter 2). Figure 4 illustrates the transformer architecture. The architecture was introduced by Vaswani et al. (2023) and inspired most of the advances in NLP pre-trained models such as BERT (Devlin et al., 2019) and GPT (Radford et al., 2018). This is because the proposed architecture can be

adapted for various types of NLP computational learning tasks, from text classification, generation, and translation. The impact is evident in the development of large pre-trained models that leverage different components of the architecture to utilise their strengths by training on large linguistic corpora. As a result, downstream computational tasks can be adapted through fine-tuning without requiring training from scratch. Within the transformer, the encoder (left side of Figure 4) produces contextualised representations of input tokens that can be used for classification or other predictive tasks, while the decoder (right side of Figure 4) supports generative tasks and serves as the backbone of models like GPT.

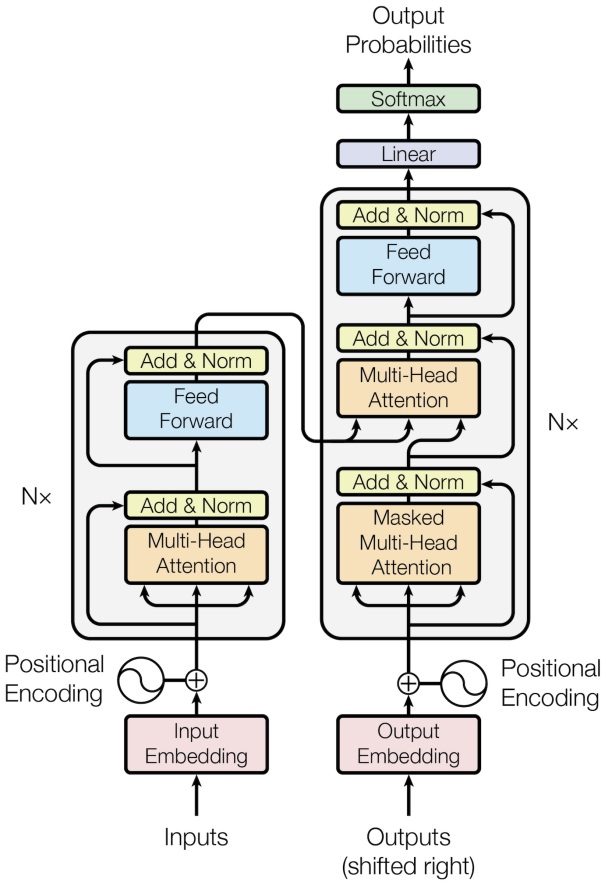


Figure 4: The transformer architecture obtained from the original paper by Vaswani et al. (2023).

CROSS-VALIDATION METHOD

K-fold cross-validation is a procedure commonly implemented during the evaluation phase of a machine learning model. The use of this procedure is to provide a more reliable estimate of the model's generalisation error. The procedure consists of the following steps:

1. Compute the class proportions for the entire dataset.
2. Divide the dataset into K partitions while preserving the class proportions calculated in Step 1.
3. For each trial, designate one of the K partitions as the test set and train the model using the remaining $K - 1$ partitions combined.
4. Calculate the evaluation metrics across the K trials and report the aggregated performance.

The procedure yields a reliable estimate of generalisation error for supporting the evaluation of hyperparameter quality during training. Figure 5 represents an example of the explained K-fold cross-validation.

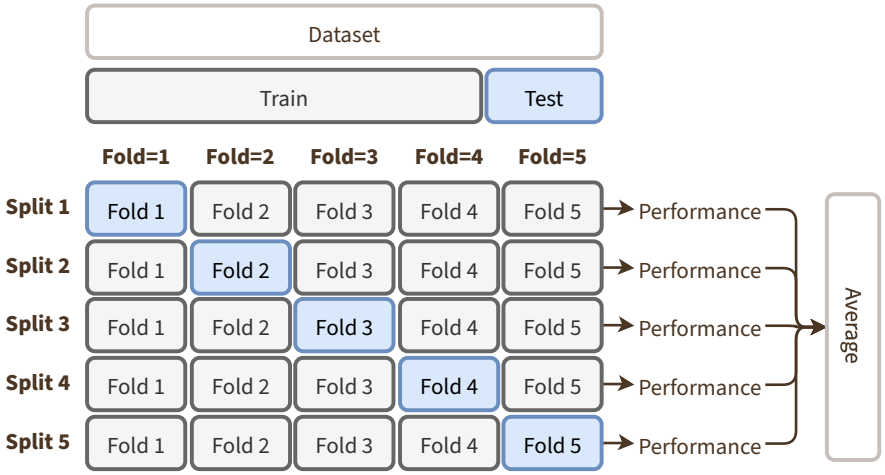


Figure 5: An example of k-fold cross-validation where $k = 5$.

EVALUATION METRICS

The performance of the classification models is then evaluated using a range of widely adopted metrics to illustrate a model's performance in computational learning tasks. These metrics are obtained from the confusion matrix that summarises the relationship between predicted and actual class labels. The core elements of a confusion matrix are:

- **True Positives (TP)**: instances correctly predicted as belonging to a given class.
- **False Positives (FP)**: instances incorrectly predicted as belonging to a given class when they actually belong to another class.
- **True Negatives (TN)**: instances correctly predicted as not belonging to a given class.
- **False Negatives (FN)**: instances incorrectly predicted as not belonging to a given class when they actually do.

Table 2 presents an example confusion matrix for a three-class classification task with classes A, B, and C. This example illustrates how the main evaluation metrics are computed from the core elements of a confusion matrix.

		Predicted class	
	A	**B**	**C**
Actual A	TP_A	$FN_{A \to B}$	$FN_{A \to C}$
Actual B	$FP_{B \to A}$	TP_B	$FN_{B \to C}$
Actual C	$FP_{C \to A}$	$FP_{C \to B}$	TP_C

Table 2: Example confusion matrix for a three-class classification task with classes A, B, and C. The arrows $\to$ indicate the direction of misclassification.

A computational task for classifying text is trained on a set of labelled items (the training set) and then used to classify unseen items (the test set). For evaluating them, the predictions are compared

against the ground truth labels. The following performance metrics are then computed:

- **Accuracy** measures the proportion of correctly predicted instances within all predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision** measures the proportion of correctly predicted positive instances within all instances predicted as positive. For a given class, it reflects the ability of the classifier to avoid false positives:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall** measures the proportion of correctly predicted positive instances out of all actual positive instances. For a given class, it reflects the ability of the classifier to avoid false negatives:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-score** is the mean of precision and recall for providing a balanced measure between the two. It penalises imbalances by making sure that reported performance takes into account the balance of precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

For multi-class classification tasks, which is the case of this research, accuracy alone does not account for potential biases introduced during training. While accuracy provides a quick overview of performance, it can be misleading in cases where the model may perform well simply by favouring the majority class. Hence, F1-score metrics are commonly computed for each class individually and then averaged because they provide a better indication of whether performance is balanced across classes, since low recall or precision for one class will penalise the overall F1 score. This can be presented through *macro-averaging* (treating all classes equally) or *weighted-averaging* (weighting by class frequency).

3.2.2 *Connecting computational insights with design*

Previous research has argued for grounding interactive design studies in existing computational limitations to better approximate real-world use cases and system performance (Baldassarre et al., 2024; Shneiderman, 2022). This way, design concepts can be perceived as more reliable regarding challenges that are likely to align closely with the post-evaluation of a deployed system. Similarly, grounding a design concept in actual NLP model outputs provides a level of veracity to the evaluation process for ensuring that user insights are relevant to both current and near-future technical realities. This thesis adopts this approach to establish a relationship between computational findings and the design rationale (see Chapter 5).

The aim is not only to motivate the conceptual formulation but also to ground it in the capabilities and constraints of NLP methods. With that said, the design is made realistic rather than hypothetical, positioning the empirical evaluation with users within a concrete technical context. This integration also enhances ecological validity by avoiding overly idealised representations of a system (Holleman et al., 2020). Considering technical feasibility within the design further addresses a common concern in HCI research about the risk of evaluating interfaces that are disconnected from realistic system behaviour, causing misleading conclusions about user experience (Ackerman, 2000).

3.2.3 *Prototyping*

Artefact, a term originating in anthropology, refers to a man-made item. In HCI, artefact items "made by designers" are commonly used in RtD approaches (Zimmerman, Forlizzi, and Evenson, 2007). Within this framing, prototypes serve as artefacts that facilitate interaction and experience with a concept. A prototype is a product-like construction allowing users to *engage with* and *respond to* design ideas

(Lim, Stolterman, and Tenenberg, 2008). The prototyping process is not solely for producing artefacts for validation. It also explores how users might interact with computational components and interface features. In turn, it supports iterative refinement of both the computational pipeline and the logic flow of interface design to maintain alignment with user needs (Vermeeren et al., 2010).

Besides, prototyping is understood both as a process of making and as a mode of investigation in user studies for critical reflection with real-world practices (Lim, Stolterman, and Tenenberg, 2008; Zimmerman, Forlizzi, and Evenson, 2007). Typically, the visual and interactive elements of a prototype progress from ideation wireframes to designs informed by visual analytics and preliminary insights. Following approaches in interactive machine learning (Amershi et al., 2014; Doshi-Velez and Kim, 2017), features of machine learning outputs were represented through sortable and filterable visualisations, contextual views, and colour-coded labels to help users make sense of complex data.

In this research, prototyping functioned as a methodological component to translate computationally driven conceptual designs into interactive artefacts. The prototype will not only act as a conceptual representation of the final product, but also as a working representation during the interactive user studies. They represent early design hypotheses to facilitate iterative engagement with prospective users. By placing the prototype within the broader research process, the research maintained a close alignment between computational innovation and human-centred values. This situates the resulting design as a solution to the problem of social media analysis, but also a research artefact that represents methodological commitments and empirical insights generated throughout user studies.

3.3 DATA COLLECTION FOR EVALUATION

Regardless of a design concept's computational affordances, its contribution to the literature depends on how well it addresses real-world user needs. Conducting user studies is therefore a fundamental part of this thesis, providing empirical insights as proof of concept while integrating human factors throughout the design process (Hartson and Pyla, 2012). There is a distinction in HCD research between expert users, who have specialised domain knowledge and practical experience, and general end-users, whose interactions are typically influenced by broader, more general concerns (Sharp, Rogers, and Preece, 2007). The following section outlines how each type of user data was collected and how it will be analysed to address the research questions set out in this thesis.

3.3.1 *Mixed-method user study*

The prototype was developed based on a concept grounded in design rationale from a computational learning exploration. In line with participatory design principles, the study then sought to evaluate how well the design aligned with broader user expectations and needs (Schuler and Namioka, 1993). While the intended users of the system are industry experts, initial validation was conducted with general users as participants to test the balance of the design rationale from additional perspectives. This step aimed to uncover potential mismatches between a researcher's assumptions and practical interaction patterns for addressing the gap between theoretical design intentions and practical use highlighted in prior work (Smith, 2006). Uncovering these overlooked details during early evaluation is important to ensure the design is sound and understandable before engaging targeted expert users.

A mixed-method evaluation approach was adopted, combining both qualitative and quantitative techniques to capture diverse facets

of user interaction and response (Creswell, 1999; Edmonds and Kennedy, 2016; Östlund et al., 2011). Qualitative methods reveal how users made sense of the interface and its outputs, helping to uncover reasons behind any confusion, hesitation, or misinterpretation (Lazar, Feng, and Hochheiser, 2017; Seaman, 1999). Quantitative data provides measurable insights into usage patterns and the perceived effectiveness of design elements (Nardi, 2018). Combining these approaches provides an all-rounded understanding of the prototype's strengths and limitations during initial development. These are valuable in validating basic usability to make sure that the interface does not rely on one-sided assumptions. Early identification and documentation of these observations help to guide the next phase of the research, which involves more targeted evaluation with expert users. Above all, it provides additional perspectives as evidence to ground the design concept during its formative stages.

3.3.1.1 *Interviews*

A semi-structured style of questions was used to gather insights into participants' perceptions and experiences while interacting with the prototype. This approach combines open-ended questioning with guidance from how questions are crafted in a way that relates to the research objectives (Kvale and Brinkmann, 2009). Not only does it allow participants to express their perspectives freely, but it also ensures that the data collected is relevant to the study objectives. The flexibility of this format supports researchers to probe interesting or unexpected responses while maintaining a consistent structure across sessions (Lazar, Feng, and Hochheiser, 2017). Also, studies with general users can accommodate more complex instructed tasks. They reveal insights that are not easily observed through task performance alone. To avoid the risk of one-sided findings, a consistent core set of design questions was repeated across different tasks and conditions. This acts as a control measure commonly adopted in HCI methods to ensure insights remain focused on each task, while supporting accurate cross-condition comparisons across participants.

3.3.1.2 *Surveys*

Questionnaires using Likert-scale items are commonly adopted within HCI as part of a survey to collect measurable data on participants' perceptions (Kaptein, Nass, and Markopoulos, 2010). While interviews provided an in-depth understanding, surveys captured quantifiable evaluations of the design's perceived usefulness, all of which are common approaches in HCI (Sauro and Lewis, 2016). Based on the nature of the mixed-method approach, all survey respondents also participated in the interview process to ensure qualitative reflections and quantitative ratings are reliable. This approach supports the triangulation of findings by combining structured responses with open-ended insights for a more credible analysis (Golafshani, 2003).

3.3.2 *Consideration of expert practices*

Incorporating real-world considerations is a key perspective for bridging the design concept of analytical tools intended for practical use. Prior studies have investigated how industry experts' workflows reflect on HCI efforts to support specific domains by involving expert participants. These studies include experts such as journalists (Atreja et al., 2023), healthcare practitioners (Patel, Kaufman, and Kannampallil, 2021), medical specialists (Lindgren, Winnberg, and Winnberg, 2011) etc.. These experts' investigations have demonstrated their impact within research in design. This thesis therefore values this perspective as an important part of the design process for proposing an interactive design within social media analysis. Engaging expert users who are familiar with the workflows this research aims to support ensures the inclusion of highly relevant knowledge and first-hand experience. Their insights are invaluable for informing the design of tools facilitating complex analytical workflows aligned with real-world professional practices.

3.3.2.1 *Target expert user group*

This thesis focuses on professional communicators who use social media data to inform strategic decisions aimed at influencing audiences within organisational contexts. This group includes experts in brand and reputation management, public communication, journalism, marketing, public health communication, and policy-related communication (see details in Section 2.1.2). Their roles typically involve monitoring, interpreting, and responding to social media discourses to guide organisational messaging, shape public perception, and achieve communicative objectives. By defining this target user group, the thesis situates its design and evaluation within professional practices where social media functions as a tool for influence. This scoping is supported by established communication studies that emphasise the role of social media in shaping public understanding and organisational decision-making (Jiang, Luo, and Kulemeka, 2016; Pekkala and Zoonen, 2022).

3.3.2.2 *Expert interviews*

Conducting expert interviews has long been established as a method of qualitative empirical research to explore or collect data about a specific field of interest (Van Audenhove and Donders, 2019). Meuser and Nagel (2009) describes an expert interview as a qualitative interview guided by a topical framework that focuses on the expert's knowledge. Within the specific file of interest, this type of qualitative approach can be broadly characterised as specialised knowledge within a domain of action. Semi-structured interviews were employed in this thesis to incorporate open-ended questions while staying aligned with the research objectives (Lazar, Feng, and Hochheiser, 2017). However, these interviews differ from the previous general users' mixed-method semi-structured interviews. The emphasis here is on expert participants, whose professional roles involve analysing social media discourse and who therefore bring specialised knowledge and insights. Unlike interviews with general users, these ses-

sions were designed to capture the specific expertise that only such experts can provide.

Participants were recruited from relevant domains where understanding social media discourse is important in their decision-making workflows. This approach not only grounded the study in real-world practice but also highlighted the kinds of interpretive tasks experts perform, the types of analysis aspects they look at and the concerns from applied settings. The engagement with experts will inform both the design requirements for professional contexts and frame the evaluation criteria that align with how the system would be judged in practice. This approach aligns with the broader literature on user-centred design and participatory methods in HCI for interactive design for addressing real-world needs while remaining methodologically rigorous (Blandford, Furniss, and Makri, 2016).

3.4 DATA ANALYSIS

3.4.1 *Statistical analysis*

Quantitative analysis was used to evaluate the design components across different experimental conditions (Dragicevic, 2016). As described in the previous section, participant responses were collected through structured questionnaires as part of a mixed-methods study. These data were analysed using a combination of descriptive and inferential statistics (Müller, Sedley, and Ferrall-Nunge, 2014). Descriptive measures, including means and variability, were used to summarise trends in participant perceptions. Given the ordinal nature of Likert-scale data and the potential for non-normal distributions, non-parametric statistical methods were selected for inferential analysis (Lazar, Feng, and Hochheiser, 2017). Specifically, the commonly used Wilcoxon Signed-Rank Test (Woolson, 2007) was applied to compare paired responses across conditions, while the Friedman Test (Friedman, 1937) was used to examine differences in perceptions

across multiple analytical features. These were chosen according to informed literature in HCI methods to ensure that comparisons are appropriate for the scale and type of data collected.

3.4.2 *Thematic Analysis*

To systematically analyse participants' transcriptions after interviews, a thematic analysis was followed to gather qualitative perceptions on the design. In social science, the term "thematic analysis", has been widely used to refer to the methods researchers use to analyse qualitative data into informative themes. A widely recognised approach was introduced in a paper named "Using thematic analysis in psychology" (Braun and Clarke, 2006). The authors categorised the process into the following steps:

1. Familiarise with the data through repeated reading of interview transcripts.
2. Generation of initial codes to capture features within the data.
3. Collation of codes into potential themes to represent broader patterns and relationships.
4. Review and refine the themes to relate them to the research objectives.
5. Define and name the themes to articulate the findings and their implications for evaluation.

Following the steps outlined above for thematic analysis allows this research to collect qualitative data from participants and systematically capture and analyse insights. This is because the approach was designed to ensure that the process from data collection to analysis can be accurately documented to support research and produce knowledge. This thesis, therefore, adopts this approach to present the resulting themes, which ground the design concept in empirical evidence collected through interviews.

3.5 CHAPTER SUMMARY

This chapter introduces the approach that informed the research planning to answer the research question established in Chapter 1. These include how the research approaches the development of a design concept, conducting data collection and analysis for the empirical work. This research adopts an RtD methodology to conduct research through a prototype as an interactive artefact. This overview approach positions design practice as a way of asking questions and generating knowledge through the development and evaluation of design solutions. This chapter then elaborates on the prototype design process, highlighting the role of computational learning in establishing the technical foundations of the design (see Chapter 4), which later informs the design rationale (see Chapter 5). The detailed steps outlined in this chapter were followed in this thesis for experimentation with algorithmic approaches to translate computational learning insights into a visual interactive format. These formats were developed into a prototype that serves both as a research instrument and as a contribution to the design itself.

The empirical evaluations adopt a multi-phase approach to gather supporting evidence for the design concept. In the first phase, a user study employing scenario-based tasks was conducted to examine user interactions and extend the researcher's design assumptions. This phase reveals usability patterns and validates core design concepts through controlled evaluation. The subsequent phase focuses on expert interviews, bridging the design with real-world professional practices. A semi-structured interview approach was used to explore existing workflows, capture experts' perceptions, and assess how the design aligns with professional needs. By combining these evaluation phases, this thesis ensures that both general usability and domain-specific requirements are systematically addressed. The methodological approaches described in this chapter provide a structured foundation for how each subsequent empirical chapter addresses specific aspects of the research questions. Adapted from

widely recognised literature, these complementary methods allow this thesis to compile a set of empirical evidence for its contributions in the chapters that follow.

Part II

EMPIRICAL WORK

4

STANCE DETECTION: AN EXPLORATORY STUDY

This chapter presents a study on stance detection, a computational learning task that identifies the opinions or attitudes expressed in a text towards a given topic. The aim is to conduct a computational learning process to support and partially automate the traditionally manual and labour-intensive process of analysing social media discourse. This allows a practical implementation of how computational models perform when introducing discourse-level context and identifies their limitations. These will pinpoint areas where human judgment remains essential and serve as evidence for the opportunities in interactive design within an analysis process.

Building on the approach discussed in Chapter 3, this study investigates stance detection as a problem targeting the limitation of data scarcity. The experiments focus on applying data augmentation techniques using LLM to strengthen a model's capability to recognise stances when the data is limited. The findings reveal both the potential and the limitations of current computational approaches. The results demonstrate that incorporating LLM-generated synthetic data can improve model performance in stance detection tasks. However, the more important part of the study is the reflections on the limitations of these computational approaches. The findings reveal indicative insights to support the motivation of our interactive design. Through an exploratory investigation into the practical implementation of computational methods, this thesis positions them as segmented evidence that, when synthesised, contributes to a broader picture that guides decisions around the development of an interactive design. These methodological insights directly guide the design decisions in the following chapter.

4.1 INTRODUCTION

Research into the design of analytical tools for social media has primarily focused on well-defined system functionalities (as discussed in Chapter 2). Existing literature fails to connect interactive design with the broader computational constraints of machine-generated analysis (Stieglitz et al., 2018). This chapter explores computational methods not only to examine practical implementation but also to identify areas where human input can shape the design of interactive systems. Conducting this computationally grounded and exploratory study allows this research to evaluate computational learning in practice. More importantly, these outline its strengths and limitations. The exploration follows three structured components outlined in Chapter 3: (1) defining the computational task, (2) identifying limitations in current approaches, (3) selecting and adapting modelling techniques.

Firstly, this study focuses on *stance detection* as a representative computational task within the broader context of social media analysis. Stances, commonly known as opinions, reveal how individuals who express textual interaction on social media are "in favour" or "against" a target topic. This perspective demonstrated its value in facilitating more effective information management strategies (Küçük and Can, 2021). Stances contribute to the understanding of misinformation spread as they reveal how narratives amplify or are manipulated regardless of factual accuracy (Lewandowsky et al., 2012). Therefore, monitoring them became significant as they influential indicator for organisations in addressing social media crises. Researchers commonly adopt stance detection as a supportive indicator when examining misinformation in social media (De Magistris et al., 2022; Küçük and Can, 2021; Lillie and Middelboe, 2019; Mohammad et al., 2016). Hence, in this thesis, stance detection serves as a significant proxy for grounding the exploration of computational methods within the broader context of social media analysis (see Chapter 2 for detailed justification).

TARGET	TEXT	STANCE
Vaccines	"COVID-19 vaccines are a microchip used for monitoring people."	*Against*
Atheism	"If you're superstitious, your thinking may be impaired."	*In favour*

Table 3: Examples of stance labels towards a target from a text.

Secondly, existing stance detection studies predominantly focused on isolated texts, overlooking the evolving and discursive nature of social media discourse (see Section 1.1 & Section 2.2). Despite its importance, a significant lack of publicly available annotated datasets remains that accurately reflect the structure and depth of real-world discourses for the task of stance detection. This limitation constrains research aiming to account for the dynamics of discourse and highlights a clear gap in current approaches. Thirdly, this study investigates *data augmentation* as a strategy to address challenges from limited data availability (see Section 3.2.1.2). Recent advances in LLMs have inspired the generation of synthetic data through prompt-based methods (see Section 2.2). These methods have shown value in other domains (Nadăş, Dioşan, and Tomescu, 2025). At the same time, they offer an opportunity to create synthetic structured discourses for stance detection tasks in this research's problem. This is because it remains unclear whether the benefits of such synthetic augmentation are consistent across varying discourse depths. This navigates an important consideration given the complex and unpredictable nature of social media discourse. To this end, this chapter develops a framework to examine the following research questions:

RQ1 Can existing discourse context approaches to stance detection be enhanced using data augmentation strategies under a data-limited scenario?

RQ2 Can prompt-based generation of synthetic discourses outperform traditional data augmentation techniques?

RQ3 What is the impact based on the depth of discourses on stance detection performance when integrating them as context?

Each of these research questions guides the analysis in this chapter by providing insights from a technical standpoint for connecting to the interactive design in the subsequent chapter. This chapter demonstrates the practical application of computational approaches to social media analysis. It is also important to clarify the context of this chapter within the timeframe in which the experiments were conducted during the early stages of this research, when the field of NLP was significantly different from its current state. Since then, NLP has advanced, reshaping how these problems are approached. Despite these changes, the work in this chapter remains definitive for this thesis, as it serves as evidence of practical implementation supporting the foundation for informing interactive design decisions.

4.2 PROBLEM DEFINITION

Given a dataset of social media texts, $D = \{T_1, T_2, \ldots, T_N\}$ where each T_i represents either a root-level or a child-level comment within a discourse (see Section 2.1.1.1 for a detailed definition of terminology used). The task is to predict the stance label $y_i \in$ {Against, In Favour, Neutral} for each text T_i. This prediction should reflect the opinion expressed towards a predefined target topic. The objective is to predict the stance label y_i for each text T_i by leveraging the context within the discourse. To formulate this, all preceding utterances within the same discourse are retrieved up to and including T_i to form a sequence $C_i = \{T_1, T_2, \ldots, T_i\}$ to capture the discourse dependencies for a context-aware prediction of y_i.

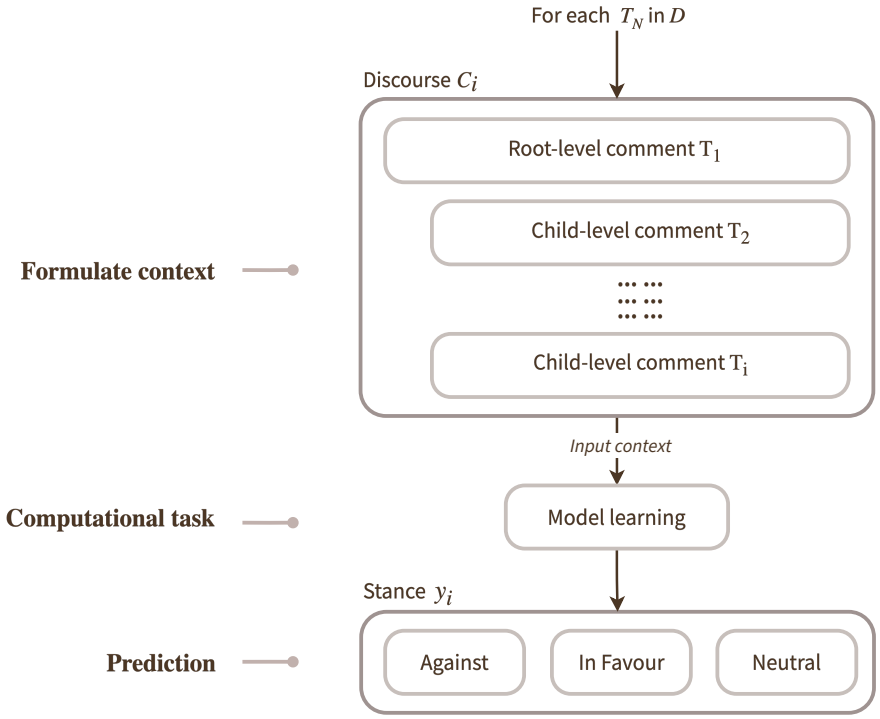


Figure 6: Problem definition of stance detection.

4.3 METHODOLOGY

This section presents the proposed framework, which consists of three core components: contextualisation, generation, and learning. The framework is designed to investigate how using discourse as context and synthetic data can support stance detection in social media under data-limited scenarios. As shown in Figure 7, the framework first contextualises social media discourses, then generates additional training examples if needed, and finally applies a learning component for the prediction of stance labels.

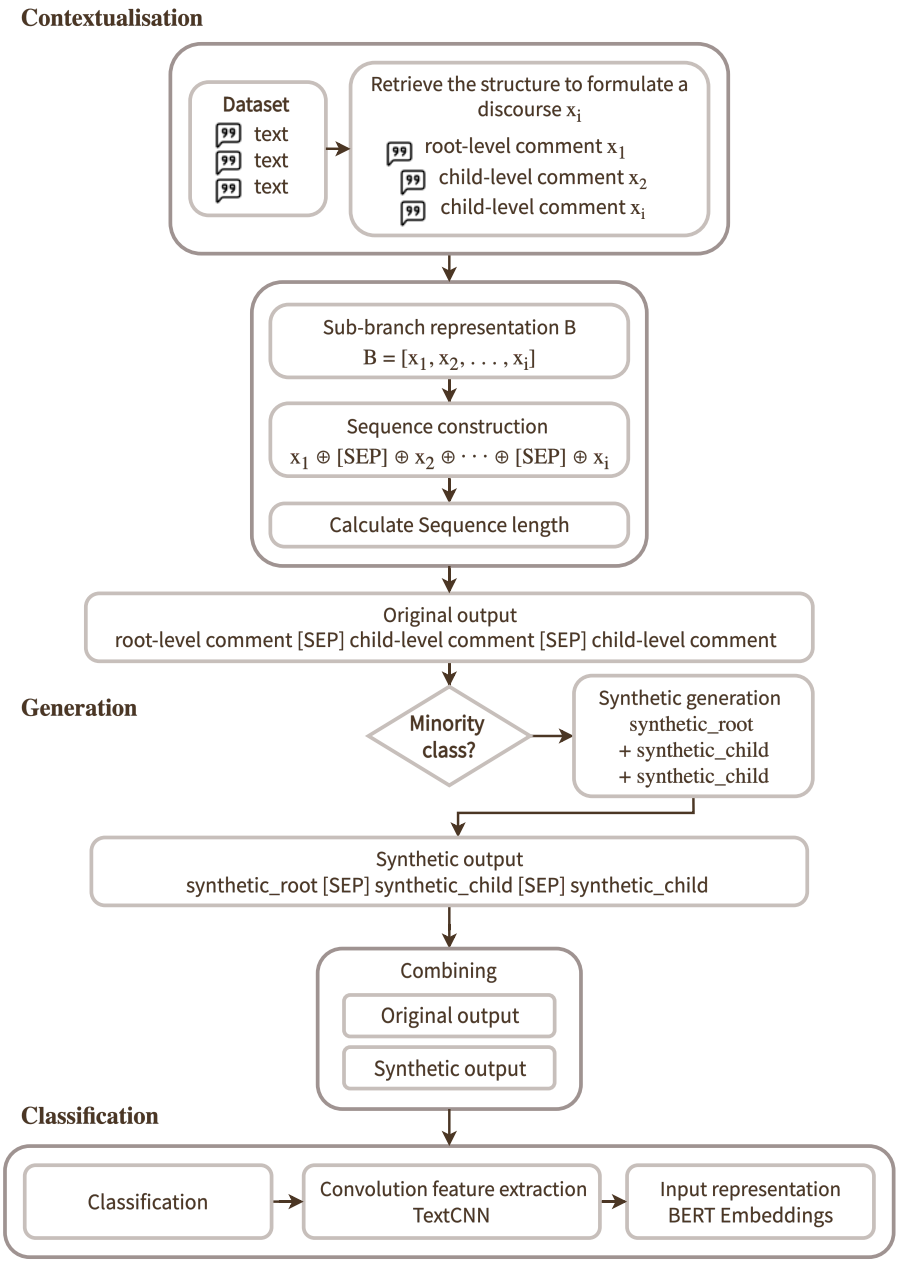


Figure 7: Overview of the proposed framework for stance detection.

4.3.1 *Contextualise*

The first component of the framework retrieves the discursive context surrounding each social media text. Each social media text T_i in the dataset is situated within a structured discourse, either as a root-level or child-level comment. To construct the context for T_i, the framework

gathers preceding comments in the discourse and concatenates them. Following prior work by Zhang et al. (2020) & Zhang, Culbertson, and Paritosh (2017), the structure is organised by linking each child-level comment back to its originating root-level comment or preceding child-level comment. This ensures that the model is fed with the knowledge of the entire discourses that are relevant to T_i.

4.3.1.1 *Sub-branch representation*

For a given text T_i, define a sub-branch $B = [x_1, x_2, \ldots, x_i]$, where each element x_u contains n_u tokens:

- x_1 corresponds to the root-level text P_0.
- x_u ($u = 2, \ldots, i$) represents the child-level comment at varying depths.
- $w_{u,v}$ denotes the v-th token in the u-th element.

This hierarchical representation captures sequential dependencies and relational context within the discourse (Li et al., 2022; Sun et al., 2018; Wei, Xu, and Mao, 2019).

4.3.1.2 *Sequence construction*

Then, all elements resulting from the sub-branch are concatenated into a single sequence to form a connection of this relation. Meanwhile, a special separator token `[SEP]` will be inserted into each of the individual elements for the model to distinguish them within the concatenated sequence. The resulting sequence is constructed as follows:

$$x_1 \oplus \texttt{[SEP]} \oplus x_2 \oplus \cdots \oplus \texttt{[SEP]} \oplus x_i, \tag{1}$$

where $\oplus$ denotes concatenation. This method of structuring the sequence ensures that the model preserves the contextual relationships between the different parts for the interpretation of the discourse's dynamics (Zhang et al., 2019).

4.3.1.3 *Sequence length*

The total sequence length l is computed as:

$$l = \sum_{u=1}^{i} n_u + (i-1), \tag{2}$$

where:

- $\sum_{u=1}^{i} n_u$ represents the total number of tokens across all elements in the sub-branch.
- $(i-1)$ accounts for the separator tokens `[SEP]` inserted between elements of the sequence.

This step ensures that the sequence length accounts for both the content of the discourse and the structural markers (separator tokens `[SEP]`) during the encoding process of formulating the structure of discourse as context.

4.3.2 *Generation*

The second component generates synthetic discourse using structured prompts to an LLM. In this case, prompts are used to produce variations that augment underrepresented stances while preserving coherence with the original discourse. The resulting synthetic samples are combined with the original dataset D. Preliminary experiments with word-level data augmentation (see Appendix A.1) revealed that traditional methods help only on small datasets and require extensive task-specific tuning. Prompt-based approaches using LLM offer a scalable alternative. This is because LLM is capable of generating relevant synthetic data within a guided context by leveraging patterns learned from existing samples (Nadăş, Dioşan, and Tomescu, 2025).

4.3.2.1 *Minority samples identification*

Minority classes in the dataset D are first identified. The set of data points belonging to a minority class c_{min} is defined as:

$$S_{min} = \{T_i \in D \mid \text{label}(T_i) = c_{min}\} \tag{3}$$

The first step is to ensure that the augmentation only targets the underrepresented classes for balancing the samples.

4.3.2.2 *Synthetic root-level comment generation*

The goal of this step is to generate synthetic discourses that replicate the structure and dynamics of minority-class samples. For each sample $T_i \in S_{min}$, the `GENERATE-DISCOURSES` function (Algorithm 1) is applied to generate the augmented samples that reflect the existing samples. The function takes as input the sub-branch structure B, target topic t, stance label s, and the desired number of augmented samples num_aug. A structured prompt is constructed for the root-level comment:

$$P_p = \text{InitialPrompt}(B, t, s) \tag{4}$$

The language model M then produces the synthetic root-level comment taking into consideration the given contexts:

$$\text{synthetic_root} = M(P_p; \phi) \tag{5}$$

This synthetic root-level comment serves as the starting point for generating subsequent child-level comments. This generation ensures that this starting point anchors the existing dataset samples by maintaining their coherence using as much context from the existing dataset as possible.

4.3.2.3 *Subsequent synthetic child-level comments generation*

After generating the root-level comment, synthetic child-level comments are generated iteratively based on the preceding synthetic_root. Each comment is generated using a dynamically constructed prompt

P_k, which incorporates the target topic t, stance label s, and the most recent generated comment in the discourse:

$$P_k = \text{CommentPrompt}(t, s, \text{discourse}[-1]) \tag{6}$$

The language model M is then used to generate the k-th comment:

$$\text{synthetic_child} = M(P_k; \phi) \tag{7}$$

where:

- P_p and P_k are distinct, with P_p used only for the root-level comment and P_k used iteratively for child-level comments.
- discourse$[-1]$ represents the most recent comment (root or child) in the discourse for the coherence of these contexts.
- ϕ denotes the parameters of the language model M.

4.3.2.4 *Combining augmented outputs to reflect a discourse*

Once all comments (root and child) are generated, they are concatenated using `[SEP]` tokens to maintain the sequential nature of the discourse contexts (similar to the contextualise step on the original dataset). The detailed procedure for generating synthetic discourses is formalised in Algorithm 1.

The synthetic comments generated for each discourse are aggregated to form an extended branch context T':

$$T' = \text{synthetic_root}_i \text{ [SEP] synthetic_child}_1 \text{ [SEP]} \ldots \text{ [SEP] synthetic_child}_k \tag{8}$$

Informed by prompt engineering practices (Nadăş, Dioşan, and Tomescu, 2025), this procedure generates synthetic discourses that preserve the target topic t, stance s, and the hierarchical discourse structure B. The resulting augmented dataset manages class imbalance by upsampling relevant samples within guided contexts for the downstream stance detection in D. This approach aligned with the broader goal of encoding discourses that capture their discursive representations for model learning.

Algorithm 1 Using LLM prompt to generate synthetic discourses

```
function GENERATE-DISCOURSE(B, t, s_i, num_aug)
    inputs:
    B, a branch of discourse
    t, a target topic
    s, a stance
    num_aug, a number of augmented samples
    local variables:
    num_comments, a number ▷ calculated based on [SEP] tokens in B
    synthetic_discourses, an empty list
    for i ← 1 to num_aug do
        P_p ← InitialPrompt(t, s, B)
        synthetic_root ← M(P_p; ϕ)
        discourse ← [synthetic_root]                ▷ Initialise a discourse
        for k ← 1 to num_comments do
            P_k ← CommentPrompt(t, s, discourse[−1])
            synthetic_child ← M(P_k; ϕ)
            discourse.append(synthetic_child)
        end for
        discourse ← join discourse with [SEP] token
        synthetic_discourses.append(discourse)
    end for
    return synthetic_discourses
end function
```

4.3.3 *Learning*

In this component, a standard model training process is employed to capture the established contextual information for the stance detection model to learn its features. Following prior work (Li et al., 2023a), BERT embeddings were used for input representation, and TextCNN was used for feature extraction. These choices are motivated by their demonstrated effectiveness in modelling textual features and stance detection tasks outlined in Section 2.2.

4.3.3.1 *Input representation*

First, BERT (Devlin et al., 2019) is used to generate deep contextualised embeddings of the input sequences. Each input X is tokenised using the BERT tokeniser, producing embeddings E that combine token embeddings E_t, segment embeddings E_s, and positional embeddings E_p:

$$E = E_t + E_s + E_p = [e_{1,1}, \ldots, e_{1,n_1}, \ldots, e_{i,1}, \ldots, e_{i,n_i}] \tag{9}$$

where $e_{i,j}$ denotes the token embedding for $w_{i,j}$. The resulting BERT hidden states are represented as $H \in \mathbb{R}^{l \times h}$, where h is the hidden size and l is the sequence length. For each sub-unit x_i, the corresponding subsequence $[h_{i,1}, \ldots, h_{i,n_i}]$ serves as its representation. Zero-padding or truncation is applied to ensure uniform sequence length. The final representation is formulated as $H_{x_i} \in \mathbb{R}^{d \times h}$, where d is the fixed sequence length.

4.3.3.2 *Convolution feature extraction*

The TextCNN (Kim, 2014) was applied as a feature extractor to capture semantic patterns across phrases and token sequences. This complements BERT embeddings by preserving relevant information that may be distributed across multiple tokens. Convolution is achieved using a filter $w \in \mathbb{R}^{h \times k}$, which is applied to a window of k token vectors from H_{x_i} to produce a feature f_j:

$$f_j = \text{ReLU}(w \cdot H_{x_i,j} + b_1) \tag{10}$$

Applying this filter across the entire sequence generates a feature map $F = [f_1, f_2, \ldots, f_{n_i - k + 1}$. Then, a max-pooling layer is applied to extract the most significant feature:

$$\hat{f} = \max(F) \tag{11}$$

The goal here is to capture diverse local patterns within these text fractions using multiple filters with varying window sizes. The outputs from all filters are concatenated to form the final feature vector:

$z = [\hat{f}^1, \hat{f}^2, \ldots, \hat{f}^p]$, where p is the number of filters. This vector z serves as the feature representation for each input to be fed to the classification layer.

4.3.3.3 *Loss function*

The model takes further steps using a weighted loss function to balance learning across classes (Fernando and Tsokos, 2022). Each class was assigned a weight w_{y_i}, and the loss function was defined as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} w_{y_i} \log p(y_i|x_i; \theta) \tag{12}$$

This adjustment ensures that underrepresented classes contribute proportionally to the overall loss while preserving the natural distribution of the dataset.

4.3.3.4 *Classification*

A dropout layer was applied to the extracted feature vector z, producing $\hat{z}$ to reduce overfitting (Srivastava et al., 2014). The vector $\hat{z}$ was then passed through a fully connected layer followed by a softmax activation to produce the probability distribution p over stance labels:

$$p = \text{softmax}(W \cdot \hat{z} + b_2) \tag{13}$$

The predicted label corresponds to the highest probability:

$$\hat{y} = \arg\max(p) \tag{14}$$

This classification step ensures that the model outputs a clear prediction based on the learned features.

4.4 EXPERIMENTAL SETUP

4.4.1 *Datasets*

Our experiments were conducted on a benchmark dataset which provides the discourse structure for retrieving the discursive nature of

discourses on social media. The **COVID-19 Vaccine Discussions on Reddit** (Brambilla and Kharmale, 2022) dataset contains 1,727 Reddit root-level comments and 12,578 corresponding child-level comments in the CovidVaccine subreddit from April 2020 to May 2021. The dataset includes varying stances related to the COVID-19 vaccine public health discourses. The dataset is divided into 80% training, 16% test, and 4% validation sets. Given the size of the entire dataset, this split aligns with common practices of a standard machine learning 80/20 training-test ratio. Additionally, a smaller subset of the validation set within the test set is used for hyperparameter tuning and model evaluation. This dataset was selected because it exhibits key characteristics essential for our research objectives: (1) it reflects real-world social media discourses, for evaluating a stance detection model; (2) it represents a limited number of stance annotations to simulate data scarcity problems in practical applications; and (2) appropriate stance categorisation labels directly align with the selected task for a focused investigation.

4.4.2 *Baseline justification*

The selected baselines provided points of comparison aligned with our research questions. The *Base* model implements a standard procedure proposed in existing literature that incorporates contextualise and learning components in Section 4.3 to capture the discursive nature of social media discussions without generating any augmented samples. In earlier experiments, we evaluated the *Base* model with the addition of *Target topics* provided for each datapoint in the dataset to observe whether explicit topic information improves performance. To illustrate the evolution of augmentation techniques, we first included Easy Data Augmentation (EDA) (Wei and Zou, 2019). EDA represent a low-cost approach for generating augmented samples to demonstrate improvements achievable with traditional methods in our computational task.

To incorporate all components outlined in Section 4.3 and evaluate synthetic data generation from LLMs beyond traditional approaches, *GPT-2* (Kumar, Choudhary, and Cho, 2020; Papanikolaou and Pierleoni, 2020; Sawai, Paik, and Kuwana, 2021) was then included as an early exploration of lightweight LLM-based augmentation. This supports the earlier hypothesis by evaluating whether an LLM with a smaller and less resource-intensive language model can synthetically generate relevant discourse samples. This is because discourse-level generative augmentation had not previously been applied to the targeted task in this chapter during the formative stages of this idea. This allowed us to examine the practical benefits of lightweight LLM-based augmentation before moving to more sophisticated models. Finally, *Gemma-2* (Team et al., 2024) represents the more advanced LLM available at the time of the experiment (June 2024) for this study to adopt in generating synthetic data. Including both lightweight and more advanced LLM models in this study supported the goal of this chapter to demonstrate the feasibility and limitations of LLM-generated synthetic data for downstream stance detection tasks.

4.4.3 *Evaluation details*

A standard 5-fold cross-validation technique was used to evaluate the model's performance by partitioning the dataset into five non-overlapping subsets (see Section 3.2.1.3). This approach reduces the risk of biased results from a single train-test split when evaluating a model's performance (Kohavi, 1995). This will result in our reported performance that reflects generalisation across different portions within the dataset. Each fold maintains a balanced representation of classes to account for potential imbalances in stance labels. The performances were measured using F1 metrics to support the final synthesis of the findings (see Section 3.2.1.3).

4.4.4 *Implementation details*

The models were implemented and fine-tuned using the Hugging Face `transformers` library (Wolf et al., 2020). Training was performed with a batch size of 8 to accommodate the size of the dataset and available computational resources. Using a smaller batch size introduces beneficial variance into the gradient estimates during model learning, which can improve generalisation in limited-data settings (Qian and Klabjan, 2020). The Adam optimiser (Kingma and Ba, 2017) was used to adjust the learning rate based on the historical gradient for each parameter. A learning rate scheduler was used to improve convergence stability. Early stopping was applied to prevent overfitting by terminating training when validation performance plateaued. All experiments were conducted on an NVIDIA GPU. These implementation details are standard procedures following practices in NLP for fine-tuning models.

4.5 FINDINGS

This section presents the findings of our experiments, organised according to the three research questions introduced earlier in this chapter. It examines the impact of the proposed framework using LLMs to generate synthetic data that reflects the discursive structure of social media discourse. The detailed performance analysis provides key insights that inform the subsequent design-oriented Chapter 5 of this thesis.

4.5.1 *Macro-averaged F1 scores*

RQ1: Can existing discourse context approaches to stance detection be enhanced using data augmentation strategies under a data-limited scenario? Table 4 presents the macro-averaged F1 scores of 5-folds across the strategies outlined in Section 4.4.2. All proposed methods in the evo-

lution of the experiment improved the *Base* model (macro F1 = 0.643). First, adding the target context increased F1 to 0.658 (+ 1.5%). Meanwhile, incorporating data augmentation from traditional EDA and LLM-based augmentation (GPT-2 and Gemma-2) yielded larger gains. Gemma-2 achieved the highest performance (macro F1 = 0.696) by indicating a relative improvement over the Base with low variance across folds (±0.009). These indicate that not only can data augmentation strategies yield improvements in learning stances, but adopting advanced LLMs in applying the augmentation strategies performed even better under limited data conditions. In contrast, EDA and GPT-2 showed higher variability, suggesting less consistent outcomes despite moderate improvements. These observations are presented in Figure 8 where the width of each violin reflects the variability of scores, with the spread corresponding to the standard deviation across folds. The value of incorporating discourse context with LLM augmentation strategies relatively improved the computational tasks of this study.

These findings strengthened the point that computational learning tasks not only model their learning from individual texts but are shaped by how the surrounding child-level comments relate to them. Meaning that, the discursive nature of discourses forms a basis of contextual knowledge for a computational model by directly impacting how it learn. This establishes the need for models analysing social media discourse to treat this discourse nature as connected sequences rather than as disconnected inputs. It is therefore important to also articulate this practical implementation within this thesis. Doing so allows us to inform our interactive design through evidence drawn from the thesis's computational learning about the design decisions. To bridge the gap between algorithmic performance and real-world applicability, interactive designs should visualise the discourse structures learned by models during stance prediction. This approach highlights how contextual knowledge from surrounding discourses can improve model performance and support human understanding of the reasoning behind automated analyses. Considering this within

an interactive design could bridge between users and computational outputs by allowing them to interpret these automated results concurrently with human expertise.

APPROACHES	MACRO-AVERAGED F1
Base	0.643±0.005
+Target topics	0.658±0.016
+EDA	0.676±0.029
+GPT2	0.666±0.031
+Gemma2	**0.696±0.009**†

Table 4: Summary of macro-averaged F1 scores (mean ± standard deviation) across varying approaches evaluated on the dataset. Best performance is in bold and marked with †.

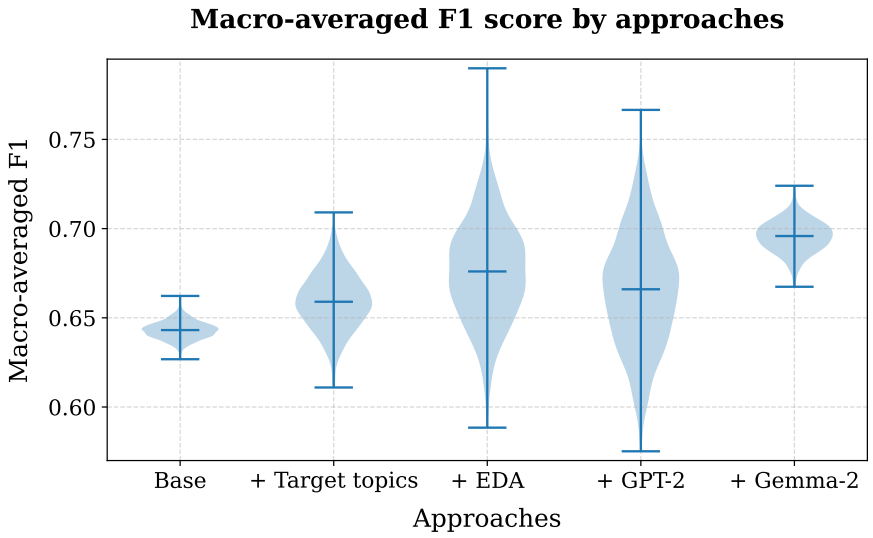


Figure 8: Violin plot of macro-averaged F1 scores for different approaches corresponding to the standard deviation across evaluation folds.

4.5.2 *Prompt-based augmentation*

RQ2: Can prompt-based generation of synthetic discourses outperform traditional data augmentation techniques? According to the same Table 4,

we compared different data augmentation strategies to assess their impact on stance detection performance. EDA slightly outperformed GPT-2 (macro F1 = 0.676 vs. macro F1 = 0.666) but showed higher variability across folds. This indicates that its performance is less consistent depending on the training subset. In contrast, Gemma-2 achieved the highest performance (macro F1 = 0.696) with very low variability (±0.009). This demonstrates its stability in performance across folds. All in all, these results suggest that advanced LLMs with prompt-based synthetic generation can produce informative training samples than rule-based or older LLMs (Nadăş, Dioşan, and Tomescu, 2025).

To translate these results into design, it is suggested that carefully calibrated prompts enable LLMs to generate synthetic discourses that reflect the structure of real-world discourse. This further supports the broader goal of our research to improve social media analysis by retrieving discourses as a structured context for computational learning. However, engineering these prompts required multiple rounds of iterative refinement to make sure that outputs maintained both correct label alignment and realistic tone and structure. Bauer, Preisig, and Volk (2024) arrived at a similar conclusion in their work around detecting offensiveness within texts. Meaning that, the observed improvements were shaped not only by model evaluation but also significantly by the design of the prompt itself. Other prior work also highlights the sensitivity of LLM outputs to subtle changes in prompt phrasing (Mizrahi et al., 2024).

4.5.3 *Impact based on the depth of discourses*

RQ3: What is the impact based on the depth of discourses on stance detection performance when integrating them as context? Table 5 presents the depth-specific macro-averaged F1 scores. The results indicate that the influence of discourse depth on model performance varies across models. At depth 3, Gemma-2-based augmentation achieved a significant F1 improvement of 0.691. This suggests that models perform

better when they have access to those contexts revolved around the discourse. However, performance gains level off or slightly drop at depth 3. This pattern suggests a trade-off where additional context can clarify stance in longer discourses, but it can also introduce irrelevant or conflicting information that reduces model accuracy. The distribution of performance across approaches and depths is shown in Figure 9 to highlight these trends. This confirmed prior work on hierarchical models noted that deeper context should be used carefully to avoid introducing noise (Young et al., 2018). Although computational learning can improve by incorporating patterns across these discourse depths, balancing the trade-off between exhaustivity and specificity requires significant computational resources. Without human expertise, achieving this balance remains challenging for automated approaches attempting to scale.

	DEPTH=1	DEPTH=2	DEPTH=3	DEPTH=4	DEPTH=5
Base	0.528±0.105	0.629±0.024	0.681±0.029	0.658±0.057	0.645±0.045
+Target	0.648±0.024	0.635±0.014	0.655±0.026	0.647±0.056	0.623±0.058
+EDA	0.648±0.036	0.663±0.043	**0.695±0.058**†	0.627±0.078	0.636±0.033
+GPT2	0.634±0.049	0.656±0.045	0.673±0.048	0.665±0.037	0.626±0.056
+Gemma2	**0.674±0.036**†	**0.678±0.020**†	0.691±0.023	**0.671±0.063**†	**0.663±0.038**†

Table 5: Depth-specific macro-averaged F1 scores (mean ± standard deviation) for each approach. Best performance in each column is in bold and marked with †.

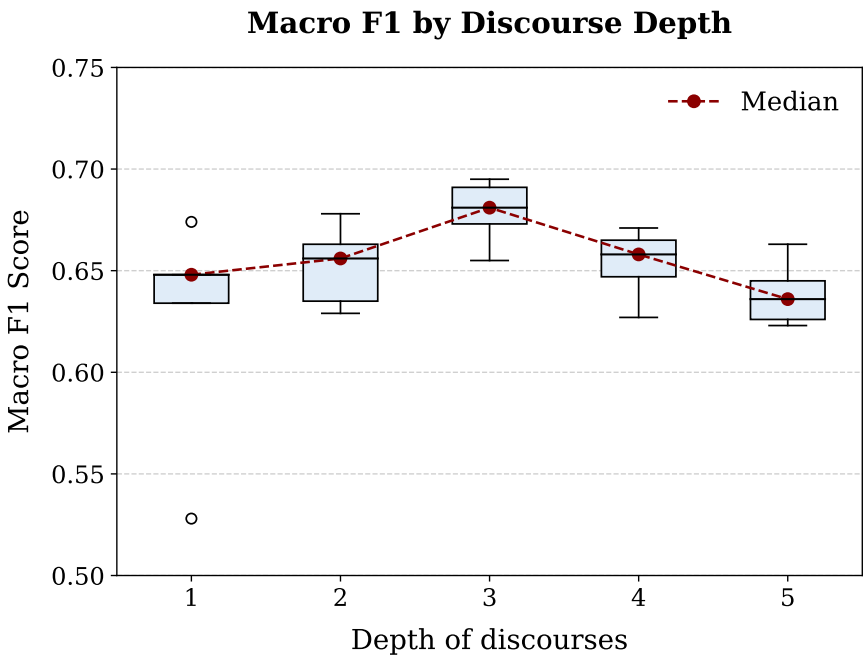


Figure 9: Boxplot illustrating the macro-averaged F1 scores across approaches for each discourse depth. The dashed line indicates the trend of the median.

4.5.4 *Limitations and future work*

The evaluation of LLM-generated synthetic discourses in this study relied on manual inspection of a subset of samples to verify label alignment and structural plausibility. While this approach provided preliminary confidence, it highlighted the challenge of scaling such evaluations to larger datasets. Future work should integrate automated quality assurance mechanisms, such as gating systems or layered filtering pipelines (Nadăş, Dioşan, and Tomescu, 2025) to manage risks of bias, harmful content, and structural inconsistencies in LLM-generated synthetic data. Besides, social media discourse is noisy, and this noise can propagate during augmentation with LLMs. Future work should prioritise adaptive augmentation procedures that tailor synthetic data generation to the characteristics of the dataset. For instance, numerical features such as discourse depth, number of conversational turns, or overall sample size could guide the genera-

tion process and help determine limits to prevent the introduction of irrelevant information. Finally, the focus on a domain-focused dataset that allows the experiment to build the discursive nature of social media discourses underscores the need for caution when generalising findings across other domains. While LLM-generated synthetic data holds promises, its reliability and applicability depend on careful validation and refinement. Future work should leverage more advanced models and refine their focus on quality and scope for engineering prompts when generating synthetic data to reflect social media discourses.

4.5.5 *Relating findings to the interactive design*

This chapter's computational learning exploration provides a foundation for understanding how computational methods can support social media analysis. Examining model capabilities positions the research within a practical perspective to inform how the subsequent design emerges from concrete experimental insights (see detailed table demonstrating the relation in Chapter 5). The overall impact of incorporating discourse context on model performance inspired an interactive design that enables users to explore discourses in depth and reveals the reasoning behind the model's decisions. Observing model limitations further emphasised the importance of human oversight in this analysis process. This indirectly reinforces interactive design principles and underscores the broader relevance of this thesis for creating an interactive design that supports analysis of social media. Overall, the experiment on stance detection, conducted to address the research questions, demonstrated the value of grounding interactive design in technical realities. The chapter has presented how this thesis bridges the gap between algorithmic capability and human-centred design by showing that effective social media analysis depends on integrating computational methods with a design focused on human interpretation.

4.6 CHAPTER SUMMARY

This chapter presents an empirical investigation into computational approaches for stance detection on social media by modelling social media discourses as context. The experiments centred around data augmentation methods and demonstrated that prompt-based augmentation using advanced LLMs, such as Gemma-2, improved model performance under data-limited conditions. Traditional augmentation methods, such as EDA and earlier LLM models like GPT-2, also improved performance, but the gains were smaller and less stable. The experiment demonstrates how the evolution of strategies to improve baseline: first by incorporating additional context from the target topic, followed by traditional data augmentation, then low-cost LLM-based augmentation, and finally higher-capacity LLM-based augmentation.

The goal of this experiment started with an investigation into the benefits of discourses when introduced to a computational learning model. The findings consequently revealed limitations of such automated output approaches and highlighted the impact of discourses as contextual information on a model's predictive performance. The main contribution of this chapter is that it provides technical and conceptual guidance for designing a social media analytical tool presented in the next Chapter 5. To briefly summarise, the results in this chapter suggest that interactive tools should: (1) present discourse context to support human interpretation of automated predictions, (2) make model limitations visible by enabling navigation across discourses reflecting the depths of real-time social media, and (3) encourage users to engage critically with outputs rather than treating them as fully reliable.

5

CREATING *DISCOURSEDASH*: AN INTERACTIVE PROTOTYPE DESIGN

This chapter introduces *DiscourseDash*, an interactive prototype inspired by the opportunities and limitations of a computational learning task for social media analysis identified in Chapter 4. While computational models demonstrated potential in leveraging discourse structures to improve stance detection analysis tasks, they also revealed important limitations. These limitations include the handling of linguistic complexity and variable performance that depends on the depth of conversational context. From a high-level perspective, these limitations emphasise the continuing importance of human expertise in interpreting computational outputs by situating them within broader discursive environments from which they are drawn.

DiscourseDash was designed as a research probe to form the foundation for facilitating the evaluation studies in Chapters 6 and 7. The prototype seeks to bridge computational analysis with human interpretation by combining automated predictions with discourse-level insights through an interactive design. What distinguishes the design is its deliberate shift away from treating automated analysis as a final output. Instead, *DiscourseDash* foregrounds interpretability, uncertainty, and contextual grounding for aligning with the broader aim of this thesis to support human expertise in understanding the complexities of social media discourses. The chapter begins by translating key findings from the exploratory computational study into design implications, then presents the system architecture, interface design, and implementation. By the end of this chapter it should be clear how users would hypothetically interact with the prototype integrating computational analysis with discourse-level insights.

5.1 INTRODUCTION

The exploratory stance detection study presented in Chapter 4 revealed both the potential and limitations of automated approaches to social media analysis using discourse for context learning. While the study demonstrated promising capabilities in performing computational learning by unfolding the discursive nature of social media discourses, it also presented the model's uncertainties when handling linguistic complexities, and performance varies according to the depth of discourses. These findings underscore the importance of human expertise in complementing the automated analysis of social media discourses.

DiscourseDash was developed in response to these challenges. The prototype is designed as an interactive dashboard that bridges computational analysis with human interpretation through visualisation and exploration. The aim is to support human capabilities by providing structured access to analytical outputs alongside the discourse contexts that shaped them. The benefit of this approach is that it acknowledges that experts require more than visual summarised metrics; it allows them to interrogate these outputs by investigating the discourse as a form of tracing the reasoning behind predictions. This will balance their own interpretation through understanding where computational methods are limited or incorrect. Table 6 summarises how the findings from Chapter 4 were translated into specific design implications for *DiscourseDash*. The mapping demonstrates how insights from computational experiments informed practical design choices of *DiscourseDash*. This goal here is to show how the *DiscourseDash* design positions in this research were inspired by the practical exploration of a computational learning study.

FINDINGS	DESIGN IMPLICATIONS
Retrieving discourse-level context provided a better representation during the learning of a computational task for social media analysis.	Visualise the same contextual information that informed the model's analytical output by introducing the same assessment for human users when using these automated analyses.
Performance improvements varied across different discourse depths, which indicates that optimal context windows are task-specific.	Allow users to navigate discourse depths, explore contextual framings and evaluate the reliability of automated analyses at different scales.
Complex linguistic phenomena and computational resource constraints limit the scope of automated analysis.	Prioritise interpretative flexibility, encouraging users to question, refine and reinterpret automated analyses instead of treating them as definitive answers.

Table 6: Translation of key findings from Chapter 4 into specific design implications for *DiscourseDash*.

The design of *DiscourseDash* is structured around three central implications that inform its goals. First, to integrate discourse context as a supplementary component to the presentation of analysis outputs. This will enable users to investigate the analysis in-depth within the social environments in which they were formed. Second, to be transparent around model limitations by allowing navigation across different levels of discursive depth. With this, users can interpret the analysis outputs with their own expertise while exploring how social media discourses evolve alongside the associated analytical labels. Third, to encourage critical engagement with automated analysis through interactive affordances. This can be achieved through interactive methods that prompt users to treat outputs as interpretive suggestions rather than definitive results.

5.2 DESIGN RATIONALE

Building on the design goals outlined earlier, the development of *DiscourseDash* is motivated by the findings of the computational learn-

ing experiments to provide structured insights into social media discourses. Drawing from the literature review in Chapter 2, four key analytical features, i. e., stance detection, sentiment analysis, harmfulness detection, and popularity analysis were all considered useful in a design for social media analysis. As a brief reminder, these features were commonly selected to support decision-making in areas such as public health, communication, and policy response, where understanding the dynamics of social media discourses is important (see Chapter 2). Based on the design directions established so far, a simple visual analytical tool for formulating the initial concept should provide a high-level summarised overview and a detailed analysis. This means presenting discourses in detail to explore how the broader narrative landscape shapes analytical insights. This is organised into two primary tabs: an *Overview* tab for the initial identification and prioritisation of key topics, and a *Detailed Analysis* tab for in-depth exploration of the specific discourses surrounding these topics.

5.2.1 *Providing a visual overview*

The *Overview* tab provides a high-level summary of trending discourse topics and their key characteristics. Its components include:

- **Trending Topics** This should reflect clustering techniques to identify the most actively discussed topics within a selected time window. This surface-level aggregation enables users to quickly understand which issues are dominating public conversation and assess their alignment or conflict with organisational goals.
- **Harmfulness** Each topic from the *Trending Topics* is assessed for its potential level of harm, based on the proportion of posts containing toxic, hateful, or otherwise harmful language. This classification helps users allocate priorities by identifying those that may require immediate attention or intervention.

- **Stance** This displays the distribution of stances (favour, neutral, against) within each topic. We hypothesise that understanding the stance distribution on topics previously categorised as harmful can help gauge the urgency of a topic. For instance, if a highly harmful topic receives predominantly favourable stances, it can inform the prioritisation of further investigation.

- **Sentiment** Sentiment (positive, neutral, negative) provides an additional emotional context to each topic. This helps users understand public emotions and make informed decisions about countermeasures.

- **Popularity** Engagement metrics over time reveal the relevance and impact of topics. Visual graphs showing the popularity of each topic help users understand the dynamics of discourses and identify emerging trends or sudden spikes that may indicate concerning impacts.

5.2.2 *Further detail of the analysis*

Building on the *Overview*, users can then select specific topics to investigate in detail. This is to facilitate the exploration of in-depth analyses through the following features:

- **Summary** An in-depth view of the selected topic, including visual graphs of stances and sentiments. This provides clarity and supports users in comprehending the progression and influence of different viewpoints within the discourse.

- **Discourses** This presents all the discourses on a selected social media platform that reflect the topic under investigation. The goal here is to support effective countermeasures by providing a detailed view of the stances and sentiments labelled within each text in the discourse. This also facilitates the extraction of specific contexts and dynamics revealed through the analysis.

Presenting these detailed discourses alongside visuals and interactive elements aligns with the established goals.

5.3 INTERFACE DESIGN

According to the simple tabs described in the previous section, the core dashboard is divided into two main pages: (1) the *Overview* Page and (2) the *Detailed Analysis* Page. The final prototype used for this research is publicly accessible here: *DiscourseDash*.

5.3.1 *Overview Page*

The *Overview* Page provides users with a high-level summary of trending discourse topics and their key analytical insights. The page consists of the following interactive components:

A. USER INPUTS

Users specify a date range and optionally the number of topics to display. The dashboard then updates components B through E accordingly.

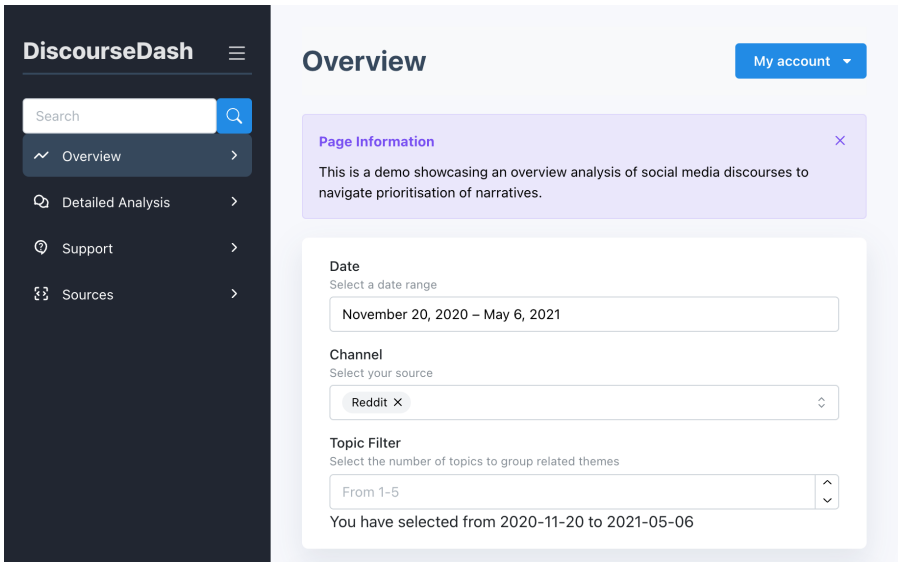


Figure 10: *Overview* page showing (A) the *User Inputs* component located at the top of the page.

B. TRENDING TOPICS

Displayed in a table summarising the results of the analysed discourses. Each *topic* includes a unique identifier (Topic_ID), the topic label (Topics), counts of root-level comments (Total volume), the number of associated child-level comments (Total engagement), and a harmfulness label (Harmfulness), categorised as low, medium, or high.

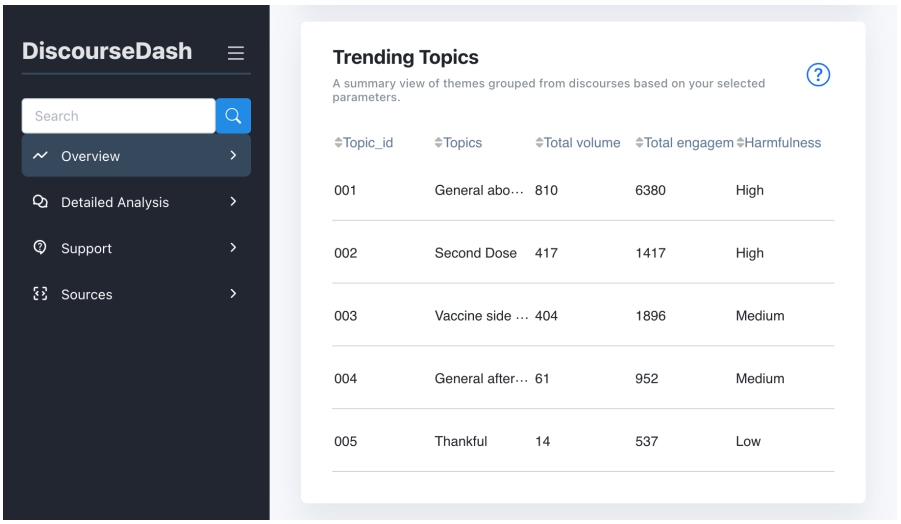


Figure 11: *Overview* page showing (B) the *Trending Topics* table directly below the *User Inputs* component.

C. STANCE DISTRIBUTION BY TOPIC

Displayed as a bar chart, this component shows the distribution of stances (favour, none, against) across all discourses within each topic from B. The y-axis lists topics by Topic_ID, while the x-axis shows the aggregated count of stance labels from both root-level and child-level comments.

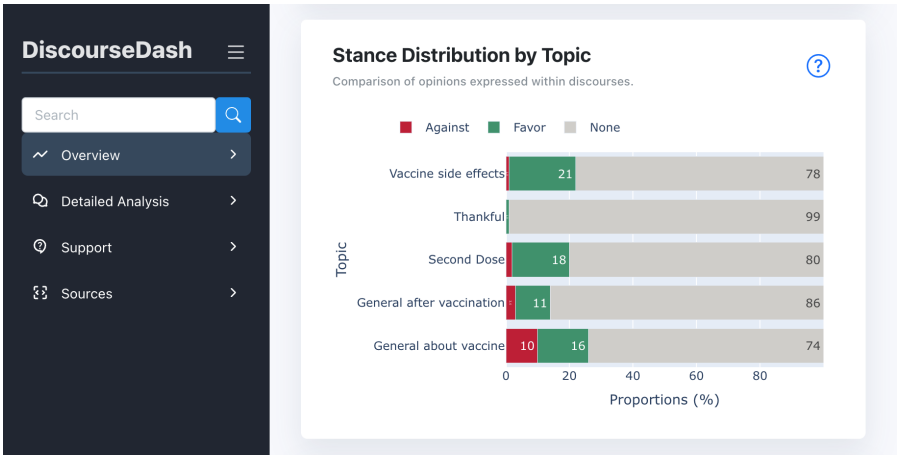


Figure 12: *Overview* page showing (C) *Stance Distribution by Topic* charts directly below the *Trending Topics* table.

D. SENTIMENT DISTRIBUTION BY TOPIC

This bar chart presents the distribution of sentiments (positive, neutral, negative) associated with each topic from B. The y-axis lists topics by Topic_ID, while the x-axis shows the aggregated counts of sentiment labels from both root-level and child-level comments.

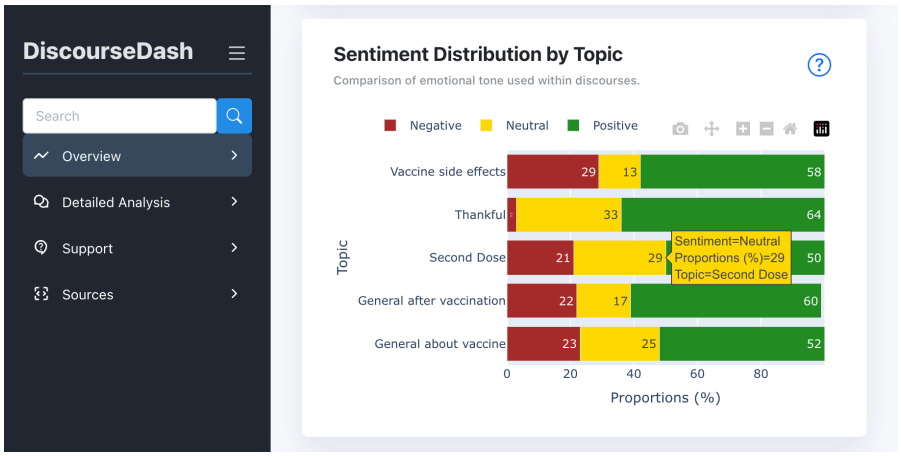


Figure 13: *Overview* page showing (D) *Sentiment Distribution by Topic* charts directly below the *Stance Distribution by Topic* chart.

E. POPULARITY OVER TIME

This line graph visualises user engagement dynamics, showing trends and peaks in public interest for each topic from B, with date shown on the x-axis as Date, and number of engagements on the y-axis as Popularity.

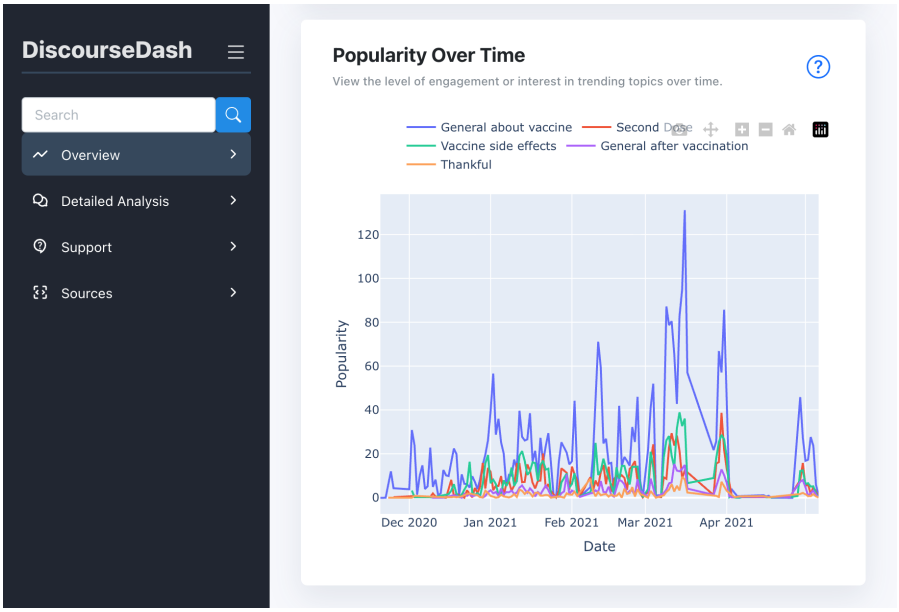


Figure 14: *Overview* page showing (E) the *Popularity Over Time* graph directly at the bottom of the page as the final component.

5.3.2 *Detailed Analysis Page*

This page presents the detailed discourses corresponding to the topics of the previous *Overview* Page. It allows users to dive deeper into specific topics they wish to investigate around targeted issues based on actual discussions.

A. USER INPUTS

Users specify a date range and select a targeted topic(s) for detailed investigation. The dashboard then updates components B and C accordingly.

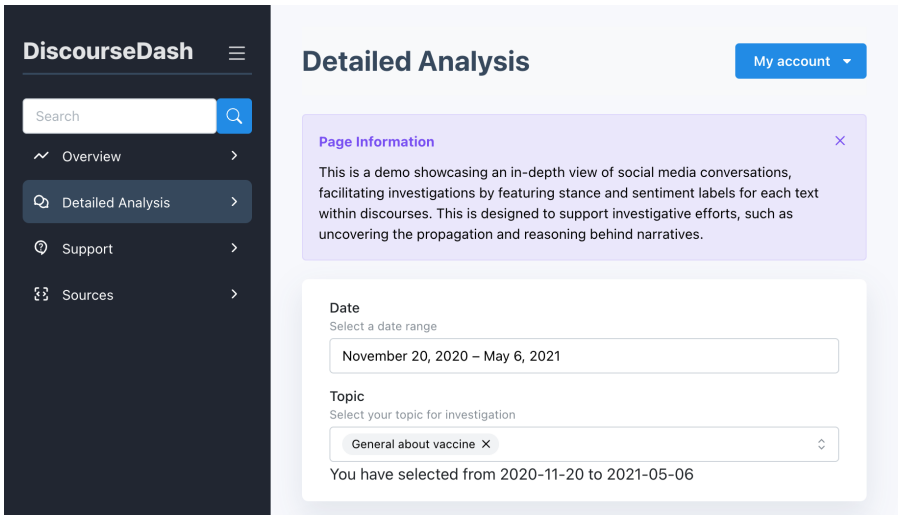


Figure 15: *Detailed Analysis* Page showing (A) the *User Inputs* component located at the top of the page.

B. SUMMARY

This component provides a summary of a discourse selected by the user to investigate. By topic, a bar chart on the left illustrates the distribution of stances (favour, none, against) across the discourse. Similarly, another bar chart on the right presents the sentiment distribution (positive, neutral, negative) across the discourse.

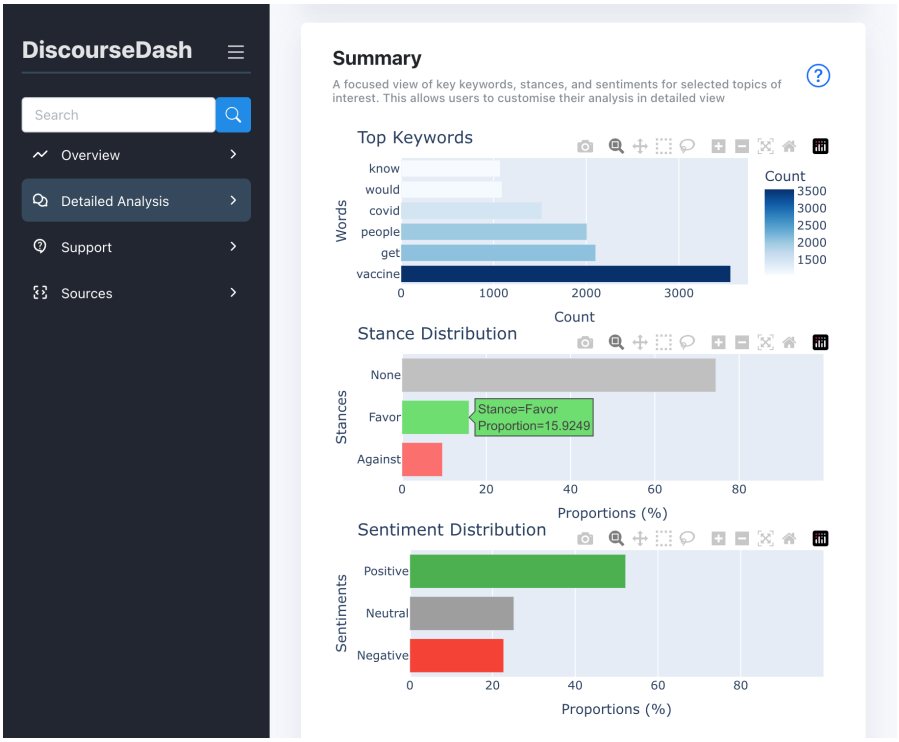


Figure 16: *Detailed Analysis* Page showing (B) the *Summary* combining *Top Keywords*, *Stance Distribution* and *Sentiment Distribution* for the selected discourse directly below the *User Inputs* component.

C. DISCOURSES

Displayed in a collapsible table format, allowing users to expand each discourse to view its associated child-level comments. Each row, representing root/child level comments, is accompanied by labels of the corresponding stance (favour, none, against) and sentiment (positive, neutral, negative). This is designed to facilitate a more detailed analysis within the broader context of the selected topics. Additional functionalities include filters that allow users to refine results based on specific stances, sentiments, and keyword searches.

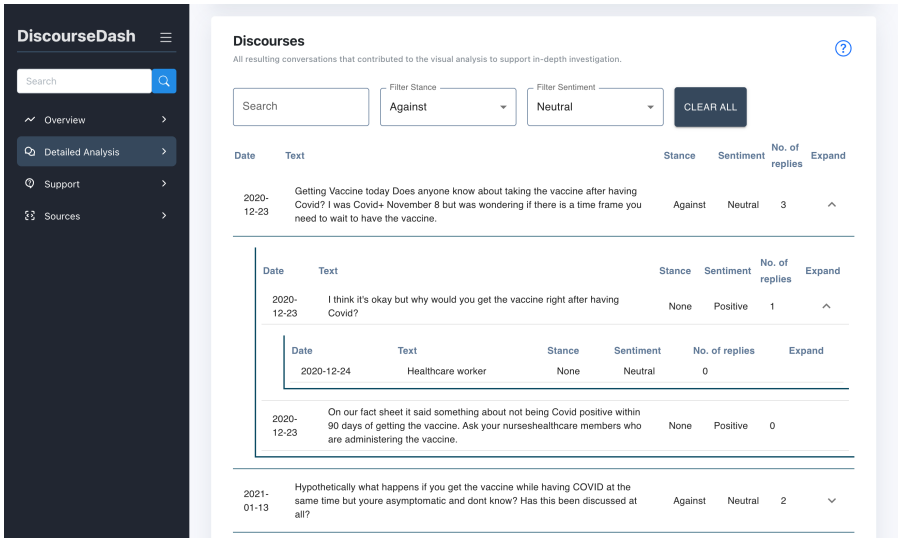


Figure 17: *Detailed Analysis* Page showing (C) the *Discourses* in collapsible tables to present each comment with stance and sentiment annotations and filter functionalities directly at the bottom of the page as the final component.

5.4 SYSTEM ARCHITECTURE

This section describes the architecture of the *DiscourseDash* prototype, including its functional requirements, data management, visualisation, and interaction components. The architecture was formulated to make sure that it allows interactive user studies to reflect the design concept outlined.

5.4.1 *Basic requirements*

Since the prototype for *DiscourseDash* was developed to support user studies, its functional requirements were defined to communicate the core components and basic interactions sufficient for gathering user insights. The system must be implemented as a web-based application with the following pages:

- ***Overview* Page:** Provides a high-level summary of the resulting analysis of social media discourses.

- ***Detailed Analysis* Page:** Displays the discourses that contribute to the visual summaries on the *Overview* Page.
- **Support & Sources Page:** Contains basic forms and details about the open-access datasets used in the prototype to simulate a realistic dashboard environment. These pages will not be further detailed in this Chapter as they were implemented with no functions for the user studies.

The system must present analysis results based on the following high-level system requirements:

- Visualisations in multiple predefined formats, including tables, bar charts, and line graphs.
- User inputs for filtering data by date range, social media channels (for demonstration), and a number for the topic filter to display.
- Dynamic update of visualisations and tables based on user selections.
- Ability to explore and inspect the displayed data with interactive details.

5.4.2 *Database*

The system uses Supabase, an open-source backend platform for storing and managing the data that will be displayed in the prototype. This was selected as it provides a relational database that supports structured storage and updates. Key functionalities it facilitates include:

- Storing datasets with text and hierarchical identifiers in separate fields for retrieving root and child-level comments.
- Allows developers to add, edit, or correct data for maintenance around accuracy and scalability.

- Supporting queries to filter and aggregate data for the visualisations.

5.4.3 *Visualisation*

The visualisation component presents social media discourse data clearly and intuitively. Key requirements include:

- Displaying different types of analysis in separate visual components for clarity, including stance, sentiment, harmfulness, popularity, and discourses.
- Using visual distinctions such as colours or labels to differentiate categories.
- Supporting multiple formats, including tables, bar charts, and line graphs.
- Enabling users to identify trends, patterns, and key insights efficiently.

5.4.4 *Interaction*

The interaction component allows users to engage with the data and visualisations effectively. Key requirements include:

- User inputs parameters must reflect an accurate display of analyses.
- Navigation links on the hosted web pages should function in line with the proposed design.
- Expanding and collapsing discourse discursive nature to unfold each related child-level comment.
- Hovering over visual elements to display detailed information.

5.5 USER FLOW DIAGRAM

Figure 18 illustrates the intended interactive flow for *DiscourseDash* by showing the sequence of interactions between a user and the prototype across its key components.

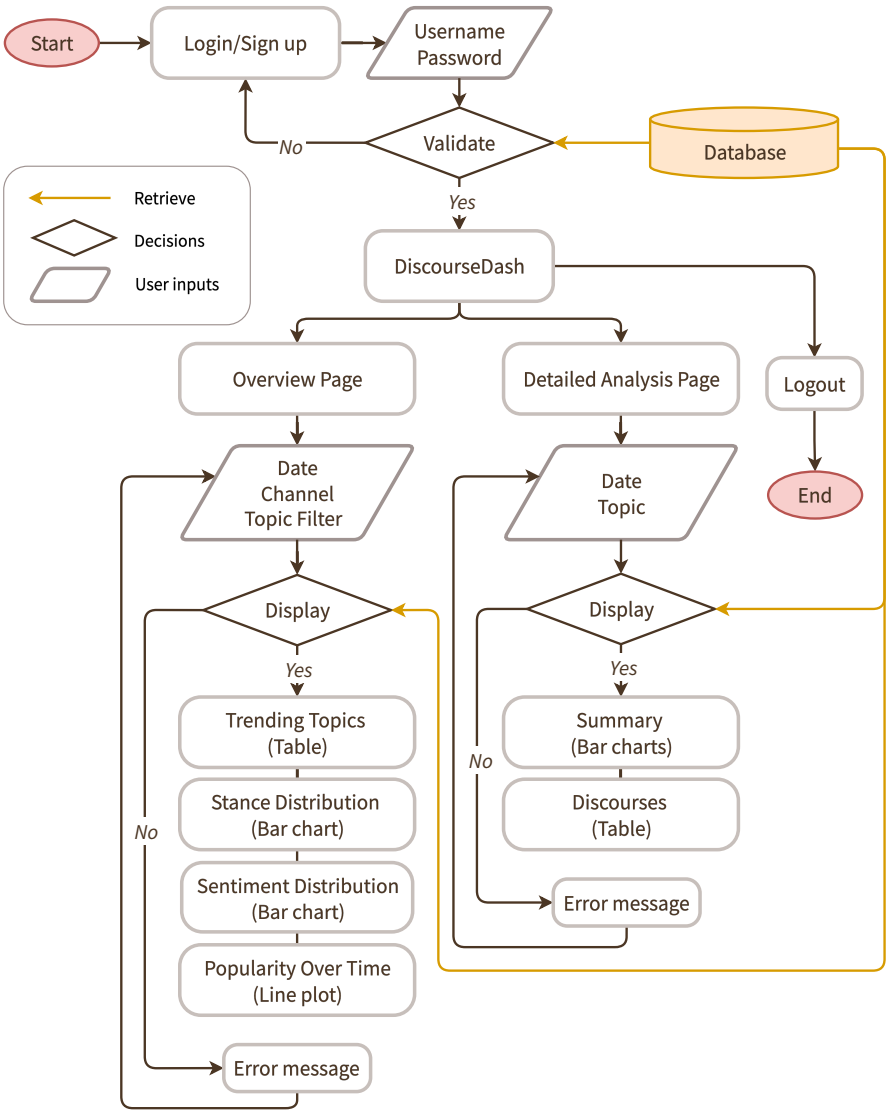


Figure 18: A user flow diagram for *DiscourseDash*.

5.6 SYSTEM OVERVIEW

Figure 19 illustrates the client-server architecture of the developed web application prototype following a Model-View-Controller (MVC)

architecture. For accessibility and ease of deployment, *DiscourseDash* was hosted on Heroku, a cloud-based platform-as-a-service (PaaS). This facilitated a web-based access to the prototype for convenience throughout the user evaluation.

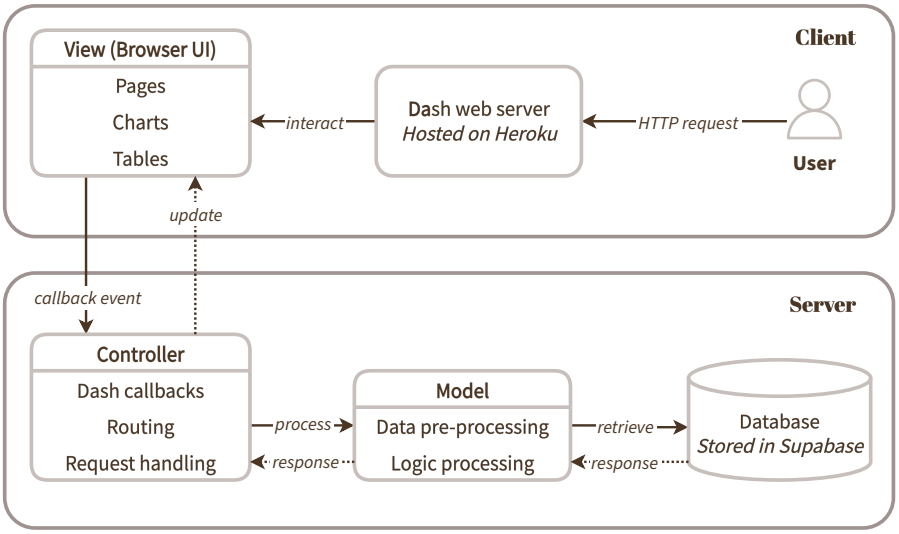


Figure 19: A system overview of *DiscourseDash* presented based on a Model-View-Controller (MVC) architecture.

5.6.1 *Client*

The client-side begins with the **User**, who initiates *interaction* by sending an *HTTP request* to the **Dash web server**. The **Dash web server** acts as the entry point for handling the initial request to load the application as well as subsequent events triggered during interaction. Then, the **View** represents the graphical user interface displayed in the user's web browser. This includes dynamic pages, charts and tables that reflect the processed data (rendered via Plotly). All user interactions, such as input changes within the **View**, generate a *callback event* which is then passed to the server for processing. These callbacks ensure that the interface dynamically updates in response to user interactions.

5.6.2 *Server*

The server side manages the application logic, data processing, and connection with the database. The process begins with the **Controller** as the coordinator of server-side logic. It receives *callback events* from the client's side for request handling, routing, and executing Dash callbacks. Based on user interactions, the Controller issues *process* commands to the **Model** by specifying the computations or data retrieval required. Then, the **Model** captures the core logic and data management of the system. It handles data pre-processing, executes predefined algorithmic logic, and communicates with the database. When the **Model** receives a *process* command from the Controller, it can issue a *retrieve* request to the **Database**. The **Model** interacts directly with the Database, which responds with the requested data upon successful query execution. Once the Controller and **Model** complete their processing, the **Dash web server** transmits the corresponding *update* instructions back to the **View**.

5.7 PROTOTYPE IMPLEMENTATION

The interactive *DiscourseDash* was developed using the Dash Open Source Framework by Plotly. This was chosen for its robust integration with Python libraries and its flexible integration with machine learning frameworks in future iterations. This framework allowed us to customise functionalities to meet our research requirements. For instance, we implemented a custom collapsible table using React boilerplate JavaScript UI libraries to represent the hierarchical structure of discourses. To simulate a realistic scenario within the prototype, we used a dataset of open-access *Reddit* discourse on COVID-19 vaccination for the evaluation studies (Brambilla and Kharmale, 2022). As of its retrieval in October 2023, this dataset consisted of 1,726 *Reddit* posts and 11,189 corresponding replies, posted between April 2020 and May 2021. Each entry was labelled with a stance (favour,

none, against) towards the topics of "General about vaccine", "Second Dose", "Vaccine side effects", "General after vaccination", "Thankful." Additionally, they provided labels of sentiments (positive, neutral, negative). We included all entries containing one or more replies to ensure every data point represents a discourse structure. The labels for stance and sentiment were directly taken from the dataset. Since they did not include labels for harmfulness, we manually and randomly assigned harmfulness values in the prototype. This means that harmfulness labels were distributed without any specific pattern or criteria solely to display the feature within the *DiscourseDash*. This approach is commonly seen in HCI research, where placeholder or synthetic data is employed when the focus is on evaluating the conceptual design prototype Pollack et al., 2019. Each row in the dataset includes timestamps and the total number of child-level comments for each row belonging to a root-level comment. These columns are used to illustrate the popularity of topics in the prototype.

5.8 CHAPTER SUMMARY

This chapter presents an artefact through the development of *DiscourseDash*. *DiscourseDash* serves both as a research probe and a functional prototype to evaluate how users interact with it during decision-making in the analysis of social media discourses. The chapter details the process of translating the design goals into a working prototype. It begins by establishing the design rationale, grounding development in the considerations and practical insights obtained from Chapter 4. The interface design is presented to illustrate the core components and their intended functionality. The chapter then introduces the system architecture and defines the structures supporting prototype development. A user flow diagram demonstrates how interactions unfold from a user's perspective, and a system overview connects these elements into a unified framework. Finally, the implementation of the prototype as a functional artefact is described. These

sections inform the conceptual and technical foundations of the prototype to prepare an interactive design for user evaluations. In the next Chapter 6, the thesis will build on this by examining user interactions with the prototype and generating insights into its effectiveness in supporting decision-making workflows.

6

A MIXED-METHOD STUDY WITH GENERAL USERS

This chapter presents the findings of a user study investigating how general users (non-experts) interact with *DiscourseDash* in a decision-making scenario. The aim is to explore the potential of the design in supporting the analysis of social media discourses with general users. We conducted a task-based evaluation using (1) an interactive prototype, (2) a realistic decision-making scenario, (3) structured tasks reflecting real-world practices, and (4) guiding sub-questions.

The insights from Chapter 4 highlighted the opportunities and limitations of a computational learning task. These insights inspired the need to extend the visual analytical output towards interpretable discourse-level insights. Building on this, Chapter 5 described the design of *DiscourseDash*, which integrates multiple analytical outputs into an interactive interface with an emphasis on unfolding the discursive nature of social media discourses. The design aimed to support decision-making workflow by allowing navigation towards interpretive insights from discourses.

This study examines how individuals without prior expertise in social media analysis perceive the conceptual flow of *DiscourseDash* while completing a set of designed tasks. A participatory approach was adopted to evaluate the design, gather evidence from user interaction experiences, and identify areas for improvement. Given a decision-making scenario, participants assessed whether *DiscourseDash*'s analytical features supported them in interpreting complex social media discourses compared to navigating an unstructured social media platform such as Reddit. This chapter presents empirical evidence on how discourse-level context supports decision-making for general users.

6.1 INTRODUCTION

Existing social media analytics tools typically focus on quantitative indicators such as sentiment, opinions, or engagement metrics (see Chapter 2). These are frequently presented individually or in combination through visual summaries. While these measures provide useful information, they fail to capture the underlying discourse that gives meaning to these metrics. This creates a significant challenge because decision-making support remains incomplete when high-level analyses cannot be interpreted with human input. This limitation is relevant in situations where a timely understanding of how and when discourses unfold is essential for informed decision-making (see Section 2.1.2). Interpreting discourse directly from unstructured platforms can be inefficient and cognitively demanding as users are required to extract relevant insights from these large volumes of information.

The design of *DiscourseDash* aimed to address these challenges by combining computational outputs with discourse-level insights presented through an interactive design. Chapters 4 and 5 outlined the opportunities and limitations of computational methods and the rationale behind the design of *DiscourseDash*. The interactive design differentiates itself by emphasising discourse-level insights through linking visual summarised outputs back to the underlying discourse. This chapter progresses to evaluating the design by examining how these ideas work in practice when applied to user interactions.

To investigate this, we conducted a mixed-method study with general users (non-experts) through a task-based evaluation approach. Participants were presented with a realistic decision-making scenario and asked to complete structured tasks using *DiscourseDash* as well as an unstructured social media platform for comparison. This design investigated how users perceive the design in contrast to manually navigating social media discourses. Through non-expert users, the study examines how accessible and useful discourse-level analytics are to users without prior experience in social media analysis.

The purpose of this evaluation is therefore twofold: first, to assess the efficiency and relevance of *DiscourseDash* in supporting user interpretation of complex social media discourses; and second, to generate empirical evidence about the potential and limitations of analysing social media through discourse-level insights in practice. The study adopts a participatory approach, combining qualitative insights from user experiences with quantitative measures of task performance. The study was guided by the following research questions:

RQ1 *How do users perceive* DiscourseDash *in analysing social media discourses in the context of making informed decisions?* This question explores user interactions and behaviours to assess the effectiveness of *DiscourseDash* in supporting decision-making.

RQ2 *To what extent do the proposed features in* DiscourseDash*, such as stance analysis, sentiment analysis, harmfulness labels, Popularity Over Time, influence or support decision-making workflows?* This involves evaluating the usefulness of individual features in supporting tasks.

RQ3 *How can the design of* DiscourseDash *be improved and refined through a user-centred approach?* This question identifies considerations for enhancing usability and the overall user experience in future iterations.

These questions are structured for the chapter that aims to demonstrate the value of *DiscourseDash* in practice and guide future designs of discourse-level social media analytics tools.

6.2 STUDY DESIGN

The study employed a within-subjects design where participants completed three predefined tasks using two experimental conditions in counterbalanced order: (1) *Reddit* as a baseline, and (2) *DiscourseDash* as our proposed interactive design. Each participant completed three predefined tasks using each condition. After completing each task, we

interviewed each participant using a semi-structured format to gather qualitative data. At the end of each condition, participants completed a questionnaire to quantify their experiences and perceptions. The sequence was then repeated for the second condition. Following interactions with both conditions, participants took part in a final semi-structured interview. This concluding interview allowed participants to compare the effectiveness of *DiscourseDash* with Reddit in performing the same tasks and to suggest potential improvements. This approach allows this research to assess the *DiscourseDash*'s effectiveness in supporting decision-making, the usefulness of its features, and opportunities for user-centred enhancements.

6.2.1 *Baseline*

Baseline selection should reflect realistic and representative user practices (Purchase, 2012). Manual exploration of unstructured social media simulates a journalist's workflow (Hellmueller et al., 2024; Humayun and Ferrucci, 2022). This makes an unstructured platform chosen as an experimental condition a suitable setting. For this, Reddit was selected as the baseline due to its well-established open discourse structure, as it presents a large amount of threaded comments, subreddits, upvotes, and downvotes (Hollender et al., 2010). Compared to *Twitter/X* or *Facebook*, *Reddit* supports longer discussions, clear hierarchical threads, and topic-focused communities, making it appropriate for comparison in this study. Although Reddit may not capture all interaction types present on other platforms, its open, community-driven nature provides a reasonable representation of unstructured social media. For the clarity of this study, Reddit *posts* are treated as *root-level comments*, and Reddit *replies* as *child-level comments*, in which all link back to the initial established definitions for this thesis (see Section 2.1.1.1).

6.2.2 *Recruitment and participants*

The study recruited participants from university research groups, including professors, researchers, and students. Further screenings were not applied to filter the participants (see Appendix A.4.1). The study assigned participants to interact with the conditions in a counterbalanced order. To respect participant privacy and because demographic details were not relevant to the study's aims, the study did not collect detailed demographic data. This approach aligns with prior evidence suggesting that participants may feel uncomfortable sharing such information (Frederick, 2021). The study used participant codes (e. g., G1) as anonymised identifiers, where "G" denotes a general user. A complete list of participant codes, their assigned counterbalanced condition orders, interview dates, and anonymised excerpts from the transcripts is available in the Nottingham Research Data Management Repository.

6.2.3 *Procedure*

The study provided participants with an initial briefing on the research topic and the study's objectives, which focused on evaluating the effectiveness of an interactive design in analysing social media discourses. To minimise bias, the study did not disclose the comparison between conditions. Each participant completed one of two conditions first, followed by the other, in a counterbalanced approach: (1) baseline Reddit or (2) *DiscourseDash*. Participants began by familiarising themselves with the assigned condition. Drawing from virtual experience methodologies, a realistic narrative was provided to enhance ecological validity (Gorini et al., 2011). Participants assumed the role of a fact checker tasked with identifying newsworthy topics and crafting impactful articles, inspired by journalistic practices (Hellmueller et al., 2024; Humayun and Ferrucci, 2022). This scenario helped link the tool's metrics with its intended impact (Lam et al., 2011). Then,

participants completed three separate, predefined tasks designed to reflect these realistic workflows. The tasks are detailed as follows.

- **Task 1:** Topic to prioritise for investigation. Participants needed to select a discussion topic to prioritise based on their interaction with the assigned condition. The objective was to select a topic that stands out based on its potential impact on current trends. The participants had to justify why the selected topic was more significant than the others.
- **Task 2:** Extract some key points of stance arguments based on the topic evident in the discourse. Participants were asked to investigate the discourse on the chosen topic from Task 1 to extract the reasoning behind various stances. For instance, they had to identify supporting arguments for the "favour" stance related to the topic.
- **Task 3:** Describe the overall sentiment present in the discourse. Similarly, participants were asked to describe the emotions reflected in the discourse surrounding the topic they were investigating in Task 2.

Throughout the tasks, participants used a think-aloud protocol to articulate their thought processes. After completing each task on the first assigned condition, participants were guided in a semi-structured interview and completed a questionnaire to evaluate their experience. The same procedure was then repeated with the second condition. The study concluded with a final semi-structured interview in which participants compared the two given conditions, identified their preferences, and suggested improvements. Each participant received a £10 Amazon voucher in recognition of approximately 60 minutes of participation. Figure 20 illustrates the study conditions and procedure.

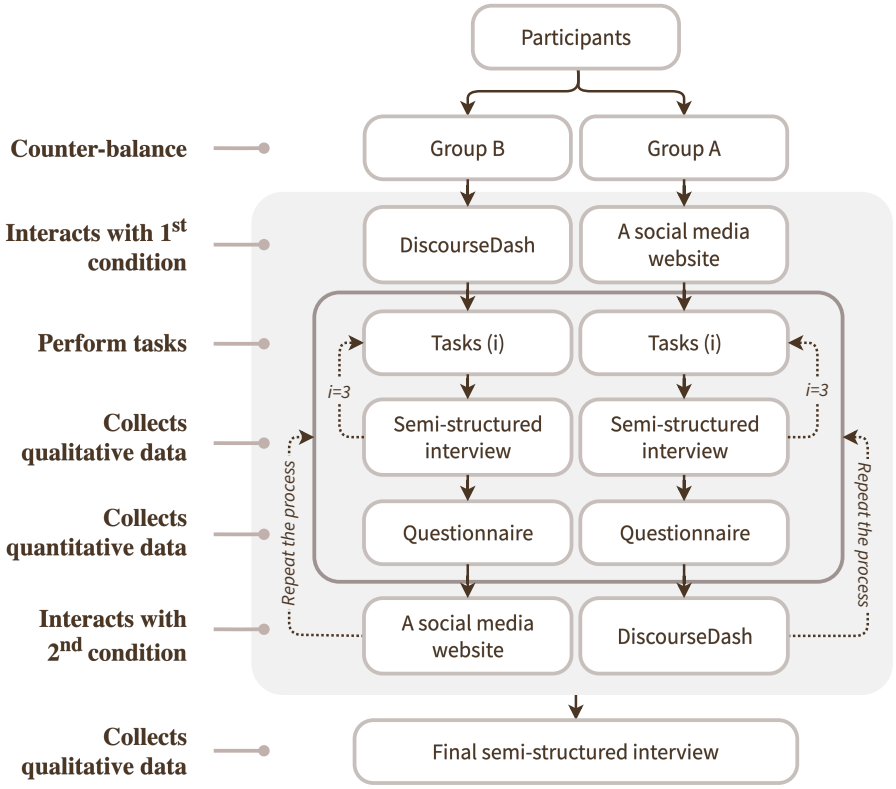


Figure 20: Overview of the study procedure involving general users.

6.2.4 *Data collection*

This study employed a mixed-method approach (Creswell, 2021) to collect both quantitative and qualitative data. Quantitative data captured measurable aspects of participants' interactions and perceptions, while qualitative data provided insight into their reasoning. Microsoft Teams automatically transcribes all audio recordings, and the transcripts were manually reviewed for accuracy. The transcripts were then organised by interview questions in Microsoft Excel, resulting in the dataset presented in the Nottingham Research Data Management Repository. This allowed a structured mapping of participant responses for a systematic and rigorous analysis. The study was conducted in English via Microsoft Teams between June and July 2024.

6.2.4.1 *Interview*

We conducted semi-structured interviews to collect qualitative data. The first round of interviews took place after each task for both conditions. To gain insights into participants' interactions while completing the tasks, we asked the following sub-questions (SQ):

SQ1 Describe the process you followed to complete the task.

SQ2 What are some key indicators from the provided condition that guided your decision-making process

SQ3 Are there any difficulties or challenges encountered while completing the task? If so, please describe them

SQ4 How intuitive do you think the provided condition guided you through the task?

Another semi-structured interview will be conducted after participants interact with both conditions. These questions will focus on comparing the conditions and gathering insights to identify improvements for *DiscourseDash*. These questions include:

Q1 Which condition did they prefer to complete the task and why?

Q2 What improvements would they suggest for *DiscourseDash*?

Q3 Are there any additional functionalities that they think would be useful to have implemented in *DiscourseDash*?

6.2.4.2 *Questionnaire*

Participants completed a questionnaire using a Likert scale. The first question, asked for both conditions, was: *"How difficult or easy was it to identify a topic to prioritise using the provided condition?"* This was followed by a set of additional Likert-scale questions specific to *DiscourseDash*:

1. The harmfulness label was a useful indication.

2. The stance analysis helped me understand the importance of each topic.
3. The sentiment analysis helped me understand the importance of each topic.
4. The popularity analysis helped me prioritise one topic over another.
5. Investigating discourses with labels helped me identify key factors for different stances and sentiments on a specific topic.

In addition, the System Usability Scale (SUS) (Brooke et al., 1996) provided a quantitative measure of the usability of each condition (see Appendix A.2).

6.2.5 *Analysis*

6.2.5.1 *Interview*

First, the interview transcripts were summarised and analysed to understand decision-making processes, the key indicators they found useful, the challenges encountered, and the overall experiences with each condition. Then, a thematic analysis (Braun and Clarke, 2012) was conducted to systematically identify, organise, and interpret patterns of meaning within the qualitative data. This method was selected for its ability to capture themes from participant narratives to understand user interactive experiences. Initially, the analysis was performed by summarising interview responses and performing an initial open coding to detect emerging patterns. These codes were iteratively refined through regular discussions with the supervisory team to improve consistency and alignment with the study's research questions and goals. Illustrative participant quotes are included to contextualise and substantiate the findings.

6.2.5.2 *Questionnaire*

Quantitative data from the questionnaire were collected after each interaction with a condition. These were then analysed using statistical methods to evaluate topic prioritisation, feature usefulness, and overall usability. To compare ratings between paired conditions, we used the Wilcoxon Signed-Rank Test. This non-parametric test is suitable for comparing two related samples as it effectively handles non-normally distributed data. In addition to that, a Friedman Test was used to assess perceptions across multiple features. This test is appropriate for ordinal data and allows us to compare more than two related samples. These allow the measurements of the relative performance of *DiscourseDash* compared to the baseline condition.

6.3 FINDINGS

This section presents both quantitative and qualitative analyses of our study. The quantitative results (Section 6.3.1) examine questionnaire ratings between conditions using statistical tests. The qualitative results (Section 6.3.2) summarise participant responses from the semi-structured interviews for comparing their experiences and challenges with both conditions for each task. Then, thematic analysis identifies and compares common themes within conditions in this study to provide insights into user perceptions and suggestions for improvement.

6.3.1 *Quantitative results*

The following results are based on all responses collected from the Likert scale questionnaire after each interaction with a condition. The Wilcoxon Signed-Rank Test was chosen to compare ratings between two paired conditions by each participant, as it is suitable for non-parametric, non-normally distributed data. We also selected the Friedman Test to compare perceptions across multiple features because it can handle more than two related samples and is appropriate for or-

dinal data. All detailed questionnaire responses are provided in the Nottingham Research Data Management Repository.

6.3.1.1 *Comparative analysis of the ease of prioritisation*

We conducted a Wilcoxon Signed-Rank test to compare the ease of identifying a topic between the Reddit baseline and *DiscourseDash*. The analysis revealed a Z statistic of $Z = -3.928$ and a p-value of $p < 0.001$. These results were statistically significant, indicating that *DiscourseDash* is perceived to be significantly easier to prioritise a topic to create impact compared to the Reddit baseline.

6.3.1.2 *Evaluating the perceptions of the usefulness of* DiscourseDash *features*

To illustrate the Likert scale agreement by participants, we present our results in a stacked bar chart (see Figure 21). Observing the proportions of "Agree" and "Strongly Agree", we can see that these categories occupy the majority, indicating the relevance of our proposed features in achieving decision-making tasks.

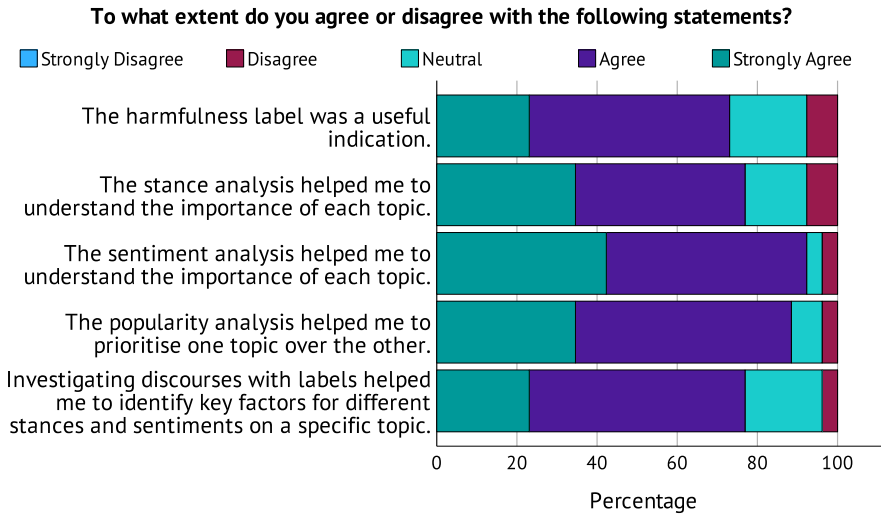


Figure 21: A stacked bar chart shows the proportion of responses from participants evaluating each feature.

6.3.1.3 *Comparative analysis of the perceptions of the usefulness of* DiscourseDash *features amongst participants*

A Friedman test was conducted to compare the participants' perceptions of the usefulness of features. The features evaluated are the harmfulness label (M = 2.56), sentiment analysis (M = 2.90), sentiment analysis (M = 3.40), popularity analysis (M = 3.25), and discourse investigation (M = 2.88). Although sentiment analysis received the highest mean rank, the test did not reveal a statistically significant difference in perceived usefulness within all the features, $\chi^2(4) = 7.295$, $p = .121$. This suggests that participants rated the usefulness of these features similarly, with no feature standing out as significantly more or less useful than the others.

6.3.1.4 *Preliminary usability testing outcomes*

DiscourseDash achieved a System Usability Scale (SUS) score of 78.9, well above the commonly accepted usability benchmark. This indicates that participants found the system both usable and accessible. The results reflect positive perceptions of the tool's ease of use and the suitability of its design for completing the assigned tasks. These preliminary usability outcomes suggest that the design flow is sufficiently intuitive for its evaluation in the subsequent study presented in Chapter 7.

6.3.2 *Qualitative results*

This section summarises the qualitative findings for each task described in Section 6.2.4.1. The results are organised according to Task 1, Task 2, and Task 3. For each task, the responses from participants using both conditions are summarised, covering the process followed, key indicators used, difficulties encountered, and the intuitiveness of both conditions. Due to the nature of this study, qualitative data were collected from a larger number of participants (n=26). The findings from participants captured a broad range of perspectives with recur-

ring and intersecting themes. To provide a clear overview, we first present a general summarised version of responses to each question. This will set the scene as context for the subsequent thematic analysis. Thematic analysis is then applied to identify and compare patterns within and across conditions for revealing insights into user perceptions and suggestions for improving *DiscourseDash*. Given the study's design, which included two conditions, multiple tasks, and several sub-questions, this structured approach ensures a coherent presentation of findings before looking into detailed thematic interpretations.

6.3.2.1 *Task 1*

Participants were asked to select a topic for prioritisation based on its relevance and potential impact. On Reddit, participants navigated their decision-making thought process through upvotes, replies, and their knowledge of current events to assess the relevance of information observed in the platform. In contrast, *DiscourseDash* offered a more structured process where the process started with the use of *Trending Topics* and the *Harmfulness* label. Then, the *Stance Distribution* and *Sentiment Distribution* graphs helped identify differing viewpoints that ultimately supported their final decision for the task. Achieving the task was more challenging on Reddit due to the overwhelming volume of irrelevant comments that lacked clear engagement indicators. On the contrary, *DiscourseDash* provided clearer summaries and graphs that facilitated their decision-making process in prioritisation.

However, participants encountered issues with the meaning of labels for *Stance Distribution* and *Sentiment Distribution* in *DiscourseDash*, and struggled to relate how these analytical features connect. Despite these limitations, *DiscourseDash* was generally perceived as more intuitive compared to manually navigating discourses through Reddit's complex structure. For a detailed comparison of each sub-question between conditions, see Table 7. Findings are presented in order of frequency, starting with the most commonly mentioned considerations by participants when prioritising a topic for investigation.

Table 7: Comparison of Task 1 summarised findings across sub-questions and conditions.

SQ	BASELINE REDDIT	*discoursedash*
1	Reviewed comments that ranked at the top, using numerical features like upvotes, replies, and dates to gauge relevance within subreddits, combined with their knowledge of current events. Some relied on prior knowledge for keyword searches and source reliability (e. g., reputable media links).	Most started from the *Trending Topics* table to generate initial ideas focusing on accompanying engagement statistics and the *Harmfulness* label. There was some uncertainty about the meaning of harmlessness. *Popularity Over Time* helped identify trends which influenced how they prioritise topics. Some further articulated the topic's significance through comprehending the analysis of *Stance Distribution* and *Sentiment Distribution*.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
2	Numerical features like upvotes & comments, and subreddit thread headings translated the relevancy of information. Some used information from the sidebar 'news' to form judgments and identify recurring themes, while others mentioned their prior knowledge and familiarity with sources as indicators of credibility.	Engagement statistics from the *Trending Topics* indicated interest levels. Some considered *Harmfulness* labels to indicate a topic's potential impact. Meaning that, they view high harmfulness as a sign of influence on a critical issue. Diverse viewpoints and emotional tones in *Stance Distribution* and *Sentiment Distribution* graphs further indicated importance due to contrasting perspectives. Additionally, *Popularity Over Time* concluded final validation based on engagement spikes and declines.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
3	Time-consuming and inefficient to sift through the large volume of replies with irrelevant information without related indicators. The nature of unstructured social media typically does not contain useful summaries or relevant decision-making indicators, making it challenging to prioritise topics efficiently. This, combined with deleted posts and trolls, further complicated the process.	Confusion about the meaning of the *Harmfulness* label and found it difficult to differentiate between *Stance Distribution* and *Sentiment Distribution*. These impacted their decision-making process as they struggled to use it to connect their thoughts and interpret insights. They suggested offering clearer explanations, addressing potential inaccuracies, and redesigning the graphs for better comparison. There were also uncertainties about the objective of the *Trending Topics*. Meaning that a topic name that is too 'short' or 'broad' with no descriptions lacks sufficient context for the user to gauge what the topic is about. Others were confused by the inconsistent ordering of topics in each component.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
4	Most found it non-intuitive due to its overwhelming volume of comments and complex interface. Upvotes were initially helpful, but having to read through all comments with no relevant indicator for this task made prioritising urgent matters challenging. Some struggled to focus on specific subtopics within broader discussions. Although engaging in discourse can be an immersive experience for readers interested in debated topics, it is not practical for professional tasks.	Most found it highly intuitive, as clear summaries and graphs made it easier to manage information for formulating informed decisions. Some participants found the tool only partially intuitive and noted that the *Stance Distribution* and *Sentiment Distribution* graphs were not well-explained as to how they handle complex social media language such as negation or sarcasm.

End of table 7.

6.3.2.2 *Task 2*

Participants were tasked with extracting key arguments representing various stances on the selected topic from Task 1. On Reddit, participants encountered challenges in navigating the high volume of discourses, relying mainly on numerical upvotes, comments, and sorting mechanisms provided by the platform. The lack of clear indicators also presented a challenge to their ability to extract key arguments

that reflect the topic. In contrast, *DiscourseDash* provided a more systematic approach. This is because the *Summary* graph and the detailed *Discourses* view were used to filter *Stance Distribution* and *Sentiment Distribution* when identifying key arguments.

Reddit was limited due to the absence of a structured overview. On the contrary, *DiscourseDash* introduced difficulties with certain discourses that extend very long and present inconsistencies in the labels that intend to support the analysis process. Even so, *DiscourseDash* was generally perceived as more intuitive and offered a more structured navigation towards understanding the discourses. For a detailed comparison of each sub-question between conditions, see Table 8. Findings are presented from most to least frequently mentioned points to highlight participant actions and interpretations during the extraction of key arguments about the topic selected in Task 1.

Table 8: Comparison of Task 2 summarised findings across sub-questions and conditions.

SQ	BASELINE REDDIT	*discoursedash*
1	Manually reviewed comments individually supported with their personal judgment in assessing stance positions. Some interpreted high upvotes as in favour positions. Others focused on analysing individual comments, mostly limited to those presented at the top by the platform's algorithm to understand underlying opinions and arguments.	The use of *Summary* graph guided and validated their findings through a brief review of the analysed *Discourses*. They filtered *Stances* from the discourse table to inspect varying stances, sometimes bypassing longer discourses. Some attempted to sort by the number of replies as an indication of the relevance of discourses, which is a function currently not supported.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
2	Upvotes were used to gauge content popularity and relevance, with high upvotes indicating widely accepted opinions. Some sorted "top" and "controversial" as a way to find relevant discourses. Others used keyword searches to gain a broader perspective on the discourse related to the topic.	The ability to filter *Stance Distribution* and *Sentiment Distribution* in the *Detailed Analysis* allowed participants to narrow down relevant information. The column next to each text within a discourse, indicating the number of replies, also directed participants' attention to the highly debated ones. Overall, having each post or reply labelled with its corresponding *Stance Distribution* and *Sentiment Distribution* in *Detailed Analysis* supported the process in extracting multiple key arguments.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
3	Irrelevant or off-topic information made it challenging to navigate the overwhelming volume of discourses and to grasp an overview stance in identifying key arguments. The combination of the discourse's hierarchical structure and the absence of indicative labels contributed to a confusing and time-consuming navigation experience.	Long texts within the *Detailed Analysis* were challenging. While participants appreciated the labels and numerical indicators, they suggested adding summaries, key themes, and highlighting keywords. Some struggled to filter discourses based on popularity and relevance. This is due to the inability to sort by the number of replies. This created uncertainty about whether the most significant discourse was being examined. There were also uncertainties in the labels, as some participants noticed discrepancies between the labels and their own interpretations.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
4	Most found it non-intuitive, with issues similar to Task 1, such as dealing with irrelevant comments that complicated their efforts to determine broader opinions and themes. Some found it partially intuitive due to familiarity with the platform, but noted a lack of structured filtering and clarity for the task.	Most found it intuitive to understand the overall stance on a topic. They appreciated how it addressed initial distrust of high-level summaries on the *Overview* page and valued the categorisation that allowed users to review different types of summaries within the underlying discourses. The process was described as a manageable approach for observing sentiments and gauging public opinions over time. However, some participants found it only partially intuitive, as certain discourses were too long and difficult to grasp.

End of table 8.

6.3.2.3 *Task 3*

Participants were asked to describe the overall sentiment in the discourse related to the selected topic. On Reddit, participants manually reviewed each comment to classify sentiment individually, relying on upvotes and filtering options. However, this process was time-consuming and did not capture the overall sentiment unless participants reviewed every single comment. In contrast, the *Summary* and *Sentiment Distribution* graphs in *DiscourseDash* provided a clearer understanding of sentiment, offering keyword filtering that allowed users to explore specific emotions, such as anger or confusion.

Despite these advantages, participants noted challenges due to the lack of an option to filter discourses by engagement numbers and inconsistencies in sentiment labels when interpreting comments themselves, which at times reduced their trust in those labels. While Reddit was considered non-intuitive for analysing the overall sentiment by topic of interest, *DiscourseDash* provided a clearer and more effective navigation through its proposed features. For a detailed comparison of each sub-question between conditions, see Table 9. Findings are presented from most to least frequently mentioned points to highlight participants' interactions aimed at understanding the overall sentiment in the discourse.

Table 9: Comparison of Task 2 summarised results across sub-questions and conditions.

SQ	BASELINE REDDIT	*discoursedash*
1	Manually inferred the overall sentiment of a discourse by manually interpreting only a few of the top comments, drawing conclusions based on their understanding from completing Task 2. Most relied on the platform's sorting options to focus on these top comments.	Reviewed the *Summary* of the sentiment graph to gauge the overall sentiment. Some further noted that during Task 2, they found a substantial volume of negative discourse. They used filtering to explore different *Sentiment Distribution* categories. This helped them to draw firm conclusions about the overall sentiment and develop an elaborate description of the sentiment types, including emotions such as confusion and anger.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
2	The use of the number of upvotes to select comments and form an understanding of sentiment, with the filtering mechanism indicating the relevance of the top comments that needed to be reviewed manually to conclude an overall sentiment.	The visual *Summary* and sentiment graphs in the *Detailed Analysis* of a topic, along with the filtering functions and the number of replies in the *Discourses* table, were all reported as helpful for identifying and narrowing specific sentiments.
3	Laborious and challenging due to the extensive volume of comments. While evaluating upvotes and shorter comments was manageable, it was insufficient to represent the complete understanding of the overall sentiment for longer or less direct comments.	Uncertain about analysing sentiment in detail due to the concerns about missing key points, because they could not filter by the engaged discourses. They also questioned the accuracy of *Sentiment Distribution* labels. They found inconsistencies between their own interpretations of labels and those shown in the *Discourses* table, which raised doubts about the visualisations and their confidence in concluding their task, relying on the presented analyses.

Continued on next page.

SQ	BASELINE REDDIT	*discoursedash*
4	Most found it non-intuitive as manually reading all posts was perceived as ineffective for an analysis task to produce informed conclusions, despite the ability to filter "top" comments.	Most participants found it intuitive and effective for gaining an overview through its visual presentation and *Detailed Analysis* of discourses. While improvements were needed, they appreciated how it combined objective analysis with subjective user input.

End of table 9.

6.3.2.4 *Final interview*

Q1: Which condition did the participant prefer to use in achieving tasks, and why? All participants preferred their experience with *DiscourseDash* for the given tasks compared to the baseline Reddit. From the given scenario (which is a fact-checker's perspective), participants compared both experiences and noted that *DiscourseDash* provided a more efficient process for addressing and creating impact on a topic of discourse. They appreciated the collaborative idea of combining supportive analysis with human input. All proposed features were considered useful for helping them focus during the investigation. Overall, participants found manually handling discourses on a social media website to be an unsatisfactory experience for extracting important information in professional settings.

Q2: What are the suggested improvements that can be done on DiscourseDash*?*

- **Explanation:** Participants suggested that the terminology used across *DiscourseDash* should be more clearly explained through interactive design to improve user interpretation. They also recommended that the naming of each topic in *Trending Topics* provide better descriptions or framing, as some participants found

the current labels too brief to convey context. Many participants struggled to understand the correlation between design components in the analysis, for instance, how the graphs convey information, how they relate to each other, and how they support comparative decision-making. Therefore, the graphs need to be revised to make it clear how they support comparative analysis. Similarly, participants recommended displaying *Harmfulness* labels as percentages to make them easier to interpret. These suggestions were brought forward from a user perspective to enhance user confidence and facilitate better-informed interaction with the design.

- **Transparency:** Some participants suggested that *DiscourseDash* could be improved by enhancing transparency in its analysis processes. They recommended providing users with detailed information about the methods used to generate the proposed labels, including how sentiment, stance, and harmfulness are calculated. In addition, indicating the confidence levels of the results was suggested as a way to help users assess the reliability of the data. Participants emphasised that these aspects are important when working with social media information, where sarcasm, trolling, and ambiguous language can create uncertainty about how accurately the system interprets discourse. Hence, improving transparency in these areas was seen as crucial for building trust in outputs designed for decision-making.

- **User Experience:** Some participants suggested that the colours used to differentiate classes in the graphs should be more contrasting to improve accessibility. They emphasised that clearer visual distinctions would make it easier to identify patterns and interpret the analyses accurately. Additionally, participants recommended improving the connectivity between topics across different parts of the analysis to create a smoother user experience. For instance, selecting a topic in one view could automatically highlight all corresponding graphs and related infor-

mation across the *Overview* and *Detailed Analysis* pages. This interactive linking would help users trace relationships, compare trends, and navigate the system by reducing cognitive load. Overall, these suggestions were brought forward to ensure that users can interact with the system intuitively.

Q3: Are there any additional functionalities they think would be useful to have implemented in DiscourseDash*?* One of the most frequently suggested improvements for the *Detailed Analysis* was a feature for *summarising* results. Participants recommended highlighting relevant keywords and generating themes from combinations of stances and sentiments to provide a clearer overview of discourse trends. For instance, integrating an LLM to support users in specifying which discourses or stances of results they wish to summarise. Given the global nature of social media, including *demographic* information was another feature considered valuable for future adaptation. Participants also suggested implementing *social network analysis* to connect information across discourses and reveal relationships or influential patterns.

From a fact-checking perspective, some participants noted that a function to *take notes* during investigations would help them track key points in cases when the investigation occurs over the span of weeks or months. For instance, they could record observations, thoughts, or references during the analysis for reference during the final review that happens later on. Additional suggestions included enhancing *filtering contents*, such as the ability to remove or tag irrelevant or sensitive items, either by providing a warning sign in the settings or by presetting a list of keywords to exclude from the presented analysis of discourses. Finally, integrating *real-time* data and providing *recommendations* for top-discussed topics based on the user's history of analysis or interested topics could support a more personalised experience within the exploration of social media discourses on *DiscourseDash*.

6.3.3 *Thematic analysis*

A thematic analysis was conducted on the qualitative data by organising it into three subsections. Each subsection highlights key themes from participant perceptions, structuring the findings by contrasting their experiences across both conditions.

6.3.3.1 *General user perceptions of* DiscourseDash

The findings in this section address **RQ1** by illustrating how users perceive the effectiveness of *DiscourseDash* in supporting decision-making through its interactive design for analysing social media discourses. Participants consistently preferred *DiscourseDash* over the baseline social media platform *Reddit* for the given tasks. They cited its efficiency in combining extensive data into useful insights. G26 highlighted this by stating:

> "This would make it easy to justify your thoughts... where you can tell him [your superior] how many thousands of people are interested in this [topic] right off the bat."

Similarly, G15 noted, "I'd say that was much more informative and easier to navigate through." Another frequently mentioned reason for preferring *DiscourseDash* was its visual clarity and ease of use. G19 pointed out that "*DiscourseDash* gave us a clear overview about where you should focus on."

The ability of *DiscourseDash* to quickly access and interpret large volumes of data allowed participants to make more informed decisions. This was perceived as valuable when decisions require justification. G25 articulated this:

> "The graphical representation and all of these graphs were telling me exactly the information required to [achieve tasks] and finding out what's going on [in the discourse]."

Similarly, G15 suggest that *DiscourseDash* is perceived as an effective means for providing objective analysis, which is important for informed decision-making.

> "[*Reddit*] gave me comments that I can read through, but I would see the statistics and graphs [from *DiscourseDash*] to make sure I make an informed decision. By doing that passively, I not only get information from all the comment sections, but I also get to read through the comments if needed."

THEME 1: EFFECTIVENESS THAT GUIDES A DECISION-MAKING PROCESS

Participants reported that *DiscourseDash* facilitated a more structured approach to decision-making by providing clear overviews and summarised insights. Several participants highlighted that the *DiscourseDash* provided a "good overview" (G16) and "helps to boil down the overall topics that are being discussed" (G6). G1 compared their experiences with Reddit, noting:

> "*DiscourseDash* gave a better overview... whereas [*Reddit*] is not conducive for me to find out the most important viewpoints and topics that I [as fact checker] should be addressing."

This structured presentation was perceived as beneficial in professional settings when decisions require concrete reasoning. Participants also emphasised that the *DiscourseDash* could save significant time. As G10 mentioned, "It's much faster and easier to identify the for and against positions that are already categorised for me." Meanwhile, G13 found the *Stance Distribution* and *Sentiment Distribution* "very useful" for gauging public perceptions of different topics.

In contrast to that, navigating *Reddit* posed considerable challenges. G3 explained, "[In Reddit] it takes ages to find what I want to do,

and if I want to determine sentiments, I need to check these replies manually." G19 shared a similar thought by saying:

> "It's like I need to read, for example, thousands of comments [...] to get an idea [from *Reddit*] while the analysis here [*DiscourseDash*] saves a lot of time [...]."

G23 added, "In *Reddit*, I have to like maybe read the comment twice to understand the sentiments and stances." G6 emphasised the importance of having support for navigating large volumes of unstructured data even before using *DiscourseDash*, stating:

> "If you want to do this work properly, you need a tool to help you [...]. Maybe a ready dashboard [...] there's too much information [on *Reddit*] that is too hard to see the big picture."

Overall, participants perceived *DiscourseDash* as more effective than manually navigating large volumes of unstructured data for the given tasks. The combination of structured outputs, visualisation of analyses, and interactive design components all supported users in making faster and more informed decisions during decision-making tasks around analysing social media discourses.

THEME 2: GUIDANCE TOWARDS ARTICULATION OF IDEAS

Participants highlighted how *DiscourseDash* facilitates understanding of discourse by revealing ideas, correlations, and trends. For instance, G15 articulated:

> "The discussion is high on the first topic, which states that there is some confusion or statements made without proper facts being provided [...]. Here [*Stance Distribution* and *Sentiment Distribution*] clearly states that they are contradictory opinions, and that's not being resolved."

Popularity Over Time was perceived as useful in revealing correlations between different topics of discourse. For instance, G16 articulated their understanding of the interconnected trends across various topics and explained:

> "Whenever there's a spike in topic one, there's a spike in topic three as well. This [particular topic] is also interesting because it follows a correlation with these [two other topics]."

Another used the *DiscourseDash* to interpret specific ideas within the discourse, where G2 noted:

> "[...] at the same time they are talking about the second dose, they are also talking about side effects, with [negative sentiments towards it] [...] they don't like vaccines and very few of them are thankful for all the work done [for vaccines] so far."

Similarly, G7 mentioned how the trends in *Popularity Over Time* graph stimulated curiosity:

> "I would tend to [investigate] what happened around that time [a decreasing point in time]. What kind of event might have occurred at that point?"

These examples demonstrate how *DiscourseDash* supports and guides participants' thought processes. Meanwhile, *Stance Distribution* and *Sentiment Distribution* shaped participants' interpretations. G18 reflected:

> "What I can make out is that people are happier in taking a vaccine rather than not taking it [...] because from both the graphs [*Stance Distribution* and *Sentiment Distribution*], [...] the number of positives is more and here the favour is more than against."

While the harmfulness label was interpreted as "high rating of harmfulness is probably more important" (G10). Adding to that, G17 emphasised that by noting:

> "harmfulness also determines newsworthiness [...] so I'm leaning [my decision] towards the first two [topics labelled as high harmfulness]."

Instead of the term harmfulness, G18 "would rather it be a saying impact, but I guess it denotes the negative sense of these topics." These observations highlight how the harmfulness label was interpreted throughout the decision-making process.

Finally, the *Detailed Analysis* facilitated participants in formulating ideas about the factors influencing different stances. For instance, G15 explained:

> "There are some people who want to take the vaccine, they show negative sentiment [...] because of limited supply, and they feel insecure as they are the essential workers, [...] people who have a neutral stance and sentiment, they are not exactly sure how the vaccination works and how it helps."

Similarly, G25 noted:

> "A lot of people are asking how the government is going to [take action towards the vaccine], [...] there's a lot of confusion."

Taken together, these participants' insights indicated that *DiscourseDash* the identification, articulation, and interpretation thought processes were guided through the proposed analyses.

THEME 3: ENHANCING TRANSPARENCY THROUGH *DETAILED ANALYSIS* OF DISCOURSES

Several participants initially expressed scepticism toward the *Overview* page due to limited transparency in the high-level sum-

maries by questioning "where are the sources from?" (G26). Similarly, G6 expressed their initial doubts while interacting with *Overview*:

> "I have to blindly trust these classifications [...] as I have no idea what data this has been generated on."

However, after exploring the *Detailed Analysis* page, G6 realised:

> "I do have access to the data [...] this addresses what I criticised [previously] about the doubts."

This contrasting view from the participant illustrates that access to underlying discourse-level details improved perceived explainability and trust. *DiscourseDash* allowed users to trace high-level summaries back to the discourses, and participants appreciated the transparency of this in supporting their understanding of the rationale behind the high-level analyses.

Another key observation highlighting how *Detailed Analysis* facilitated the investigative tasks is evident in G26's experience. They initially described having difficulty distinguishing between the meaning behind stance and sentiment. However, they noted that examining examples in the discourses helped them better understand these concepts and clarify their meaning. They expressed that:

> "It was not very clear until I toggled around. After reviewing several different combinations and permutations [of stance and sentiment pairs], I understood the correlation better."

This suggests that while the *Overview* page provides a useful high-level summary, *Detailed Analysis* played an essential part in comprehending varying stances, sentiments, and their interrelations.

THEME 4: RELIABILITY CONCERNS IN AUTOMATED ANALYSIS MEASURES

Participants also raised concerns about the accuracy of the labels presented in *Discourses*. A recurring issue was the mismatch between

their own judgements during manual inspection and the labels generated within *Discourses* in *Detailed Analysis*. For instance, G16 mentioned that they noticed "incorrectly marked sentiments," pointing to perceived inaccuracies in sentiment classification. Similarly, G6 commented:

> "When I dig into the data, which I think is important that I have the option to, I very quickly identify several [labels] that don't match up with my own judgement [...] which reduces the reliability of what I saw in *Overview*."

These observations highlight the need for clearer indications of confidence or uncertainty in the labelling process. Providing such information could support more reliable interpretation and help users calibrate their trust when using *DiscourseDash*'s outputs for decision-making.

6.3.3.2 *Usefulness of proposed features in influencing decisions*

This subsection addresses **RQ2** by grouping the proposed analytical features mentioned by participants during their interactions and examining how they supported decision-making tasks on *DiscourseDash*.

THEME 1: PERCEIVED EFFECTIVENESS OF HARMFULNESS LABELS

Harmfulness labels were seen as helpful in guiding the prioritisation of topics that were both widely discussed and considered significant due to their potential impact. For instance, G16 noted that "the harmfulness is quite high. So I'm leaning towards like the first two." Similarly, G17 articulated that "harmfulness also determines newsworthiness, so it might be important to [prioritise] more harmful news." They recognised that higher harmfulness signifies more pressing issues that deserve greater attention in decision-making. However, some participants expressed confusion about the labelling of harmfulness detection. For instance, G22 pointed out that "I don't understand

what the high harmfulness means [...]". Although without a clear indication of what it means, they were still able to interpret that as: "if the harmfulness is high, then it is important in my decision."

THEME 2: ROLES OF *stance distribution* AND *sentiment distribution*

The distribution of stances and sentiments shaped how participants perceived the importance of topics. For instance, G19 mentioned that "although people talked about it [the topic] a lot [from the *Stance Distribution*], not a lot of people have a clear opinion on that [topic]. So, that might [warrant] a discussion in the article [as a fact checker to address it]." This example shows that the presence of diverse opinions prompts deeper investigation.

Similarly, G26 noted that *Stance Distribution* helps to contextualise public opinion: "we can tell that not many people are just singing to the choir; the public seems to agree on this [topic] but may have more questions about it." This indicates that areas of both agreement and uncertainty can pinpoint issues worth further investigation.

Finally, G22 commented that "I think for me the *Stance Distribution* and *Sentiment Distribution* help shape which [topic of discourse] is important [...] because if you have [contrasting opinion like against toward a topic] it enriches the discussion and it's more exciting [to investigate]." This suggests that participants relied on the distributions of stances and sentiments not only to identify topics of interest but also to prioritise where to focus their efforts for further investigation.

THEME 3: USE OF *popularity over time*

Participants used the *Popularity Over Time* to connect shifts in public interest with specific events. For instance, G19 noted that:

> "Especially here [spike in the graph], you can see some different changes, which might correspond to specific events like news that happened at that moment."

The potential application of popularity trends for future content planning by fact-checkers was also highlighted. G26 noted:

> "While I am comparing these [topics], there is a decreasing amount of interest in the general vaccine, [...] but maybe a follow-up article like what might inform my work [as a fact checker] next would be about the second dose, which is the next trending topic [observed from the trends]."

This demonstrates how monitoring popularity trends helps guide intervention planning. Additionally, G07 emphasised the longitudinal aspect of the data by noting that:

> "[during] the investigation, focusing on specific points like maybe this part [decrease in engagement trend] would be interesting to see what kind of events happened."

Overall, "although [a topic] is less popular than [another] topic" (G07), the *Popularity Over Time* helped participants to recognise when "it's rising in importance" (G07) due to emerging relevance.

6.3.3.3 *Design improvements for a better user-centred experience*

This subsection addresses **RQ3** by synthesising the overlapping challenges participants encountered across tasks. This section identifies design refinements for enhancing usability, user confidence, and overall interactive experience with *DiscourseDash* to highlight areas for improvement.

THEME 1: DIFFICULTIES IN INTERPRETING ANALYTICAL MEASURES

Participants frequently struggled to understand the proposed analytical outputs. Confusion occurred around the definitions and metrics

underpinning harmfulness (G1, G2, G6, G9, G10, G11), as well as the distinctions and relationships between *Stance Distribution* and *Sentiment Distribution* (G1, G4, G6, G7, G18). As G1 explained, "It's unclear to me what or how they are different. So, I'm inclined to ignore it [...] but they do tell me people seem to care about topic one a lot." Such uncertainty risks reducing their value in decision-making as participants disregard features entirely. This highlights the need for clearer presentation of the analytical features, accompanied by explanations, and for redesigning the graphs to encourage more seamless comparison. Similarly, the descriptions provided in *Trending Topics* were regarded as overly general, making it difficult to cross-reference them with the visual outputs.

THEME 2: UNCERTAINTY ABOUT THE ANALYTICAL OUTPUTS

Participants raised concerns about the reliability and clarity of the analytical outputs. Several questioned how topics were extracted and clustered, with G7 expressing: "I don't know how those [*Trending Topics*] were extracted and how the objective choice of topic is." Others doubted the algorithm's ability to capture complex linguistic patterns, such as sarcasm or negation. As G6 reflected, "How advanced is the algorithm? Is it able to detect the negation of sentences? Is it able to detect irony and sarcasm? Such uncertainty illustrates that users struggled to interpret results with confidence when the processes behind labels remain opaque. As G6 concluded, "all labels seem to be ambiguous and need some explanation." These concerns underscore the importance of explaining calculations and demonstrating the confidence level of automated outputs to mitigate uncertainty in analytical features.

THEME 3: USABILITY IMPROVEMENTS FOR BETTER INTERACTIVE EXPERIENCE

Participants identified several usability challenges that affected their experience. First, inconsistencies in terminology across sections were a recurring source of confusion. For instance, G7 questioned "the difference between comments, replies, and posts," highlighting the need for more consistent labelling across a design tool. Visual design choices were also reported as difficult to distinguish due to the colour scheme. G24 suggested that "some other colour contrast would make it [the different outputs in *Stance Distribution* or *Sentiment Distribution*] more obvious and presentable." These observations demonstrate how minor details in terminology and visual design can significantly impact user experience.

Participants also suggested some relevant features for improving usability. A common request was incorporating the ability to visually highlight related data points across *DiscourseDash*'s *Overview* when a topic is selected. As G7 noted, "When you click [a topic] here [in the *Trending Topics*], it highlights everywhere on the graphs. I guess that would help in terms of User Experience (UX)." Implementing such functionality could enhance navigation experience and help users perceive connections between different analytical outputs during interactions.

6.4 CHAPTER SUMMARY

This chapter presented the findings from a mixed-method study evaluating *DiscourseDash*, a dashboard designed to support decision-making through the analysis of social media discourses. The results indicate that integrating harmfulness detection, stance detection, sentiment analysis, and popularity analysis as features provided participants with supportive insights during decision-making tasks. Participants demonstrated how they used these features to accomplish tasks

in realistic scenarios that reflect expert workflows for analysing complex social media discourses. Compared to navigating unstructured platforms such as *Reddit*, *DiscourseDash* was consistently perceived as more effective for identifying relevant information in making informed decisions.

The findings provide foundational evidence supporting the design concept presented in Chapter 5, while also highlighting areas for refinement. Participants' interactions reveal the strengths of presenting discourse-level analysis through an interactive design for decision-making, as well as the importance of transparency, usability, and confidence in interpreting automated outputs. Overall, the study demonstrated the value of communicating analysis through the lens of discourse-level insights and the role of a human-centred approach in the development of social media analysis interactive design. It also highlights the critical influence of human factors in ensuring trust, transparency, and actionable insights in automated analyses. Having established the validity of the core design, this study provides a foundation for minor refinements of design elements and for conducting a follow-up investigation with expert users.

7

AN INTERVIEW STUDY WITH INDUSTRY EXPERTS

This chapter presents the final empirical study of the thesis through semi-structured interviews with industry experts who analyse social media discourses to inform decision-making. Building on the previous mixed-method evaluation in Chapter 6, this study examines experts' current workflows, their perceptions of *DiscourseDash*, and their views on its potential integration into professional practice. This draws on methodological principles which emphasise that the most effective way to ensure the real-world applicability of *DiscourseDash* is to involve the very group of users it is designed to impact. This study explores how the tool aligns with established practices and expectations by engaging professionals directly. At the same time, questions are also focused on gathering feedback to guide further refinement of *DiscourseDash*.

The interviews focused on three main areas. These were existing practices in social media analysis, perceptions of *DiscourseDash*'s conceptual design, and reflections on its practical use. The findings highlight barriers, opportunities, and design implications, positioning *DiscourseDash* closer to professional application. This chapter extends the empirical validation of the thesis by examining how discourse-focused analytical tools align with expert practices. To this end, the study conducted in this chapter strengthens the evidence around how discourse-level insights in social media analysis support real-world decision-making.

7.1 INTRODUCTION

Industry experts who analyse social media discourses for societal impact play a central role in decision-making across public health (Lit-

man et al., 2020), brand management (Mills and Robson, 2020), and policy-making (Shahid et al., 2020). Chapter 2 highlighted that these workflows involve navigating large volumes of rapidly evolving content, interpreting diverse perspectives, and prioritising information for actionable outcomes. Despite the importance of these tasks, little is known about how experts structure their workflows, apply analytical strategies, and engage with interactive tools designed to support decision-making. In particular, it remains unclear how tools like *DiscourseDash* that emphasise discourse-level insights can be adapted to fit professional workflows.

In the previous Chapter 6, *DiscourseDash* was evaluated with non-expert users to explore its potential for supporting decision-making through the lens of discourses in social media analysis. Although the evaluation demonstrated the effectiveness and practicality of *DiscourseDash*, it did not reflect the workflows and expertise insights of industry experts. This study complements the previous evaluation by drawing on expert insights to bridge the gap between design and real-world practices. To achieve this, we conducted semi-structured interviews with 11 industry experts who are experienced in analysing social media discourses to inform decision-making. The study investigates their current workflows, approaches to social media analysis, perceptions of *DiscourseDash*, and suggestions for its practical integration in professional contexts.

Through examining expert practices, this chapter identifies factors that influence decision-making. This included the limitations of current approaches and feedback to guide future developments. The study ensures that *DiscourseDash* aligns with established workflows while revealing opportunities for refinement. The following research questions guide the empirical investigation:

RQ1 How do industry experts perceive and interact with a dashboard tool that offers a *Detailed Analysis* with discourse-level analysis to make informed decisions?

RQ2 How do industry experts currently analyse social media discourses, and how well does *DiscourseDash* align with these real-world practices?

RQ3 What key design implications can be drawn to guide future iterations of *DiscourseDash* and better support decision-making across different professional contexts?

7.2 STUDY DESIGN

We designed our interview study using an inductive approach to explore how experts interpret social media discourses and what they need from supportive tools. This approach allowed us to use predefined questions while remaining open to new themes that emerged during the interviews. We structured the questions into three parts. The first examined participants' professional roles and workflows, including the types of decisions they make, how they analyse social media discourses, and the challenges they face. The second explored their perceptions of *DiscourseDash* by interacting with the prototype, focusing on features that stood out and those that were missing. The third assessed how *DiscourseDash* fits into their workflows, inviting participants to describe use cases, potential barriers, and suggestions for future analyses.

7.2.1 *Recruitment and participants*

Participants were recruited through existing professional networks, direct outreach on *LinkedIn*, and targeted recruitment via *Prolific* (see Appendix A.4.2). Initial expert selection was based on predefined criteria (see Appendix A.4.3). Years of experience were not used as the sole indicator of expertise, since prior research indicates that experience alone does not reliably reflect skill or judgment (Ericsson et al., 2018). Instead, domain knowledge, contextual understanding, and the ability to provide reflective insights were prioritised. Participants

who did not demonstrate these qualities during the interview were excluded from the final sample. The final sample consisted of 11 participants with diverse roles, responsibilities, and sector experiences. Most participants were based in the UK, while a few had experience in Japan, Nigeria, and other European countries. Participants were assigned anonymised codes (e. g., E1), where 'E' denotes an expert. A complete list of participant codes, professions, interview dates, and anonymised excerpts from the transcripts is available at the Nottingham Research Data Management Repository.

7.2.2 *Procedure*

The interview began with a brief introduction outlining the study goals and reassuring each participant about the confidentiality of their participation. A short demographic questionnaire was administered to ensure diversity and contextualise perspectives, given the small sample size and focus on high-quality insights. The interview was structured in three stages.

7.2.2.1 *Stage 1: Baseline*

Participants were asked questions to establish a baseline regarding their engagement with social media discourses in decision-making, the tools and methods they currently use, and the limitations of these approaches. The questions for this stage are as follows:

Q1 Can you describe your role and the types of decisions you make based on analysing conversations on social media?

Q2 How do you use these social media conversations to support your decisions, or in other words, what details do you pay attention to or look for?

Q3 How do you currently navigate these social media conversations to make decisions? For example, is there any tool or method that you rely on? Or do you use Internet Search?

Q4 Do you use software or tools to help with the analysis, such as spreadsheets, social media monitoring tools, databases, or LLMs/*ChatGPT* for summarisation?

Q5 Can you describe some challenges you face when using these tools or methods to support your current workflow?

7.2.2.2 *Stage 2: Interaction with* DiscourseDash

Participants were introduced to *DiscourseDash* via a link to the prototype and instructed to share their screen while answering questions. They first interacted with the *Overview*, followed by the *Detailed Analysis page*. The same set of questions was asked for both pages to gather feedback on the relevance of *DiscourseDash* features, including which elements supported decision-making and what was missing. The questions for this stage are as follows:

Q1 Describe which information presented in *DiscourseDash* would help you decide a topic to prioritise among all the ongoing public discourses. For example, what details stand out to you when deciding what is important?

Q2 Within this process, can you think of one thing that is missing from *DiscourseDash* to help you with this specific task, or your workflow more broadly?

7.2.2.3 *Stage 3: Real-world reflections*

After interacting with *DiscourseDash*, participants were asked to reflect on their overall experience. This stage gathered insights on potential real-world use cases, adoption barriers, and integration considerations. The questions for this stage are as follows:

Q1 Can you describe some use case scenarios where you see *DiscourseDash* being useful? Examples of how you see it being used in your workflow or any other workflows you have experience with?

Q2 What potential barriers or challenges do you foresee in adopting *DiscourseDash* in practice, and how could they be addressed?

Q3 On the basis of *DiscourseDash*, can you think of any new types of analysis that you might do in the future? If so, what kind of improvements or additional features would you suggest for that?

At the end of each session, participants received a £20 Amazon voucher as a token of appreciation for 45 minutes of their time. Figure 22 illustrates the study design and procedure.

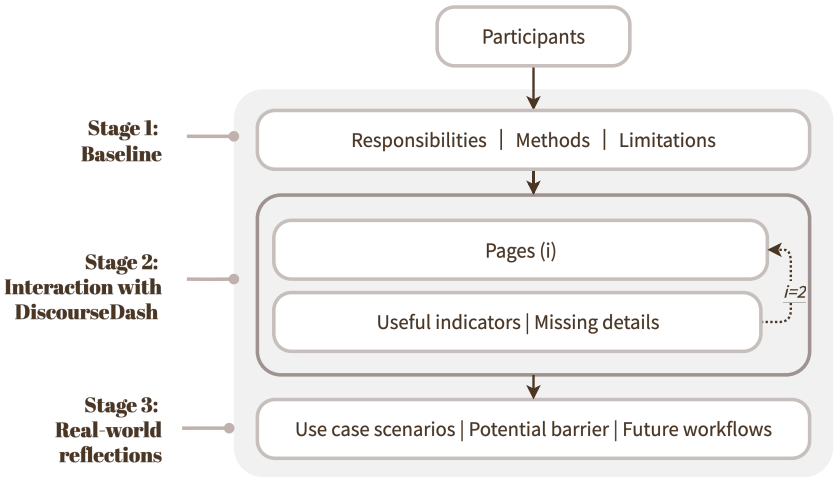


Figure 22: Overview of the study procedure with industry experts. The two key interactive pages are denoted as (i): when $i = 1$, this refers to the *Overview* Page; when $i = 2$, this refers to the *Detailed Analysis* Page.

7.2.3 *Data collection and analysis*

All interviews were transcribed using Microsoft's automated transcription immediately after each session. Interviews were conducted in English via Microsoft Teams between March and April 2025. Each transcript was reviewed manually against the original audio recording to ensure accuracy and completeness. Thematic analysis was used to analyse the data. This approach is described in Chapter 3 and was

applied previously in Chapter 6. Transcripts were read and participant responses summarised. Open coding was conducted using an inductive and interpretive approach (Braun and Clarke, 2012). Codes were organised according to sections of the interview stages to guide structured interpretation. Themes were developed iteratively through multiple review sessions to ensure consistency and coherence across the dataset. This process identified patterns that addressed the research questions and provided insights into expert workflows, perceptions of *DiscourseDash*, and considerations for practical integration.

7.3 FINDINGS

The findings are organised according to the interview stages: before use to establish a baseline, during interaction with *DiscourseDash*, and after use to capture reflections. This structure allows a systematic presentation of themes in relation to the research questions, with each section highlighting insights relevant to that stage. Compared with the findings in Chapter 6, the expert user study (n=11) generated more distinctive and focused insights, with less overlap or repetition across participants. As a result, the findings are reported concisely by highlighting the key themes directly.

7.3.1 *Understanding expert's existing workflows*

The findings in this section address **RQ2** by examining how industry experts currently analyse social media discourses. These practices and their relevance were brought forward as we establish a baseline for evaluating how *DiscourseDash* aligns with expert perceptions and real-world workflows. We examine experts' responsibilities, workflows, approaches, and challenges in analysing social media discourses to highlight how *DiscourseDash* could fit within existing practices. Understanding these workflows informs our research to iden-

tify where an interactive tool can complement an expert's decision-making.

THEME 1: USAGE OF SOCIAL MEDIA DISCOURSES IN DECISION-MAKING CONTEXTS

This theme captures how experts engage with social media discourses as part of their routine decision-making. Their practices demonstrate that these discourses are relevant within organisational workflows and priorities for shaping strategic planning and immediate operational responses.

Integrating social media insights for strategic coordination

Social media insights serve as a key resource that guides both long-term strategic planning and immediate organisational responses. Experts described social media discourses as a valuable resource for shaping brand direction, managing public perception, and aligning team efforts with broader organisational goals.

These experts analyse social media discourses to detect emerging trends, guide campaign planning, and inform a range of management activities. This strategic use supported coordinated actions and informed decision-making across teams. E5 described the use of social media insights in marketing and sales as follows:

> "to tell the insights into marketing and sales strategies, [...] crisis management of the brand, [...] specifically the ones more related to the bad reputation of the brand."

E3 highlighted the role of social media ideation and collaborative planning:

> We get the [team] working together and see how we could generate plans, generate ideas [...], push a particular brand out there [...] sometimes, we roll with the trends because it helps whatever it is you're marketing to get to the required audience."

These accounts reveal how social media insights support collaboration and inform strategic decisions that align with organisational objectives. Consulting practices extend this strategic use by applying social media analysis to benchmark industry trends and assess internal performance. For instance, E7 examined employee discourses to refine strategies in areas such as the "employee happiness index, productivity, gender equality, and internal cost optimisation."

Experts further emphasised that their analyses represent only one component within broader organisational workflows and are subsequently translated into actionable outputs such as reports or presentations. These outputs serve as important channels for conveying insights across teams that guide the implementation of strategic initiatives. As E7 explained:

> "After analysing the data, I compile the insights into a report, which is then shared with the marketing and sales teams to guide their next steps."

Overall, these findings underscore the role of social media discourses as a strategic resource that informs decision-making and coordinates actions across teams and departments within organisations.

Understanding audiences to inform communication strategies

Understanding audience perspectives was frequently highlighted as important when it comes to shaping public behaviour. This may involve encouraging people to adopt a product or service or supporting the implementation of government interventions. On social media, different audiences engage with content in distinct ways, raise their own kinds of concerns, and require tailored language depending on the platform or group. These insights help design communication strategies that align with specific audiences and reveal the potential impact of interventions.

For instance, E11 described how analysing misconceptions in vaccine-related discourses allowed them to use social media to "build some level of resilience [towards vaccines]." They explained that they

distribute factual information based on what "people at the community level" understand. This suggests that experts viewed social media discourses as a way to identify and filter communities and to design intentional communication strategies with greater impact.

Experts also emphasised the importance of platform-specific dynamics. Communication strategies on *Facebook* differ from those on *Twitter* or *Instagram*. For instance, experts used ongoing monitoring of public discourses to inform the tone, framing, and timing of their intervention messages. This shows that understanding audiences is not just about analysing content from discourses alone, but also requires attention to external factors. These factors include demographics, language and platform norms that inform the design of effective public communication.

Experts highlighted that community-building is also part of a strategic approach. This approach will focus on maintaining formulated communities that require them to understand and address their needs. This involves actively monitoring discourses as a form of feedback that guides decision-making actions. E6 described this approach as "building community [...] and to monitor the conversations about your products to gather feedback and insights."

E2 similarly highlighted the use of social media for sustaining engagement and promoting brand advocacy. They explained:

> "There's a way you could use social media to create buzz and excitement about the new product launch and also engage with customers and create a community to foster loyalty and advocacy."

These practices show that understanding audiences involves more than observing social media discourses. It requires careful interpretation to maintain relationships and build trust through responsive and intentional communication.

Responding to crisis and managing misinformation

Experts described how actively monitoring social media discourses is critical for managing crises. This is relevant in cases of misinfor-

mation that can potentially undermine public trust or compromise a reputable organisation.

The following details some examples of how proactive observation of social media discourses supports timely responses to mitigate potential negative impacts on an organisation or entity. For instance, E11 described this in the context of widespread vaccine hesitancy during the COVID-19 rollout caused by misinformation:

> "When you see some propaganda coming out from social media [...] that this vaccine causes this and that, which is all misinformation. And as a result, people tend to become sceptical [...] it's not only affecting COVID-19 vaccine uptake but also other routine immunisations that we've been doing."

Their response involved a culturally grounded investigation into the reasons behind the propaganda:

> "Once we identified the reason, we realised we need to use the traditional rulers to speak to them, we need to engage the religious leaders [...] to let them [the public that is against vaccination] know that this is a normal routine immunisation."

Through the analysis of social media discourses, they were able to design strategies that communicated messages in ways that resonated with local values and beliefs. This shows how experts analyse social media discourses to identify the factors driving opinions shaped by misinformation and use these insights to develop more effective crisis interventions.

On the other hand, E9 described a continuous approach to monitoring social media aimed at anticipating potential crises. They explained that their responsibilities include:

> "Tracking biomedical incidents in Japan and using social media discourses to identify emerging ethical concerns before they escalate"

This proactive practice demonstrates that social media supports responses to immediate crises and also guides preventive strategies by enabling the early identification of emerging issues.

Using public signals as evidence for reporting and policy shaping

In more formal contexts, experts analyse social media discourses to inform reporting and guide policymaking. These analyses generate evidence that supports decisions by government agencies or organisational leadership. For instance, E10 described how they monitor multiple platforms to compile reports that facilitate government agencies in designing appropriate interventions:

> "Based on the [analysed] narratives, I create reports allowing the governmental agencies to create interventions [...] in the context of any emergencies or events happening around the countries."

Similarly, E7 explained that the organisation uses the analysis of social media discourse as part of an approach for performing "internal assessments". Insights from public trends then guide the adaptation of company policies and strategic decisions, directly influencing operational practices. In these cases, insights from social media discourses inform the development of organisational policies.

THEME 2: GUIDING FACTORS IN DECISION-MAKING

When experts described their workflows, from data aggregation and analysis to creating roadmaps, plans, and management strategies, they outlined several factors they analysed from social media discourses that guided their decisions. These factors can be grouped into three complementary types:

1. Direct indicators from platforms, which are quantifiable signals obtained from social media. These include performance metrics such as likes, shares, comments, and reach, as well as audience characteristics like age, gender, and profession.

2. Interpretive insights from discourses, including qualitative understanding and subjective assessments of ongoing discourses.

3. Situational and contextual dependence, reflecting decisions shaped by broader organisational goals and stakeholder expectations.

Direct indicators from platforms

The integration of statistical analysis into decision-making processes was a consistent guiding factor across experts' workflows. They routinely used quantifiable signals from social media platforms to inform and validate strategic choices.

These indicators are fundamental for experts working in marketing, where activities include planning campaigns, creating content, and engaging with audiences to promote a product, service, or brand. Key metrics included engagement rates (e. g., likes, shares, comments), follower growth, reach, and longitudinal content performance. These metrics serve both as evidence to support decisions and as a way to evaluate the effectiveness of strategies over time. E6 described how these metrics directly informed decision-making for immediate actions and longer-term strategic directions:

> "[We] analyse the impression and the views and make sure that OK, it's meeting the standard and making sure that [...] we understand what works for us and then what drives more crowd and more prospects to our business."

Here, *impressions* indicate active attention or engagement, while *views* refer to the number of times content is displayed to the audience. Similarly, E5 explained how these performance metrics guided adaptive decisions for interactive strategies:

> "When we noticed that our stories [a method of content posting on *Instagram*] lacked interactivity, [statistics from the platform indicated that] people saw the content but didn't engage with it. [These direct indicative observations led us to] Adding polls with direct questions like

> 'Do you want this product to continue?' boosted engagement by 20 per cent."

In this context, *polls* are interactive components on social media platforms that allow audiences to vote on a question posed by the expert to gather immediate feedback on preferences and engagement.

These examples demonstrate that direct indicators do support experts when modifying their strategies as a form of response to audience behaviour. They can increase engagement and ensure their content resonates with the intended audience by using performance metrics as a guiding factor for decisions around marketing strategies

Audience characteristics also allowed experts to segment their strategy according to audiences' demographics, including age, gender, and profession. This segmentation was cross-mentioned multiple times across the study as highly relevant for refining strategies and tailoring communication to specific groups. Experts further explained that these insights guided adjustments to marketing content and delivery, as E3 noted:

> "Demographics matter. A message that resonates with 18–25-year-olds might not work for a 40–55 demographic. Men and women, even within the same age range, interpret and respond to content differently. We also look at professions to understand patterns of engagement."

These suggest that demographic differentiation would facilitate experts in adjusting strategies according to each audience segment for improving the relevance of their decisions.

Interpreting insights from discourses

While direct indicators provide measurable and immediate signals, experts also discussed the importance of interpreting social media discourses for qualitative insights. This process involves analysing sentiment, tone, and information on discourses to identify emerging narratives, shifting opinions, and broader public concerns.

Within these, experts also keep track of simpler signals, such as mentions, hashtags, and key phrases, which can indicate shifts in public interest or concern. These interpretive practices complement quantitative metrics and directly guide decision-making by highlighting areas that require strategic attention or intervention. For example, E9 described their workflow for exploring public understanding and perceptions of organ donation:

> "We first check very general terms like organ donation. What do people understand about the medical aspects of it? When problems arise, what kind of problems are they? If they don't want to donate, usually, why not? Those can be very subjective decisions, but also objective; it depends. It's a broad approach to try to understand what people think about organ donation, how it's done, and why they would or wouldn't do it. After that, we start doing more specific analysis."

This example demonstrates how qualitative signals provide context that informs and shapes subsequent decisions in the expert's workflow. Similarly, E10 described tracking narratives and sentiment around specific keywords to anticipate public reactions and guide strategic responses:

> "I pay attention to the narratives and sentiments among the users in the context of particular keywords or key phrases, and based on those phrases regarding known topics. For instance, for vaccines, the pandemic, flooding, or even sports events or concerts. Based on that, I [...] research on their sentiment, is it growing or going down, depending on the situation."

By tracking shifts in sentiment and narratives around specific topics, experts can adjust their decisions on the appropriate strategies for responding to emerging concerns or events. E11 highlighted the importance of monitoring narratives and sentiment during the COVID-19 vaccine rollout:

> "The information we were getting during the COVID-19 vaccine rollout was filled with claims like 'this vaccine kills,' produced by unreliable sources. To tackle this accordingly, [...] we organised strategies to counter these myths and build resilience to such misinformation."

Experts also emphasised the value of analysing discourses to support interpretation for strategic purposes, including community engagement, marketing, and product refinement. For instance, E2 described how these insights inform decisions about community building, audience engagement, and competitive positioning:

> "We use social media insights to create buzz and excitement about the new product launches, engage with customers, foster loyalty and advocacy for the products and build communities. Feedback from these [social media discourses] reveals competitive areas about the product, [inform] how our products can be better or can be made known to people, and what other products aren't doing that we could do. You know, I make the product a bit different from what is out there already."

Overall, qualitative interpretation encourages experts to move further than just quantitative metrics and uncover insights that directly inform actionable decisions. Analysing social media discourses allows them to identify emerging concerns, refine strategies, and make decisions that are contextually relevant to real-time discussion. In contrast to direct indicators, these interpretive insights provide a deeper level of reasoning that informs and corroborates expert decision-making.

Situational and contextual dependence

Although social media discourses provide valuable insights, experts emphasised that these analyses constitute only one component of a broader decision-making process. Insights from social media must be

interpreted in conjunction with institutional, situational, and contextual factors to support strategically aligned decisions.

In practice, analysing social media discourse provides an initial overview. Experts then complement these insights with additional methods such as surveys, interviews, or field studies to develop a more in-depth understanding before making decisions. For instance, during politically sensitive product launches, E8 highlighted the limitations of relying solely on public social media monitoring:

> "Not everybody is comfortable expressing themselves on social media. [...] So you have to look for the spoken words and the unspoken words."

They explained that in certain cultural contexts, members of the public may be reluctant to express criticism openly:

> "People don't always feel safe giving criticism online. Sometimes it's the body language or their language tone that says more than the words. You have to read between the lines [during interviews] to understand what people think."

These cases demonstrate that social media discourses provide only a partial perspective on public opinion. E9 similarly noted that social media alone cannot capture all perspectives necessary to inform their decisions. To obtain more accurate and targeted information to guide decision-making, they rely on controlled studies and experiments. This means that, after an initial analysis of social media discourses, supplementary methods such as surveys or direct interviews typically follow to gather more targeted information tailored to specific objectives. This iterative approach demonstrates that experts frequently need to integrate discourse-level insights with supplementary insights required to refine their understanding for decision-making.

THEME 3: NAVIGATING SOCIAL MEDIA DISCOURSES

Social media platforms generate an overwhelming volume of discourse each day. Experts tasked with interpreting these discourses face challenges in making sense of the fast-moving and ambiguous nature of social media.

This section presents an overview of how experts navigate this space and translates their practices into insights relevant to decision-making workflows. Findings are organised into two main categories: (1) manual exploration and (2) automated analysis. Although described separately, experts frequently combined these approaches within their workflows and adapted them to meet specific analytical needs.

Manual exploration

This is commonly used for exploratory analysis. Experts rely on them when engaging with unfamiliar topics or looking to understand the cultural context and underlying dynamics of a discourse.

In this type of navigation, experts begin by using platform features such as trending topics, hashtags, keywords, and brand mentions to identify discourses relevant to their analysis and decision-making. They then review the filtered discourses to obtain relevant guiding factors for their decisions. For instance, E9 described how they strategically curated a *Twitter* account to focus on credible and topic-specific content relevant to their research:

> "What we did was to train one of my accounts to recognise that kind of content [...] and then it automatically was recommending me more similar content [...] so it's created for expertise topics from, for example, newspapers. A more credible resource, right? But also a resource that a lot of people read [i. e., reputable articles]. So, that is like my starting point sometimes to find, 'oh, they are talking about [this], and [that] has organ donation in it,' or maybe

> in cases where a hospital is doing advances on something related to tissue donation."

They further explained how this guided their exploratory research:

> "It's again exploratory work, so I have a rough idea only at first what we could be checking on. We then download that data sample and perform a quick analysis of OK. Is it truly what we are looking for? And if it is, well, we continue a little bit more in-depth. If it's not, then we change some settings or maybe even keywords. Then again, we perform an analysis on it."

Similarly, E8 relied on real-time observation of trending topics and hashtags to track emerging issues:

> "I follow so many [key opinion leaders] to know what is going on [...] every day has its hashtag [...] So I just go to social media, look at the current trend and pattern [..] you just have to follow the hashtag [and it] will lead you to what is currently going on."

At times, manual processes were informed by established frameworks. E10 described how they built a custom taxonomy based on World Health Organisation (WHO) guidelines. They adapted it to specific cultural and linguistic contexts, such as youth language or regional dialects. Although coding them thematically remained manual, the taxonomy gave a consistent structure for navigating their analysis. These examples highlight the limitations of manual exploration, as it requires repeated engagement with social media discourses to generate insights for decision-making.

Adopting existing automated or software tools

While manual exploration is effective for small-scale analysis, experts turn to automated tools for managing large-scale or time-sensitive social media data. Automated tools support faster analysis and reduce cognitive load when navigating complex social media discourses. We

categorised them into three types: (1) AI tools, (2) social media monitoring and management tools, and (3) data analytics platforms.

The use of **AI tools** such as *ChatGPT*, *DeepSeek*, and *Claude* was frequently mentioned by experts to automate social media content categorisation, summarisation, and interpretation analysis. E6 explained that they used *ChatGPT* to evaluate their strategies with a focus on analysing sentiment and addressing potential biases:

> "We use *ChatGPT* to understand the sentiments [...], and to really improve on our content structures [...] for making sure that we are not producing biased content, we use it to improve on that aspect and understand the needs of our customers."

E7 reported using internal AI-based tools to summarise discourses at scale:

> "We have internal tools that work for getting all the employees' public discussions, and company internal meetings' transcripts to rationalise and summarise the content [for analysis]."

Besides, the use of **social media monitoring and listening tools** was described to manage more than one social media account of an organisation. Commonly mentioned are *Sprout Social*, *Sendible*, *Brandwatch*, *Brand24*, *CrowdTangle* and *Google Trends*. These commercial platforms were mentioned as they facilitated experts in their current multiple processes involved in decision-making. These processes range from engagement tracking, data collection, sentiment analysis, brand monitoring, to the identification of emerging trends. For instance, E6 explained how *Sprout Social* supported their content strategy:

> "We're able to [use them to] differentiate between the positive and the negative [discourses] and in the long run, we're able to understand the needs of our prospect."

E5 highlighted the tools' role in evaluating marketing performance as "it helped to compile those engagement details and the metrics."

They used these insights to assess how their campaign strategies were performing and adapt their decisions for real-time adjustments.

However, experts also mentioned that the selection of which tool to use can vary based on different goals. E10 explained that choosing between *Brand24*, *CrowdTangle*, or *Google Trends* depended on the communication channel and audience:

> "Different communities use different social media and different ways of communication [...] so most of the time we adopt [different] tools that are designed for that source of communication."

E1 highlighted *Sendible's* role in simplifying multi-platform management:

> "It helps you connect all your social media accounts [...] it tracks your social media metrics and algorithm. It kind of helps to know which posts are related."

E3 described using *Google Trends* to integrate demographic and regional insights into their workflow for monitoring global issues. They explained, "[to identify] what's happening all around the world," tracking regional trends and public sentiment. They provided the example of immigration debates:

> "Let's say right now we're talking about immigration in the [COUNTRY], and not everybody appreciates what [PERSON] is doing there. But there are also lots of people who support it [...] We're trying to look at the demographics and see how many people are talking about a particular topic [...] that's [constantly] evolving at that point, to help inform our decisions."

While these tools primarily support exploratory analysis or provide statistical insights, other tools, such as *Hootsuite*, focus on management tasks. This helps experts efficiently coordinate content and monitor multiple platforms. The summarisation of how these tools

demonstrate how experts use automation within the analysis process to navigate social media and make informed decisions.

Data analytics platforms

Data analytics platforms allow experts to conduct more specific and refined analyses tailored to specific goals. While social media tools provide real-time trends and sentiment, platforms such as *Excel*, *Google Analytics*, and *IBM Statistical Analysis* offer greater flexibility for structuring, managing, and visualising data. These platforms are useful in cases where experts want to revisit datasets, track long-term patterns, and generate customised statistical insights. E8 explained the value of such platforms in their workflow:

> "It helps when everything is arranged in Excel. If I need to extract a particular piece of data or a specific figure, I just go there. And if I need to improve on it, I use it for that too."

This demonstrates the role of analytics platforms in enabling flexible and detailed decision-making.

THEME 4: LIMITATIONS OF CURRENT APPROACHES IN WORKFLOWS

Concerns about **data-related issues**, such as *data overload* (E4), *overwhelming* (E1), *time-consuming* (E9) nature of qualitative analysis. Depending on context, experts noted that they have flexibility when things can afford more time for accurate analysis, but there are also times when these are time-sensitive. Expert also struggled with *misleading metrics*:

> "The biggest challenge you have is misleading data because in the social media space, there are times when your posts are trending for the wrong reasons. A post may receive a lot of engagement, comments, and interactions, but

> the attention could be driven by negative reactions or criticism. The challenge is to distinguish the positive signals from the negative ones."

Using an automated approach still requires a lot of corrective effort from humans due to *discrepancies in those results* (E6) provided from different tools. Concerns about the **capabilities** of these existing tools, such as *the lack of predictive capabilities* (E7) for forward-looking insights to support proactive decision-making or *limitation of integration platforms*(E7). E7 that uses an internally developed tool noted:

> "Compared with other software, currently it has only been integrated well with [a specific platform], and we cannot work it out with other social media platforms like *X*, *Instagram* or some of the conversation part."

One of the challenges raised relates to the lack of real-time data collection, which limits the ability to detect issues early and respond promptly. As E11 noted, "you don't get to know about the impact earlier[...] until it is deep into your data," explaining that their current process involves collecting and analysing data every month. For instance, in monitoring vaccine uptake, insightful details only emerge at the end of the month, delaying opportunities for timely intervention.

The **lack of qualitative analysis to support investigation**. E5 described the limitations of existing tools that only look at the statistical numerical aspects of social media and lack contextual understanding:

> "Rather than just focusing on the numeric aspects [...] should be able to identify [...] towards the sentimental aspects of [comments] that are positive, and the ones that are negative."

They further explained:

> "We have to go back and start digging deep to find out the origin of the problem [...] we try to get feedback from

> people and also we look into some of the comments [...] and compare [...] we still need an upgrade to be able to get to those problems faster."

Meaning that manually investigating these comments and gathering user feedback to trace issues is challenging and exciting. The state of their current tool lacks this supportive mechanism to help trace the origin or cause of negative virality. E9 noted that qualitative analysis yields more valuable insights when dealing with less frequent but more contextually dependent disocourses:

> "For example, with body donation, there's much less content available, but it's often much closer to what we're trying to understand. Although the volume is low, the content tends to be more valuable because it involves medical collaborators and people directly connected to the topic. In contrast, organ donation is more organised [...] So, the perspective can feel more distant. With body donation, it's often a very close, personal experience [...] In these cases, qualitative analysis is necessary, but it's also more time-consuming to do manually."

In addition, experts stressed the **accessibility** around cost, usability, and learning curves required to use automated and software tools. E8: "If I don't pay for some of these tools, I might not have access to them." E2 raised the affordability concern for small businesses:

> "Some of these tools can be expensive, you know, especially for a small business that's trying to start up, and you want to go for a large market to push a product, it would shake budgets and cause discomfort."

They further noted the steep learning curve required to adapt these tools as "some tools are difficult to learn and you know it requires some kind of training." E10 explained the lack of accessibility open for researchers as a challenge of how these tools are becoming more and more difficult or unstable to access:

> "*CrowdTangle* going down because Meta (formerly *Facebook*) decided to end that support and they are not sharing at all for free now or for the open public with that access to a API and you can't use that anymore [...] then the same is regarding the *Twitter* portal where *Twitter* before granted access to researchers, even for free or with small amounts to be in a position to check the narratives and sentiment and research that now it's all paid or behind the special requirements and and it's not that easy to use it anymore."

Apart from these, the inclusivity of these tools still lacks consideration of broader languages and cultures. For instance, E9 noted that tools developed in one setting might not apply to another setting because societal challenges are different according to the countries' dynamics:

> "There are so many things that weren't taken into consideration [...] we have over 300 languages. So imagine you're trying to conduct an analysis [...] it's impossible for me to learn all languages and cultural contexts."

When it comes to adopting the use of tools with AI, it is not unfamiliar that the research communities are still making efforts around **trust and ethical concerns**. Hence, it was no surprise that this was raised similarly when experts mentioned that they still lack confidence in trusting AI systems. Working in medical research, which deals with sensitive data, E9 noted:

> "I don't use AI [...] Because it consumes a lot of resources and energy, and also it's trained according to its bias. It's very exploitative from what I know. There are a lot questions as to how it is trained and also the algorithms they have make it a black box. It's very unclear how they work [...] it's highly unreliable."

E2 noted their primary concerns: "AI makes work faster, and reduces complexity [but] collecting and storing sensitive customer data

can also raise privacy concerns." These limitations reflect experts' perception of the current methods adopted in their workflow for analysing social media discourses to support decision-making. Some directly relate to the interactive design of *DiscourseDash*, while others point to broader considerations for the development of future tools.

7.3.2 *Perceptions during expert's interactive experience with* DiscourseDash

The findings in this section address **RQ1**, focusing on how experts perceive and interact with a dashboard tool that provides a *Detailed Analysis* with discourse-level analysis to support informed decision-making. Building on the analysis of experts' workflows, this section explores their perceptions of *DiscourseDash*. We explore experts' interactions with *DiscourseDash*, assessing alignment with their practices, support for analysis and decision-making, and opportunities for improvement.

THEME 1: SUPPORTS THE BALANCING OF TIME AND EFFORT TO ALLOCATE PRIORITIES

The *Overview* was described as helping experts make split-second decisions about which issues required immediate attention (E6). Experts emphasised prioritising social media topics based on their potential harm regarding brand damage or public response. Identifying and addressing harmful content emerged as a key requirement for supporting timely and effective decisions. E5 illustrated how a harmfulness indicator can support resource allocation across issues of varying severity:

> "The more harmful ones might bring down your brand very fast [...] so when making decisions, we prioritise those with high impact first, and the ones with low to

> medium effect come after. We can't prioritise [topics] that are less likely to have a harmful effect on the brand."

This illustrates how the harmfulness indicator actively supports operational decision-making for risk management, prioritisation, and balancing resources. Even when harmful topics lack immediate solutions, experts noted they can still be prioritised due to their potential to escalate quickly. This allows organisations to balance resources between urgent threats and less harmful, manageable topics.

Another important point mentioned by experts is that it would be better to determine relevancy and response strategies if the overview of *Popularity Over Time* was coupled with temporal dynamics. Having this feature could support experts in distinguishing short-term attention spikes from longer conversations that may signal more serious or lasting public concerns. E4 described how understanding a topic's virality can guide prioritisation by revealing its current and potential reach:

> "It's important to know what is in virality at the moment [...] to know the range. OK, is it just a temporary thing? Is it something that has been going on, getting more audience over time, or is it just [...] a momentary buzz? After knowing that, whatever is trending becomes very important."

This illustrated how tracking the momentum of virality helps experts make decisions about whether a topic needs immediate action or further monitoring. Similarly, E3 highlighted topic longevity as an indicator of importance and engagement:

> "If the topic is well spoken about, definitely we'll have to check the time at which people have been talking about it. Some topics just come, and in 10 minutes they're out [...] but topics that last for days, those are the ones we check out."

Experts perceived this feature as useful for filtering noise and focusing on persistent topics that may signal long-term risks. These demonstrate how our proposed analytical features are valuable for helping experts strategically allocate attention and resources to high-priority issues.

THEME 2: MONITORING PUBLIC DISCOURSES SUPPORTS TARGETED INTERVENTIONS

Experts highlighted how *DiscourseDash* can support them in monitoring social media discourses for supporting timely decision-making in their domains. They perceived it as more than just a monitoring tool. They mentioned how it could also support the development and adjustment of relevant interventions. For instance, E1 stressed the importance of public stance over personal opinion:

> "I won't be the one to decide if my products are what the market needs. It's actually people's opinion that matters, not mine."

Similarly, E4 described how sentiment analysis supports decision-making by revealing public preference as "[It] helps to know how people are leaning towards whatever is going on. What decisions are they making?" Building on this, E7 describes how employees' discussions, such as debates around remote work, can inform the changes within internal strategies and workplace policies:

> "If there's a topic, there's really a heated argument about the work hours for us in the corporate setting [...] we'll align a lot of [workplace] strategies and probably policies accordingly [...] and redesign our workflow setting [...] and align those benefits as well."

Besides, *DiscourseDash* was seen as relevant for tracking sentiment shifts and evolving narratives, which are important for timely interventions. E2 described how the tool could help anticipate responses

to vaccination campaign sentiment by analysing discourse across key demographic groups before and after campaign announcements:

> "Before even the information about the second dose was released to the public, I would just search the media [...] for the narratives and then based on that, for the sentiment [...] and then trace the same keywords [...] How the narratives and sentiment change over time."

They further explained how this process would unfold through different stages of the campaign:

> "From the time before the second dose information was even released, I would already be monitoring how the public was discussing it. Then, once the information became public, I would observe how anti-vaccine groups introduced their narratives and how different age groups or community segments reacted. I would continue tracking this over time, during the rollout of the intervention and afterwards, to understand how narratives and sentiments evolve within specific age groups or community groups."

This demonstrates how the expert's workflow aligns with the design of *Trending Topics* and *Sentiment Distribution*. These analytical features were perceived as helpful in informing appropriate strategies for public health messaging that address emerging concerns or misconceptions. They further noted that this kind of analysis not only supports immediate responses but also long-term planning:

> "Based on [*Trending Topics* & *Popularity Over Time*], it is now possible to create interventions for current goals. But it's also important for the next time, for future doses or other vaccines. How do we prepare communities, especially vulnerable ones, to accept the doses? Or maintain a positive response? If one community responds well according to *Sentiment Distribution*, there's an opportunity to study what makes them different."

Experts explained how the tool can inform future strategies and help tailor responses to different communities more effectively by analysing the outcomes of initial interventions through *DiscourseDash*'s outputs. Apart from health contexts, experts also highlighted how harmfulness labels support strategic communication. For instance, E6 described: "I'll be able to determine how to mitigate the harmfulness of the content and make sure I'm marketing to the right people. Because if you're marketing to the wrong people, it can actually increase the harmfulness of engagements."

THEME 3: RE-EMPHASIS ON THE NEED FOR TOOLS TO PROVIDE QUANTITATIVE ASPECTS BEYOND STATISTICAL VISUALS

Before navigating to *Detailed Analysis*, a key finding from our interviews was the consistent need amongst experts to move further than just summary metrics. This means allowing the navigation towards underlying discourses that drove these public responses. This need was expressed even before participants were introduced to the *Detailed Analysis* feature in *DiscourseDash*, which allows for deeper investigation of filtered conversations.

This emphasis on qualitative analysis validates our design decision to include capabilities of detailed analysis from the discourse perspective in *DiscourseDash*, emphasising a broader gap in existing analytic tools. As E7 noted, while the visual distribution of data points met their expectations, they also highlighted the challenge of bridging qualitative insights with quantitative analysis, emphasising that "there's a need for us to translate qualitative data into measurable terms."

We observed a similar pattern to when experts describe the limitations of their current workflow. They repeatedly noted that high-level statistical indicators are only sufficient to an extent. To further understand public sentiment, experts need to trace these high-level analyses back to the underlying conversations. For instance E1 explained:

> "You have to understand if you're actually doing better this week or you're not doing so well. You have to know the cause. This would actually help you strategise your next move."

Similarly, E5 stressed that identifying general trends is not enough: "We need something specific [...] we need to put ourselves in their shoes, so we need to get exactly what the problem is." To highlight the need for this in experts' workflow, E8 elaborated on how qualitative analysis fits into their decision-making process:

> "After the employer has [...] looked through the charts, looked through the engagement and all that, the person wants to read why people don't like this particular service."

These insights reveal the complexity of how experts interpret social media discourses, not only through metrics but by understanding narrative shifts and contextual factors. Current tools prioritise scale over interpretation of discourses, leaving a gap in addressing these needs. *DiscourseDash* was designed to bridge this gap by providing qualitative insights.

In the next theme, we examine how experts perceived and aligned the *Detailed Analysis* of discourses with their existing workflows after it was introduced. We explore how *DiscourseDash*'s ability to select topics, filter conversations, and view automatically labelled stances and sentiments aligns with their needs for understanding of public responses.

THEME 4: THE CONCEPTUAL IDEA TO PROVIDE A DETAILED ANALYSIS OF DISCOURSES ALIGNS WITH EXPERT PRACTICES

The ability to investigate not only what is happening in public discourse but also why remains a key limitation in many current analytical tools. Before introducing *DiscourseDash*'s *Detailed Analysis*, experts consistently emphasised the need for qualitative insights. One

key need was to contextualise the underlying drivers of patterns observed in visual analytics.

This need directly informed our interactive design and reinforces its relevance. Following our introduction of the *Detailed Analysis* page, experts acknowledged that it not only aligned with their current investigative practices, but also addressed their previous limitations. For instance, E1 described how being able to access actual social media conversations with structured sentiment and stance labels supports interpretation efforts:

> "Having to know words, months, people talk more about the product and the reason why it's trending and also why people talk negatively about it [...] kind of gives you a heads up to know what's coming [...] and I can choose to know why people go against the product and how many percent of the people going against the product."

Here, unfolding analysis on a discourse level becomes a means of rationalising decisions. This allows experts to stay ahead of emerging risks. E2 expanded on this by noting how the combination of stance and sentiment dimensions supports their understanding of public sentiments:

> "[It] would really work well because [...] it gives you a first-hand degree of information about these topics. You know, involving the intensity of these emotions and understanding our public opinion and sentiments."

These granular details from the fast-moving discursive nature of conversation were perceived as relevant for staying aware of the depth of emotion in situations. Besides, E5 directly pointed out how this type of analysis on discourses is lacking in their current workflow:

> "I was talking about the discourse earlier because I haven't even seen it before [...] we need to get to the root of the

> problem. We need to read what people are saying, their experiences, their concerns."

For these experts, it's important to unfold public voices to make sense of trends, controversies, or misinformation. Similarly, E6 discussed that when a potentially harmful topic begins trending, they would investigate the replies alongside stance and sentiment indicators for a more elaborate examination:

> "If something is very harmful, I want to understand what people meant by that, I want to see the replies, I want to see their stance [...] analysing those replies helps us understand if more [on why] people are 'for' or 'against' the product. That knowledge helps us get better at delivering to customers."

These signals were valued for providing evidence of real-world impact and public investment on a topic. Apart from reactive monitoring, experts picture using it for strategic applications. E4 reflected on how these discourse-level analyses could inform long-term planning:

> "We can [use this feature to] shape a general view of how people are thinking [...] and we can redefine our products and services to better align with their readiness."

With that said, the tool was perceived as supportive for experts to forecast what might be upcoming and prepare for it. As they elaborated further, they drew on examples from trade policy and climate initiatives to describe how *DiscourseDash* could support anticipating future developments and informing strategic planning. In healthcare, experts noted that understanding public beliefs, hesitations, and motivations was key to designing effective interventions. As E8 reflected on the importance of presenting these justifications through *Detailed Analysis*:

> "These are people's opinions. These are people's thoughts. It's important for those providing services to know what

> people think about what they want to present [...] What are the fears, the sentiments, reasons why they do or don't engage."

Rather than relying on abstract metrics, they valued analytic tools that support the process of unfolding the public concerns and experiences. In addition, experts also noted that this interpretive depth supported their ability to shape responsive actions. As E11 explained that attaching sentiment analysis in a discourse context format can directly support timely decision-making:

> "Having seen this, this is how we will strategise [...] If the sentiment distribution is highly negative, we restart the system [...] we can bring in some interventions to get people to accept [vaccines]."

In these demonstrations, qualitative insights functioned as inputs to human efforts in investigation and planning. Across domains, experts valued the distinct concept of *DiscourseDash* for supporting their reasoning and investigative judgment.

THEME 5: MISALIGNMENTS WITH EXISTING ANALYTICAL PRACTICES

Experts consistently highlighted missing key features essential to their daily workflows, revealing gaps in our initial design that are crucial for aligning with established practices. Addressing these misalignments will guide iterative refinement to better fit real-world organisational needs and sociotechnical contexts.

A frequently raised gap was the ability to **filter demographics** of analysed discourses. Targeted demographic details such as age, location, language, and behaviour are central for tailoring strategies and contextualising sentiment. For instance, healthcare experts have emphasised the importance of segmentation in designing interventions for specific audiences. E10 explained:

> "It would be nice to filter by, say, 30–45-year-old mothers in [country] [...] so I could create specific reports and track changes over time for preparedness and response."

Experts also noted the **absence of hashtag analysis**, which is important for monitoring trends since hashtags usually serve as entry points into the overwhelming volume of discourses.

Experts described the current **sentiment analysis as too abstract** for decision-making. Instead of just positive, negative, or neutral categories, they needed more in-depth insights into emotional intensity. E2 explained this through an analogy:

> "Imagine giving a child a biscuit. If the child doesn't want the biscuit and is forced to take it, that's not the same as the child eagerly accepting the biscuit."

This analogy illustrates the importance of understanding whether someone's emotion or reaction is genuine. To measure this, the expert suggests using a graph to visualise the intensity of emotions. For instance, a graph could show that 70 per cent of people strongly support an idea, while 20 per cent only somewhat support it. They suggest allowing for a more accurate representation of how deeply people feel about a topic of discourse. Similarly, E7 noted that they:

> "Need to understand the degree of how much people are engaged, how strong the negativity is. That helps us dig into the intensity of how people feel about certain things."

For instance, adding a 'confused' category to capture people who are undecided or unsure, which is crucial in public health contexts like vaccine uptake (E8). Tracking **sentiment over time** was also emphasised as crucial for responsive interventions. E10 noted:

> "Without that information, you can't build any future response or interventions because you don't know the changes [of sentiments] over time."

Experts also called for better **contextual and cultural relevance**. Large data volumes are less useful without alignment to specific contexts that are relevant to their decisions. Filters and indicators need to be extended for regions with diverse dialects or sensitive cultural environments. Some research experts preferred manual, qualitative interpretation over automated tools, arguing that qualitative categories must emerge through human evaluation. E9 explained:

> "Every research is different and embedded in a cultural context that is very specific [...] We have had to make the categories [analysis aspects] ourselves [...] these kinds of tools cannot give you that flexibility."

This reflects a broader gap in the current analytical tools' ability to support experts conducting targeted or high-stakes social media analysis. Although this theme has identified specific misalignments with existing analytical practices, it reveals a broader tension in designing a tool to balance specialised functionality for expert workflows with general non-expert users. The importance of HCI considerations in mediating between automated outputs and diverse user requirements is emphasised when these misalignments highlight certain expert needs that general users did not address or may contrast. The implications of this tension are explored further in Chapter 8.

7.3.3 *Expert's reflection regarding to* DiscourseDash*'s real-world applicability*

The findings in this section address **RQ3** about the key design implications that guide future iterations of *DiscourseDash* and enhancing its support for decision-making across different professional contexts. Experts' overall reflections provided insights into the practical integration of the tool, including adoption barriers, use cases, and recommendations for achieving real-world relevance of the design.

THEME 1: PRACTICAL APPLICATIONS IDENTIFIED BY EXPERTS

Experts reflected on how *DiscourseDash* could support their or their organisation's decision-making by fitting into existing workflows for monitoring conversations, informing strategies, and responding to emerging issues.

Monitoring real-time discussions

Experts emphasised *DiscourseDash*'s value in unfolding emerging conversations and reducing manual effort across platforms. This supports early detection of sentiment shifts and discourse attention for preventative actions before narratives spread. For brand or campaign monitoring, it was viewed as a diagnostic tool for detecting timely matters. As E6 described:

> "[I] need to know why people are talking [negatively] about [my brand], so if we realise it's the price, that way we can try to [fix it quickly]. If you don't know exactly the problem, it's gonna take a longer time before it comes to a solution. So [*DiscourseDash*] is gonna help us get to the root of the problem, help us find the starting point to solve these problems."

These reflections showed experts perceived *DiscourseDash* as supporting strategic analysis, moving from statistical monitoring to understanding the drivers behind public discourses. Assessing whether trends are relevant or spreading across platforms was crucial for determining response scale and urgency. In journalism and media planning, experts envisioned using *DiscourseDash* to align content strategies with public attention in real time. E4 described:

> "So I put up the date [to compare] pre-Grammy to after Grammy [...] it gives me the summary and brings out the keywords [...] the sentiment if it's +1 the charts will show this particular reading [...] I get to know the number of

> people engaging with that particular topic at a particular set time, it just makes life easy for me."

Rather than relying on assumptions about audience interest, experts perceive *DiscourseDash* in helping them navigate discourses with informative analyses that are relevant to their workflow. Across domains, they highlighted how the tool helps identify when conversations start, gain momentum, and why public attention shifts. These insights were perceived as valuable for staying informed and making rapid decisions in fast-paced settings.

Informed decision-making around planning and evaluations

Experts noted that *DiscourseDash* could support planning and evaluation throughout an intervention, not just at the start. Instead of one-time decisions, *DiscourseDash* would facilitate flexible management of ongoing processes. These processes involve refined actions on the expert's end because public narratives shift and public sentiment directly affects outcomes of organisational goals. For instance, E10 explained how they would use *DiscourseDash* across different phases of health intervention planning:

> "It's really helpful for monitoring and evaluation of [health] interventions, which gives the opportunity to scale up or scale down depending on the needs as you go. After the intervention, it helps to reflect and create preparedness plans for future interventions in the same context, so for vaccines, or for flooding"

Here, discourse insights were valued not only for real-time response but also for learning and preparing for future scenarios. Instead of waiting for traditional reports or surveys, *DiscourseDash* provides immediate indicators of public responses. Similarly, E6 highlighted how analysing public reactions to discourses could help refine product development:

> "When a new product was launched [...] we wanted to know, read their feedback [...] [with this] we're able to determine how to improve the product."

These reflections suggest experts perceived the tool as a method to monitor impact in real-time, guide adjustments around decisions, and inform future strategies. Instead of waiting for formal feedback (e. g., surveys, sales stats, delayed reports), *DiscourseDash* offers quick insights into public signals. This supports evidence-based decisions grounded in ongoing discourse, enhancing expert workflows for both real-time action and post-implementation reflection.

Managing crises and reputation risks

Experts emphasised *DiscourseDash*'s value in crises like public opinion shifts, misinformation, or reputational damage. Quickly identifying the core issue was seen as a structured alternative to current workflows for a timely response. As E4 explained: "in crisis management [...] when you know what the issue is specifically, it should be easy to turn things around." They further explained how the visualisation and sentiment indicators would allow a faster problem diagnosis for response coordination:

> "[Knowing that there is damaging] information about my brand that's going on, I'll use it [*DiscourseDash*] to understand the sentiment of the audience [...] it would help in making decisions [according to] response protocol and there's like a graph to back it up showing the range, so it's very explicit and it makes things easy to comprehend."

DiscourseDash helped navigate overwhelming discussions by providing a structured overview of concerns for quicker identification of root causes and targeted actions. For instance, E5 explained it could pinpoint a specific problem during a product backlash:

> "[I] need to know why people are talking [negatively] about [my brand], so [for example] if we realise it's the

> price, that way we can try to [fix it quickly]. If you don't know exactly the problem, it's going to take longer to find a solution. So the [*DiscourseDash*] is gonna help us get to the root of the problem, help us find the starting point to solve these problems."

These reflections illustrate that *DiscourseDash* facilitates timely understanding of the 'why' behind negative public reactions or emerging narratives, which is more valuable than merely detecting issues like sentiment shifts. In crises, fast decision-making impacts outcomes, so *DiscourseDash*'s support in navigating social media to identify root causes is crucial for effective response and strategy alignment.

Shaping decisions dependent on understanding public views

Experts also perceived *DiscourseDash* as a tool that guides reflective decisions around brand positioning, internal policy, and strategic communication. These uses focus less on immediate response and more on understanding how public perceptions evolve. Having the analysis of social media discourses in detail, experts can track shifts in tone, values, and alignment with intended identity. Hence, *DiscourseDash* becomes more than a monitoring tool—it's a resource for staying informed through broader narratives. For instance, E7 explained how discourse insights inform workplace policy design:

> "In terms of designing the perks and benefits for all our employees, that is something we can incorporate to understand the trends [...] there are a lot of culture changes [...] So we can redesign [things] with a set of trendy [ideas] that could improve health, productivity, and the happiness index."

This illustrated how *DiscourseDash* can support experts in aligning decisions with a broader understanding of employee sentiments and social media discourses. In turn, this supports a healthier workplace identity. Instead of relying only on assumptions or surveys, experts

use *DiscourseDash* to gain accurate insights by analysing employees' discussions about their work environment to inform strategic reflections.

Following public voices to shape societal values

Experts also valued *DiscourseDash*'s potential to track evolving narratives in policy debates, public health campaigns, and social movements. Apart from reactive monitoring, the support was perceived to uncover deeper layers of the public's values, concerns, and perspectives for strategic alignment with ongoing or emerging policy development. For instance, E11 described using *DiscourseDash* to inform government decision-making:

> "If the government asked me to survey what is currently going on, on ways to improve people's perception about them, I would use everything you shared with me [*DiscourseDash*] to include people's opinions, what people are trying to pass on to the government. Like, if John Doe says, 'This is a way the government can help the people [...]', it's my job to write that down and explain [to the government] what led to that suggestion."

This highlights how *DiscourseDash* acts as a tool for mediated listening, helping experts translate public voices from social media into suggestions output tailored to institutions' goals. This demonstrated the *DiscourseDash*'s potential to help governments identify emerging concerns, trace the roots of public suggestions, and communicate these strategically to decision-makers. From a broader perspective, this would support the expert's responsibility in bridging public expression and formal institutional processes.

THEME 2: ADOPTION BARRIERS IN REAL-WORLD

While experts acknowledged the value of *DiscourseDash*, they also highlighted several challenges that can limit adoption in practice.

These barriers arise from organisational constraints, resource limitations, and workflow considerations. Experts further raised concerns about data privacy, ethical use, and the contextual fit of *DiscourseDash* within professional environments. This section outlines these adoption barriers and sets the stage for considering their implications for future iterations.

Resource constraints

Many experts from smaller organisations highlighted that limited budgets and technical capacity can set back the adoption of new tools. These experts stressed that an interactive design should be easy to use for seamlessly integrated into their existing workflows. This includes requiring minimal training and avoiding substantial infrastructure modifications. Drawing on prior experience with similar tools, experts emphasised the importance of long-term reliability. This helps prevent disruptions and rising costs caused by inadequate maintenance or organisational changes. These observations underline the need for future tools to prioritise sustainability and dependability to effectively support ongoing workflows.

Accountability

Experts raised key concerns around data privacy, ethical use, and transparency in algorithmic processing. These concerns reflect broader sensitivities involved in monitoring public discourse at scale. They stressed that any system design must prioritise ethical safeguards. This is because social media discourses frequently engage with sensitive topics. It is therefore important to ensure user identities are protected. Privacy must be safeguarded, and fairness should be promoted throughout the analysis. Such considerations align with principles of responsible AI and sociotechnical accountability for the need for ethical governance throughout the tool's lifecycle.

Organisational structure

Experts in hierarchical or collaborative settings emphasised the need for role-specific permissions and access controls. They drew paral-

lels with how platforms like *Meta* manage page roles, where different users have varying levels of access and responsibilities. In practice, analysts, managers, and administrators may require different views or editing rights within the same workspace. Experts mentioned the need for tools to support diverse team workflows while preserving data security and respecting organisational protocols. Addressing these challenges influences how large teams adopt such tools by enabling collaboration without compromising sensitive information or disrupting decision-making.

Real-time integration

Experts emphasised the importance of real-time performance in time-sensitive situations. Low latency and immediate data processing were considered essential for actionable insights. They highlighted that seamless integration with their existing real-time monitoring workflows and platforms is a fundamental for adoption. Achieving this requires back-end optimisation and a flexible system design. Such interactive design ensures that tools can adapt to real-world contexts and support decision-making without delays or technical bottlenecks. These considerations are especially important when discourse-level insights directly inform operational or strategic choices.

Adaptability

Experts highlighted limitations in supporting diverse languages, regional dialects, and culturally specific slang. They explained that failing to account for these variations can lead to incomplete or biased insights for contexts within global or multicultural levels. There should be a prioritisation of inclusivity. This includes building multilingual and analysis that is aware of cultural contexts for accurately capturing the complexity of different public discourses. With that said, decisions can be implemented with more representative and actionable insights across varied populations. Hence, adaptability towards regions and contexts is important for ensuring that discourse-level analysis remains relevant and sensitive to local dynamics.

Reliability

Experts with qualitative backgrounds expressed scepticism about automated analysis that oversimplifies complex discourses. This stems from their experience where they notice these analyses frequently fail to capture subtle sentiment, irony, or specific meanings relevant to context. Hence, they suggested better communication in interactive design on how outputs are generated, including the data sources, assumptions, and confidence levels associated with analyses presented to a user. These would help users interpret results critically and mitigate over-reliance on inaccurate outputs. Without these reliable measures, experts hesitate to integrate new tools into high-stakes workflows because they do not trust them with decisions that can have social, political, or organisational consequences.

THEME 3: CONSIDERATION OF EVOLVING ANALYTICAL WORKFLOWS

Experts reflected on how their workflows might adapt to the rapidly changing nature of social media and emerging analytical practices. Their insights highlight key opportunities and set expectations for future tool development. Considering these perspectives ensures that *DiscourseDash* remains relevant and adaptable over time. It also supports decision-making workflows that account not only for current dynamics but also for near-future shifts in social media discourse.

Integration of customised AI-based features

Experts envisioned future workflows enhanced by collaborative AI to support proactive and informed decision-making. Apart from simple analysis to classify text, they mentioned the potential for AI-assisted processes such as summarising past events, outlining strategies, and responses that further guide decisions they make. These designs could include instruction-based prompts and adjustable recommendations embedded within *DiscourseDash*. For instance, a conversational agent could support users by providing specific instruc-

tions that focus on analysed discourses and generate targeted insights relevant to their domain. Some experts also suggested incorporating features inspired by *X*'s (formerly *Twitter*'s) Community Notes as an indicator for future analyses. These features could help identify misleading content, track its spread, and provide clear contextual explanations. Incorporating such indicators into *DiscourseDash* AI analyses could serve to communicate transparency and support more informed decision-making workflows.

Adapting temporal analysis beyond real-time

Experts emphasised that temporal analysis should extend further from immediate reactions. They highlighted the need to understand how social media discourses develop over time, including before, during, and after key events. Observing these patterns allows experts to detect emerging trends, anticipate potential risks, and identify misinformation before it spreads widely. For instance, one expert, E9, explained: "You want to know where people stand at the beginning, how that shifts after an announcement or crisis," demonstrating the value of comparing data over time to guide informed action. Experts also suggested that combining real-time monitoring with historical data can reveal shifts in sentiment, stance, and engagement that may not be apparent from short-term observations alone. This approach helps identify which topics are gaining traction, which communities are most affected, and where interventions may be necessary. Having the ability to track temporal dynamics, experts can plan more adaptive strategies rather than reacting solely to immediate changes.

THEME 4: USABILITY IMPROVEMENTS FOR INTERACTIVE EXPERIENCES

During the experts' interactions with *DiscourseDash*, several key areas related to usability and desirable features were brought forward to improve the interactive experiences.

Filter for engagement metrics

Experts highlighted the importance of ranking discourses by engagement metrics such as likes, shares, or replies. Discourses with higher engagement should appear first for users to focus on discourses that shape or reflect broader audience sentiment. Implementing such filters would streamline navigation and support more targeted analysis.

Support for collaborative workflows

Collaboration was a key requirement for teams analysing social media discourses. Experts suggested features such as share buttons for data export and multi-level user management to allow role-based access. These mechanisms support team members to work according to their responsibilities while maintaining data security. Integration with offline data sources was also recommended to support flexible analysis.

Mobile accessibility and visual clarity

Experts noted that *DiscourseDash* should function across multiple devices to support on-the-go monitoring. Some visualisations were described as unclear or overwhelming. Suggestions included tooltips, clearer graph legends, and simplified layouts. These enhancements would make insights more interpretable and ensure consistent usability across platforms.

User-level analysis

Distinguishing contributions from credible sources and general users was seen as important. Experts recommended analysing user behaviour, including post frequency, duration, and repetition. This type of user-level analysis could reduce noise, identify patterns, and improve the efficiency of investigative workflows.

Integration of Community Notes

Experts proposed features similar to *X*'s (formerly *Twitter*'s) Community Notes. These would help identify misleading content, track its virality, and provide clear contextual explanations. Using credible,

traceable indicators alongside verified claims would improve transparency and support informed decision-making.

7.4 CHAPTER SUMMARY

This chapter concludes the final empirical study presented in Part II of the thesis. Through an interview study with industry experts, we gathered the practical use of *DiscourseDash* in real-world contexts. The study provided a general understanding of expert workflows that further highlights the relevance of this research. More importantly, it brought valuable insights into designing interactive tools for social media analysis.

While the previous mixed-method evaluation in Chapter 6 focused on general user interaction through scenario-based tasks, this chapter grounded our investigation in the knowledge and practices of industry experts. Through the integration of expert perspectives in the interview study, we identified specific design priorities and challenges that are frequently overlooked in non-expert evaluations. These findings underscore the important role of expert perspectives in shaping tool design for the actual decision-making workflow.

These insights not only inform the future refinement of *DiscourseDash* but also contribute to broader discussions on designing social media analysis through the lens of discourse-level insights.

Part III

SYNOPSIS

8

DISCUSSION

This chapter reflects on the main contributions of the thesis by synthesising findings across studies and relating them to the main research questions. First, let's briefly recap how the empirical chapters of the thesis collectively address these questions:

RQ1 How can interactive tools be designed to enhance social media analysis to support more effective decision-making workflows?

This question is addressed through the development and evaluation of an interactive tool. First, the practical implementation of a computational method for modelling discourse context in the stance detection task informed the initial interactive design (Chapter 4). Then, this design was developed into a functional prototype (Chapter 5) that facilitated a mixed-methods evaluation with general users to ground the design rationale with empirical insights (Chapter 6).

RQ2 How do industry experts' workflows and experiences in analysing social media discourse shape the real-world applicability of *DiscourseDash*?

This question is explored through a semi-structured interview study with industry experts (Chapter 7). The study examines expert practices, challenges, and expectations, providing empirical grounding for assessing the relevance and applicability of *DiscourseDash* in real-world analytic workflows.

8.1 DESIGNING FOR SOCIAL MEDIA ANALYSIS

The problem of social media analysis established throughout the thesis is that (1) computational learning methods frequently treat text

as separate instances and (2) interactive design approaches focus on summarising outputs without preserving the discourse contexts that support user interpretation during analysis (see Section 1.1 & Literature Review in Chapter 2). Both of these aspects, i. e., computational learning and interactive design, overlook the discursive and contextual complexities of social media discourses. This disconnect motivated the central thesis argument, in that integrating discourse-level computational insights with human-centred design can better support real-world decision-making. This is achieved by *DiscourseDash* designed in this thesis to support human users in interpreting automated outputs together with the context that lies within social media discourses (see Chapter 5 for the designed prototype).

From a computational perspective, engineers focus on optimising automated outputs for performance. In contrast, a human-computer interactional approach focuses on human values and behaviours in shaping interactions. This research values both approaches by starting with a computational study to inspire the design concept and then grounding the design through empirical investigation of user perceptions (see Chapter 3 on the connection of these multidisciplinary approaches). In this thesis, the overall findings suggested the value of *DiscourseDash*, in that interactive design of automated output with discourse-level insights allows users to better interpret the analysis for informed decisions.

8.1.1 *Grounding design concepts*

The exploration study of computational methods in Chapter 4 revealed that algorithmic outputs are only as capable as the context from which they are fed to learn. Treating social media discourses as unified instances during implementation on a computational learning task showed how context shapes predictions (see Section 4.5). Hereafter, these insights from the computational learning study became central to the inspiration of the proposed interactive design

(see the relation in Table 6). The value of this computational study from a high-level perspective is that it not only grounded the design in practical realities but also made sure that it remained feasible and avoided overly optimistic assumptions. As a result of this exploration, the findings suggested that the impact on the model's performance after introducing discourses as contextual knowledge during learning shapes algorithmic outcomes. The inspiration here is that human users should also be introduced to the same discourse contexts that the machines learn from. This can be constructed as a way to communicate reasoning around the details (contexts) through discourses that supported a computational model in arriving at their final output. This approach aligns with HCI principles for transparency in AI systems because it was argued that they should be accompanied with reasoning for collaboration, not reliance (Doshi-Velez and Kim, 2017; Ehsan et al., 2021). The idea is reinforced around how interactive designs for social media analysis should present algorithmic outputs within the discursive contexts that give them meanings for informed decision-making.

This thesis addresses the points discussed above through two empirical studies: one with 26 general users performing scenario-based tasks (Chapter 6) and one with 11 industry experts reflecting real-world relevance (Chapter 7). Both user studies here suggest a consistent need for analytical depth in dealing with the process of using a tool for social media analysis. First, this was evident when experts described the limitations of their existing tools, which are limited in providing deeper contextual interpretation of the high-level summaries (see Section 7.3.1). This need for more in-depth analysis became even more apparent during their interaction with *DiscourseDash* (see Section 7.3.2). Initially, as they explored *DiscourseDash*'s *Overview* page, they acknowledged its similarity to current tools or methods they had adopted in use. However, after interacting with the *Detailed Analysis* page of *DiscourseDash*, experts noted that it effectively addressed the interpretive gaps they had consistently mentioned all before the interaction on *Detailed Analysis* in *DiscourseDash* (see Section 7.3.2).

Similar evidence was also observed when general users perceived the *Detailed Analysis* that presents discourse-level insights as highly useful during the decision-making tasks they were instructed to perform (see Section 6.3.3.2). Apart from reflecting it through statistical analysis of the questionnaire, it was further strengthened when they articulated their reasoning using these discourse-level insights during tasks (see Section 6.3.3.1). Findings from both user groups indicated that these perceptions were closely related to the ability of *DiscourseDash* that guides their understanding of complex discourses. These shaped the value around *Detailed Analysis* in facilitating users during deeper investigations into unfolding narratives.

8.1.2 *The values within the design*

Apart from validating the design concept, the findings reveal broader lessons about how social media analysis can be supported through interactive design. Chapter 1 outlined the motivation for this thesis, where existing tools frequently prioritise high-level metrics or visual summaries. This research identifies an opportunity to shift toward supporting interpretive analysis by focusing the design around the context of social media discourses. This shift highlights how interactive systems can be structured to accommodate the complexity of social media discourses (Nelimarkka, Laaksonen, and Semaan, 2018).

Synthesising observations across both studies points to a design direction in which discourse context should be treated as a core interactive component rather than an add-on to computational learning processes. This can be perceived as an extended form of support to make sense of complex social media analyses by leveraging surrounding discourses accompanied by interpretive indicators to guide investigations. To strengthen this point, empirical findings (see Section 7.3.3) showed that industry experts valued this perspective by noting how these detailed explorations of discourse-level insights ex-

tend high-level visual summaries towards a more active approach in supporting their interpretation process.

A concrete example of these insights is the role of the harmfulness indicator. During the design phase, we formulated an initial hypothesis that such an indicator can support decision-making grounded in insights from prior literature. However, the concept remains vague due to its subjectivity (see Section 2.2.1.3). The performed evaluations with non-expert users showed that this feature prompted questions about its meaning and relevance within social media analysis. In contrast, expert users did not surface these questions around the harmfulness indicator. Some expert, in fact, articulated their interpretation of them in ways consistent with their existing workflows. This suggests that the initial hypothesis was relevant in real-world contexts, but must be clearly defined and presented in a way that still allows users to integrate their own interpretations. These findings reinforce the value of the feature while pointing to the need for further studies to refine its definition and presentation for effective analytical use.

These findings align with broader perspectives in HCI that emphasise the interpretive and situated nature of user engagement (Dourish, 2006; Ehsan et al., 2021). They highlight frequently overlooked user priorities such as the need to contextualise information within broader narratives (see Section 6.3.3.1 & Section 7.3.1) and to investigate the meaning behind computational outputs (see Section 7.3.2). Future work should consider both computational effectiveness and how design can support users. This means interpreting discourse contexts with supportive indicators to guide the understanding of complex social media environments.

Prior work in HCI and visual analytics has acknowledged the challenge of making complex systems understandable (De Choudhury et al., 2009; Heer and Agrawala, 2007; Majeed et al., 2020; Zubiaga et al., 2016). This thesis's contribution fits within the limited studies that have provided empirical insight into how discourse contexts shape the process of social media analysis. Through the design of *DiscourseDash*, this thesis bridges perspectives on designing both *with*

and *for* user practices by demonstrating how the interpretation of social media discourses can be integrated into the interaction itself rather than being appended as a context to improve computational outputs.

8.1.3 *Connecting the value to future directions*

This thesis encourages future interactive system designs to move forward from limiting social media analysis to a single dimension of quantifiable outputs. This is because our empirical findings appear to validate the value by providing multiple forms of evidence in designing interfaces that engage users with the interpretive insights of social media discourses. These evaluations of *DiscourseDash* provided practical demonstrations of how computational insights can be presented through interactive design to support users making high-stakes decisions in fast-paced real-world environments. For instance, this includes making appropriate intervention decisions to counter a public health crisis or decisions to manage an organisation's reputation following unforeseen incidents. More detailed examples are provided in Section 2.1.2.

The grounding of our research in empirical user engagement extends ongoing HCI discussions on transparency and the collaborative construction of meaning between humans and computational systems (Amershi et al., 2019; Shneiderman, 2020). Rather than treating computational outputs just as automated measures, these findings encourage a shift in design priorities toward supporting investigative reasoning when navigating complex social media phenomena. This highlights not only the role of computation in supporting human judgment but also the importance of moving past simple analytical interfaces toward designs that encourage exploration and reflection through discourse-level analytical insights.

The approach by understanding a design through human interaction experiences reinforces core HCI principles of participatory de-

sign (Maguire, 2001; Spinuzzi, 2005). Not only does it reveal the potential of this underexplored area, but it also demonstrates that grounding analysis in situated user practices strengthens the relevance of computational insights. By articulating these design considerations, this research lays a foundation for future research to improve the bridge between computational methods and practical users.

More broadly, the values within the findings of Chapter 6 extend the recognition of mixed methods in HCI research (Oghaz et al., 2020). This was demonstrated by how combining qualitative insights with quantitative metrics provides a complementary perspective for synthesising the values within a design. On one hand, users provided distinct ratings for the specific features under evaluation, which allows a quantitative measure of perceived usefulness. On the other hand, qualitative findings explained why features were evaluated in particular ways. E. g., it revealed how users connected the design's affordances to their goals and expectations. These demonstrated that an interactive design of a social media analytical tool can benefit from considering both quantitative outcomes and qualitative interpretations.

8.2 UNDERSTANDING THE REAL-WORLD APPLICABILITY

Building on earlier discussions, this section illustrates how *DiscourseDash* directly supports expert decision-making processes. This connection directly addresses RQ2 as to how industry experts can shape the real-world applicability of *DiscourseDash*. The literature review (see Section 2.1.2) initially established a conceptual link between professional responsibilities and decision-making processes by drawing on documented practices in fields such as policy, communication, and brand management. These helped frame the potential use cases for *DiscourseDash* within expert workflows for outlining design rationale. As part of the empirical findings from interviews with experts (see Section 7.3.3), this research contributes to documenting

how these conceptual ideas are applied in practice. The findings presented various forms of analysis in real-world settings, with the focus on analysing social media discourses. The following sections discuss how the findings revealed that experts see social media analysis as not just a straightforward task, but as part of broader workflows that interpret discourse meanings to guide planning, review, and strategic decision-making.

8.2.1 *Shaping values with experts*

Engagement with industry experts revealed that the process of analysing social media discourse is a strategic activity. Decision-making is not just based on conducting retrospective evaluations, e. g., assessing public reactions to past events; the context embedded within discourses is part of the process to perform proactive planning, monitoring, and strategising (see Section 7.3.1). These practices emphasise identifying emergent narratives, tracking shifts in public sentiment, and understanding the evolution of social media discourses. With that said, phenomena such as misinformation, ambiguity, and rapid change cannot be fully understood through simplistic or purely quantitative representations, e. g., categories of sentiment analysis.

Through interactions with experts, it became clear that although there is an increase in the use of existing analytical tools, they are limited in supporting these strategic practices (see Section 7.3.1). These analytical tool was described to prioritise aggregate metrics at the expense of contexts or explanations. These findings align with a prior study from policymaking that emphasised increasing engagement through HCI and social computing (Spaa et al., 2019). There were prior findings that also suggested that mismatches between tool capabilities and the interpretive demands of expert analysts would limit effectiveness when decisions require evidence-based justifications (Seidel et al., 2018). Thus, industry experts need tools that not just visualise what is happening with descriptive visualisation, but

also reason about why certain discourses unfold as they do. This indicates an underexplored gap in the interactive design of analytical tools that has not been sufficiently addressed.

Building on ongoing calls in HCI, our expert discussions highlight the importance of prioritising explanatory insights over purely predictive outputs (Hofman et al., 2021). Chapter 7 aligned how experts perceive *DiscourseDash* and outlined their current analysis processes. These findings clarified the types of support that interactive design must consider to be valuable in practice when designing for social media analysis. It has become evident that narrative correlation and interpretability of automated analysis are invaluable in practice (see Section 7.3.1 & Section 7.3.2). Overly abstract summaries lack contextual grounding, not only constraining their analytical processes but also raising concerns about potential unintended consequences arising from computational outputs. With all these insights, the thesis encourages that design approaches must consider the discursive nature of how online communication is formulated (O'Halloran et al., 2018).

8.2.2 *The evolution of perception*

The analysis in Chapter 7 uncovers an evolution of perception that substantiates the core argument of this thesis in emphasising discourse-level insights in interactive design. Before engaging with *DiscourseDash*, experts described their workflows as fragmented and labour-intensive. This means that the process requires a combination of multiple analytics platforms, manual data curation, and informal reasoning to make sense of social media discourse (see Section 7.3.1). These descriptions show that industry experts rely on ad hoc methods to bridge the gap between computational outputs and the external insights needed to support their organisational goals. Our findings reaffirm that social media problems cannot be explained or analysed through a one-dimensional design output (i. e., high-level visual

summaries of analysis), but rather to uncover more by synthesising further interpretive insights (i. e., investigation into social media discourses with supportive indicators).

Upon interaction with *DiscourseDash*, experts quickly recognised how features such as analysing discourse-level insights support unfolding narrative aligned with their existing analytical approaches and needs. They pointed out that these analytical features, i. e., stance analysis, sentiment analysis, harmfulness, trending topics, popularity over time, were considered useful not in isolation, but as entry points for exploring how certain topics gained traction, how public opinions evolved, and what strategies can better frame their goals. Rather than viewing them as passive summaries, these features were seen as an indicator to reason through complex navigation within discourses. Hence, *DiscourseDash* provides additional insights that naturally support navigation of discourses, as such discourse-level insight is highly valued. It became evident that explanations are most effective when they align with a user's reasoning process rather than purely offering technical transparency.

Post-interaction reflections further synthesise these findings. Experts articulated specific scenarios in which *DiscourseDash* could support their strategic workflows, from reputation monitoring to early warning for emergent issues (see Section 7.3.3). These reflections suggest that *DiscourseDash* has the potential not only to support expert judgment, but also to serve as a tool for navigating shifts in social media discourses over time. This aligns with the growing emphasis in HCI and AI research on developing human-centred systems that prioritise meaning for supporting human expertise, over fully automation in high-stakes domains (Amershi et al., 2019; Green and Chen, 2019; Leal et al., 2025), e. g., policy making, healthcare industries, government bodies (see Section 2.1.2).

All in all, many of the findings show that experts' perceptions of *DiscourseDash* were shaped through their interactions with the tool and their evaluation of its practical usefulness. This can be observed from the illustration of how these evolving perceptions support the

thesis argument, which is to facilitate the navigation of social media analysis through discourse contexts. This thesis consequently contributes to the broader conversations about the design of interactive tools for social media analysis by aligning with experts' workflows. The findings further highlight the need to design tools that reflect the practical realities of expert practices and decision-making processes.

8.2.3 *Key takeaways from experts' perspectives*

Engaging with industry experts revealed how context around discourses with indicators such as in *DiscourseDash* shapes the use of social media analytical tools. Apart from documenting their workflow practices, this engagement study highlighted what matters in experts' workflows. E. g., navigating overwhelming social media environments, addressing conflicting indicators, or making informed decisions under uncertainty. These show that designing for trust, flexibility, and support for interpretive reasoning is not an abstract principle, but a requirement illustrated through how expert articulated their workflow (Amershi et al., 2019).

Rather than limiting an interactive design to a set of predefined analytical features, experts emphasised the need for flexibility to accommodate evolving needs. Feedback from industry experts (see Section 7.3.3) highlighted forward-looking perspectives that account for anticipated shifts in foreseeable changes or expectations. These not only grounded the design in the lived realities of expert users but also informed how the interactive design can adapt to the evolving social media environments. Incorporating expert perspectives in *DiscourseDash* reflects a design that values expertise, encourages interpretive reasoning, and anticipates the evolving needs of real-world social media analysis.

8.3 BRIDGING HUMAN JUDGMENT WITH ANALYTICAL TOOLS

The previous section linked the findings to the design concept and how it supports analytical perspectives. However, the context around social or cultural settings requires decisions to balance experts' qualitative judgments with algorithmic outputs. A key challenge that emerged in this research is that social media discourses are often open-ended, with meanings shaped by the surrounding discussion context. The interactive design presented in the thesis provides a foundational step toward bridging computational analysis and human reasoning; i. e., it shapes automated analysis through interactive design in decision-making by balancing the view of social media discourses in a way to encourage human judgments. However, achieving this balance depends on refining how analytical tools accommodate and respect human judgment. This emphasis extends calls within HCI to design systems for human–AI collaboration (Amershi et al., 2019).

In Chapter 6, our qualitative analysis highlighted the importance of each feature based on the participants' rankings (see Section 6.3.1). Participants also provided detailed justifications for their preferences during decision-making tasks (see Section 6.3.2). Synthesising these findings showed how general users perceived these features as useful for navigating complex social media content. In Chapter 7, these design features were examined to align with real-world workflows. Before interacting with *DiscourseDash*, experts described how they routinely integrated various forms of analysis to guide their decision-making processes, aligning closely with the features proposed in our tool (see Section 7.3.1). Thematic findings from their interactions with *DiscourseDash* further illustrated how perceptions of feature usefulness influenced their evaluations of the tool's relevance to the intended context (see Section 7.3.2).

However, in sensitive or high-stakes contexts such as healthcare, where decisions can directly impact people's well-being, experts emphasised the ongoing importance of human interpretation along-

side automated insights. Although they acknowledged the value of these analytical tools, they emphasised the reliance on human judgment. Factors such as trust, cultural norms, and established decision-making protocols shape the extent to which machine analyses are adopted (Kocielnik, Amershi, and Bennett, 2019). Additional concerns about data privacy, bias, and ethical accountability vary between domains (see Section 7.3.3). These findings align with prior HCI work that highlights how perceptions of fairness influence willingness to rely on algorithmic systems (Binns et al., 2018). Depending on the context, we found that user reliance and trust exist on a spectrum, ranging from those who prefer extensive AI support to those who favour human judgment with minimal algorithmic input (see Section 7.3.3). This reinforces that the interpretation of meaning in the process of analysing social media discourses should not be seen as a one-stop solution delivered by analytical tools, but rather as something heavily shaped by specific practices or expectations.

To summarise this, although our findings align with user validation in supporting social media analysis, the extent to which these tools can be impactful depends on their ability to balance automation with transparency, flexibility, and responsiveness to human interpretation. This work encourages design goals to foreground it as essential to positioning interpretability in support of human involvement in real-world decision-making. Future work should examine specific contexts to illustrate these differences, compare how human judgment and expectation vary, and establish distinct design guidelines that are responsive to the nature of expert practices.

8.4 IMPLICATIONS FOR DESIGN

These empirical findings inform several design implications grounded in the feedback and interactions of both experts and general users with *DiscourseDash*. The insights highlight key considerations for improving the relevance of such tools in real-world settings

that guide the future development of similar systems (see Section 7.3.3).

8.4.1 *Improving analysis outputs*

There is a clear indication for a more detailed and accurate interpretation of the analysis output. On one end, some general users struggled to understand the meaning behind specific labels such as harmfulness, stance, or sentiment. This highlights the importance of providing concise and accessible *explanations*, such as through tooltips (Dai et al., 2015) or in-context guides, to help users quickly grasp the meaning of each metric. Consistent with broader eXplainable AI (XAI) research that emphasised that analytical results must resonate with users' cognitive frameworks (Doshi-Velez and Kim, 2017; Miller, 2019). Experts, on the other hand, expressed interest in more *refined categorisations* that could better reflect the complexity of social media discourse. For instance, rather than using broad categories, future iterations could explore hierarchical or layered classification systems that first identify the presence of an issue and then allow for deeper investigation into its nature (e. g., types of harm, intensity of emotions).

These insights suggest two main directions for future work: (1) to improve what these varying computational output means and how they relate to each other for non-expert users, and (2) to offer more granular analytical options for advanced users. These gaps should be addressed so that these analysis not only integrates easily into interaction experiences, but also become more valuable in experts' contexts. In other words, further targeted investigations should explore designs that seamlessly translate these insights for non-experts. Additionally, research should examine which types of granular analysis outputs experts across different domains prioritise and whether their needs align or differ.

8.4.2 *Handling variability in data sources*

There needs to be flexibility in the analytical data sources to support real-time and offline inputs for diverse and dynamic workflows. Similar emphasis in crisis management studies that also highlighted the importance of accommodating diverse data contexts and situational demands (Jin, Liu, and Austin, 2014; Saroj and Pal, 2020). The current prototype presented an analysis of vaccine-related topics from the COVID-19 pandemic, which, while suitable for initial prototype testing, does not fully represent the evolving nature of social media.

Some participants struggled to relate the data for the delegated tasks to current or personally relevant topics. However, experts' feedback highlighted the importance of flexible data input methods. Some experts work with curated or sensitive datasets, such as anonymised surveys, interview transcripts, or internal organisational data, which are not presented on public social media platforms. Meaning that, relying solely on online data in contexts where public discourse is culturally or politically restricted may fail to provide accurate insights for high-stakes decisions. They emphasised that allowing user-uploaded data would better align with practical adoption.

This flexibility is critical in workflows, where decisions rely on understanding refined perspectives that are not always apparent in public discourses. Findings indicated that public expression is sometimes limited due to cultural norms, but accurate community insights are necessary for informed policy-making decisions (see Section 7.3.2). This suggests that supporting real-time integration and secure offline data upload can better accommodate different use cases. These features can strengthen the practical use and ethical performance of interactive platforms across different sociocultural contexts.

On that note, another option could be to include more relevant user inputs for filtering the data to be analysed from populated social media discourses. This is because the current prototype only provides a few basic options such as date, channel, and topic filters (see Chapter 5). The empirical findings from this thesis suggest adding de-

mographic filters (discussed further in the next section). However, a deeper understanding could be gained by investigating which types of filters different user groups would find most valuable.

8.4.3 *Incorporating demographics*

There is a strong emphasis on the importance of demographic context, which is consistent with prior work that emphasises its influence on social media analysis (Sadah et al., 2016). The empirical findings suggest that demographic data is better for improving the contextualisation of analysis. This was noted to facilitate a deeper understanding of how different groups perceive and engage with specific topics for informing the development of more targeted interventions. For instance, categorising responses by demographic factors such as community affiliation, cultural background, or belief systems could offer valuable insight into varied behavioural patterns to address diverse perspectives. This is evident from experts suggesting that the lack of demographic insights limits the tool's practical relevance (see Section 7.3.2). For many experts, demographic segmentation is a core component in their decision-making workflows because accurate and culturally sensitive strategies are required.

These findings suggest that presenting demographic categorisation remains a foundational element in social media analysis. Building on that, future integration of social network analysis would benefit from incorporating the understanding of user interactions and information dynamics. This consideration also highlights the value of supporting relational analysis (Peng et al., 2018). The ability to visualise patterns of interaction, influence, and community formation illuminates how narratives propagate across groups, offering a more integrated understanding of social media discourses. However, it is important to recognise that representing user relationships introduces significant privacy considerations when it comes to sensitive topics. Future work

should investigate how to balance the analytical benefits of social network visualisation with ethical and responsible design practices.

8.4.4 *Balancing the needs of different users*

Although the design of *DiscourseDash* takes a step in proposing a set of analytical features to support decision-making, our findings suggest that more flexibility is needed to address the diverse needs and preferences of users. There is a clear tension indicated from both user groups that suggested *DiscourseDash*'s features should incorporate varying features that guide actions specific to the user's context. With the advantages of AI (Phillips et al., 2017), several intersecting suggestions were unfolded in these empirical findings. Firstly, participants consistently emphasised the value of *temporal prediction* as an extension of the ability to anticipate emerging trends based on ongoing conversations or past events. While monitoring provides clear benefits for early issue detection, incorporating predictive functionality that learns from past events and applies to future predictions would further support a shift from reactive to proactive decision-making in terms of enhancing the tool's strategic value.

There is also an observation where users would like to engage with discourse at varying levels of detail through optional *summarisation* features. Amongst general users, the findings suggested that some valued access to *Detailed Analysis* of discourses, while others preferred concise summaries for quicker insights. Based on this, it can be argued that experts can also take advantage of such flexibility to adjust their engagement according to the specific demands of their investigations. This would support decision-making across a spectrum of time sensitivities, urgencies, and depths of analysis. However, experts emphasised the importance of transparency in design for sensitive domains and preferred to validate AI-generated summaries against authentic data. This indicates that while summarisation features accommodate certain general user needs, a better design implied from this

tension is to allow user control over the level of summarised detail presented. This, in turn, balances rapid insights with opportunities for deeper exploration. This suggestion aligns with HCI literature on designing for user preferences and flexible data presentation to meet the needs of diverse user groups (Goodman, Langdon, and Clarkson, 2007; Perer and Shneiderman, 2008).

Besides, additional support could be provided to guide users towards more effective actions. One possible feature would be to *generate suggestions* informed by the analytical output. For instance, the system could reference a repository of predefined knowledge or guidelines that reflect established practices, or recommend interventions based on historical events. Several experts noted that while identifying and understanding issues is an important part of the process, the value would increase when accompanied by practical recommendations that help guide subsequent decisions. This potential is foreseeable given recent advances in LLMs and Retrieval-Augmented Generation (RAG) systems that have been widely adopted across industries seeking to utilise such capabilities to improve traditional workflows (Singh et al., 2025).

Experts benefit from highly tailored features that align with their decision-making processes, while general users require simpler, more intuitive interfaces. While generic solutions provide baseline support for expert decision-making workflows, there is significant scope for future work to address more complex and varied analytical needs. It is important to note that some users prefer automation to streamline decision-making, while others are sceptical of AI's reliability and prefer to retain greater control. This tension underscores the role of HCI in translating complex computational outputs into actionable insights for diverse user groups. Future work should prioritise flexibility while also being subject to ongoing evaluation and refinement to maintain relevance across diverse domains. Incorporating HCI principles such as progressive disclosure and adaptability can help ensure that discourse-level insights remain purposeful and actionable for both expert and general users (Feigh, Dorneich, and Hayes, 2012).

8.4.5 *Inclusivity and representation*

It is also important to consider accessibility for a wide range of users, regardless of their linguistic, cultural, or resource constraints (Abascal and Nicolle, 2005). An important point raised by our experts was the accessibility of these tools. There is a preference for open-sourced solutions, which were seen as a more practical way to promote broader use, encourage collaboration, and reduce cost-related barriers. These tools should be *affordable* to accommodate user groups such as small teams or researchers with limited budgets.

These findings extend the ongoing discussions in NLP about *language diversity* and *multilingual support* (Hershcovich et al., 2022; Joshi et al., 2021). Participants highlighted the importance of designing tools that support languages other than English. To put this into perspective, the global pandemic was a shared crisis that affected people worldwide. However, if analytical tools do not support multilingual analysis, they risk excluding researchers and communities who work in languages other than English. To promote fairness and broader adoption, future tools should support diverse linguistic contexts for underrepresented local dialects and languages in multilingual regions where dialectal variations are common.

Furthermore, social media reflects only a partial view of societal realities, shaped by *cultural and contextual* factors. Experts in our findings highlighted that the usefulness of *DiscourseDash* can vary depending on cultural settings. Without such considerations, these tools risk misinterpretation or overlooking important contextual meanings, which can lead to limited applicability in certain settings. For broader relevance and effective use across diverse user groups, future efforts should be grounded in underrepresented local practices and informed by cultural and linguistic sensitivity.

8.4.6 *Layering multi-levels of trust*

Establishing trust is a fundamental requirement for the adoption of AI-driven analytical tools. The findings suggested that system transparency is important for presenting how the tool generates output, processes data, and ensures fairness. This reflects existing HCI research that emphasises the role of accountability in AI system design to support user trust (Abdul et al., 2018; Binns et al., 2018). Future designs of such analytical tools must carefully address this by making each analytical process understandable that involves sensitive or high-stakes decision-making.

Experts within these domains tend to express scepticism toward current AI-based solutions for qualitative analysis. They argued that such approaches are not yet sufficiently validated or trustworthy to use in practice. Hence, this reflects broader concerns about the early application of analytical tools in socially complex environments (Ehsan and Riedl, 2020; Suresh and Guttag, 2021). This need for more systems to incorporate transparency and flexibility that allow users to provide feedback on the output when needed or maintain overall control (Amershi et al., 2019).

These point towards future efforts in integrating mechanisms for explaining system logic (e. g., how classifications are made). E. g., design better components to communicate how data is sourced and protected and allow users to engage with AI outputs at varying levels of interpretive depth. Following these findings and suggestions, designs are strengthened and evolve with system use and user feedback (Miller, 2019). Future work should consider how to address limitations in AI-generated insights, for instance, by presenting the confidence level of the output (Rechkemmer and Yin, 2022), or explainable methods (Danilevsky et al., 2020; Liu, Yin, and Wang, 2019). Trust can be encouraged where users can regulate their reliance adequately in domains where judgment of qualitative insights is subjective.

Besides, there should be transparent documentation that outlines how analytical tools are intended to evolve, be maintained, and be

supported over time. When users place trust and reliance on such systems, there is a reasonable expectation that these tools will remain accessible and functional. The absence of clarity or the sudden withdrawal of these tools undermines confidence in their long-term adoption since experts have previously encountered how such disruptions can impact their workflows. It was noted in our expert findings that the discontinuation of *Meta*'s *CrowdTangle* and the removal of research access by *X* (formerly *Twitter*) caused considerable setbacks for many researchers who had come to depend on these platforms. These highlight the importance of sustained transparency and clear communication around long-term support and platform commitments when tools become embedded in critical analytical workflows.

8.5 CHAPTER SUMMARY

This chapter synthesises the key findings from the empirical chapters of this thesis, which focused on the design, implementation, and evaluation of a human-centred interactive system for social media analysis to support decision-making. The discussion reflects on how the insights gathered from multiple empirical chapters of this thesis support the central argument of this thesis, in that incorporating discourse-level analysis into interactive systems for social media analysis can better support decision-making processes. The value of discourse-level insights in allowing users to understand patterns and relationships within social media discourse can be observed based on how participants articulated their perception and positioned themselves using *DiscourseDash* to make informed decisions. The empirical work in this thesis examined two distinct user groups to capture varying perspectives: (1) general users and (2) industry experts. Although users differed in experience and expectations, the findings consistently strengthen the point that the interactive system's design supports discourse-level interpretation

The discussion chapter drew on the research questions set out in Chapter 1 to synthesise the findings according to how different user groups perceive and interact with automated analysis. These observations underscore the importance of balancing automated analyses by presenting discourse-level insights as proposed throughout this thesis. This motivates further studies within interactive systems to provide context around discourses as a method of explanation and allow interpretive control for decision-making. The discussion not only validates the proposed interactive design but also offers broader implications for the development of computationally informed tools that integrate automated analysis with human expertise. More importantly, the chapter outlined valuable guidance for future work and design refinement in this area of research.

9

CONCLUSION

This chapter presents the conclusion of this thesis by providing a summary of its key findings, insights, and contributions to the field. The chapter begins with a detailed overview of the research, restating the goal and the experimental studies conducted, while relating these elements to the broader aims and significant findings of the research. The chapter then elaborates on the contributions of this research, highlighting their significance to demonstrate how they advance understanding in the field. Following this, the chapter critically reflects on the limitations of the research that may have influenced the outcomes. Finally, the chapter identifies and discusses potential directions for future research on the application of interactive tools for social media analysis by emphasising how future studies can build upon the foundations established in this thesis.

9.1 SUMMARY

This thesis investigated the role of discourses by combining computational methods with HCI approaches for social media analysis in decision-making contexts. The research set out to answer the following questions: (1) How can computational methods inform the design of interactive tools for social media analysis? and (2) How do users engage with such designs in practice, and what insights can be drawn to improve decision-making workflows? To address these questions, this thesis proposes a practical approach in which computational methods are first applied to social media data to inspire an interactive design. Subsequent user studies were then conducted to validate and refine this design by evaluating human experiences within realistic tasks and real-world perspectives.

The core findings from the computational experiments highlighted the importance of discourse context in task-specific learning, such as stance detection. While LLMs offer valuable predictions, they remain limited when applied to navigating the meaning behind discourse-specific insights (see Chapter 4). The main contribution of this exploration study is in demonstrating how the practical implementation of computational methods can inform and inspire the design of interactive tools. The observation in the experiment directly guided the design goals of the interactive prototype (see Chapter 5).

Because there is limited documentation on how users interact with tools that provide discourse-level qualitative analyses for decision-making, the thesis drew inspiration from journalists' workflows to design the first user study with general user participants. This mixed-method evaluation tested the rationale of the design by gathering user experiences alongside the initial conceptual formulation. Results showed that the design provided more effective support for navigating unstructured social media discourses than traditional approaches. At the same time, the study surfaced aspects that were not initially anticipated, such as the need for clearer definitions in certain analyses and stronger alignment between visualisations and user expectations.

The thesis then examined the bridge between design and real-world application by engaging industry experts. This investigation validated the practical value of the design by highlighting how experts interacted with the prototype in the context of their professional workflows. Before the study, there were no clear pathways for anticipating how the design would be perceived. Surprisingly, the conceptual design aligned closely with the limitations experts described in their existing tools and methods. Observing experts interact with the prototype allowed the research to confirm that the grounded design effectively addressed real-world needs, while also providing insights for further refinement.

Overall, this research established a framework for grounding design in practical implementation, linking computational methods with HCI principles, and validating through two complementary

user groups. By integrating computational learning exploration with human-centred evaluations to conclude a design, the thesis provides evidence for the value of emphasising discourse-level insights in social media analysis in decision-making contexts. The work contributes both a methodological approach and a validated interactive prototype that demonstrates how computational methods and human expertise can collaborate to support complex analytical tasks.

9.2 MAIN CONTRIBUTIONS

This thesis makes the following contributions:

1. *Bridging NLP computational methods with HCI design principles*. The conceptualisation of a design was achieved through obtaining how the discursive nature of social media and the limitations of NLP models identified through a stance detection study can inform the design of computational tool to better support decision-making.

2. *An interactive design prototype* – DiscourseDash. This thesis presents the design and implementation of a working interactive prototype that integrates HCI principles with computational methods to operationalise transparency and accountability in social media analysis. The novelty of the design lies in its emphasis on discourse-level analysis to improve contextual understanding, in contrast to existing tools that primarily rely on aggregated visual summaries.

3. *Evidence-based validation through human experiences for general applicability*. Using established HCI evaluation methods, this work provides empirical evidence on how access to discourse-level context supports users' decision-making processes. A user study with general users offers a comparative baseline for understanding how a design for decision-making shapes interactions with discourse-level analytical tools.

4. *Real-world applicability implications for discourse-level analytical tools*. This thesis identifies key insights into the alignment and misalignment between automated analytical outputs and experts' real-world practices. These findings inform: (1) design implications to better support the decision-making needs of professional communicators, as defined in Section 2.1.2; and (2) broader methodological and design insights for the research communities.

These contributions are summarised as follows.

9.2.1 *Bridging NLP computational methods with HCI design principles*

The thesis demonstrates how technical limitations, e.g., the challenges of data scarcity and model uncertainty, can inform user-centred design decisions. This was achieved by first gaining a practical understanding of how computational learning performs in practice to establish a framework that incorporates the discursive nature of social media into the model's learning process. By bridging the gap between HCI design principles and computational methods, this exploration aims to devise a design motivated by a practical implementation of computational methods. This approach prioritises the purposeful application of these ideas over simple technical iteration to better serve the needs of practitioners.

This resulted in the development and evaluation of a computational framework for stance detection under data-limited conditions, which are commonly observed in practical computational learning scenarios (see Section 3.2.1.2). The study compared several data augmentation strategies, ranging from target-topic-specific augmentation to traditional word-level approaches and an LLM-based prompt-generation method using both a low-cost GPT-2 model and the more advanced Gemma-2 model. The research highlighted the importance of contextual knowledge from social media discourses by demonstrating its impact on the performance of NLP models for stance detection.

The main contribution of this practical implementation lies in its integration of computational exploration with design-oriented considerations. Having grounded the interactive design in empirical evidence, we demonstrated that the design is scoped within practical considerations rather than relying on theoretical assumptions. This approach also presents transparency in the design process by showing how data-driven insights can inform design decisions. Within the scope defined by this thesis, we have demonstrated that approaches which emphasise the surrounding discourse context can be iteratively refined in response to emerging patterns. This addresses the difference in how they are better suited to capturing the dynamic nature of social media.

However, the use of LLMs requires refinement and adaptation when applied to domains where there is a lack of sensitivity to the evolving linguistic practices that characterise social media (Huang et al., 2025). It should be noted, however, that the aspect addressed in this thesis represents only a foundational investigation into this broader problem. Through this study, we confirmed that when designing interactive systems, it is important to account for computational learning tasks that involve complex linguistic patterns and advanced language capabilities. Building on these findings, the study uses them to inform the design goals for an interactive social media analysis tool. So first, the findings highlight the value of discourse-level information in computational social media analysis. Secondly, the insights observed from the study motivated the development of an interactive design aimed at supporting decision-making.

9.2.2 *An interactive design prototype* - DiscourseDash

The computational learning studies presented in this thesis (Chapter 4) directly informed the design of an interactive tool for analysing social media discourses. The resulting prototype uses the NLP study as a bridge toward the design's rationale to emphasis the focus on

a transparent view of automated analysis. As most exising design focuses on technical affordably, our approach in deisgn instead, directly addresses the ambition of shifting from achieving sophisticated model outputs alone to the design of human-centred interfaces that better support the interpretation and application of those outputs.

The prototype and walkthrough of the design *DiscourseDash* is presented in Chapter 5. In that same chapter, we outlined how insights from the computational study in Chapter 4 were translated into design goals that shaped the analytical components and design rationale of the interactive tool. These components were expanded into visual and functional prototypes guided by principles from user-centred design.

The prototype design was then evaluated through two distinct studies targeting different user groups with different goals and experimental designs. The first study (Chapter 6) involved non-expert university participants with no prior experience in social media analysis. This study examined their experiences in using *DiscourseDash* and engaging with unstructured social media data for decision-making tasks to provide an initial foundation for the design. It validated the prototype by extending the researcher's initial conceptualisation through user feedback supported by mixed-method analysis. In the respect of design, the study has shown that: general users preferred to navigate social media analysis through *DiscourseDash* complemented by analytical insights from discourses because it supported their trust and articulation of decisions around observations within discourse-level insights.

The second study (Chapter 7) involved participants who were industry experts with experience in social media analysis to explore how *DiscourseDash* aligns with their experiences and workflows. This not only bridged *DiscourseDash* with real-world applicability, but it also documented experts' current perceptions of their workflow. This study focused on social media discourses to highlight their significance for future research in social media analysis. In the respect of design, the study has shown that: the study revealed that experts val-

ued *DiscourseDash* because it complemented their existing workflows and highlighted the limitations of these workflows, which rely on a combination of manual and automated methods to obtain discourse-informed insights.

Overall, *DiscourseDash*'s design offers concise and summarised views that enable users to interpret information effectively and make informed decisions. Even more so, presenting discourse-level analysis in depth enhances transparency and reasoning, allowing users to form their own interpretations and identify potential discrepancies. The contribution of this design is significant because it demonstrates how computational insights can inspire a practical design and be appreciated by human users. *DiscourseDash* achieves a balance between analytical depth and enables users to engage with discourse-level analysis with high-level summarised visualisations. This research also provides a foundation for future interactive systems in social media analysis by illustrating how empirical evidence can guide design iterations and ensure that tools remain aligned with user needs and real-world workflow limitations.

9.2.3 *Evidence-based validation through human experiences for general applicability*

The thesis formulated a conceptual design focused on discourse-level contexts, which is an area previously unexplored through direct user interaction, can facilitate complex decision-making processes. By first conducting a foundational study with general users (see Chapter 6), this research establishes a comparative baseline for the research communities. The user study was designed according to established qualitative and quantitative methods; the derived evidence-based validation grounds the findings of this thesis within human experiences. This contribution makes sure that the design is driven by human needs rather than technical capability alone through perceptions of real-world users.

In HCI research, evaluating a system with general users during early stages of development helps determine whether the design aligns with its intended decision-making purposes for social media analysis. This study provides a supplementary perspective to a researcher's formulated design. The prior computational study on stance detection informed the pre-design phase, while this mixed-method evaluation study represents a prerequisite post-design evaluation before expert evaluation.

As mentioned previously, these general users were not experienced in any social media analysis work. The experiment was designed based on use-case scenarios inspired by journalistic workflows to guide interactions on two conditions (1) *DiscourseDash*, (2) an unstructured social media platform. The purpose of this study design is to present the measured and compared effectiveness and interactive experiences around the given decision-making tasks for social media analysis of discourses. Through this approach, the study investigated how individuals without prior expertise in social media analysis engaged with the *DiscourseDash* and performed tasks that simulated real-world decision-making processes.

The integration of quantitative and qualitative methods from this study eventually generated empirical evidence that grounded the design rationale in actual user experiences. The findings indicate that *DiscourseDash*, which integrates analyses of harmfulness, stance, sentiment, and popularity, effectively supports decision-making by analysing social media discourses. Furthermore, our study emphasises that trust and transparency remain crucial in exploring the roles and influence of human-AI collaborations. The findings also underscored how general users interpreted different features to guide their decision-making processes. At the same time, the evaluation outlined implications that guide future iterations of the design in interpreting automated outputs.

These findings contributed to HCI and social media research by showing how computationally informed design can bridge the gap between complex analytical models and end-user decision-making.

We confirmed the potential of discourse-level insights in supporting social media analysis around decision-making. A broader contribution of the study is how it demonstrated the importance of user-centred evaluation in the design of interactive systems. First, it validates the initial hypothesised features that were compiled together in *DiscourseDash*, second, it also provides implications for future refinements and innovation.

9.2.4 *Real-world applicability implications for discourse-level analytical tools*

This thesis emphasises the application of existing approaches by focusing on the real-world decision-making needs of practitioners such as professional communicators. It addresses critical challenges in social media analysis by examining how discourse-level insights can be operationalised within interactive analytical tools to better support practical decision-making.

The main contribution lies in synthesising empirical findings into a set of actionable recommendations for future analytical systems. These implications are grounded in a expert user study on top with a previous general user study to reveal that current analytical efforts often fail to facilitate actionable support. By addressing these gaps, this work provides a roadmap for translating HCI research into practical impact within professional communication and decision-making contexts.

Extending the previous investigation, this contribution includes an evaluation of *DiscourseDash* in a real-world professional context. We conducted structured interviews with industry experts in social media analysis to align *DiscourseDash* more closely with real-world workflows. This study was designed to qualitatively assess the tool's relevance for professional practice, focusing on its alignment with existing workflows and potential areas for improvement. Whereas the previous study (see Chapter 6) established a foundational understand-

ing of how general users interact with discourse-level features for scenario-based decision-making tasks, this study focused on experts, who routinely face the challenges of managing large-scale social media analysis. The purpose of the study was to obtain the perceived value and limitations of *DiscourseDash* within a real-world perspective for outlining implications for the design.

This study revealed that most experts were optimistic about the potential of discourse-level insights for supporting their analysis workflow. These experts demonstrated how the core concept of *DiscourseDash* aligned with their workflows due to the value of contextualising summarised visual analyses with discourse insights, and how the demand for it better supports actionable outputs in their workflow. However, we also identified several design considerations that need to be addressed before the tool can achieve real-world applicability. These findings not only reinforce the tool's relevance but also help synthesise practical expectations for its future development through formulated implications from experts. The perspectives gathered from industry participants were particularly valuable in guiding refinements to *DiscourseDash*. This is because experts were able to highlight aspects of the tool that aligned with their existing practices, confirming the applicability of discourse-level analysis for supporting strategic decision-making. A more important aspect is how the gathered perceptions help this research identify challenges in integrating discourse-level analyses into established workflows.

This contribution has broader implications for the design of real-world analytical systems. The knowledge and insights rooted within these domain expertise introduced multiple overlooked perspectives during the design process. Hence, these experts' insights were a continuous source of knowledge that shaped the development of *DiscourseDash*. From a broader perspective within HCI, the study demonstrates the value of combining human expertise with computational insights in interactive systems by bridging the gap between computational development and professional application. All being said, it reinforces the central argument of this thesis in that discourse-level

perspectives provide a better support for decision-making in social media contexts.

9.3 CRITICAL REFLECTIONS

While our study provided valuable insights for developing the *DiscourseDash*, several limitations must be acknowledged.

First, the general user study primarily involved university staff and students. While this group provided useful feedback on the overall design concept, it may not fully reflect the perspectives of broader general user populations, such as users outside academia or those differing in age, profession, or technical experience. Similarly, the expert study included a small sample relative to the wider field of social media analysis. These limitations restrict the generalisability of the findings and highlight the need for broader participant recruitment to capture a more diverse range of experiences and perspectives.

Second, the general user study participants' familiarity with *Reddit* was not measured. This may have influenced their ability to objectively assess the features and benefits of *DiscourseDash*. Even so, *Reddit* was selected as a baseline platform due to its widespread use and recognisable threaded structure, which participants commonly encounter in personal and professional contexts. Hence, this limitation is worth noting but does not undermine the study's focus on comparative perceptions and user experiences. The observed differences likely reflect natural variability in engagement with unstructured social media content in real-world settings.

Third, using *Reddit* as a proxy for unstructured social media discourse limits the study because the platform's community norms, interaction styles, and user demographics may not represent the diversity of behaviours across other social media platforms. Additionally, practical data constraints such as deleted or private posts, API rate limits, and incomplete historical data restrict access to the full corpus. These limitations, however, do not affect the main findings of our user

studies, as the transferable contributions relate to demonstrating the broader relevance of the design and HCI principles explored in this thesis.

Finally, our research focused only on text-based interactions that naturally guided the selection of *Reddit* as the baseline platform due to its primarily threaded discourse structure. Although this focus aligns with the analysis of unstructured textual discourses, it does not fully represent the broader nature of social media. This is because when considering the broader nature, content modalities such as images, videos, and other multimodalities play a significant role as well in social media analysis.

9.4 FUTURE WORK

Building on the reflections and limitations discussed in the previous section, several areas for future work are apparent. These recommendations are scoped to the current proof-of-concept study while suggesting opportunities for broader investigation.

- *Conducting ethnographic studies* to observe how users interact with social media analysis in real-world decision-making contexts. Such studies can uncover detailed routines, strategies, and challenges that may be missed in controlled experiments. By understanding expert users' day-to-day practices, researchers can explore how *DiscourseDash* performs within existing workflows, identify potential friction points, and improve understanding as to how it supports decision-making tasks in ways that align with the natural daily routines of expert users.

- *Conducting longitudinal studies within real-world deployments* to examine sustained user engagement and evolving experiences with *DiscourseDash*. This includes investigating long-term usage patterns and actual workflow integration to assess the tool's ongoing relevance and practical impact. Iterative design processes involving continuous stakeholder collaboration can refine fea-

tures and adapt the tool to evolving needs for documenting its continued intended impact in supporting decision-making over time.

- *Incorporating multimodal content* to identify patterns and insights that are not obvious from text alone. The current study focused on text-based interactions, but the nature of social media is multimodal. Future research should extend analyses to include images, videos, and other media. This will inform the design of tools capable of integrating heterogeneous social media content.

- *Cross-platform evaluation* can reveal variations in user interactions and outcomes by providing an extended understanding than focusing on a single platform such as *Reddit*. An additional note to that is that measuring participants' prior experience with different platforms can help account for variability in engagement and ensure that observed outcomes reflect the tool's design with prior familiarity.

- *Expanding recruitment profiles* to include diverse user groups, professions, sectors, and geographical locations. This will support a better understanding by performing a comparison across different contexts of use and improving the generalisability of design decisions.

These recommendations were devised based on the feedback and review received throughout this research. They provide a foundation for future work in areas of research that can extend the focus of this multidisciplinary investigation.

Part IV

APPENDIX

A

APPENDIX

A.1 PRELIMINARY DATA AUGMENTATION EXPERIMENTS

A.1.1 *Motivation*

A preliminary experiment was conducted to investigate the impact of word-level data augmentation on stance detection. All annotations are against (0), neutral (1), and in favour (2). The goal is to examine how traditional augmentation performs in different subsampled dataset sizes.

A.1.2 *Data Augmentation Methods*

We focused on widely used word-level augmentations, which transform an original text x into augmented samples x_{aug}. The methods included:

- **Random Insertion (RI):** Insert a synonym of a randomly selected word using WordNet.
- **Random Deletion (RD):** Randomly remove words from the text.
- **Random Swapping (RS):** Swap positions of words randomly.
- **Synonym Replacement (SR):** Replace words with WordNet synonyms.
- **Contextual Substitution (CS) and Contextual Insertion (CI):** Replace or insert words using a pre-trained BERT model.

- **Easy Data Augmentation (EDA):** Combines RI, RD, RS, and SR (Wei and Zou, 2019).
- **Back-Translation (BT):** Translate to German and back to English (Tiedemann and Thottingal, 2020).

A.1.3 *Datasets*

We evaluated augmentation methods on three publicly available stance detection datasets:

- COVID-19 Vaccination Stance Detection (Cotfas et al., 2021). Hydrated via Twitter API in 2022; some tweets were unavailable due to deletion.
- SemEval-2016 Stance Dataset (Mohammad et al., 2016), covering multiple topics including Hillary Clinton, Feminist Movement, and Legalisation of Abortion. The Climate Change is a Real Concern topic was excluded due to severe class imbalance.
- Twitter Stance Election-2020 (Kawintiranon and Singh, 2021), containing tweets related to the 2020 U.S. Election candidates Joe Biden and Donald Trump.

A.1.4 *Results*

To answer our research questions, we evaluated performance changes in F1-score across dataset sizes (75, 300, and 600 samples) between these augmentation methods and the dataset's topic. Figure 23 compares the performance of each augmentation method against no augmentation. Results indicate that augmentation generally improves performance for smaller datasets (75 and 300 samples). For larger datasets (600 samples), improvements are minimal for some methods, and in certain cases, performance decreases.

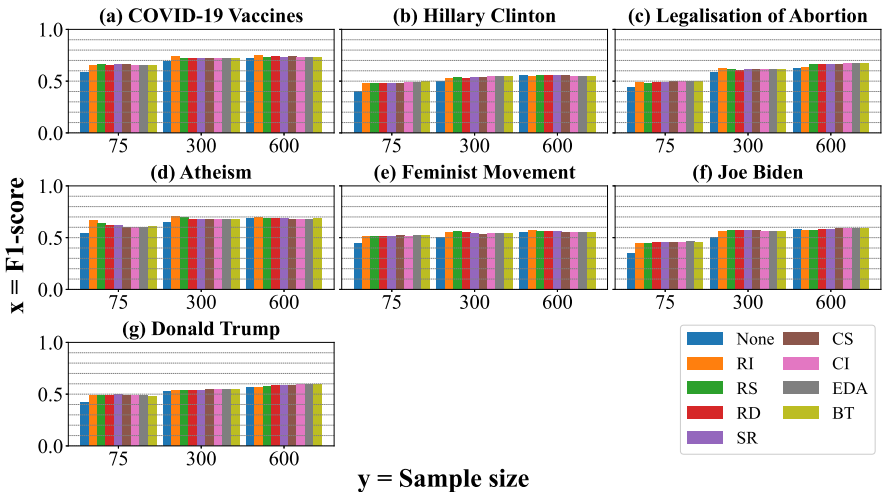


Figure 23: Bar graph comparing the impact of augmentation methods when applied versus when none are applied.

Figure 24 presents boxplots of performance improvements by topic. We observed that smaller datasets benefit more substantially from augmentation, while larger datasets experience diminished gains. This aligns with the expectation that augmentation is most effective when training data is limited.

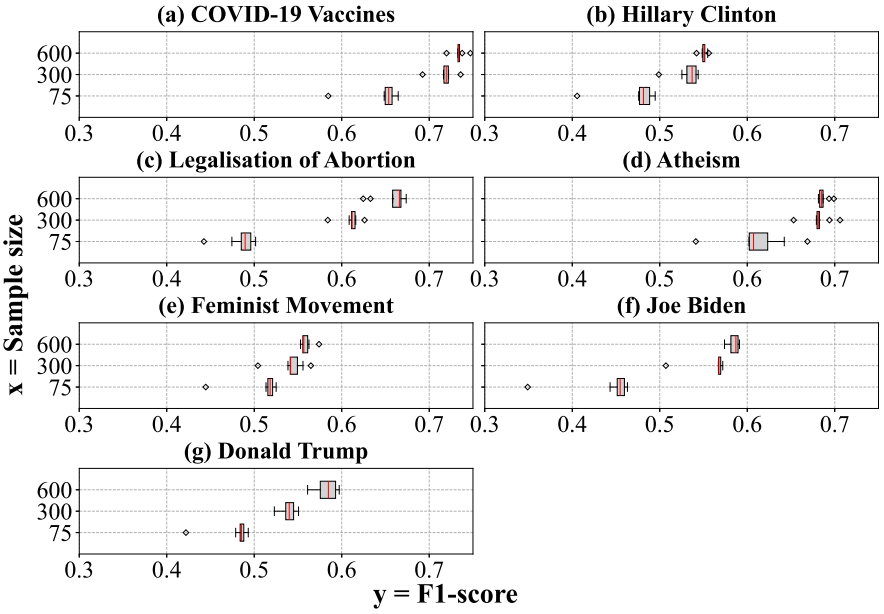


Figure 24: Box plots comparing the performance improvements of different augmentation methods across training set sizes.

A.1.5 *Conclusion*

Across datasets and sample sizes, EDA consistently achieves competitive performance. While methods like BT can sometimes outperform EDA, EDA is simpler and computationally efficient. Its combination of four basic word-level operations (RI, RD, RS, SR) makes it representative of the traditional augmentation method in Chapter 4. This serves as a baseline for comparing more advanced augmentation techniques in stance detection tasks.

A.2 QUESTIONNAIRES

These questionnaires were used in Chapter 6 for collecting quantitative data from participants.

- *DiscourseDash* condition (see Figure 25, 26).
- Reddit condition (see Figure 27, 28)

Questionnaire for DiscourseDash

Please answer the following questions as accurately as possible.

* Required

1. How easy or difficult was it to identify a topic to prioritize using the provided method? *

	Very easy	Easy	Neutral	Difficult	Very difficult
Rate the question.	◯	◯	◯	◯	◯

2. To what extent do you agree or disagree with the following statements? *

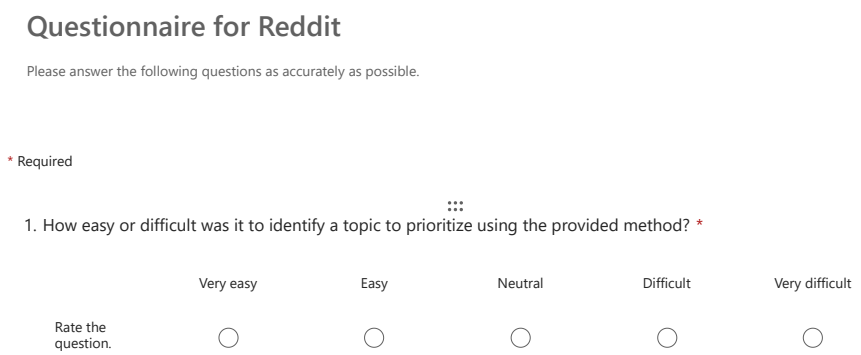
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The harmfulness label was a useful indication.	◯	◯	◯	◯	◯
The stance analysis helped me to understand the importance of each topic.	◯	◯	◯	◯	◯
The sentiment analysis helped me to understand the importance of each topic.	◯	◯	◯	◯	◯
The popularity analysis helped me to prioritise one topic over the other.	◯	◯	◯	◯	◯
Investigating discourses with labels helped me to identify key factors for different stances and	◯	◯	◯	◯	◯

Figure 25: Questionnaire for *DiscourseDash* condition (page 1).

3. To what extent do you agree or disagree with the following statements about the usability of the dashboard? *Evaluate in terms of the dashboard's usability not towards achieving the task. *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think that I would like to use this system frequently.	◯	◯	◯	◯	◯
I found the system unnecessarily complex.	◯	◯	◯	◯	◯
I thought the system was easy to use.	◯	◯	◯	◯	◯
I think that I would need the support of a technical person to be able to use this system.	◯	◯	◯	◯	◯
I found the various functions in this system were well integrated.	◯	◯	◯	◯	◯
I thought there was too much inconsistency in this system.	◯	◯	◯	◯	◯
I would imagine that most people would learn to use this system very quickly.	◯	◯	◯	◯	◯
I found the system very cumbersome to use.	◯	◯	◯	◯	◯
I felt very confident using the system.	◯	◯	◯	◯	◯
I needed to learn a lot of things before I could get going	◯	◯	◯	◯	◯

Figure 26: Questionnaire for *DiscourseDash* condition (page 2).



Questionnaire for Reddit

Please answer the following questions as accurately as possible.

* Required

1. How easy or difficult was it to identify a topic to prioritize using the provided method? *

	Very easy	Easy	Neutral	Difficult	Very difficult
Rate the question.	◯	◯	◯	◯	◯

Figure 27: Questionnaire for *Reddit* condition (page 1).

2. To what extent do you agree or disagree with the following statements about the usability of the Reddit website? *Evaluate based on the website's usability in general, not towards achieving the described task. *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think that I would like to use this system frequently.	◯	◯	◯	◯	◯
I found the system unnecessarily complex.	◯	◯	◯	◯	◯
I thought the system was easy to use.	◯	◯	◯	◯	◯
I think that I would need the support of a technical person to be able to use this system.	◯	◯	◯	◯	◯
I found the various functions in this system were well integrated.	◯	◯	◯	◯	◯
I thought there was too much inconsistency in this system.	◯	◯	◯	◯	◯
I would imagine that most people would learn to use this system very quickly.	◯	◯	◯	◯	◯
I found the system very cumbersome to use.	◯	◯	◯	◯	◯
I felt very confident using the system.	◯	◯	◯	◯	◯
I needed to learn a lot of things before I could get going with this system.	◯	◯	◯	◯	◯

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

Figure 28: Questionnaire for *Reddit* condition (page 2).

A.3 CONSENT PROCEDURES

All participants provided informed consent before their participation. For general user participants in Chapter 6, consent was obtained

through a digital document exchange via email or Microsoft Teams messages. In industry expert participants in Chapter 7, consent was obtained as part of the scheduling form before interview confirmation. Figure 29 presents an example of the consent forms. Additionally, verbal confirmation of consent was obtained at the beginning of each study session.

School of Computer Science
University of Nottingham

Section B. Information to be provided to research participants

PROJECT TITLE: User Study on Prototype Tool

1. The research
a) Aims and objectives of the research
As part of my PhD research at the University of Nottingham, I have developed an initial prototype for a dashboard tool. The goal is to gather insights from experts on their existing workflows and obtain feedback on the tool to enhance its effectiveness and alignment with real-world needs.
b) Funder information
Engineering and Physical Sciences Research Council [grant number EP/V00784X/1]
c) Governance
This research has been approved by the School of Computer Science Research Ethics Committee (CS REC), ethics application ID ***CS-2023-R61***

2. Taking part in the research
Participants will be instructed to complete a series of simple tasks on the designed prototype. Following this, I will ask a set of questions aimed at evaluating the participants' experiences interacting with the prototype. Additionally, I will collect any feedback participants may have regarding their interaction. The interview will be **screen and audio recorded** for analysis, which is necessary to accurately capture and report the interview results in my research. All data collected will be anonymised to protect participants' identities. The interview is expected to take **approximately 45 minutes** and will be conducted online via **Microsoft Teams**.

3. Risks of participation
a) Risks
There is always a risk of unauthorised access to data.
b) Mitigation of risks
See section 5 for the measures we put in place to mitigate the risk of unauthorised access.

4. Purpose of data processing
a) Data collected
We collect the following categories of data during your participation in the research:

(v.9 August 2022)

Figure 29: An example of the consent form (page 1).

Interview Data: This includes audio recordings of the interviews. Screen Recorded Data: This involves screen share recordings of the tool in use, captured via Teams calls.
b) Specific purposes for which the data are processed
Data collected during the research that identifies you may be: • Analysed to meet the aims and objectives described in Section 1. • Reviewed and discussed in supervision sessions between researchers and their supervisors or in research meetings between members of the research team, including project partners. • If audio recordings are collected during the research, these may be transcribed and **anonymous** quotations of your spoken words may be used in scientific works, including presentations, reports and publications stored in databases and posted online, and in marketing materials that promote the research and its findings.
c) Automated decision-making and profiling
N/A
d) Legal basis for processing your data
We collect personal data under the terms of the University of Nottingham's Royal Charter and in our capacity as a teaching and research body to advance education and learning. We thus process your data on the legal basis that our research is in the public interest, we have legitimate interests and / or that you consent to data processing in freely and voluntarily participating in our research activities.

5. Storage and retention of your data
a) Data protection measures
We put the following organisational and / or technical safeguards in place to protect your data and your identity to the best of our ability: i) All data stored digitally will be encrypted and password protected and all physical data will be stored in a secure location. ii) Participants will only be identified by a unique code. Interview recordings will be transcribed. Any identifiable information will be removed from the transcripts before analysis. All original recordings will be deleted after analysis.
b) Retention period
Data protection law allows us to retain personal data for an indefinite period and use it in future for public interest, scientific or historical research purposes or statistical purposes, subject to the implementation of technical and organisational measures that safeguard your data, your legal rights and your freedoms. These safeguards include the storage measures described above to protect your data against unauthorised access, and de-identification (anonymisation or pseudonymisation) of your data wherever possible and practicable. Data that identifies or could identify you will not be made public without your consent.

6. Third party recipients, services and data transfers
a) Project partners
Your data will not be shared with others
b) Third-party services
N/A

(v.9 August 2022)

Figure 30: An example of the consent form (page 2).

c) Data transfers
Your data will not be transferred to another country

7. Your legal rights
Data protection rights (Data Protection Act 2018) You have the right: • To be informed about the collection and use of personal data (as per this document). • To access and receive a copy of your personal data, and other supplementary information, on request. • To object to and restrict data processing if you think we are not complying with data protection law, and to rectify inaccuracies. • To be forgotten, i.e., to have your personal data erased. • To data portability and to obtain your data in an accessible and machine-readable format if appropriate, or to transfer your data to another organisation if technically feasible. • To complain to about the way we process your personal data to our ethics committee (cs-ethicsadmin@cs.nott.ac.uk), our Data Protection Officer (dpo@nottingham.ac.uk) or the Information Commissioner's Office (https://ico.org.uk/make-a-complaint).*

8. Your ethical rights
a) Right to withdraw
You have the right to withdraw from the research at any time without explanation. You also have the right to request that your data be deleted if you do withdraw.
b) Handling of 'mixed' data
If the data is 'mixed' data – i.e., if it involves other people (not including the researchers), e.g., a conversation or video of multi-party interaction – it cannot be deleted unless **all parties** request it. However, any mixed data involving you will be redacted wherever possible, with the exception of scientific works produced prior to your notification of withdrawal.
c) Withdrawal procedure
If you wish to withdraw, please notify ***researcher insert your name and email address*** If you do not receive confirmation of withdrawal from the research, please email cs-ethicsadmin@cs.nott.ac.uk

9. Consent to participate	
a) I consent to participate in the research and my signature or mark confirms the following: • I understand the aims and objectives of the research • I understand what the research requires me to do • I accept the risks of participation	***Signature or mark:***

* Our DPO's postal address is Data Protection Officer, Legal Services, A5 Trent Building, University of Nottingham, University Park, Nottingham NG7 2RD.

(v.9 August 2022)

Figure 31: An example of the consent form (page 3).

• I understand what data will be collected and the purposes for which the data will be used • I understand safeguards will be put in place to protect my data and my legal rights • I understand I will not be identified unless the use of identifiable data has been requested and I consent to it • I understand that I can withdraw at any time without explanation • I have been able to ask questions about the research, my participation, and my data and my questions have been answered satisfactorily • I agree to participate and my participation is voluntary	
b) Opt out. I do not consent to use of my visual image in scientific works or materials that promote the research and its results	*Signature or mark:*
Name of participant(s)	
Date	
Witness *If participant(s) cannot sign*	*In signing I confirm the participant(s) named above have been fully informed about the research, have been able to ask questions, and consent freely.*

Figure 32: An example of the consent form (page 4).

A.4 RECRUITMENT PROCEDURES

A.4.1 *General users*

- **Platform:** Microsoft Teams (institutional research groups).
- **Eligibility:** Open to all interested users; no screening applied.
- **Process:**
  1. Recruitment post shared internally (see Figure 33).
  2. Participants completed an interest form (see Figure 36 & 37).
  3. All respondents were invited to schedule an interview via email/Teams' message.

A.4.2 *Industry experts*

- **Platforms:** LinkedIn and Prolific.

- **Eligibility:** Participants were required to demonstrate relevant professional experience through a screening questionnaire.
- **Process:**
    1. Recruitment materials were disseminated via a LinkedIn post (see Figure 34) and a targeted survey on Prolific (see Figure 35).
    2. Interested individuals were invited to complete a screening questionnaire for assessing their eligibility (see Figure 38 & 39).
    3. Respondents meeting the eligibility criteria outlined in Section A.4.3 were invited for participation.
    4. Eligible participants received personalised invitations to schedule an interview session (see Section 43).

A.4.3 *Conditions to filter experts*

1. Must have professional experience in analysing social media discourses, either
2. independently or as part of a team that uses such analysis for decision-making.
3. Must be involved directly in analysing social media discourses in their role to achieve specific objectives within their respective industries.
4. Must have at least two or more of the following objectives in their role:
    a) To monitor and manage brand reputation;
    b) To evaluate audience sentiment and engagement;
    c) To identify emerging trends or topics for strategy development;
    d) To assess the effectiveness of marketing or communication campaigns;

e) To consult decision-making in policy, public relations, or media strategies;

f) For crisis management;

A.4.4 *Recruitment materials*

Platform-specific recruitment materials were designed and shared to target appropriate participant groups.

- Microsoft Teams post (see Figure 33) was used for Chapter 6.
- LinkedIn post (see Figure 34) and Prolific (see Figure 35) ad were used for Chapter 7.

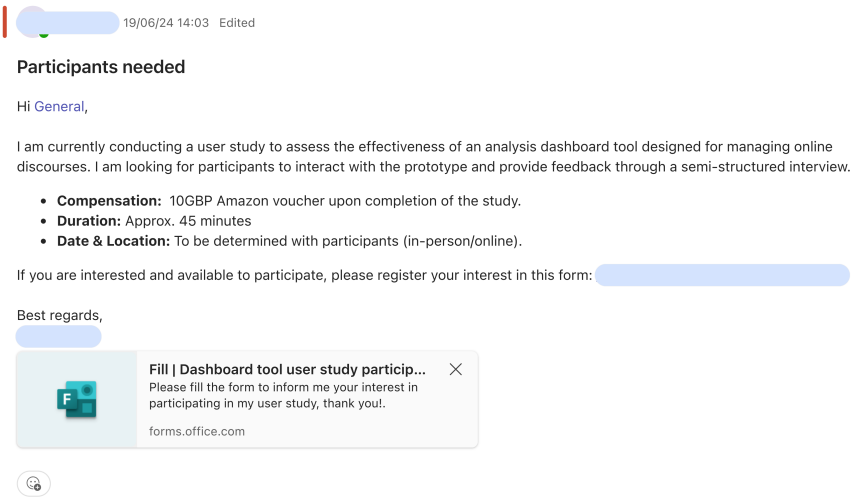


Figure 33: Recruitment post for general users shared via Microsoft Teams.

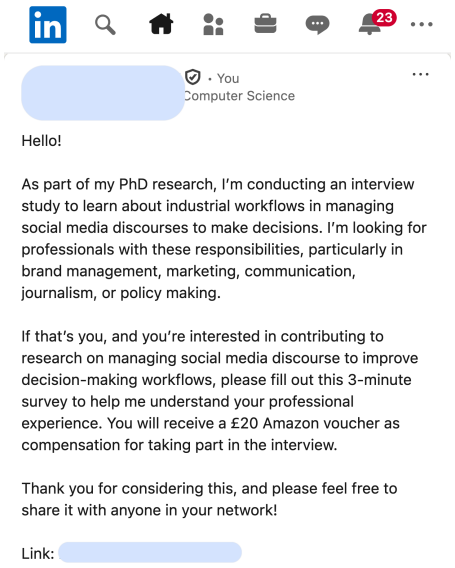


Figure 34: Recruitment post targeting experts shared on LinkedIn.

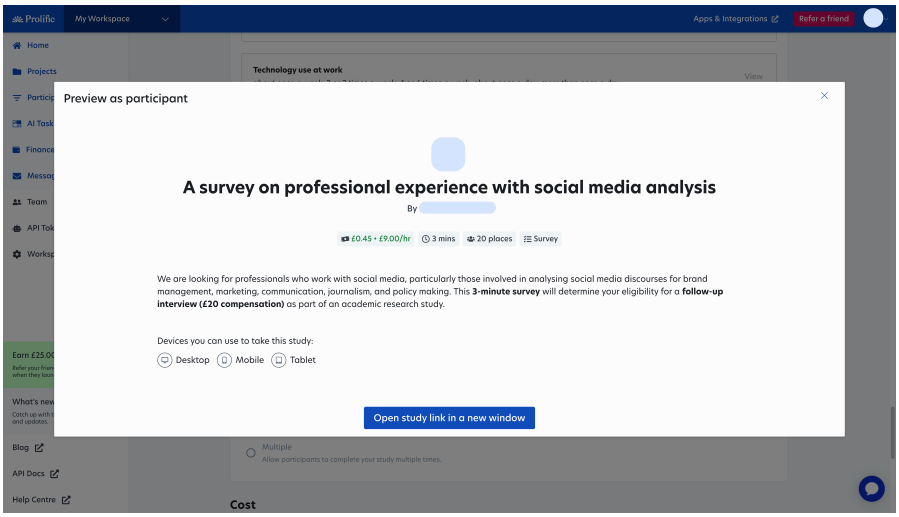
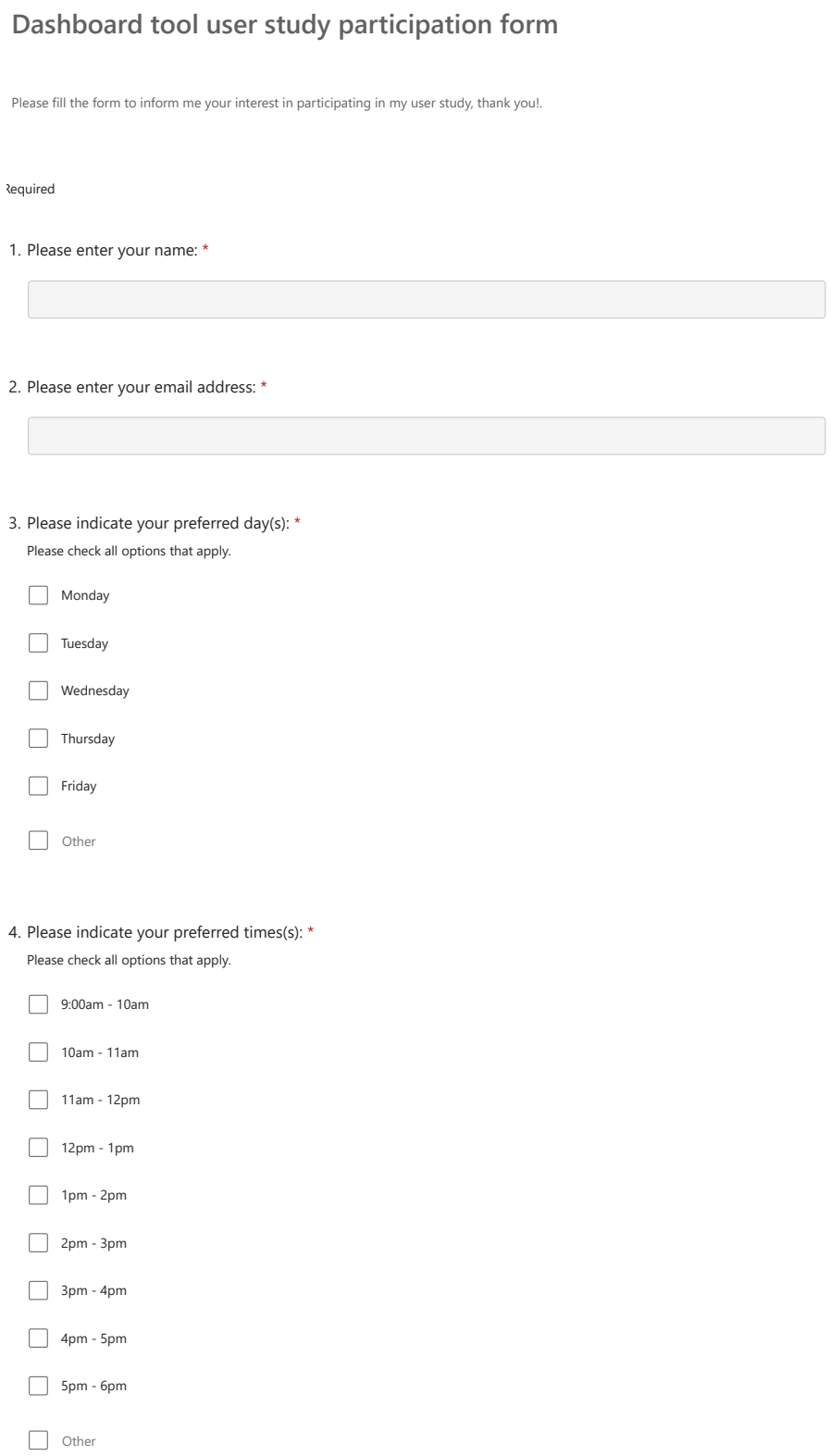


Figure 35: Prolific recruitment advertisement for experts.

A.4.5 *Interest forms*

This section includes two interest forms for recruiting participants:

- Figure 36, 37 was used in Chapter 6.
- Figure 38, 39, 40, 41, 42 was used in Chapter 7.



Dashboard tool user study participation form

Please fill the form to inform me your interest in participating in my user study, thank you!.

Required

1. Please enter your name: *

2. Please enter your email address: *

3. Please indicate your preferred day(s): *

Please check all options that apply.

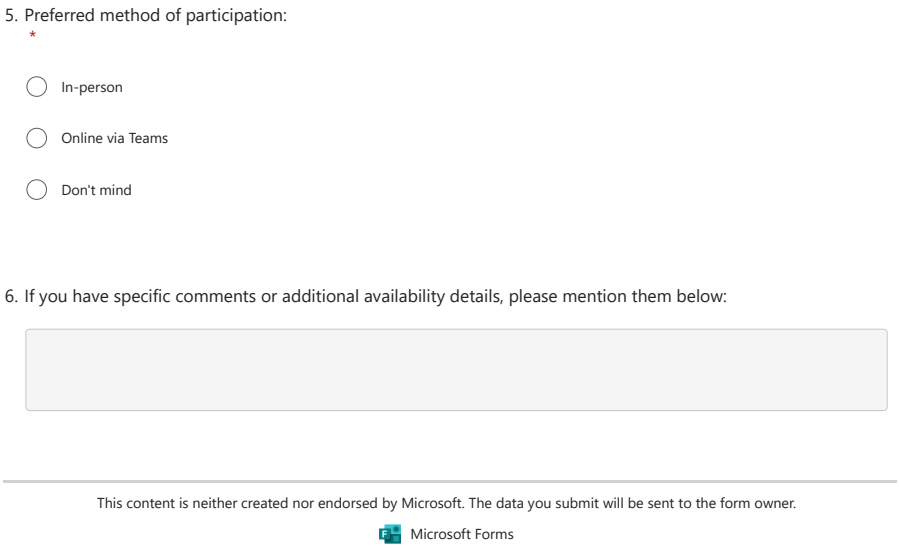
- [ ] Monday
- [ ] Tuesday
- [ ] Wednesday
- [ ] Thursday
- [ ] Friday
- [ ] Other

4. Please indicate your preferred times(s): *

Please check all options that apply.

- [ ] 9:00am - 10am
- [ ] 10am - 11am
- [ ] 11am - 12pm
- [ ] 12pm - 1pm
- [ ] 1pm - 2pm
- [ ] 2pm - 3pm
- [ ] 3pm - 4pm
- [ ] 4pm - 5pm
- [ ] 5pm - 6pm
- [ ] Other

Figure 36: Interest form for general user participants (page 1).


5. Preferred method of participation:
*

In-person

Online via Teams

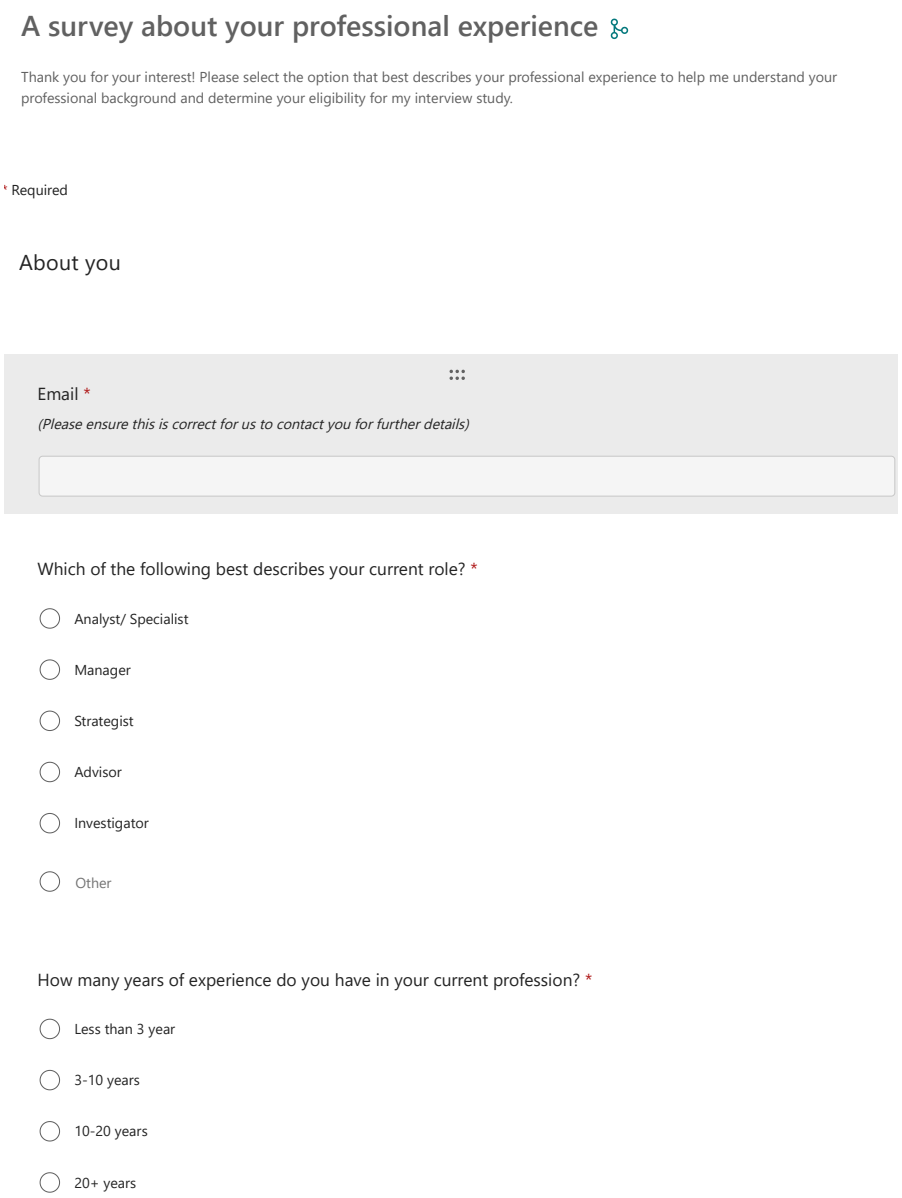
Don't mind

6. If you have specific comments or additional availability details, please mention them below:

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

Figure 37: Interest form for general user participants (page 2).



A survey about your professional experience

Thank you for your interest! Please select the option that best describes your professional experience to help me understand your professional background and determine your eligibility for my interview study.

* Required

About you

Email *

(Please ensure this is correct for us to contact you for further details)

Which of the following best describes your current role? *

- Analyst/ Specialist
- Manager
- Strategist
- Advisor
- Investigator
- Other

How many years of experience do you have in your current profession? *

- Less than 3 year
- 3-10 years
- 10-20 years
- 20+ years

Figure 38: Interest form for industry experts participants (page 1).

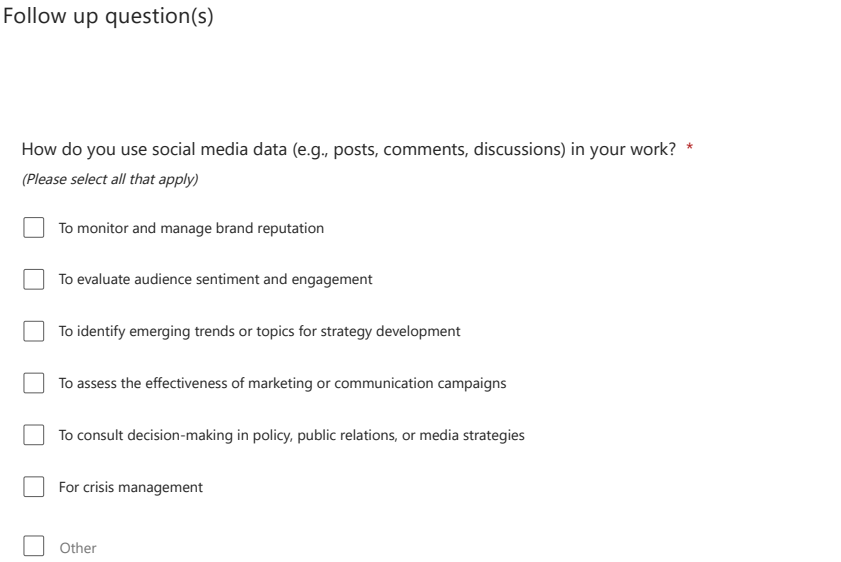

Which industry or sector do you work in? *

- ◯ Marketing, Branding, and Public Relations
- ◯ Journalism, Media, Communications
- ◯ Policy, Government, Politics
- ◯ Healthcare
- ◯ Technology
- ◯ Other

Does your role involve analysing social media discussions to support decision-making or achieve specific objectives? *

- ◯ Yes, my role is focused on analysing social media discussions for decision-making.
- ◯ Yes, I occasionally analyse social media discussions for decision-making.
- ◯ No, analysing social media discussions is not part of my role.

Figure 39: Interest form for industry experts participants (page 2).


Follow up question(s)

How do you use social media data (e.g., posts, comments, discussions) in your work? *

(Please select all that apply)

- ☐ To monitor and manage brand reputation
- ☐ To evaluate audience sentiment and engagement
- ☐ To identify emerging trends or topics for strategy development
- ☐ To assess the effectiveness of marketing or communication campaigns
- ☐ To consult decision-making in policy, public relations, or media strategies
- ☐ For crisis management
- ☐ Other

Figure 40: Interest form for industry experts participants (page 3).

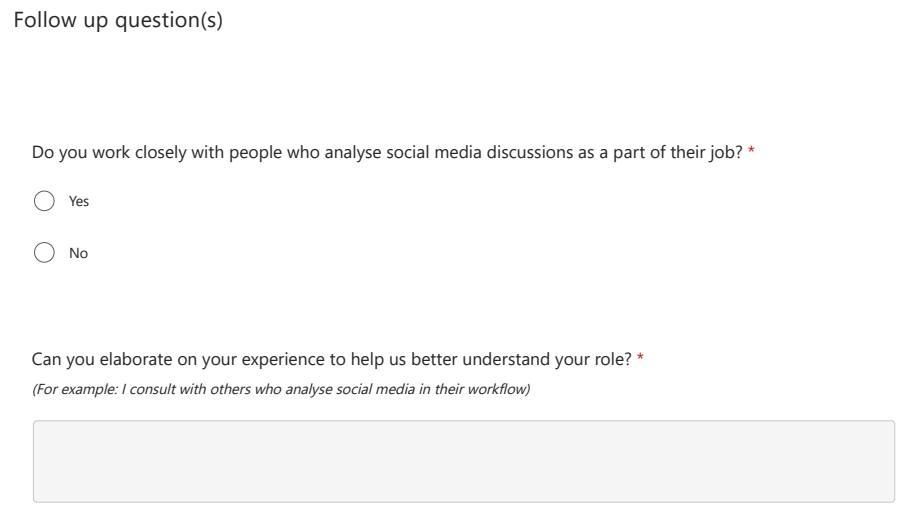


Figure 41: Interest form for industry experts participants (page 4).

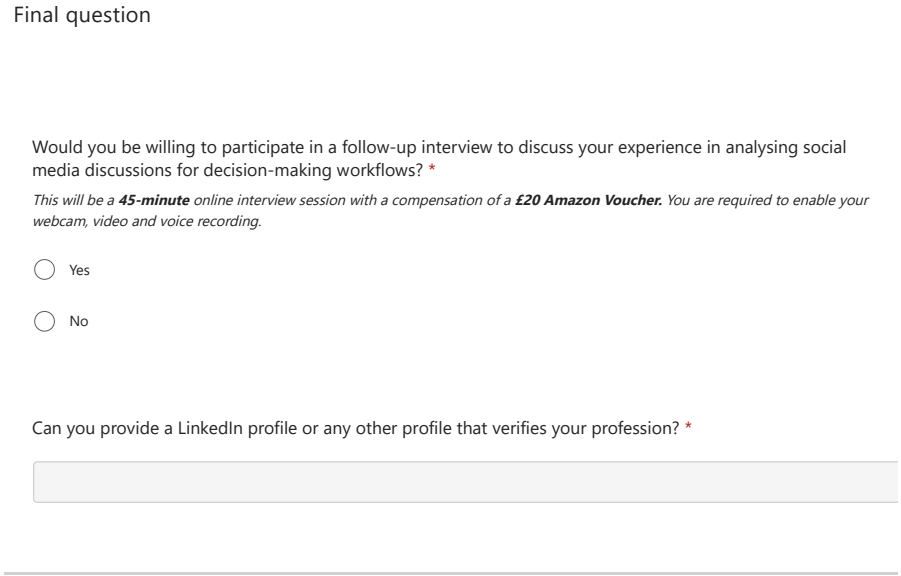
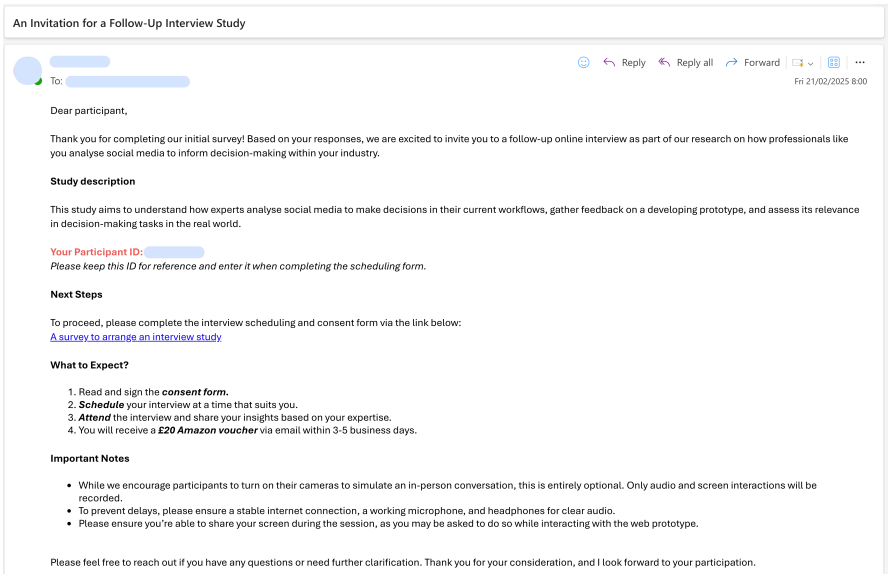


Figure 42: Interest form for industry experts participants (page 5).

A.4.6 *Scheduling and follow-up procedures*

Following the expression of interest or successful screening, participants were guided through a scheduling process according to the recruitment platform. Communications included consent information and instructions for selecting an interview slot. General user participants were contacted via Microsoft Teams or email to arrange inter-

view times. Selected industry expert participants were contacted via personalised email with a scheduling link (see Figure 43). For Prolific, qualified experts received an invitation to complete a follow-up survey with scheduling instructions (see Figure 44, 45).



An Invitation for a Follow-Up Interview Study

Reply Reply all Forward

To:

Fri 21/02/2025 8:00

Dear participant,

Thank you for completing our initial survey! Based on your responses, we are excited to invite you to a follow-up online interview as part of our research on how professionals like you analyse social media to inform decision-making within your industry.

Study description

This study aims to understand how experts analyse social media to make decisions in their current workflows, gather feedback on a developing prototype, and assess its relevance in decision-making tasks in the real world.

Your Participant ID:
Please keep this ID for reference and enter it when completing the scheduling form.

Next Steps

To proceed, please complete the interview scheduling and consent form via the link below:
A survey to arrange an interview study

What to Expect?

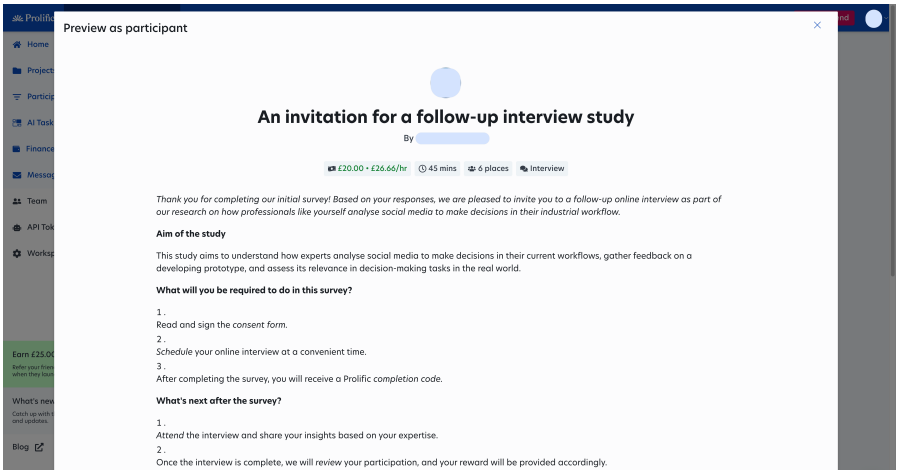
1. Read and sign the ***consent form.***
2. ***Schedule*** your interview at a time that suits you.
3. ***Attend*** the interview and share your insights based on your expertise.
4. You will receive a ***£20 Amazon voucher*** via email within 3-5 business days.

Important Notes

- While we encourage participants to turn on their cameras to simulate an in-person conversation, this is entirely optional. Only audio and screen interactions will be recorded.
- To prevent delays, please ensure a stable internet connection, a working microphone, and headphones for clear audio.
- Please ensure you're able to share your screen during the session, as you may be asked to do so while interacting with the web prototype.

Please feel free to reach out if you have any questions or need further clarification. Thank you for your consideration, and I look forward to your participation.

Figure 43: Example of email invitation sent to qualified LinkedIn experts.



Preview as participant

An invitation for a follow-up interview study

By

£20.00 • £26.66/hr | 45 mins | 6 places | Interview

Thank you for completing our initial survey! Based on your responses, we are pleased to invite you to a follow-up online interview as part of our research on how professionals like yourself analyse social media to make decisions in their industrial workflow.

Aim of the study

This study aims to understand how experts analyse social media to make decisions in their current workflows, gather feedback on a developing prototype, and assess its relevance in decision-making tasks in the real world.

What will you be required to do in this survey?

1. Read and sign the *consent form.*
2. *Schedule* your online interview at a convenient time.
3. After completing the survey, you will receive a Prolific *completion code.*

What's next after the survey?

1. *Attend* the interview and share your insights based on your expertise.
2. Once the interview is complete, we will *review* your participation, and your reward will be provided accordingly.

Figure 44: An invitation survey posted via Prolific to selected experts (part 1).

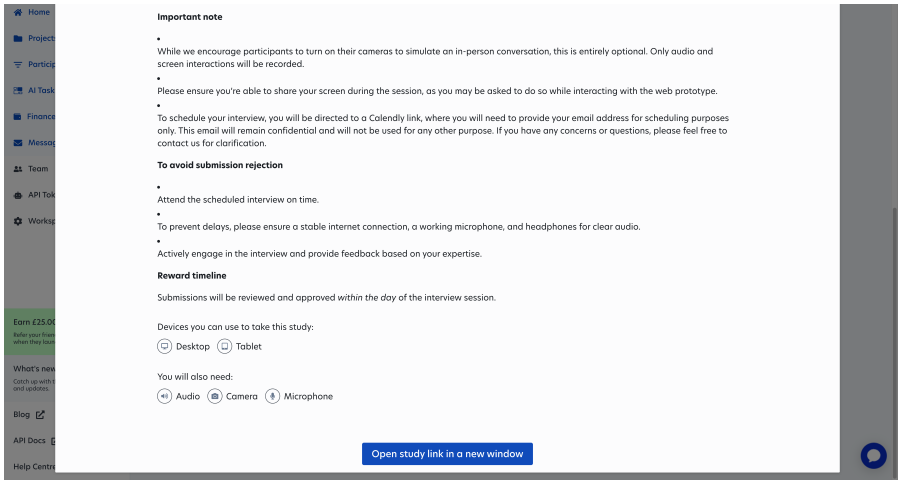


Figure 45: An invitation survey posted via Prolific to selected experts (part 2).

BIBLIOGRAPHY

ALDayel, Abeer and Walid Magdy (July 2021). "Stance detection on social media: State of the art and trends." In: *Inf. Process. Manage.* 58.4. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2021.102597.

Abascal, Julio and Colette Nicolle (Sept. 2005). "Moving towards inclusive design guidelines for socially and ethically aware HCI." en. In: *Interacting with Computers* 17.5, pp. 484–505. ISSN: 09535438. DOI: 10.1016/j.intcom.2005.03.002.

Abdul, Ashraf, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli (2018). "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, pp. 1–18. ISBN: 9781450356206. DOI: 10.1145/3173574.3174156.

Ackerman, Mark S. (Sept. 2000). "The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility." en. In: *Human-Computer Interaction* 15.2-3, pp. 179–203. ISSN: 0737-0024, 1532-7051. DOI: 10.1207/S15327051HCI1523_5.

Aïmeur, Esma, Sabrine Amri, and Gilles Brassard (2023). "Fake news, disinformation and misinformation in social media: a review." In: *Social Network Analysis and Mining* 13.1, p. 30. DOI: 10.1007/s13278-023-01028-5.

Akter, Shahriar, Yogesh K. Dwivedi, Shahriar Sajib, Kumar Biswas, Ruwan J. Bandara, and Katina Michael (May 2022). "Algorithmic bias in machine learning-based marketing models." en. In: *Journal of Business Research* 144, pp. 201–216. ISSN: 01482963. DOI: 10.1016/j.jbusres.2022.01.083.

Al Sharou, Khetam, Zhenhao Li, and Lucia Specia (2021). "Towards a better understanding of noise in natural language processing."

In: *Proceedings of the International conference on recent advances in natural language processing (RANLP 2021)*, pp. 53–62. URL: https://aclanthology.org/2021.ranlp-1.7/.

Alam, Firoj, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov (2022). *A Survey on Multimodal Disinformation Detection*. arXiv: 2103.12541 [cs.MM].

Albladi, Aish, Minarul Islam, Amit Das, Maryam Bigonah, Zheng Zhang, Fatemeh Jamshidi, Mostafa Rahgouy, Nilanjana Raychawdhary, Daniela Marghitu, and Cheryl Seals (2025). "Hate speech detection using large language models: A comprehensive review." In: *IEEE Access*. DOI: 10.1109/ACCESS.2025.3532397.

Alkomah, Fatimah and Xiaogang Ma (2022). "A literature review of textual hate speech detection methods and datasets." In: *Information* 13.6, p. 273. DOI: 10.3390/info13060273.

Alonso, Miguel A, David Vilares, Carlos Gómez-Rodríguez, and Jesús Vilares (2021). "Sentiment analysis for fake news detection." In: *Electronics* 10.11, p. 1348. DOI: 10.3390/electronics10111348.

Alsaif, Hissa F and Hmood D Aldossari (2023). "Review of stance detection for rumor verification in social media." In: *Engineering Applications of Artificial Intelligence* 119, p. 105801. DOI: 10.1016/j.engappai.2022.105801.

Amershi, Saleema, Maya Cakmak, William Bradley Knox, and Todd Kulesza (2014). "Power to the people: The role of humans in interactive machine learning." In: *AI magazine* 35.4, pp. 105–120. DOI: 10.1609/aimag.v35i4.2513.

Amershi, Saleema et al. (2019). "Guidelines for Human-AI Interaction." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, pp. 1–13. ISBN: 9781450359702. DOI: 10.1145/3290605.3300233.

Anthropic (2025). *Claude*. URL: https://claude.ai.

Appel, Gil, Lauren Grewal, Rhonda Hadi, and Andrew T Stephen (2020). "The future of social media in marketing." In: *Journal of the*

Academy of Marketing science 48.1, pp. 79–95. DOI: 10.1007/s11747-019-00695-1.

Atreja, Shubham, Shruthi Srinath, Mohit Jain, and Joyojeet Pal (2023). "Understanding Journalists' Workflows in News Curation." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery. ISBN: 9781450394215. DOI: 10.1145/3544548.3581542.

Ausat, Abu Muna Almaududi (Aug. 2023). "The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions." In: *Technology and Society Perspectives (TACIT)* 1.1, pp. 35–44. ISSN: 3025-5104. DOI: 10.61100/tacit.v1i1.37.

Azzaakiyyah, Hizbul Khootimah (2023). "The Impact of Social Media Use on Social Interaction in Contemporary Society." In: *Technology and Society Perspectives (TACIT)* 1.1, pp. 1–9. DOI: 10.61100/tacit.v1i1.33.

Bakir, Vian and Andrew McStay (2018). "Fake news and the economy of emotions: Problems, causes, solutions." In: *Digital journalism* 6.2, pp. 154–175. DOI: 10.1080/21670811.2017.1345645.

Baldassarre, Brian, Giulia Calabretta, Ingo Oswald Karpen, Nancy Bocken, and Erik Jan Hultink (Nov. 2024). "Responsible Design Thinking for Sustainable Development: Critical Literature Review, New Conceptual Framework, and Research Agenda." en. In: *Journal of Business Ethics* 195.1, pp. 25–46. ISSN: 0167-4544, 1573-0697. DOI: 10.1007/s10551-023-05600-z.

Banerjee, Debanjan and TS Sathyanarayana Rao (2020). "Psychology of misinformation and the media: Insights from the COVID-19 pandemic." In: *Indian Journal of Social Psychiatry* 36.Suppl 1, S131–S137. DOI: 10.4103/ijsp.ijsp_112_20.

Barel, Guy, Oren Tsur, and Dan Vilenchik (Jan. 2025). "Acquired taste: Multimodal stance detection with textual and structural embeddings." In: *Proceedings of the 31st International Conference on Computational Linguistics*. Abu Dhabi, UAE: Association for Compu-

tational Linguistics, pp. 6492–6504. URL: https://aclanthology.org/2025.coling-main.433/.

Bauer, Nikolaj, Moritz Preisig, and Martin Volk (May 2024). "Offensiveness, Hate, Emotion and GPT: Benchmarking GPT3.5 and GPT4 as Classifiers on Twitter-specific Datasets." In: *Proceedings of the Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024*. Torino, Italia: ELRA and ICCL, pp. 126–133.

Benamara, Farah, Diana Inkpen, and Maite Taboada (Dec. 2018). "Introduction to the special issue on language in social media: Exploiting discourse and other contextual information." In: *Comput. Linguist.* 44.4, pp. 663–681. ISSN: 0891-2017. DOI: 10.1162/coli_a_00333.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell (Mar. 2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" en. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM, pp. 610–623. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445922.

Bessi, Alessandro, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi (2015). "Trend of narratives in the age of misinformation." In: *PloS one* 10.8, e0134641. DOI: 10.1371/journal.pone.0134641.

Bhuiyan, Md Momen, Michael Horning, Sang Won Lee, and Tanushree Mitra (Oct. 2021). "NudgeCred: Supporting News Credibility Assessment on Social Media Through Nudges." en. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW2, pp. 1–30. ISSN: 2573-0142. DOI: 10.1145/3479571.

Bhutani, Bhavika, Neha Rastogi, Priyanshu Sehgal, and Archana Purwar (2019). "Fake news detection using sentiment analysis." In: *2019 twelfth international conference on contemporary computing (IC3)*. IEEE, pp. 1–5. DOI: 10.1109/IC3.2019.8844880.

Binns, Reuben, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt (2018). "'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions." In:

Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18. Montreal QC, Canada: Association for Computing Machinery, pp. 1–14. ISBN: 9781450356206. DOI: 10.1145/3173574.3173951.

Blandford, Ann, Dominic Furniss, and Stephann Makri (2016). *Qualitative Hci Research: Going Behind the Scenes*. Morgan & Claypool Publishers. ISBN: 1627057595. DOI: 10.5555/3019315.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation." In: *Journal of machine Learning research* 3.Jan, pp. 993–1022. DOI: 10.5555/944919.944937.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov (2017). *Enriching Word Vectors with Subword Information*. arXiv: 1607.04606 [cs.CL].

Bonta, Venkateswarlu, Nandhini Kumaresh, and N. Janardhan (Jan. 2019). "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis." In: *Asian Journal of Computer Science and Technology* 8.S2, pp. 1–6. DOI: 10.51983/ajcst-2019.8.S2.2037.

Bouvier, Gwen and David Machin (2020). "Critical discourse analysis and the challenges and opportunities of social media." In: *Critical Discourse Studies and/in Communication*, pp. 39–53. DOI: 10.1080/15358593.2018.1479881.

Brambilla, Marco and Kalyani Kharmale (2022). *COVID-19 Vaccine Discussions on Reddit with Sentiment, Stance, Topics, and Timing*. Version V1. DOI: 10.7910/DVN/XJTBQM.

Braun, Virginia and Victoria Clarke (2006). "Using thematic analysis in psychology." In: *Qualitative research in psychology* 3.2, pp. 77–101. ISSN: 1478-0887. DOI: 10.1191/1478088706qp063oa.

Braun, Virginia and Victoria Clarke (2012). *Thematic analysis*. American Psychological Association.

Brooke, John et al. (1996). "SUS-A quick and dirty usability scale." In: *Usability evaluation in industry* 189.194, pp. 4–7. DOI: 10.1201/9781498710411-35.

Brown, Peter F., Vincent J. Della Pietra, Peter V. Desouza, Jennifer C. Lai, and Robert L. Mercer (1992). "Class-based n-gram models of natural language." In: *Computational linguistics* 18.4, pp. 467–480.

Bu, Yuyan, Qiang Sheng, Juan Cao, Peng Qi, Danding Wang, and Jintao Li (2023). "Combating online misinformation videos: Characterization, detection, and future directions." In: *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8770–8780. DOI: 10.1145/3581783.3612426.

Caled, Danielle and Mário J Silva (2022). "Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation." In: *Journal of Computational Social Science* 5.1, pp. 123–159. DOI: 10.1007/s42001-021-00118-8.

Carroll, John M (2003). *Making use: scenario-based design of human-computer interactions*. MIT press. ISBN: 0-262-51388-9.

Caruana, Rich and Alexandru Niculescu-Mizil (2006). "An empirical comparison of supervised learning algorithms." en. In: *Proceedings of the 23rd international conference on Machine learning - ICML '06*. Pittsburgh, Pennsylvania: ACM Press, pp. 161–168. DOI: 10.1145/1143844.1143865.

Cavaliere, Paolo (2020). "From journalistic ethics to fact-checking practices: defining the standards of content governance in the fight against disinformation." In: *Journal of media law* 12.2, pp. 133–165. DOI: 10.1080/17577632.2020.1869486.

Ceron, Andrea, Luigi Curini, Stefano M. Iacus, and Giuseppe Porro (2014). "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France." In: *New media & society* 16.2, pp. 340–358. DOI: 10.1177/14614448134804.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (June 2002). "SMOTE: Synthetic Minority Over-sampling Technique." In: *Journal of Artificial Intelligence Research* 16, pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953.

Chen, Li, Shifeng Shang, and Yawen Wang (2025). "Bridging resource gaps in cross-lingual sentiment analysis: adaptive self-alignment

with data augmentation and transfer learning." In: *PeerJ Computer Science* 11, e2851. DOI: 10.7717/peerj-cs.2851.

Chen, Siming, Lijing Lin, and Xiaoru Yuan (June 2017). "Social Media Visual Analytics." en. In: *Computer Graphics Forum* 36.3, pp. 563–587. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.13211.

Chhabra, Anusha and Dinesh Kumar Vishwakarma (Jan. 2023). "A literature survey on multimodal and multilingual automatic hate speech identification." In: *Multimedia Syst.* 29.3, pp. 1203–1230. ISSN: 0942-4962. DOI: 10.1007/s00530-023-01051-8.

Cho, HyunYi, Wenbo Li, and Rachel Lopez (Aug. 2024). "A Multidimensional Approach for Evaluating Reality in Social Media: Mixed Methods Study." In: *J Med Internet Res* 26, e52058. ISSN: 1438-8871. DOI: 10.2196/52058.

Cho, Kyunghyun, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio (2014). "On the properties of neural machine translation: Encoder-decoder approaches." In: *arXiv preprint*. DOI: 10.3115/v1/W14-4012.

Choi, Daejin, Jinyoung Han, Taejoong Chung, Yong-Yeol Ahn, Byung-Gon Chun, and Ted Taekyoung Kwon (2015). "Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors." In: *Proceedings of the 2015 ACM on Conference on Online Social Networks*. COSN '15. Palo Alto, California, USA: Association for Computing Machinery, pp. 233–243. ISBN: 9781450339513. DOI: 10.1145/2817946.2817959.

Cinelli, Matteo, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo (2021). "Dynamics of online hate and misinformation." In: *Scientific reports* 11.1, p. 22083. DOI: 10.1038/s41598-021-01487-w.

Cinelli, Matteo, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala (2020). "The COVID-19 social media infodemic." In: *Scientific reports* 10.1, p. 16598. DOI: 10.1038/s41598-020-73510-5.

Cock Buning, Madeleine de (2018). *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation*. Publications Office of the European Union.

Cohen, Seth (May 2020). *'For Once, Don't Do It': The Powerful Idea Behind Nike's New Anti-Racism Ad*. Publication Title: Forbes. URL: https://www.forbes.com/sites/sethcohen/2020/05/30/for-once-dont-do-it---the-powerful-idea-behind-nikes-new-anti-racism-ad/.

Cotfas, Liviu-Adrian, Camelia Delcea, Ioan Roxin, Corina Ioanas, Dana Simona Gherai, and Federico Tajariol (2021). "The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement." In: *IEEE Access* 9, pp. 33203–33223. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3059821.

Creswell, John W (1999). "Mixed-method research: Introduction and application." In: *Handbook of educational policy*. Elsevier, pp. 455–472. DOI: 10.1016/B978-012174698-8/50045.

Creswell, John W (2021). *A concise introduction to mixed methods research*. SAGE publications.

D'Sa, Ashwin Geet, Irina Illina, and Dominique Fohr (Feb. 2020). "BERT and fastText Embeddings for Automatic Detection of Toxic Speech." In: *2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA)*. Tunis, Tunisia: IEEE, pp. 1–5. ISBN: 978-1-7281-6403-8. DOI: 10.1109/OCTA49274.2020.9151853.

Dai, Yibo, George Karalis, Saba Kawas, and Chris Olsen (2015). "Tipper: Contextual Tooltips that Provide Seniors with Clear, Reliable Help for Web Tasks." In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '15. Seoul, Republic of Korea: Association for Computing Machinery, pp. 1773–1778. ISBN: 9781450331463. DOI: 10.1145/2702613.2732796.

Danilevsky, Marina, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen (2020). "A Survey of the State of Explainable AI for Natural Language Processing." In: *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, pp. 447–459. DOI: 10.18653/v1/2020.aacl-main.46. URL: http://dx.doi.org/10.18653/v1/2020.aacl-main.46.

De Choudhury, Munmun, Hari Sundaram, Ajita John, and Dorée Duncan Seligmann (2009). "What makes conversations interesting? themes, participants and consequences of conversations in online social media." In: *Proceedings of the 18th international conference on World wide web*, pp. 331–340. DOI: 10.1145/1526709.1526754.

De Magistris, Giorgio, Samuele Russo, Paolo Roma, Janusz T Starczewski, and Christian Napoli (2022). "An explainable fake news detector based on named entity recognition and stance classification applied to covid-19." In: *Information* 13.3, p. 137. DOI: 10.3390/info13030137.

Dekavalla, Marina (2022). "Understanding online safety through metaphors: UK policymakers and industry discourses about the internet." In: *Television & New Media* 23.8, pp. 855–873. DOI: 10.1177/15274764211042077.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (May 2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. DOI: 10.48550/arXiv.1810.04805.

Di Domenico, Giandomenico, Jason Sit, Alessio Ishizaka, and Daniel Nunan (2021). "Fake news, social media and marketing: A systematic review." In: *Journal of Business Research* 124, pp. 329–341. DOI: 10.1016/j.jbusres.2020.11.037.

Diaz-Garcia, J. Angel and Joao Paulo Carvalho (June 2025). "A Literature Review of Textual Cyber Abuse Detection Using Cutting-Edge Natural Language Processing Techniques: Language Models and Large Language Models." In: *WIREs Data Mining and*

Knowledge Discovery 15.3. ISSN: 1942-4795. DOI: 10.1002/widm.70029.

Dieng, Adji B, Francisco JR Ruiz, and David M Blei (2020). "Topic modeling in embedding spaces." In: *Transactions of the Association for Computational Linguistics* 8, pp. 439–453. DOI: 10.1162/tacl_a_00325. URL: https://aclanthology.org/2020.tacl-1.29/.

Digital Culture, Media & Sport Department for and Home Office (2020). *Online harms white paper: Full government response to the consultation*. URL: https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response.

Dix, Alan, ed. (2010). *Human-computer interaction*. 3. ed., [6. Nachdr.] Harlow: Pearson Prentice-Hall. ISBN: 978-0-13-046109-4.

Domingos, Pedro (Oct. 2012). "A few useful things to know about machine learning." en. In: *Communications of the ACM* 55.10, pp. 78–87. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2347736.2347755.

Doshi-Velez, Finale and Been Kim (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. DOI: 10.48550/ARXIV.1702.08608.

Dourish, Paul (2006). "Implications for design." In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 541–550. DOI: 10.1145/1124772.1124855.

Dragicevic, Pierre (2016). "Fair statistical communication in HCI." In: *Modern statistical methods for HCI*. Springer, pp. 291–330. DOI: 10.1007/978-3-319-26633-6_13.

Drouin, Michelle, Brandon T McDaniel, Jessica Pater, and Tammy Toscos (2020). "How parents and their children used social media and technology at the beginning of the COVID-19 pandemic and associations with anxiety." In: *Cyberpsychology, Behavior, and Social Networking* 23.11, pp. 727–736. DOI: 10.1089/cyber.2020.0284.

Edmonds, W Alex and Thomas D Kennedy (2016). *An applied guide to research designs: Quantitative, qualitative, and mixed methods*. Sage Publications. DOI: 10.4135/9781071802779.

Egelhofer, Jana Laura and Sophie Lecheler (2019). "Fake news as a two-dimensional phenomenon: A framework and research

agenda." In: *Annals of the International Communication Association* 43.2, pp. 97–116. DOI: 10.1080/23808985.2019.1602782.

Ehsan, Upol, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz (2021). "Expanding Explainability: Towards Social Transparency in AI systems." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery. ISBN: 9781450380966. DOI: 10.1145/3411764.3445188.

Ehsan, Upol and Mark O Riedl (2020). "Human-centered explainable ai: Towards a reflective sociotechnical approach." In: *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, pp. 449–466. DOI: 10.1007/978-3-030-60117-1_33.

Elmas, Tuğrulcan, Selim Stephane, and Célia Houssiaux (2023). "Measuring and Detecting Virality on Social Media: The Case of Twitter's Viral Tweets Topic." In: *Companion Proceedings of the ACM Web Conference 2023*. WWW '23 Companion. Austin, TX, USA: Association for Computing Machinery, pp. 314–317. ISBN: 9781450394192. DOI: 10.1145/3543873.3587373.

Endert, Alex, William Ribarsky, Cagatay Turkay, BL William Wong, Ian Nabney, I Díaz Blanco, and Fabrice Rossi (2017). "The state of the art in integrating machine learning into visual analytics." In: *Computer Graphics Forum*. Vol. 36. Issue: 8. Wiley Online Library, pp. 458–486. DOI: 10.1111/cgf.13092.

Epstein, Ziv, Matthew Groh, Abhimanyu Dubey, and Alex Pentland (Oct. 2021). "Social Influence Leads to the Formation of Diverse Local Trends." In: *Proc. ACM Hum.-Comput. Interact.* 5.CSCW2. DOI: 10.1145/3479553.

Ericsson, K Anders, Robert R Hoffman, Aaron Kozbelt, and A Mark Williams (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press. DOI: 10.1017/9781316480748.

Evolvi, Giulia (2018). "Hate in a tweet: Exploring internet-based islamophobic discourses." In: *Religions* 9.10, p. 307. DOI: 10.3390/rel9100307.

Feigh, Karen M., Michael C. Dorneich, and Caroline C. Hayes (Dec. 2012). "Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers." en. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54.6, pp. 1008–1024. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720812443983.

Fernando, K. Ruwani M. and Chris P. Tsokos (July 2022). "Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks." In: *IEEE Transactions on Neural Networks and Learning Systems* 33.7, pp. 2940–2951. ISSN: 2162-237X, 2162-2388. DOI: 10.1109/TNNLS.2020.3047335.

Flintham, Martin, Christian Karner, Khaled Bachour, Helen Creswick, Neha Gupta, and Stuart Moran (Apr. 2018). "Falling for Fake News: Investigating the Consumption of News via Social Media." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, pp. 1–10. DOI: 10.1145/3173574.3173950.

Fortuna, Paula and Sérgio Nunes (July 2018). "A Survey on Automatic Detection of Hate Speech in Text." In: *ACM Comput. Surv.* 51.4. ISSN: 0360-0300. DOI: 10.1145/3232676.

Frederick, Jennifer K. (Jan. 2021). *When to ask (or not ask) demographic questions*. Available at: https://sr.ithaka.org/blog/when-to-ask-or-not-ask-demographic-questions/. (accessed May 16, 2025).

Friedman, Milton (Dec. 1937). "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." en. In: *Journal of the American Statistical Association* 32.200, pp. 675–701. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1937.10503522.

Garmur, Matt, Gary King, Zagreb Mukerjee, Nate Persily, and Brandon Silverman (2019). *CrowdTangle Platform and API*. DOI: 10.7910/DVN/SCCQYD. (Visited on 05/20/2025).

Giacomin, Joseph (2014). "What Is Human Centred Design?" In: *The Design Journal* 17.4, pp. 606–623. ISSN: 1460-6925. DOI: 10.2752/175630614X14056185480186.

Golafshani, Nahid (2003). "Understanding reliability and validity in qualitative research." In: *The qualitative report* 8.4, pp. 597–607. DOI: 10.46743/2160-3715/2003.1870.

Goldzycher, Janis and Gerold Schneider (Oct. 2022). "Hypothesis Engineering for Zero-Shot Hate Speech Detection." In: *Proceedings of the Third Workshop on Threat, Aggression and Cyberbullying (TRAC 2022)*. Gyeongju, Republic of Korea: Association for Computational Linguistics, pp. 75–90.

Goodman, Joy, Patrick Langdon, and P. John Clarkson (2007). "Formats for user data in inclusive design." In: *Proceedings of the 4th International Conference on Universal Access in Human Computer Interaction: Coping with Diversity*. UAHCI'07. Beijing, China: Springer-Verlag, pp. 117–126. ISBN: 9783540732785.

Gorini, Alessandra, Claret S Capideville, Gianluca De Leo, Fabrizia Mantovani, and Giuseppe Riva (2011). "The role of immersion and narrative in mediated presence: the virtual hospital experience." In: *Cyberpsychology, behavior, and social networking* 14.3, pp. 99–105. DOI: 10.1089/cyber.2010.0100.

Graves, Alex (2012). "Long short-term memory." In: *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 37–45. ISBN: 978-3-642-24797-2. DOI: 10.1007/978-3-642-24797-2_4.

Green, Ben and Yiling Chen (Nov. 2019). "The Principles and Limits of Algorithm-in-the-Loop Decision Making." In: *Proc. ACM Hum.-Comput. Interact.* 3.CSCW. DOI: 10.1145/3359152.

Güner, Samet, Halil Ibrahim Cebeci, and Emrah Aydemir (2025). "How popular is a topic on social media? A multi-criteria

decision-making framework based on user engagement." In: *Kybernetes* 54.1, pp. 414–430. DOI: 10.1108/K-05-2023-0884.

Gunson, Nancie, Weronika Sieińska, Yanchao Yu, Daniel Hernandez Garcia, Jose L. Part, Christian Dondrup, and Oliver Lemon (Sept. 2021). "Coronabot: A Conversational AI System for Tackling Misinformation." In: *Proceedings of the Conference on Information Technology for Social Good*. Roma Italy: ACM, pp. 265–270. DOI: 10.1145/3462203.3475874.

Guo, Bao, Chunxia Zhang, Junmin Liu, and Xiaoyi Ma (2019). "Improving text classification with weighted word embeddings via a multi-channel TextCNN model." In: *Neurocomputing* 363, pp. 366–374. DOI: 10.1016/j.neucom.2019.07.052.

Guo, Yujuan, Di Fan, and Xiao Zhang (2020). "Social media-based customer service and firm reputation." In: *International Journal of Operations & Production Management* 40.5, pp. 575–601. DOI: 10.1108/IJOPM-04-2019-0315.

Gupta, Aditi, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier (2014). "Tweetcred: Real-time credibility assessment of content on twitter." In: *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings 6*. Springer, pp. 228–243. DOI: 10.48550/arXiv.1405.5490.

Haleem, Abid, Mohd Javaid, and Ravi Pratap Singh (2022). "An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges." In: *BenchCouncil transactions on benchmarks, standards and evaluations* 2.4, p. 100089. DOI: 10.1016/j.tbench.2023.100089.

Hamm, Andrea (Sept. 2024). "New Objects, New Boundaries: How the "Journalism of Things" Reconfigures Collaborative Arrangements, Audience Relations and Knowledge-Based Empowerment." en. In: *Digital Journalism* 12.8, pp. 1077–1096. ISSN: 2167-0811, 2167-082X. DOI: 10.1080/21670811.2022.2096088.

Han, Lei, Rudra Sawant, Shaoyang Fan, Glenn Kefford, and Gianluca Demartini (Dec. 2021). "An Analysis of the Australian Political

Discourse in Sponsored Social Media Content." In: *Australasian Document Computing Symposium*. Virtual Event Australia: ACM, pp. 1–5. DOI: 10.1145/3503516.3503533.

Hardalov, Momchil, Arnav Arora, Preslav Nakov, and Isabelle Augenstein (July 2022). "A Survey on Stance Detection for Mis- and Disinformation Identification." In: *Findings of the Association for Computational Linguistics: NAACL 2022*. Seattle, United States: Association for Computational Linguistics, pp. 1259–1277. DOI: 10.18653/v1/2022.findings-naacl.94.

Hartson, Rex and Pardha Pyla (2012). *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 0123852412.

Hatebase (2022). *Hatebase*. English. URL: https://hatebase.org/ (visited on 09/19/2025).

Haupt, M.R., J. Li, and T.K. Mackey (2021). "Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on twitter using unsupervised machine learning." In: *Big Data and Society* 8.1. DOI: 10.1177/20539517211013843.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). "Deep Residual Learning for Image Recognition." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

Heer, Jeffrey and Maneesh Agrawala (2007). "Design considerations for collaborative visual analytics." In: *2007 IEEE symposium on visual analytics science and technology*. IEEE, pp. 171–178. DOI: 10.1145/1391107.1391112.

Hellmueller, Lea, Lindita Camaj, Sebastián Vallejo Vera, and Peggy Lindner (Oct. 2024). "The Impact of Journalistic Cultures on Social Media Discourse: US Primary Debates in Cross-Lingual Online Spaces." en. In: *Digital Journalism*, pp. 1–21. ISSN: 2167-0811, 2167-082X. DOI: 10.1080/21670811.2024.2402371. (Visited on 05/21/2025).

Hershcovich, Daniel et al. (2022). *Challenges and Strategies in Cross-Cultural NLP*. arXiv: 2203.10020 [cs.CL].

Hirschberg, Julia and Christopher D Manning (2015). "Advances in natural language processing." In: *Science* 349.6245, pp. 261–266. DOI: 10.1126/science.aaa8685.

Hofman, Jake M. et al. (July 2021). "Integrating explanation and prediction in computational social science." en. In: *Nature* 595.7866, pp. 181–188. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03659-0.

Holleman, Gijs A., Ignace T. C. Hooge, Chantal Kemner, and Roy S. Hessels (Apr. 2020). "The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity." In: *Frontiers in Psychology* 11, p. 721. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2020.00721.

Hollender, Nina, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz (2010). "Integrating cognitive load theory and concepts of human–computer interaction." In: *Computers in human behavior* 26.6, pp. 1278–1288. DOI: 10.1016/j.chb.2010.05.031.

Hsu, Tiffany (Jan. 2019). *Gillette Ad With a #MeToo Edge Attracts Support and Outrage*. Publication Title: The New York Times. URL: https://www.nytimes.com/2019/01/15/business/gillette-ad-men.html.

Huang, Lei et al. (Mar. 2025). "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." en. In: *ACM Transactions on Information Systems* 43.2, pp. 1–55. ISSN: 1046-8188, 1558-2868. DOI: 10.1145/3703155.

Humayun, Muhammad Fahad and Patrick Ferrucci (2022). "Understanding social media in journalism practice: A typology." In: *Digital Journalism* 10.9, pp. 1502–1525. DOI: 10.1080/21670811.2022.2086594.

Jahng, M Rosie (2021). "Is fake news the new social media crisis? Examining the public evaluation of crisis management for corporate organizations targeted in fake news." In: *International Journal*

of Strategic Communication 15.1, pp. 18–36. DOI: 10.1080/1553118X.2020.1848842.

Jiang, Hua, Yi Luo, and Owen Kulemeka (May 2016). "Leading in the digital age: A study of how social media are transforming the work of communication professionals." In: *Telematics and Informatics* 33.2, pp. 493–499. ISSN: 0736-5853. DOI: 10.1016/j.tele.2015.10.006. URL: https://www.sciencedirect.com/science/article/pii/S0736585315001264.

Jin, Ruidong, Xin Liu, and Tsuyoshi Murata (2024). "Predicting popularity trend in social media networks with multi-layer temporal graph neural networks." In: *Complex & Intelligent Systems* 10.4, pp. 4713–4729. DOI: 10.1007/s40747-024-01402-6.

Jin, Yan, Brooke Fisher Liu, and Lucinda L. Austin (Feb. 2014). "Examining the Role of Social Media in Effective Crisis Management: The Effects of Crisis Origin, Information Form, and Source on Publics' Crisis Responses." en. In: *Communication Research* 41.1, pp. 74–94. ISSN: 0093-6502, 1552-3810. DOI: 10.1177/0093650211423918.

Johnson, Neil F., Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu (June 2020). "The online competition between pro- and anti-vaccination views." en. In: *Nature* 582.7811, pp. 230–233. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2281-1.

Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury (2021). *The State and Fate of Linguistic Diversity and Inclusion in the NLP World*. arXiv: 2004.09095 [cs.CL].

K, Anandhan., Damodharan. D, Anupam Lakhanpal, K. Manoj Sagar, K. Murugan, and Ajay Shanker Singh (Mar. 2021). "Discover Pretend Disease News Misleading Data in Social Media Networks Using Machine Learning Techniques." In: *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. Greater Noida, India: IEEE, pp. 784–788. DOI: 10.1109/icacite51222.2021.9404648.

Kalantari, Niloofar, Duoduo Liao, and Vivian Genaro Motti (2021). "Characterizing the online discourse in Twitter: Users' reaction to misinformation around COVID-19 in Twitter." In: *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 4371–4380. DOI: 10.1109/BigData52589.2021.9671740.

Kaplan, Andreas M. and Michael Haenlein (Jan. 2010). "Users of the world, unite! The challenges and opportunities of Social Media." en. In: *Business Horizons* 53.1, pp. 59–68. ISSN: 0007-6813. DOI: 10.1016/j.bushor.2009.09.003.

Kaptein, Maurits Clemens, Clifford Nass, and Panos Markopoulos (2010). "Powerful and consistent analysis of likert-type rating scales." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: Association for Computing Machinery, pp. 2391–2394. ISBN: 9781605589299. DOI: 10.1145/1753326.1753686.

Karande, Hema, Rahee Walambe, Victor Benjamin, Ketan Kotecha, and TS Raghu (2021). "Stance detection with BERT embeddings for credibility analysis of information on social media." In: *PeerJ Computer Science* 7, e467. DOI: 10.7717/peerj-cs.467.

Kawintiranon, Kornraphop and Lisa Singh (2021). "Knowledge Enhanced Masked Language Model for Stance Detection." en. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, pp. 4725–4735. DOI: 10.18653/v1/2021.naacl-main.376.

Kemp, Simon (Feb. 2024). *Digital 2024: The United Kingdom*. Available at: https://datareportal.com/reports/digital-2024-united-kingdom. (accessed May 15, 2025).

Kemp, Simon (Mar. 2025). *The time we spend on social media - datareportal - global digital insights*. URL: https://datareportal.com/reports/digital-2024-deep-dive-the-time-we-spend-on-social-media.

Kennedy, Jay P, Melissa Rorie, and Michael L Benson (2021). "COVID-19 frauds: An exploratory study of victimization during a global

crisis." In: *Criminology & Public Policy* 20.3, pp. 493–543. DOI: 10.1111/1745-9133.12554.

Kheiri, Kiana and Hamid Karimi (2023). "Sentimentgpt: Exploiting gpt for advanced sentiment analysis and its departure from current machine learning." In: *arXiv preprint*. DOI: 10.48550/arXiv.2307.10234.

Khyani, Divya, B. S. Siddhartha, N. M. Niveditha, and B. M. Divya (2021). "An interpretation of lemmatization and stemming in natural language processing." In: *Journal of University of Shanghai for Science and Technology* 22.10, pp. 350–357.

Kikkisetti, Dhanush, Raza Ul Mustafa, Wendy Melillo, Roberto Corizzo, Zois Boukouvalas, Jeff Gill, and Nathalie Japkowicz (2024). *Using LLMs to discover emerging coded antisemitic hate-speech in extremist social media*. arXiv: 2401.10841 [cs.CL].

Kim, Ji Won (2018). "Rumor has it: The effects of virality metrics on rumor believability and transmission on Twitter." In: *New Media & Society* 20.12, pp. 4807–4825. DOI: 10.1177/1461444818784945.

Kim, Yoon (2014). "Convolutional Neural Networks for Sentence Classification." In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics. DOI: 10.3115/v1/d14-1181.

Kingma, Diederik P. and Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs.LG].

Kizgin, Hatice, Bidit L Dey, Yogesh K Dwivedi, Laurie Hughes, Ahmad Jamal, Paul Jones, Bianca Kronemann, Michel Laroche, Lisa Peñaloza, Marie-Odile Richard, et al. (2020). "The impact of social media on consumer acculturation: Current challenges, opportunities, and an agenda for research and practice." In: *International Journal of Information Management* 51, p. 102026. DOI: 10.1016/j.ijinfomgt.2019.10.011.

Kocielnik, Rafal, Saleema Amershi, and Paul N. Bennett (2019). "Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19.

Glasgow, Scotland Uk: Association for Computing Machinery, pp. 1–14. ISBN: 9781450359702. DOI: 10.1145/3290605.3300641.

Koh, Hyukhun, Dohyung Kim, Minwoo Lee, and Kyomin Jung (2024). *Can LLMs Recognize Toxicity? A Structured Investigation Framework and Toxicity Metric*. arXiv: 2402.06900 [cs.CL].

Kohavi, Ron (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection." In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'95. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., pp. 1137–1143. ISBN: 1558603638.

Kolko, Jon (2011). "Wicked Problems." en. In: *Thoughts on Interaction Design*. Elsevier, pp. 96–111. ISBN: 978-0-12-380930-8. DOI: 10.1016/B978-0-12-380930-8.50007-3.

Krugmann, Jan Ole and Jochen Hartmann (2024). "Sentiment analysis in the age of generative AI." In: *Customer Needs and Solutions* 11.1, p. 3. DOI: 10.1007/s40547-024-00143-4.

Kumar, Varun, Ashutosh Choudhary, and Eunah Cho (Dec. 2020). "Data augmentation using pre-trained transformer models." In: *Proceedings of the 2nd Workshop on Life-long Learning for Spoken Language Systems*. Suzhou, China: Association for Computational Linguistics, pp. 18–26. DOI: 10.18653/v1/2020.lifelongnlp-1.3.

Kvale, Steinar and Svend Brinkmann (2009). *Interviews: Learning the craft of qualitative research interviewing*. sage. ISBN: 0-7619-2542-2.

Küçük, Dilek and Fazli Can (Jan. 2021). "Stance Detection: A Survey." en. In: *ACM Computing Surveys* 53.1, pp. 1–37. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3369026.

Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman (2017). "Building machines that learn and think like people." en. In: *Behavioral and Brain Sciences* 40. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/s0140525x16001837.

Lam, Heidi, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale (2011). "Empirical studies in information visualization: Seven scenarios." In: *IEEE transactions on visualiza-*

tion and computer graphics 18.9, pp. 1520–1536. DOI: 10.1109/TVCG.2011.279.

Lama, Charles, Brian Leungb, Cora Yipb, and Jason Yungb (2020). "A linguistic approach to misinformation in chinese." In: *Proceedings http://ceur-ws. org ISSN* 1613, p. 0073. URL: https://ceur-ws.org/Vol-2723/short29.pdf.

Lan, Xiaochong, Chen Gao, Depeng Jin, and Yong Li (May 2024). "Stance Detection with Collaborative Role-Infused LLM-Based Agents." In: *Proceedings of the International AAAI Conference on Web and Social Media* 18.1, pp. 891–903.

Lazar, J., J.H. Feng, and H. Hochheiser (2017). *Research Methods in Human-Computer Interaction*. Morgan Kaufmann. ISBN: 978-0-12-809343-6.

Leal, Hugo, Stefanie Felsberger, Timothy Charlton, and Gina Neff (2025). *Enabling Data-Driven Research for Evidence-Based Interventions*. en. Tech. rep. Minderoo Centre for Technology and Democracy. DOI: 10.17863/CAM.115665.

Lee, Seongmin, Sadia Afroz, Haekyu Park, Zijie J. Wang, Omar Shaikh, Vibhor Sehgal, Ankit Peshin, and Duen Horng Chau (2022). "MisVis: Explaining Web Misinformation Connections via Visual Summary." In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA '22. New York, NY, USA: Association for Computing Machinery. ISBN: 978-1-4503-9156-6. DOI: 10.1145/3491101.3519711.

Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook (Dec. 2012). "Misinformation and Its Correction: Continued Influence and Successful Debiasing." en. In: *Psychological Science in the Public Interest* 13.3, pp. 106–131. ISSN: 1529-1006, 1539-6053. DOI: 10.1177/1529100612451018.

Li, Chen, Hao Peng, Jianxin Li, Lichao Sun, Lingjuan Lyu, Lihong Wang, Philip S. Yu, and Lifang He (June 2022). "Joint Stance and Rumor Detection in Hierarchical Heterogeneous Graph." In: *IEEE Transactions on Neural Networks and Learning Systems* 33.6,

pp. 2530–2542. ISSN: 2162-237X, 2162-2388. DOI: 10.1109/tnnls.2021.3114027.

Li, Yupeng, Haorui He, Shaonan Wang, Francis CM Lau, and Yunya Song (2023a). "Improved target-specific stance detection on social media platforms by delving into conversation threads." In: *IEEE Transactions on Computational Social Systems*. DOI: 10.48550/arXiv.2211.03061.

Li, Yupeng, Dacheng Wen, Haorui He, Jianxiong Guo, Xuan Ning, and Francis CM Lau (2023b). "Contextual Target-Specific Stance Detection on Twitter: Dataset and Method." In: *2023 IEEE International Conference on Data Mining (ICDM)*. IEEE, pp. 359–367. DOI: 10.1109/ICDM58522.2023.00045.

Lillie, Anders Edelbo and Emil Refsgaard Middelboe (2019). *Fake News Detection using Stance Classification: A Survey*. arXiv: 1907.00181 [cs.CL].

Lim, Youn-Kyung, Erik Stolterman, and Josh Tenenberg (July 2008). "The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas." en. In: *ACM Transactions on Computer-Human Interaction* 15.2, pp. 1–27. ISSN: 1073-0516, 1557-7325. DOI: 10.1145/1375761.1375762.

Lindgren, Helena, Patrik J Winnberg, and Peter Winnberg (2011). "Domain experts tailoring interaction to users–an evaluation study." In: *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part III 13*. Springer, pp. 644–661. DOI: 10.1007/978-3-642-23765-2_44.

Litman, Leib, Zohn Rosen, Cheskie Rosenzweig, Sarah L Weinberger-Litman, Aaron J Moss, and Jonathan Robinson (2020). "Did people really drink bleach to prevent COVID-19? A tale of problematic respondents and a guide for measuring rare events in survey data." In: *Medrxiv*, pp. 2020–12. DOI: 10.1371/journal.pone.0287837.

Liu, Bing (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Liu, Hui, Qingyu Yin, and William Yang Wang (2019). *Towards Explainable NLP: A Generative Explanation Framework for Text Classification*. arXiv: 1811.00196 [cs.CL].

Lovejoy, Kristen and Gregory D. Saxton (Apr. 2012). "Information, Community, and Action: How Nonprofit Organizations Use Social Media*." en. In: *Journal of Computer-Mediated Communication* 17.3, pp. 337–353. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2012.01576.x.

Lundberg, Scott M. and Su-In Lee (2017). "A unified approach to interpreting model predictions." In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., pp. 4768–4777. ISBN: 9781510860964. DOI: 10.5555/3295222.3295230.

Lupinacci, Ludmila (2021). "'Absentmindedly scrolling through nothing': liveness and compulsory continuous connectedness in social media." In: *Media, Culture & Society* 43.2, pp. 273–290. DOI: 10.1177/0163443720939454.

Ma, Junxia, Changjiang Wang, Hanwen Xing, Dongming Zhao, and Yazhou Zhang (2024). "Chain of stance: Stance detection with large language models." In: *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer, pp. 82–94. DOI: 10.1007/978-981-97-9443-0_7.

Madhav, Nita, Ben Oppenheim, Mark Gallivan, Peter Mulembakani, Elyse Rubin, and Nathan Wolfe (2017). "Pandemics: Risks, Impacts, and Mitigation." In: *Disease Control Priorities: Improving Health and Reducing Poverty*. Ed. by Dean T. Jamison, Hellen Gelband, Susan Horton, Prabhat Jha, Ramanan Laxminarayan, Charles N. Mock, and Rachel Nugent. 3rd ed. Washington, DC: The International Bank for Reconstruction and Development / The World Bank. Chap. 17. DOI: 10.1596/978-1-4648-0527-1_ch17.

Maguire, Martin (2001). "Context of use within usability activities." In: *International journal of human-computer studies* 55.4, pp. 453–483. DOI: 10.1006/ijhc.2001.0486.

Maharana, Kiran, Surajit Mondal, and Bhushankumar Nemade (June 2022). "A review: Data pre-processing and data augmentation techniques." en. In: *Global Transitions Proceedings* 3.1, pp. 91–99. ISSN: 2666285X. DOI: 10.1016/j.gltp.2022.04.020.

Mahoney, L Meghan and Tang Tang (2024). *Strategic social media: From marketing to social change*. John Wiley & Sons.

Majchrzak, Ann, Samer Faraj, Gerald C. Kane, and Bijan Azad (Oct. 2013). "The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing." en. In: *Journal of Computer-Mediated Communication* 19.1, pp. 38–55. ISSN: 1083-6101. DOI: 10.1111/jcc4.12030.

Majeed, Sadia, Muhammad Uzair, Usman Qamar, and Aftab Farooq (2020). "Social Network Analysis Visualization Tools: A Comparative Review." In: *2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE, pp. 1–6. DOI: 10.1109/INMIC50486.2020.9318162.

Mandl, Thomas, Sandip Modha, Prasenjit Majumder, Daksh Patel, Mohana Dave, Chintak Mandlia, and Aditya Patel (2019). "Overview of the HASOC track at FIRE 2019: Hate Speech and Offensive Content Identification in Indo-European Languages." In: *Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation*. FIRE '19. Kolkata, India: Association for Computing Machinery, pp. 14–17. ISBN: 9781450377508. DOI: 10.1145/3368567.3368584.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy (2014). "Sentiment analysis algorithms and applications: A survey." In: *Ain Shams engineering journal* 5.4, pp. 1093–1113. DOI: 10.1016/j.asej.2014.04.011.

Meikle, Graham (Feb. 2016). *Social Media*. en. 1st Edition. Routledge. ISBN: 978-1-134-66096-4. DOI: 10.4324/9781315884172.

Meuser, Michael and Ulrike Nagel (2009). "The expert interview and changes in knowledge production." In: *Interviewing experts*. Springer, pp. 17–42. DOI: 10.1057/9780230244276_2.

Miah, Md Saef Ullah, Md Mohsin Kabir, Talha Bin Sarwar, Mejdl Safran, Sultan Alfarhood, and M. F. Mridha (2024). "A multi-

modal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM." In: *Scientific Reports* 14.1, p. 9603. DOI: 10.1038/s41598-024-60210-7.

Micallef, Nicholas, Mihai Avram, Filippo Menczer, and Sameer Patil (Apr. 2021). "Fakey: A Game Intervention to Improve News Literacy on Social Media." en. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1, pp. 1–27. ISSN: 2573-0142. DOI: 10.1145/3449080.

Mickoleit, Arthur (Dec. 2014). *Social Media Use by Governments: A Policy Primer to Discuss Trends, Identify Policy Opportunities and Guide Decision Makers*. OECD Working Papers on Public Governance 26. OECD Publishing. DOI: 10.1787/5jxrcmghmk0s-en.

Microsoft (2025). *Microsoft Copilot*. URL: https://copilot.microsoft.com.

Mikolov, Tomás, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Efficient Estimation of Word Representations in Vector Space." In: *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. URL: http://arxiv.org/abs/1301.3781.

Miller, Tim (2019). "Explanation in artificial intelligence: Insights from the social sciences." In: *Artificial intelligence* 267, pp. 1–38. DOI: 10.1016/j.artint.2018.07.007.

Mills, Adam J and Karen Robson (2020). "Brand management in the era of fake news: narrative response as a strategy to insulate brand value." In: *Journal of Product & Brand Management* 29.2, pp. 159–167. DOI: 10.1108/JPBM-12-2018-2150.

Mirzaei, Abas, Dean C. Wilkie, and Helen Siuki (Feb. 2022). "Woke brand activism authenticity or the lack of it." en. In: *Journal of Business Research* 139, pp. 1–12. ISSN: 01482963. DOI: 10.1016/j.jbusres.2021.09.044.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru (Jan. 2019). "Model Cards for Model Reporting." en. In: *Proceedings of the Conference on Fairness, Ac-*

countability, and Transparency. Atlanta GA USA: ACM, pp. 220–229. ISBN: 978-1-4503-6125-5. DOI: 10.1145/3287560.3287596.

Mizrahi, Moran, Guy Kaplan, Dan Malkin, Rotem Dror, Dafna Shahaf, and Gabriel Stanovsky (2024). "State of what art? a call for multi-prompt llm evaluation." In: *Transactions of the Association for Computational Linguistics* 12, pp. 933–949. DOI: 10.1162/tacl_a_00681.

Modgil, Sachin, Rohit Kumar Singh, Shivam Gupta, and Denis Dennehy (2024). "A confirmation bias view on social media induced polarisation during Covid-19." In: *Information Systems Frontiers* 26.2, pp. 417–441. DOI: 10.1007/s10796-021-10222-9.

Moe, Wendy W. and David A. Schweidel (Sept. 2017). "Opportunities for Innovation in Social Media Analytics." en. In: *Journal of Product Innovation Management* 34.5, pp. 697–702. ISSN: 0737-6782, 1540-5885. DOI: 10.1111/jpim.12405.

Mohammad, Saif, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry (2016). "Semeval-2016 task 6: Detecting stance in tweets." In: *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pp. 31–41. DOI: 10.48550/arXiv.1606.03784.

Moss, Jonathan, Emily Robinson, and Jake Watts (Nov. 2020). "Brexit and the Everyday Politics of Emotion: Methodological Lessons from History." en. In: *Political Studies* 68.4, pp. 837–856. ISSN: 0032-3217, 1467-9248. DOI: 10.1177/0032321720911915.

Mumuni, Alhassan and Fuseini Mumuni (Dec. 2022). "Data augmentation: A comprehensive survey of modern approaches." en. In: *Array* 16, p. 100258. ISSN: 25900056. DOI: 10.1016/j.array.2022.100258.

Muric, Goran, Yusong Wu, and Emilio Ferrara (Nov. 2021). "COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies." In: *JMIR Public Health and Surveillance* 7.11, e30642. ISSN: 2369-2960. DOI: 10.2196/30642.

Müller, Hendrik, Aaron Sedley, and Elizabeth Ferrall-Nunge (2014). "Survey research in HCI." In: *Ways of Knowing in HCI*, pp. 229–266.

Nadăş, Mihai, Laura Dioşan, and Andreea Tomescu (2025). "Synthetic Data Generation Using Large Language Models: Advances in Text and Code." In: *IEEE Access* 13, pp. 134615–134633. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2025.3589503.

Nardi, Peter M (2018). *Doing survey research: A guide to quantitative methods*. Routledge. DOI: 10.4324/9781315172231.

Neff, Gina, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn (2017). "Critique and contribute: A practice-based framework for improving critical data studies and data science." In: *Big data* 5.2, pp. 85–97. DOI: 10.1089/big.2016.0050.

Nelimarkka, Matti, Salla-Maaria Laaksonen, and Bryan Semaan (June 2018). "Social Media Is Polarized, Social Media Is Polarized: Towards a New Design Agenda for Mitigating Polarization." en. In: *Proceedings of the 2018 Designing Interactive Systems Conference*. Hong Kong China: ACM, pp. 957–970. ISBN: 978-1-4503-5198-0. DOI: 10.1145/3196709.3196764.

Neubaum, German and Nicole C Krämer (2017). "Monitoring the opinion of the crowd: Psychological mechanisms underlying public opinion perceptions on social media." In: *Media psychology* 20.3, pp. 502–531. DOI: 10.1080/15213269.2016.1211539.

Ngai, Cindy Sing Bik, Rita Gill Singh, and Le Yao (2022). "Impact of COVID-19 vaccine misinformation on social media virality: content analysis of message themes and writing strategies." In: *Journal of medical Internet research* 24.7, e37806. DOI: 10.2196/37806.

Nihlén Fahlquist, Jessica (2021). "The moral responsibility of governments and individuals in the context of the coronavirus pandemic." In: *Scandinavian Journal of Public Health* 49.7, pp. 815–820. DOI: 10.1177/14034948219902.

Niu, Fuqiang, Min Yang, Ang Li, Baoquan Zhang, Xiaojiang Peng, and Bowen Zhang (2024). *A Challenge Dataset and Effective Models for Conversational Stance Detection*. arXiv: 2403.11145 [cs.CL].

O'Shea, Keiron and Ryan Nash (2015). *An Introduction to Convolutional Neural Networks*. Version Number: 2. DOI: 10.48550/ARXIV.1511.08458.

Ofcom (Sept. 2020). *Half of UK adults exposed to false claims about coronavirus*. URL: https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-adults/half-of-uk-adults-exposed-to-false-claims-about-coronavirus (visited on 09/19/2025).

Ofcom (2024). *Adults' Media Use and Attitudes Report 2024*. https://www.ofcom.org.uk/__data/assets/pdf_file/0027/285336/adults-media-use-and-attitudes-report-2024.pdf. Accessed: 2025-07-23.

Oghaz, Toktam A., Ece Çiğdem Mutlu, Jasser Jasser, Niloofar Yousefi, and Ivan Garibay (2020). "Probabilistic Model of Narratives Over Topical Trends in Social Media: A Discrete Time Model." In: *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. New York, NY, USA: Association for Computing Machinery, pp. 281–290. ISBN: 978-1-4503-7098-1. DOI: 10.1145/3372923.3404790.

OpenAI (2025). *ChatGPT*. URL: https://chat.openai.com/ (visited on 09/30/2025).

O'Halloran, Kay L., Sabine Tan, Duc-Son Pham, John Bateman, and Andrew Vande Moere (Jan. 2018). "A Digital Mixed Methods Research Design: Integrating Multimodal Analysis With Data Mining and Information Visualization for Big Data Analytics." en. In: *Journal of Mixed Methods Research* 12.1, pp. 11–30. ISSN: 1558-6898, 1558-6901. DOI: 10.1177/1558689816651015.

Palmer, David D. (2000). "Tokenisation and sentence." In: *Handbook of natural language processing*, p. 11.

Pamungkas, Endang Wahyu, Valerio Basile, and Viviana Patti (2019). *Stance Classification for Rumour Analysis in Twitter: Exploiting Affective Information and Conversation Structure*. arXiv: 1901.01911 [cs.CL].

Papanikolaou, Yannis and Andrea Pierleoni (2020). *DARE: Data Augmented Relation Extraction with GPT-2*. arXiv: 2004.13845 [cs.CL].

Pascual-Ferrá, Paola, Neil Alperstein, Daniel J Barnett, and Rajiv N Rimal (2021). "Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic." In: *Big Data & Society* 8.1, p. 20539517211023533. DOI: 10.1177/20539517211023533.

Patel, Vimla L, David R Kaufman, and Thomas Kannampallil (2021). "Human-computer interaction, usability, and Workflow." In: *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Springer, pp. 153–175. DOI: 10.1007/978-3-030-58721-5_5.

Patro, S. Gopal Krishna and Kishore Kumar Sahu (2015). *Normalization: A Preprocessing Stage*. Version Number: 1. DOI: 10.48550/ARXIV.1503.06462.

Pekkala, Kaisa and Ward van Zoonen (Feb. 2022). "Work-related social media use: The mediating role of social media communication self-efficacy." In: *European Management Journal* 40.1, pp. 67–76. ISSN: 0263-2373. DOI: 10.1016/j.emj.2021.03.004. URL: https://www.sciencedirect.com/science/article/pii/S0263237321000499.

Peng, Sancheng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia (Mar. 2018). "Influence analysis in social networks: A survey." en. In: *Journal of Network and Computer Applications* 106, pp. 17–32. ISSN: 10848045. DOI: 10.1016/j.jnca.2018.01.005.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning (2014). "Glove: Global vectors for word representation." In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

Perer, Adam and Ben Shneiderman (Jan. 2008). "Systematic yet flexible discovery: guiding domain experts through exploratory data analysis." en. In: *Proceedings of the 13th international conference on Intelligent user interfaces*. Gran Canaria Spain: ACM, pp. 109–118. ISBN: 978-1-59593-987-6. DOI: 10.1145/1378773.1378788.

Persily, N., J.A. Tucker, and J.A. Tucker (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform*. SSRC Anxieties of Democracy. Cambridge University Press. ISBN: 9781108835558. DOI: 10.1017/9781108890960. URL: https://books.google.co.uk/books?id=NEjzDwAAQBAJ.

Phan, Thanh Cong, Thanh Tam Nguyen, Matthias Weidlich, Hongzhi Yin, Jun Jo, and Quoc Viet Hung Nguyen (2022). "exrumourlens: Auditable rumour detection with multi-view explanations." In: *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, pp. 3174–3177. DOI: 10.1109/ICDE53745.2022.00291.

Phillips, Lawrence, Chase Dowling, Kyle Shaffer, Nathan Hodas, and Svitlana Volkova (2017). *Using Social Media to Predict the Future: A Systematic Literature Review*. DOI: 10.48550/ARXIV.1706.06134.

Piccolo, Lara, Azizah C. Blackwood, Tracie Farrell, and Martino Mensio (July 2021). "Agents for Fighting Misinformation Spread on Twitter: Design Challenges." In: *CUI 2021 - 3rd Conference on Conversational User Interfaces*. Bilbao (online) Spain: ACM, pp. 1–7. DOI: 10.1145/3469595.3469628.

Poddar, Lahari, Wynne Hsu, Mong Li Lee, and Shruti Subramaniyam (2018). "Predicting Stances in Twitter Conversations for Detecting Veracity of Rumors: A Neural Approach." In: *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 65–72. DOI: 10.1109/ICTAI.2018.00021.

Pollack, Ari H., Tamara D. Simon, Jaime Snyder, and Wanda Pratt (2019). "Creating synthetic patient data to support the design and evaluation of novel health information technology." In: *Journal of biomedical informatics* 95, p. 103201. DOI: 10.1016/j.jbi.2019.103201.

Preston, Stephanie, Anthony Anderson, David J. Robertson, Mark P. Shephard, and Narisong Huhe (2021). "Detecting fake news on Facebook: The role of emotional intelligence." In: *PloS one* 16.3, e0246757. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0246757.

Purchase, Helen C (2012). *Experimental human-computer interaction: a practical guide with visual examples*. Cambridge University Press.

Puri, Neha, Eric A Coomes, Hourmazd Haghbayan, and Keith Gunaratne (2020). "Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases." In: *Human vaccines & immunotherapeutics* 16.11, pp. 2586–2593. DOI: 10.1080/21645515.2020.1780846.

Purificato, Erasmo, Saijal Shahania, and {Ernesto William} {De Luca} (Nov. 2022). "Tell Me Why It's Fake: Developing an Explainable User Interface for a Fake News Detection System." In: *Proceedings of the 3rd Italian Workshop on Explainable Artificial Intelligence co-located with 21th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022)*. CEUR Workshop Proceedings, pp. 51–63. URL: https://aixia2022.uniud.it/.

Qaiser, Shahzad and Ramsha Ali (2018). "Text mining: use of TF-IDF to examine the relevance of words to documents." In: *International journal of computer applications* 181.1, pp. 25–29. DOI: 10.5120/ijca2018917395.

Qian, Xin and Diego Klabjan (2020). *The Impact of the Mini-batch Size on the Variance of Gradients in Stochastic Gradient Descent*. arXiv: 2004.13146 [math.OC].

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever (2018). "Improving language understanding by generative pre-training." In: *Open AI*.

Rathore, Ashish K., Arpan K. Kar, and P. Vigneswara Ilavarasan (Dec. 2017). "Social Media Analytics: Literature Review and Directions for Future Research." en. In: *Decision Analysis* 14.4, pp. 229–249. ISSN: 1545-8490, 1545-8504. DOI: 10.1287/deca.2017.0355.

Rechkemmer, Amy and Ming Yin (2022). "When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models." In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New Orleans, LA, USA: Association for Computing Machinery. ISBN: 9781450391573. DOI: 10.1145/3491102.3501967.

Reisach, Ulrike (2021). "The responsibility of social media in times of societal and political manipulation." In: *European journal of opera-*

tional research 291.3, pp. 906–917. DOI: 10.1016/j.ejor.2020.09.020.

Reuter, Christian, Amanda Lee Hughes, and Marc-André Kaufhold (2018). "Social media in crisis management: An evaluation and analysis of crisis informatics research." In: *International Journal of Human-Computer Interaction* 34.4, pp. 280–294. DOI: 10.1080/10447318.2018.1427832.

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386. DOI: 10.1037/h0042519.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986). "Learning representations by back-propagating errors." In: *nature* 323.6088, pp. 533–536. DOI: 10.1038/323533a0.

Sadah, Shouq A, Moloud Shahbazi, Matthew T Wiley, and Vagelis Hristidis (June 2016). "Demographic-Based Content Analysis of Web-Based Health-Related Social Media." en. In: *Journal of Medical Internet Research* 18.6, e148. ISSN: 1438-8871. DOI: 10.2196/jmir.5327.

Saffer, Dan (2010). *Designing for interaction: creating innovative applications and devices*. 2nd ed. Voices that matter. Berkeley, CA: New Riders. ISBN: 978-0-321-64339-1.

Salamkar, Muneer Ahmed (2024). "Data Visualization: AI-enhanced visualization tools to better interpret complex data patterns." In: *Journal of Bioinformatics and Artificial Intelligence* 4.1, pp. 204–226. DOI: 10.21275/SR240210115511.

Salminen, Joni, Sercan Sengün, Juan Corporan, Soon-gyo Jung, and Bernard J Jansen (2020). "Topic-driven toxicity: Exploring the relationship between online toxicity and news topics." In: *PloS one* 15.2, e0228723. DOI: 10.1371/journal.pone.0228723.

Saroj, Anita and Sukomal Pal (Sept. 2020). "Use of social media in crisis management: A survey." en. In: *International Journal of Disaster Risk Reduction* 48, p. 101584. ISSN: 22124209. DOI: 10.1016/j.ijdrr.2020.101584.

Saud, Muhammad, Musta'in Mashud, and Rachmah Ida (2020). "Usage of social media during the pandemic: Seeking support and awareness about COVID-19 through social media platforms." In: *Journal of public affairs* 20.4, e2417. DOI: 10.1002/pa.2417.

Sauro, J. and J.R. Lewis (2016). *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann. ISBN: 978-0-12-802548-2.

Sawai, Ranto, Incheon Paik, and Ayato Kuwana (2021). "Sentence augmentation for language translation using gpt-2." In: *Electronics* 10.24, p. 3082. DOI: 10.3390/electronics10243082.

Schuler, Douglas and Aki Namioka (1993). *Participatory Design: Principles and Practices*. USA: L. Erlbaum Associates Inc. ISBN: 0805809511. DOI: 10.5555/563076.

Seaman, Carolyn B. (1999). "Qualitative methods in empirical studies of software engineering." In: *IEEE Transactions on software engineering* 25.4, pp. 557–572. DOI: 10.1109/32.799955.

Sehat, Connie Moon, Ryan Li, Peipei Nie, Tarunima Prabhakar, and Amy X. Zhang (Apr. 2024). "Misinformation as a Harm: Structured Approaches for Fact-Checking Prioritization." In: *Proc. ACM Hum.-Comput. Interact.* 8.CSCW1. DOI: 10.1145/3641010.

Seidel, Stefan, Leona Chandra Kruse, Nadine Székely, Michael Gau, and Daniel Stieger (Mar. 2018). "Design principles for sensemaking support systems in environmental sustainability transformations." en. In: *European Journal of Information Systems* 27.2. Ed. by Ken Peffers, Tuure Tuunanen, Björn Niehaves, and Kalle Lyytinen, pp. 221–247. ISSN: 0960-085X, 1476-9344. DOI: 10.1057/s41303-017-0039-0.

Shahid, Farhana, Shahinul Hoque Ony, Takrim Rahman Albi, Sriram Chellappan, Aditya Vashistha, and A. B. M. Alim Al Islam (May 2020). "Learning from Tweets: Opportunities and Challenges to Inform Policy Making During Dengue Epidemic." In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW1. DOI: 10.1145/3392875.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer (2016). "Hoaxy: A platform for tracking on-

line misinformation." In: *Proceedings of the 25th international conference companion on world wide web*, pp. 745–750. DOI: 10.1145/2872518.2890098Focustolearnmore.

Sharevski, Filipo, Peter Jachim, and Kevin Florek (2020). "To tweet or not to tweet: Covertly manipulating a Twitter debate on vaccines using malware-induced misperceptions." In: *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pp. 1–12. DOI: 10.48550/arXiv.2003.12093.

Sharma, Karishma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu (2019). "Combating fake news: A survey on identification and mitigation techniques." In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.3, pp. 1–42. DOI: 10.48550/arXiv.1901.06437.

Sharma, Karishma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu (2020). *COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations*. arXiv: 2003.12309 [cs.SI].

Sharp, Helen, Yvonne Rogers, and Jenny Preece (2007). *Interaction design: beyond human-computer interaction*. Wiley.

Sheth, Amit, Valerie L Shalin, and Ugur Kursuncu (2022). "Defining and detecting toxicity on social media: context and knowledge are key." In: *Neurocomputing* 490, pp. 312–318. DOI: 10.1016/j.neucom.2021.11.095.

Shneiderman, Ben (2020). "Human-centered artificial intelligence: Reliable, safe & trustworthy." In: *International Journal of Human–Computer Interaction* 36.6, pp. 495–504. DOI: 10.1080/10447318.2020.1741118.

Shneiderman, Ben (2022). *Human–centered AI*. Oxford scholarship online. Oxford: Oxford University Press. ISBN: 978-0-19-284529-0 978-0-19-193758-3. DOI: 10.1093/oso/9780192845290.001.0001.

Shu, Kai, H Russell Bernard, and Huan Liu (2019). "Studying fake news via network analysis: detection and mitigation." In: *Emerging research challenges and opportunities in computational social net-*

work analysis and mining, pp. 43–65. DOI: 10.1007/978-3-319-94105-9_3.

Silva, Mirela, Fabrício Ceschin, Prakash Shrestha, Christopher Brant, Juliana Fernandes, Catia S Silva, André Grégio, Daniela Oliveira, and Luiz Giovanini (2020). "Predicting misinformation and engagement in covid-19 twitter discourse in the first months of the outbreak." In: *arXiv preprint*. DOI: 10.48550/arXiv.2012.02164.

Singh, Aditi, Abul Ehtesham, Saket Kumar, and Tala Talaei Khoei (2025). *Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG*. DOI: 10.48550/ARXIV.2501.09136.

Singh, Jasmeet and Vishal Gupta (Sept. 2017). "Text Stemming: Approaches, Applications, and Challenges." en. In: *ACM Computing Surveys* 49.3, pp. 1–46. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/2975608.

Smith, Matthew Longshore (2006). "Overcoming theory-practice inconsistencies: Critical realism and information systems research." In: *Information and organization* 16.3, pp. 191–211. DOI: 10.1016/j.infoandorg.2005.10.003.

Solovev, Kirill and Nicolas Pröllochs (2022). "Moral Emotions Shape the Virality of COVID-19 Misinformation on Social Media." In: *Proceedings of the ACM Web Conference 2022*. WWW '22. Virtual Event, Lyon, France: Association for Computing Machinery, pp. 3706–3717. ISBN: 9781450390965. DOI: 10.1145/3485447.3512266.

Spaa, Anne, Abigail Durrant, Chris Elsden, and John Vines (May 2019). "Understanding the Boundaries between Policymaking and HCI." en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, pp. 1–15. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300314.

Spinuzzi, Clay (2005). "The methodology of participatory design." In: *Technical communication* 52.2, pp. 163–174. URL: https://api.semanticscholar.org/CorpusID:38600380.

Spradling, Matthew, Jeremy Straub, and Jay Strong (2021). "Protection from 'fake news': the need for descriptive factual labeling

for online content." In: *Future Internet* 13.6, p. 142. DOI: 10.3390/fi13060142.

Srivastava, Akash and Charles Sutton (2017). "Autoencoding Variational Inference For Topic Models." In: URL: https://api.semanticscholar.org/CorpusID:29842525.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (Jan. 2014). "Dropout: a simple way to prevent neural networks from overfitting." In: *J. Mach. Learn. Res.* 15.1, pp. 1929–1958. ISSN: 1532-4435. DOI: 10.5555/2627435.2670313.

Stappers, Pieter Jan (Dec. 2007). "Doing Design as a Part of Doing Research." In: *Design Research Now*. Ed. by Ralf Michel. DE GRUYTER, pp. 81–91. ISBN: 978-3-7643-8471-5. DOI: 10.1007/978-3-7643-8472-2_6.

Statista (2025). *Social Media Advertising Worldwide - Statistics & Facts*. https://www.statista.com/statistics/271406/advertising-revenue-of-social-networks-worldwide/. Accessed: 2025-07-23.

Stephanie Soo (n.d.). *Rotten Mango*. English. URL: https://www.rottenmangopodcast.com/ (visited on 09/19/2025).

Stieglitz, Stefan, Milad Mirbabaie, Björn Ross, and Christoph Neuberger (Apr. 2018). "Social media analytics - Challenges in topic discovery, data collection, and data preparation." en. In: *International Journal of Information Management* 39, pp. 156–168. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2017.12.002.

Subrahmanyam, Kaveri, Stephanie M Reich, Natalia Waechter, and Guadalupe Espinoza (2008). "Online and offline social networks: Use of social networking sites by emerging adults." In: *Journal of applied developmental psychology* 29.6, pp. 420–433. DOI: 10.1016/j.appdev.2008.07.003.

Sun, Qingying, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou (Aug. 2018). "Stance Detection with Hierarchical Attention Network." In: *Proceedings of the 27th International Conference on Computational Linguistics*. Ed. by Emily M. Bender, Leon Derczyn-

ski, and Pierre Isabelle. Santa Fe, New Mexico, USA: Association for Computational Linguistics, pp. 2399–2409. URL: https://aclanthology.org/C18-1203/.

Suresh, Harini and John Guttag (2021). "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle." In: EAAMO '21. DOI: 10.1145/3465416.3483305.

Taranukhin, Maksym, Vered Shwartz, and Evangelos Milios (2024). *Stance Reasoner: Zero-Shot Stance Detection on Social Media with Explicit Reasoning*. arXiv: 2403.14895 [cs.CL].

Team, Gemma, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, and Alexandre Ramé (2024). "Gemma 2: Improving open language models at a practical size." In: *arXiv preprint arXiv:2408.00118*.

Tiedemann, Jörg and Santhosh Thottingal (2020). "OPUS-MT–Building open translation services for the World." In: *Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation, pp. 479–480.

Treem, Jeffrey W. and Paul M. Leonardi (2013). "Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association." In: *Annals of the International Communication Association* 36.1, pp. 143–189. DOI: 10.1080/23808985.2013.11679130.

Underwood, Lauren (Feb. 2024). *Digital 2024 United Kingdom report shows strong growth in social media*. Available at: https://wearesocial.com/uk/blog/2024/02/digital-2024-united-kingdom-report-shows-strong-growth-in-social-media/.(accessed May 16, 2025).

Van Audenhove, Leo and Karen Donders (2019). "Talking to people III: Expert interviews and elite interviews." In: *The Palgrave handbook of methods for media policy research*. Springer, pp. 179–197.

Vasist, Pramukh Nanjundaswamy, Debashis Chatterjee, and Satish Krishnan (Apr. 2024). "The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configural Narrative."

en. In: *Information Systems Frontiers* 26.2, pp. 663–688. ISSN: 1387-3326, 1572-9419. DOI: 10.1007/s10796-023-10390-w.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2023). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].

Vedula, Nikhita and Srinivasan Parthasarathy (Mar. 2021). "FACE-KEG: Fact Checking Explained using KnowledgE Graphs." In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. Virtual Event Israel: ACM, pp. 526–534. DOI: 10.1145/3437963.3441828.

Vermeeren, Arnold P. O. S., Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila (2010). "User experience evaluation methods: current state and development needs." In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. NordiCHI '10. Reykjavik, Iceland: Association for Computing Machinery, pp. 521–530. ISBN: 9781605589343. DOI: 10.1145/1868914.1868973.

Vicario, Michela Del, Walter Quattrociocchi, Antonio Scala, and Fabiana Zollo (2019). "Polarization and fake news: Early warning of potential misinformation targets." In: *ACM Transactions on the Web (TWEB)* 13.2, pp. 1–22. DOI: 10.1145/3316809.

Visentin, Marco, Gabriele Pizzi, and Marco Pichierri (2019). "Fake news, real problems for brands: The impact of content truthfulness and source credibility on consumers' behavioral intentions toward the advertised brands." In: *Journal of Interactive Marketing* 45.1, pp. 99–112. DOI: 10.1016/j.intmar.2018.09.001.

Wagner, Stefan Sylvius, Maike Behrendt, Marc Ziegele, and Stefan Harmeling (2024). *SQBC: Active Learning using LLM-Generated Synthetic Data for Stance Detection in Online Political Discussions*. arXiv: 2404.08078 [cs.CL].

Wagner, Stefan Sylvius, Maike Behrendt, Marc Ziegele, and Stefan Harmeling (2025). *The Power of LLM-Generated Synthetic Data for*

Stance Detection in Online Political Discussions. arXiv: 2406.12480 [cs.CL].

Wahlström, Mattias and Anton Törnberg (May 2021). "Social Media Mechanisms for Right-Wing Political Violence in the 21st Century: Discursive Opportunities, Group Dynamics, and Co-Ordination." en. In: *Terrorism and Political Violence* 33.4, pp. 766–787. ISSN: 0954-6553, 1556-1836. DOI: 10.1080/09546553.2019.1586676.

Wang, Jiawen, Longfei Zuo, Siyao Peng, and Barbara Plank (Nov. 2024). "MultiClimate: Multimodal stance detection on climate change videos." In: *Proceedings of the Third Workshop on NLP for Positive Impact*. Miami, Florida, USA: Association for Computational Linguistics, pp. 315–326. DOI: 10.18653/v1/2024.nlp4pi-1.27.

Wang, Yuxi, Martin McKee, Aleksandra Torbica, and David Stuckler (2019). "Systematic literature review on the spread of health-related misinformation on social media." In: *Social science & medicine* 240, p. 112552. DOI: 10.1016/j.socscimed.2019.112552.

Waseem, Zeerak, Thomas Davidson, Dana Warmsley, and Ingmar Weber (2017). "Understanding Abuse: A Typology of Abusive Language Detection Subtasks." en. In: *Proceedings of the First Workshop on Abusive Language Online*. Vancouver, BC, Canada: Association for Computational Linguistics, pp. 78–84. DOI: 10.18653/v1/W17-3012.

Weeks, Brian E. (2015). "Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation." In: *Journal of Communication* 65.4, pp. 699–719. DOI: 10.1111/jcom.12164.

Wei, Jason and Kai Zou (Nov. 2019). "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks." In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong,

China: Association for Computational Linguistics, pp. 6382–6388. DOI: 10.18653/v1/D19-1670.

Wei, Penghui, Nan Xu, and Wenji Mao (Sept. 2019). *Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity*. DOI: 10.48550/arXiv.1909.08211.

Weinzierl, Maxwell, Suellen Hopfer, and Sanda M Harabagiu (2021). "Misinformation adoption or rejection in the era of covid-19." In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 15, pp. 787–795. DOI: 10.1609/icwsm.v15i1.18103.

Weiss, Joseph W (2021). *Business ethics: A stakeholder and issues management approach*. Berrett-Koehler Publishers.

Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang (Dec. 2016). "A survey of transfer learning." en. In: *Journal of Big Data* 3.1, p. 9. ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6.

Wolf, Thomas et al. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. arXiv: 1910.03771 [cs.CL].

Woolson, Robert F. (2007). "Wilcoxon signed-rank test." In: *Wiley encyclopedia of clinical trials*, pp. 1–3.

Xu, Zhan and Hao Guo (2018). "Using text mining to compare online pro-and anti-vaccine headlines: Word usage, sentiments, and online popularity." In: *Communication studies* 69.1, pp. 103–122. DOI: 10.1080/10510974.2017.1414068.

Yan, Ming, Tianyi Zhou Joey, and W. Tsang Ivor (2024). "Collaborative Knowledge Infusion for Low-Resource Stance Detection." In: *Big Data Mining and Analytics* 7.3, pp. 682–698. DOI: 10.26599/BDMA.2024.9020021.

Yang, Fan, Shiva K. Pentyala, Sina Mohseni, Mengnan Du, Hao Yuan, Rhema Linder, Eric D. Ragan, Shuiwang Ji, and Xia (Ben) Hu (May 2019a). "XFake: Explainable Fake News Detector with Visualizations." In: *The World Wide Web Conference*. San Francisco CA USA: ACM, pp. 3600–3604. DOI: 10.1145/3308558.3314119.

Yang, Qian, Justin Cranshaw, Saleema Amershi, Shamsi T. Iqbal, and Jaime Teevan (May 2019b). "Sketching NLP: A Case Study of Exploring the Right Things To Design with Language Intelligence."

en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, pp. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300415.

Yang, Qian, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman (Apr. 2020). "Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design." en. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, pp. 1–13. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376301.

Yeung, Douglas (2018). "Social media as a catalyst for policy action and social change for health and well-being." In: *Journal of medical Internet research* 20.3, e94. DOI: 10.2196/jmir.8508.

Yildirim, Nur et al. (May 2024). "Sketching AI Concepts with Capabilities and Examples: AI Innovation in the Intensive Care Unit." en. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, pp. 1–18. ISBN: 979-8-4007-0330-0. DOI: 10.1145/3613904.3641896.

Yin, Wenpeng (2020). *Meta-learning for Few-shot Natural Language Processing: A Survey*. Version Number: 1. DOI: 10.48550/ARXIV.2007.09604.

Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria (2018). "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]." In: *IEEE Computational Intelligence Magazine* 13.3, pp. 55–75. DOI: 10.1109/MCI.2018.2840738.

Yu, Erxin, Jing Li, and Chunpu Xu (2024). *PopALM: Popularity-Aligned Language Models for Social Media Trendy Response Prediction*. arXiv: 2402.18950 [cs.CL].

Yu, Jianfei, Jing Jiang, Ling Min Serena Khoo, Hai Leong Chieu, and Rui Xia (Nov. 2020). "Coupled Hierarchical Transformer for Stance-Aware Rumor Verification in Social Media Conversations." In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, pp. 1392–1401. DOI: 10.18653/v1/

2020.emnlp-main.108. URL: https://aclanthology.org/2020.emnlp-main.108.

Zampieri, Marcos, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar (Jan. 2019). "Predicting the Type and Target of Offensive Posts in Social Media." In: *Proceedings of the 2019 Conference of the North*, pp. 1415–1420. DOI: 10.18653/v1/N19-1144.

Zhang, Amy, Bryan Culbertson, and Praveen Paritosh (May 2017). "Characterizing Online Discussion Using Coarse Discourse Sequences." In: *Proceedings of the International AAAI Conference on Web and Social Media* 11.1, pp. 357–366. ISSN: 2334-0770, 2162-3449. DOI: 10.1609/icwsm.v11i1.14886.

Zhang, Bowen, Daijun Ding, Liwen Jing, Genan Dai, and Nan Yin (2024). *How would Stance Detection Techniques Evolve after the Launch of ChatGPT?* arXiv: 2212.14548 [cs.CL].

Zhang, Lei, Shuai Wang, and Bing Liu (2018). "Deep learning for sentiment analysis: A survey." In: *Wiley interdisciplinary reviews: data mining and knowledge discovery* 8.4, e1253. DOI: 10.1002/widm.1253.

Zhang, Yin, Rong Jin, and Zhi-Hua Zhou (Dec. 2010). "Understanding bag-of-words model: a statistical framework." en. In: *International Journal of Machine Learning and Cybernetics* 1.1-4, pp. 43–52. ISSN: 1868-8071, 1868-808X. DOI: 10.1007/s13042-010-0001-0.

Zhang, Yizhe, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan (May 2020). *DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation*. DOI: 10.48550/arXiv.1911.00536.

Zhang, Zhen, Jiajun Bu, Martin Ester, Jianfeng Zhang, Chengwei Yao, Zhi Yu, and Can Wang (Dec. 2019). *Hierarchical Graph Pooling with Structure Learning*. DOI: 10.48550/arXiv.1911.05954.

Zhao, Sihong, Simeng Hu, Xiaoyu Zhou, Suhang Song, Qian Wang, Hongqiu Zheng, Ying Zhang, Zhiyuan Hou, et al. (2023). "The prevalence, features, influencing factors, and solutions for

COVID-19 vaccine misinformation: systematic review." In: *JMIR public health and surveillance* 9.1, e40201. DOI: 10.2196/40201.

Zia, Haris Bin, Ignacio Castro, Arkaitz Zubiaga, and Gareth Tyson (May 2022). "Improving Zero-Shot Cross-Lingual Hate Speech Detection with Pseudo-Label Fine-Tuning of Transformer Language Models." In: *Proceedings of the International AAAI Conference on Web and Social Media* 16, pp. 1435–1439. ISSN: 2334-0770, 2162-3449. DOI: 10.1609/icwsm.v16i1.19402.

Zimmerman, Eric and Katie Salen (2003). *Rules of play: game design fundamentals*. Cambridge, Mass: MIT Press. ISBN: 978-0-262-24045-1.

Zimmerman, John, Jodi Forlizzi, and Shelley Evenson (Apr. 2007). "Research through design as a method for interaction design research in HCI." en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, pp. 493–502. ISBN: 978-1-59593-593-9. DOI: 10.1145/1240624.1240704.

Zimmerman, John, Erik Stolterman, and Jodi Forlizzi (2010). "An analysis and critique of Research through Design: towards a formalization of a research approach." In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. DIS '10. Aarhus, Denmark: Association for Computing Machinery, pp. 310–319. ISBN: 9781450301039. DOI: 10.1145/1858171.1858228.

Zubiaga, Arkaitz, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein (2018). "Discourse-aware rumour stance classification in social media using sequential classifiers." In: *Information Processing Management* 54.2, pp. 273–290. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2017.11.009.

Zubiaga, Arkaitz, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie (2016). "Analysing how people orient to and spread rumours in social media by looking at conversational threads." In: *PloS one* 11.3, e0150989. DOI: 10.1371/journal.pone.0150989.

Östlund, Ulrika, Lisa Kidd, Yvonne Wengström, and Neneh Rowa-Dewar (2011). "Combining qualitative and quantitative research within mixed method research designs: a methodological review." In: *International journal of nursing studies* 48.3, pp. 369–383. DOI: 10.1016/j.ijnurstu.2010.10.005.