

2024

Navigating Democracy's Challenges: A Review of Research Projects on False Information and Hate Speech

Isabel Sophia Bezzaoui

FZI Research Center for Information Technology, bezzaoui@fzi.de

Kai Schewina

University of Potsdam, Germany, kai.schewina@uni-potsdam.de

Georg Voronin

University of Potsdam, Germany, georg.voronin@uni-potsdam.de

Follow this and additional works at: <https://aisel.aisnet.org/wi2024>

Recommended Citation

Bezzaoui, Isabel Sophia; Schewina, Kai; and Voronin, Georg, "Navigating Democracy's Challenges: A Review of Research Projects on False Information and Hate Speech" (2024). *Wirtschaftsinformatik 2024 Proceedings*. 122.

<https://aisel.aisnet.org/wi2024/122>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2024 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Navigating Democracy's Challenges: A Review of Research Projects on False Information and Hate Speech

Workshop Paper

Isabel Bezzaoui¹, Georg Voronin², and Kai Schewina²

¹ FZI Forschungszentrum Informatik, Information Process Engineering, Berlin, Germany
bezzaoui@fzi.de

² Universität Potsdam, Wirtschaftsinformatik und Digitale Transformation, Potsdam, Germany
{georg.voronin,kai.schewina}@uni-potsdam.de

Abstract. In recent years, society has faced a polycrisis, characterized by intertwined challenges such as the climate crisis, global conflicts, and the COVID-19 pandemic. During these crises, digital platforms have been exploited to disseminate disinformation and hate, which can harm society. These issues contribute to societal polarization and erode trust within communities. To build resilient democracies, it is imperative to conduct research on false information and hate speech to identify mechanisms and evaluate countermeasures. This study aims to provide an overview of the current publicly funded research in these areas by examining projects supported by the German Federal Ministry of Education and Research (BMBF), the German Research Foundation (DFG), and the European Union (EU). The findings reveal several gaps in current research that need to be addressed by federal and international organizations to ensure the resilience of democratic societies.

Keywords: False Information, Hate Speech, Project Review, Digital Democracy.

1 Introduction

In recent years, society has experienced what can be considered a polycrisis (Henig & Knight 2023) - while the climate crisis leads to natural hazards, there are multiple global wars such as the Russian invasion of Ukraine in 2022 and the Israel-Hamas war since 2023. Meanwhile, social media platforms are utilized to spread disinformation and hate (e.g., Shahi et al. 2024). This might cause harm to society, e.g., due to false health advice such as in the COVID-19 pandemic (Naeem, Bhatti and Khan 2021), placing our democracies under significant strain. Further, hate speech poses risks for individuals psychologically (Bilewicz and Soral 2020). Both issues relate to polarizing societies (Vasist et al. 2024) and therefore constitute a threat to the trust in society (Weinhardt et al. 2024).

To address the challenges facing the public sphere in the digital age, it is essential for researchers to critically engage with the design, governance and regulation of digital platforms. This includes analyzing algorithmic biases, handling information manipulation, fostering trust in digital artifacts, and proposing design principles that align with democratic values. Today, however, large platform providers such as X (formerly Twitter)

increasingly restrict the possibility of researching platform mechanisms and collecting data, thus making it more difficult for researchers to access the domain (Ledford 2023). Suggesting the establishment of six research areas for Digital Democracy research, Weinhardt et al. (2024) call for information systems researchers to engage in research exploring how platforms influence human behavior and social cohesion in order to recognize their broader impact beyond business models and interfaces. As networks originally meant to inform and connect individuals are now increasingly being used to spread hate and disinformation (Aïmeur et al. 2023), interdisciplinary research across information systems, computer science, political science, sociology, communication science, law, and others has become crucial (Sample et al. 2020). Information systems researchers are called upon to prioritize transparency, inclusion, and literacy, focusing on innovative ways to preserve and promote democracy. To build resilient democracies, research is essential in the areas of disinformation and hate speech to identify mechanisms and evaluate countermeasures (Bennett and Livingston 2018). One research area introduced by Weinhardt et al. (2024) focuses on the foundation of democratic engagement: trust. It examines how various forms of misinformation, disinformation, malinformation, and hate speech influence the political landscape and trust. It is critical to assess and map out the current efforts within the discipline of information systems research regarding the impact of misinformation, disinformation, and hate speech on democracy. For this reason, we formulate the following research question:

RQ: How do current publicly funded research projects in Germany and the EU address the impact of false information and hate speech on (digital) democracies, and what gaps exist that information systems researchers can fill to enhance the resilience of democratic societies in the digital age?

By understanding what research is currently being undertaken, we can identify gaps and areas that require further exploration. This evaluation can help create future projects, ensuring they address the most pressing issues and contribute effectively to preserving and promoting democratic values in the digital age. Therefore, we aim to provide an overview of the current state of publicly funded research on these topics. To do so, we consider all projects that are currently funded by the German Federal Ministry of Education and Research (BMBF), the German Research Foundation (DFG), and projects sponsored by the European Union (EU). These three are among the most important sources for third-party funding in Germany Hornbostel (2001). We identify several gaps in current research that need to be addressed by federal and international organizations to ensure the resilience of our democratic society.

The remainder of the paper is structured as follows: Section 2 provides the theoretical background, exploring the relevancy of false information and hate speech in the context of digital democracy. Section 3 details the methodology for systematically reviewing ongoing research projects. Section 4 presents the results, starting with a descriptive analysis followed by a qualitative content analysis to synthesize the key findings. Section 5 discusses the role of Information Systems (IS) research in addressing these issues, highlighting the interdisciplinary potential of IS to contribute to the understanding and mitigation of false information and hate speech. Finally, Section 6 concludes the paper with a summary of the findings and suggestions for future research directions.

2 Theoretical Background

In today's digital age, the rapid proliferation of information has transformed the way individuals communicate and access news. This chapter delves into the critical theoretical notions necessary to understand the phenomena of false information and hate speech, two pervasive issues that significantly impact societal discourse and public opinion.

2.1 False Information.

The contemporary capability for virtually anyone to publish and share content online not only enhances opportunities for social participation but also generates new avenues for the dissemination of false information (Appel 2020; Shu et al. 2017). Presently, the research on detecting manipulated information is a rapidly expanding domain that spans multiple disciplines, including computer science, information systems, media studies, and social sciences (Kapantai et al. 2021; Mahyoob et al. 2020; Verma et al. 2021; Yu and Lo 2020). It is critical to distinguish between the terms false information, misinformation, disinformation, and malinformation. False information pertains to "verifiably false information", with disinformation and misinformation being subcategories dependent on the intent. While misinformation refers to "false information that is shared without the intention to mislead or to cause harm" (Aimeur et al. 2023), disinformation is defined as "false information that is shared to intentionally mislead". Further, malinformation is defined as "genuine information that is shared with an intent to cause harm" (Aimeur et al. 2023), therefore differentiating itself from the other terms by the genuine property of its authenticity. These concepts are crucial as they relate to the potential erosion of trust in society (Weinhardt et al. 2024), which can be severely undermined by negative experiences, such as deception through disinformation (Schwerter and Zimmermann 2020).

The spread of misleading or deceptive information can be supported by the use of technology. Social bots offer the opportunity to spread news with high frequency. However, it is often humans who voluntarily spread false information, especially via social media such as X (formerly known as Twitter) or Facebook (Wardle and Derakshan 2017). In this context, the question also arises as to who is particularly vulnerable to deceptive information. Some studies suggest that, rather than partisan bias, too little analytical thinking is a significant risk factor. The higher the ability to think critically, the less individuals appear to believe in false news (Bronstein et al. 2019; Faragó et al. 2023; Pennycook and Rand 2019) Therefore, it is essential to develop a comprehensive understanding of the mechanisms and dissemination of the phenomena related to false information, while simultaneously devising systematic methods to counteract them (Bezzaoui et al. 2022).

2.2 Hate Speech.

Kansok-Dusche et al. (2023) define hate speech as derogatory expressions based on assigned group characteristics, intended to harm, and capable of causing harm on multiple levels (individual, communal, societal). This includes negative stereotyping,

dehumanization, and expressions of violence (Paasch-Colberg et al. 2021). Bäumler et al. (2024) add that unlike cyberbullying, hate speech can be subtle or humorous, targeting individuals and social groups vicariously. Online hate speech significantly impacts democracy by polarizing society and undermining democratic discourse (Weinhardt et al., 2024).

The public sphere, as described by Habermas (1962), is a space for rational discourse and public opinion formation. Social media platforms have the potential to be such spheres. However, hate speech on platforms often excludes marginalized groups from the dominant public sphere, leading them to form counterpublics—alternative spaces for expressing experiences and advocating for change (Fraser 1990). While online hate speech normalizes discriminatory behavior and increases societal polarization (Cialdini et al. 1990; Soral et al. 2020), counterpublics provide platforms for marginalized groups to organize, support each other, and engage in activism, fostering collective resilience and challenging discriminatory norms (Eckert et al. 2021). A democratic discourse that includes marginalized individuals is crucial, as the discourse in the public sphere underpins common social values of coexistence and democratic legal norms. Excluding social groups means that these values and norms may no longer be supported by all parts of society, potentially leading to discrimination against minorities. Addressing online hate speech and including minorities from counterpublics is essential for maintaining democratic discourse and societal cohesion. Research on the mechanisms of hate speech dissemination and the effectiveness of counter-narratives is vital to ensure the resilience of democratic societies.

3 Methodology

Although there is ample methodological guidance for conducting structured literature reviews, limited instruction is available on how to review practical artifacts such as research projects. For this reason, we make use of Gnewuch and Mädche's (2022) approach to reviewing software artifacts and adapt their seven-step method to our context of a structured project review. We adapt their seven steps as follows:

Problem formulation. The review's main objectives are determined, focusing on the project's characteristics, properties, or features central to the review. Additionally, it is crucial to establish the scope of the review. The scope is defined by the inclusion of three project sponsors and a focus on currently ongoing projects. The focus of this study is on research projects in the EU and Germany as an example of investigating research projects on a federal level. The EU is one of the largest political and economic entities globally, comprising 27 member states with a combined population of over 440 people. Its policies and regulations often set standards that influence global norms, particularly in digital governance, data protection, and media regulation. Germany is not only the largest economy in the EU but also a key player in shaping EU policies. Its actions and approaches often have a significant impact on the direction and effectiveness of EU-wide initiatives (European Union 2024).

Software Artifact Search. Relevant projects are searched for via the internet and relevant databases, and decisions are made about their suitability for the review. The pre-defined keywords for projects on false information were "disinformation", "Desinfor-

mation", "fake news", "Falschinformation", "false information", and "misinformation". For Projects regarding hate speech we searched for "hate speech" and "Hassrede", respectively. We extracted data from the BMBF, DFG, and EU websites. For DFG, we conducted a search in the database GEPRIS for the pre-defined search terms and filtered for ongoing projects. In the second step, the details of the consortium and further information on the identified projects were conducted through an additional web search. For the EU, we searched the database of the Community Research and Development Information Service (CORDIS) for the defined search string and filtered for ongoing projects. Subsequently, the project consortium and individual members were identified in order to further categorize the projects based on their relation to the field of information systems. For BMBF, as there is no central database that lists and categorizes projects, we use a search engine as well as the website search functionality to identify disinformation and hate speech-related projects. Further, once identified, we consider the respective line of funding.

Screening for Inclusion. Projects are screened based on predetermined criteria to determine their relevance, resulting in a list of 79 eligible projects. All projects were screened in terms of the project title, project focus, project description, involved countries, sponsors, consortium, duration, involved disciplines, and target groups. We only included projects that are currently running and whose main object of research is either false information or hate speech.

Quality Assessment. The quality of the selected projects may be assessed based on practical relevance or target group feedback. As this step explicitly does not include scientific quality (Gnewuch and Madche 2022) and the analysis' scope is of an empirical rather than normative nature, we exclude this step from our review.

Data Extraction. Applicable information is extracted from each project by examining the information provided by the relevant databases and search results based on our predetermined criteria.

Documentation and Archiving. The project information and any related material used as an additional source of information in the review are documented, stored, and archived in an Excel sheet.

Data Analysis and Synthesis. The evidence extracted from the included projects is collated, summarized, aggregated, organized, and compared, with the findings presented in a consequential manner. We aggregate related target groups to higher orders of abstraction (e.g., "scientists" and "researchers", or "users", "citizens" and "general society" to "users"), as well as for disciplines ("natural language processing", "computer and information science", and "computational linguistics" to "computer science and adjacent"). Further, we classify the non-research consortial partners according to NGOs and other non-profit organizations, for-profit organizations, and public bodies, drawing from the classification by the EU CORDIS database. Through an additional qualitative content analysis after Mayring (2015), the projects' main focal points, as addressed in their descriptions, are analyzed and compared. Proceeding inductively during the empirical analysis, relevant categories are derived directly from the project descriptions. This approach follows a conventional content analysis in which codes are defined during data analysis. The main focus lies on a synthetic creation of categories displaying complex content-related evidence instead of only functioning as markers for certain

Table 1. Category system for content analysis following Mayring (2015). Categories are sorted by frequency

No	Category	Definition	Anchor Example
1	Tool Development	The project involves the development of a practical tool.	"The aim of the joint project [...] is to develop a software-based analysis tool that helps experts to better assess disinformation campaigns."
2	Digital Platforms	The project focuses on digital platforms and the online realm as a subject of research.	"Second, the project will address criminal liability in digital networks, not only for users but also for hosts and service providers."
3	Machine Learning	The project focuses on employing methods of machine learning.	"[The project] will integrate the structured knowledge provided by social and human sciences into natural language processing tools and deep learning algorithms, so as to develop new hybrid intelligence systems."
4	Policy Advice and Frameworks	The project focuses on formulating policy advice or theoretical (legal) frameworks for implementation.	"Finally, the project results will be made available in concrete recommendations for action for citizens, the media and politicians."
5	Social Media	The project involves research on social media or uses social media data.	"This project deals with punishable behaviors in social media, with a focus on expression offenses."
6	Qualitative and Mixed Methods Research	The project conducts qualitative research or mixed methods approaches.	"It will adopt a mixed-methods approach using archival and social media analysis, interviews, cross-sectional and longitudinal surveys, and experiments [...]."
7	Open Access	The project actively makes (part of) its results accessible to the general public.	"The dataset, the model, the training workflow, and the software for operating the service will be made openly available whenever possible and thus can be utilized for other subject areas as well."
8	Fact-Checking	The project focuses on applying or developing methods of fact-checking.	"For this purpose, they are trained in scientific and ethical working methods for well-founded fact-checking."
9	Science and Healthcare	The project focuses on topics such as health, medicine, and science.	"Thus, the project combines methods of transfer learning, information extraction, and fact-checking for the biomedical domain."

passages. By going through the material, former categories are either subsumed or a new category is formulated. After working through 50 percent of the data, all categories are revised and eventually reduced to main categories. Following Mayring's method of summary content analysis, the original material is summarized. The aim is to demarcate text elements without distorting the textual core of the data. Through this kind of reduction, more transparency shall be created that still corresponds to the material's basic form (Mayring 2015). Table 1 displays the final category system applied for qualitative data analysis with distinct definitions of each code and respective anchor examples. The data for our analysis is available via OSF [\[clickable\]](#).

4 Results

This chapter provides an examination of the primary findings from our study, focusing on the analysis of 79 identified projects that address false information and hate speech. The investigation is divided into two sections, Descriptive Analysis and Qualitative Content Analysis, each utilizing a different analytical approach to uncover key insights.

4.1 Descriptive Analysis

The following section presents a summary of the primary findings derived from a descriptive analysis of the characteristics of the 79 identified projects addressing false information and hate speech.

False Information. Overall, we identified 60 ongoing research projects regarding disinformation and related constructs. Of those, eight projects involve information systems researchers (i.e., include professors or doctoral employees that have a degree or PhD in information systems and/or work at an information systems institute), and further 23 projects involving researchers from adjacent disciplines such as computer science, data science, information science or computational linguistics. Of the eight projects involving information systems, six are funded by the BMBF, one by the EU, and one by the DFG. Correspondingly, most of the involved institutions stem from Germany, the EU project covers 15 countries. The projects run for three (BMBF, DFG) to five years (EU). The target groups of the involved projects are diverse, including authorities and organizations with security tasks (3), healthcare workers and the healthcare system (2), users (4), researchers and innovators (1), and platforms (1). Involved disciplines include information systems (8), computer science and adjacent (4), communication science (2), information science (1), sociology (1), economics (1), law (1) and ethics (1). The projects involve overall six non-profit organizations and eight for-profit organizations, most of which are software development or consulting companies, about half of which are part of one EU project, and the remaining from different BMBF projects. Of those eight projects involving information systems researchers, seven (87.5%) are interdisciplinary projects, involving multiple of the disciplines outlined above.

Figure 1 shows the distribution of false information projects across the EU, the DFG and the BMBF according to the involvement of disciplines related to information systems (computer science, data science, information science or computational linguistics). We find 23 ongoing research projects from related fields. Of those, four are funded by the BMBF, 17 by the EU, and two by the DFG. The European projects cover more than 30 countries. The projects span two to five years and target researchers and innovators (10), citizens and the general public (9), human resources (2), health care workers (1), data analysts (1), journalists (1), news institutions (1), and authorities and organizations with security tasks (1). The projects involve 12 NGOs and 16 public, non-research organizations, many of which are public news institutions, organizations or public bodies such as ministries of interior or police, and NGOs for gender and sexual diversity organizations. Further, there are 54 for-profit organizations involved, many of which are

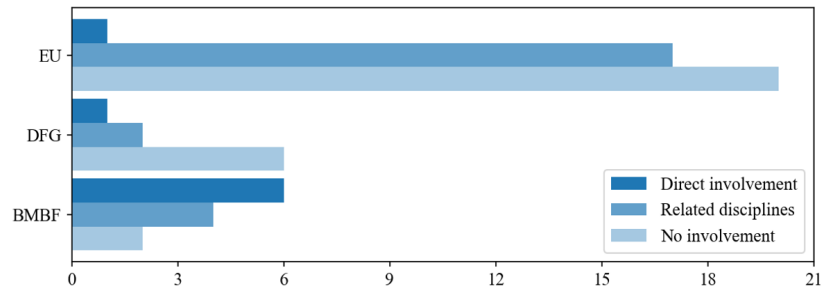


Figure 1. Sponsors of Projects in the False Information Dataset by Involvement of Information Systems.

private news institutions. Most non-research partners are involved in European projects. Of those 23 research projects, 17 (73.9%) are interdisciplinary.

Hate Speech. Through our analysis, we determine 19 ongoing projects connected to hate speech. Of those, only one includes information systems researchers and seven adjacent disciplines. The IS-related project is funded by the BMBF, takes action for three years until July 2026, and specifically targets investigative and law enforcement authorities. They interdisciplinarily combine information systems with computer science researchers and involve one for-profit organization for software development. Further seven projects include researchers from adjacent disciplines. One is funded by the DFG and seven by the EU. They span from 1.5 (EU) to 5 years (EU) and involve researchers from 14 European countries. They target users (3), authorities (2), research (2), and community managers (1). Researchers stem from a variety of disciplines such as computer science and similar fields (7), communication science (1), politics (2), linguistics (2), human-computer interaction (1), and humanities (1). Overall, there are two NGOs, nine for-profit organizations, and five public bodies involved. Out of those seven projects, five (71.4%) are interdisciplinary projects involving the disciplines listed above.

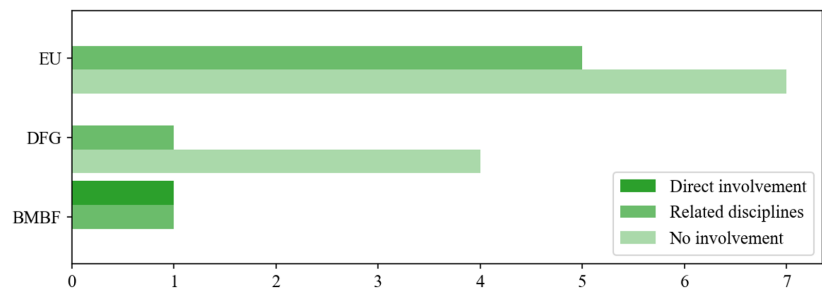


Figure 2. Sponsors of Projects in the Hate Speech Dataset by involvement of Information Systems.

Figure 2 depicts the distribution of hate speech projects among the EU, the DFG, and the BMBF according to the involvement of disciplines related to information systems. Of

the eight ongoing projects from the field of information systems and adjacent fields, five are funded by the EU, two by the BMBF, and one by the DFG. Two of the EU-funded projects are registered only in Germany, one only in Italy, and the other two span 12 other European countries, targeting scientists (3), investigative and law enforcement authorities (3), social media users (2), online community and comment section managers (1), the general public (1), police authorities (1), and minority language users (1).

4.2 Qualitative Content Analysis

The following section presents a summary of the primary findings derived from a content analysis (Mayring 2015) of the descriptions of the 79 identified projects addressing false information and hate speech.

False Information. Out of the 60 identified research projects, 21 projects focus on formulating policy advice and/or theoretical (legal) frameworks for implementation. Specifically, nine projects develop policy recommendations for national and international legislators and create new legal frameworks. The other 12 projects propose theoretical models or frameworks addressing notions of disinformation, related phenomena, and educational concepts. Additionally, 21 projects concentrate on developing practical tools. These include mobile applications for detecting manipulated content, analysis tools for experts, dashboards for discourse tracking, and collaborative platforms. Digital platforms are a common focus, with 19 projects targeting them and specifically investigating social media (14). Here, the primary aim is to analyze the spread of disinformation, moderate digital networks, and detect manipulative content on online platforms and messenger services. Machine learning methods are employed in 18 projects to develop tools or analyze data, frequently using natural language processing for text categorization and information extraction systems. These approaches often include solutions for human-machine interaction. Public accessibility is a key concern for eight projects, which make their tools available via APIs and consider users with diverse backgrounds. Fact-checking is a focus for seven projects, combining automated and human-based methods. Another seven projects specifically target disinformation in science and healthcare, particularly concerning COVID-19, vaccinations, and pseudoscientific conspiracy theories. Lastly, six projects utilize qualitative methods or mixed-methods approaches, predominantly through expert interviews as well as content and discourse analyses. These qualitative methods are often combined with quantitative, computational approaches for comprehensive insights. Figure 3 depicts the frequency of codes applied in the dataset of projects on false information, offering a glimpse into the most prominent focal points within this area of research.

In examining the role of the Information Systems (IS) discipline within this research area, we observed that out of nine projects on false information, the majority focus on developing tools (8) and applying machine learning methods (8), rather than creating theoretical frameworks or policy advice (1). These projects often investigate digital realms (5) and social media (4), with some effort to make results open access (4). Fact-checking methods (0) and qualitative or mixed methods approaches (1) are rarely



Figure 3. Distribution of Codes in the False Information Dataset.

included. While two projects focus on science and health, most (7) adopt a holistic, domain-independent perspective on false information.

Hate Speech. Among the 19 research endeavors focusing on the topic of hate speech, eight projects employ machine learning methods, primarily using natural language processing and deep learning for detecting hate speech and analyzing digital hate. Seven projects focus on digital platforms, with three of them specifically targeting social media. These studies primarily analyze the occurrence and spread of digital hate and political hostility, as well as their implications for criminal liability, frequently mentioning Facebook, Telegram, and X (formerly Twitter). Six projects involve developing tools such as AI-based tools for managing online communities, and dashboards as well as browser extensions for analyzing cyber abuse content. Five projects apply qualitative or mixed-methods approaches, using interviews and discourse analysis, often combined with computational analysis. Two projects aim to make their results accessible to the general public, offering them free of charge and focusing on “low-resource” countries. Finally, one project focuses on creating policy advice, proposing a model of accountability mechanisms guided by a civic code of conduct. Figure 4 displays the frequency of codes applied in the dataset of projects on hate speech, providing insights into the most prevalent focal points within this area of research.



Figure 4. Distribution of Codes in the Hate Speech Dataset.

Among hate speech research projects, the only one involving IS researchers focuses on digital platforms and social media, developing a tool for detecting and addressing

cyberbullying and hate speech. Unlike other projects that use machine learning and qualitative or mixed methods, this project lacks specific methodological details, though it mentions a participatory development process.

5 Discussion

Comparing IS projects to the broader landscape of initiatives addressing false information and hate speech in our dataset reveals distinct trends and gaps within the discipline. IS research prominently addresses these issues by developing digital tools and focusing on digital environments. This technological focus has led to the creation of various digital artifacts, such as applications and dashboards, designed to detect and mitigate the spread of false information and hate speech. However, this emphasis on practical, digital solutions has the potential to overshadow the development of theoretical outcomes, such as policy advice or educational frameworks, which are crucial for a holistic approach to these problems. Moreover, the methodological approaches within the IS discipline show a clear preference for quantitative, macro-level studies, frequently employing analysis of big data. This preference results in a limited adoption of qualitative methods, which are essential for understanding the nuanced, human aspects of how false information and hate speech propagate and affect individuals and communities. Our examination of ongoing projects in Germany and the European Union highlights that while there are numerous initiatives addressing false information and hate speech, the involvement of IS research remains relatively limited. Instead, many of these projects are driven by the field of computer science, with a strong emphasis on algorithm development. This indicates a significant opportunity for IS researchers to contribute more robustly to the current discourse and efforts against false information and hate speech. The interdisciplinary nature of IS, which inherently blends technological and social perspectives, positions it uniquely to address these complex issues. This is underlined by our identified IS projects being more frequently interdisciplinary projects than those involving related disciplines, although the sample size is small. By incorporating socio-technical perspectives, IS research can bridge the gap between purely technical solutions and the broader societal implications. This involves integrating insights from ethics, law, and other relevant fields to evaluate and implement mechanisms and countermeasures effectively in real-world applications, particularly within governmental and regulatory authorities.

Despite the current limitations, the projects addressing false information and hate speech cover a wide variety of target groups and countries, underlining the global importance of these issues. This diversity in focus underscores the need for comprehensive solutions that are adaptable to different cultural and social contexts. The IS discipline's strong focus on technological solutions provides valuable tools for combating false information and hate speech. However, to enhance the impact of this research, there is a critical need to integrate theoretical frameworks, policy advice, and qualitative methods. By embracing a more balanced and interdisciplinary approach, IS researchers can make significant contributions to building resilient democracies. These democracies would be better informed, more inclusive, and more capable of countering the challenges posed by false information and hate speech in the digital age. Eventually, the IS discipline should feel encouraged to heed the call for action, particularly in the area of hate speech, where its contributions have been sparse. By leveraging its interdisciplinary strengths and adopting a socio-technical perspective, IS research can not only advance the understanding of false information and hate speech but also develop more effective strategies to combat these issues, ultimately fostering a more informed and cohesive society.

6 Conclusion

To build and preserve resilient democracies, it is essential to evaluate the current state of publicly funded research on false information and hate speech. By mapping out existing efforts, we can identify gaps and areas requiring further exploration. This evaluation may guide future projects, ensuring they address the most pressing issues and contribute effectively to preserving and promoting democratic values. Our project review presented in this paper reveals that the IS discipline's current research landscape on false information and hate speech while interdisciplinary is heavily oriented toward technological solutions, with an emphasis on digital tools and machine learning. While this reflects the discipline's strengths, there is a notable gap in theoretical, policy-oriented, and qualitative research. Addressing these gaps could lead to more comprehensive strategies for combating false information and hate speech, ultimately fostering a more informed and safe digital democracy. Additionally, information systems as a discipline is underrepresented in projects funded by the DFG and the EU, implying there are still opportunities for IS to be involved in other types of projects. Finally, hate speech is rarely researched in projects by information systems researchers, although as a discipline we might be able to provide valuable insights for theory and practice.

The valuable insights provided by this research have some minor limitations. Due to practical reasons, only publicly funded projects, which were listed in the BMBF, GEPRIS and DFG databases, could be taken into consideration. Still, these funding sources cover the most important organizations Hornbostel (2001). Additionally, this research adopted a particular emphasis on Germany and the EU. Expanding the geographic focus, especially towards the global south, would be beneficial in capturing a more diverse range of projects and insights.

Reflecting on the call by Weinhardt et al. (2024) to establish novel areas for Digital Democracy research, there is a clear need for IS researchers to broaden their focus beyond technological solutions to include the exploration of how digital platforms influence human behavior and social cohesion. Interdisciplinary research across information systems, computer science, political science, sociology, communication science, and law is crucial to understanding and mitigating the broader impacts of platforms in our democracies. IS researchers are encouraged to prioritize transparency, inclusion, and literacy, developing innovative ways to preserve and promote democratic values. By focusing on trust, the foundation of democratic engagement, researchers can examine how misinformation, disinformation, malinformation, and hate speech influence the political landscape and public trust.

References

- Appel, M. (2020), Innovation and intellectual property rights, in M. Appel, ed., 'Die Psychologie des Postfaktischen: Über Fake News, Lügenpresse, Clickbait & Co.', Springer, pp. 1–7.
- Aïmeur, E., Amri, S. & Brassard, G. (2023), 'Fake news, disinformation and misinformation in social media: a review', *Social Network Analysis and Mining* **13**(1), 1–36.
- Bennett, W. & Livingston, S. (2018), 'The disinformation order: Disruptive communication and the decline of democratic institutions', *European Journal of Communication* **33**(2), 122–139.
- Bezzaoui, I., Fegert, J. & Weinhardt, C. (2022), 'Truth or fake? developing a taxonomical framework for the textual detection of online disinformation', *International Journal on Advances in Internet Technology* **15**(3/4), 53–63.
- Bilewicz, M. & Soral, W. (2020), 'Hate speech epidemic. the dynamic effects of derogatory language on intergroup relations and political radicalization', *Political Psychology* **41**, 3–33.
- Bronstein, M. (2018), 'Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking', *Journal of Applied Research in Memory and Cognition* **8**(1), 108–117.
- Bäumler, J., Kaufhold, M.-A., Voronin, G. & Reuter, C. (2024), Towards an online hate speech classification scheme for german law enforcement and reporting centers: Insights from research and practice, in 'MCI-WS101: Workshop Mensch-Maschine-Interaktion in sicherheitskritischen Systemen'.
- Cialdini, R., Reno, R. & Kallgren, C. (1990), 'A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places', *Journal of Personality and Social Psychology* **58**(6), 1015.
- Eckert, S., Metzger-Riftkin, J., Kolhoff, S. & O'Shay-Wallace, S. (2021), 'A hyper differential counterpublic: Muslim social media users and islamophobia during the 2016 us presidential election', *New Media & Society* **23**(1), 78–98.
- Faragó, L., Krekó, P. & Orosz, G. (2023), 'Hungarian, lazy, and biased: the role of analytic thinking and partisanship in fake news discernment on a hungarian representative sample', *Scientific Reports* **13**(1), 178.
- Fraser, N. (1990), 'Rethinking the public sphere: A contribution to the critique of actually existing democracy', *Social Text* **25**(26), 56–80.
- Gnewuch, U. & Mädche, A. (2022), Toward a method for reviewing software artifacts from practice', in A. Drechsler, A. Gerber & A. Hevner, eds, 'The Transdisciplinary Reach of Design Science Research', Springer International Publishing, Cham, pp. 337–350.
- Habermas, J. (1962), 'Structural transformation of the public sphere'.
- Henig, D. & Knight, D. (2023), 'Polycrisis: Prompts for an emerging worldview', *Anthropology Today* **39**(2), 3–6.
- Hornbostel, S. (2001), 'Third party funding of german universities. an indicator of research activity?', *Scientometrics* **50**(3), 523–537.

- Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann-Herz, L., Wachs, S. & Bilz, L. (2023), 'A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena', *Trauma, Violence, and Abuse* **24**(4), 2598–2615.
- Kapantai, E. (2021), 'A systematic literature review on disinformation: Toward a unified taxonomical framework', *New Media & Society* **23**(5), 1301–1326.
- Ledford, H. (2023), 'Researchers scramble as twitter plans to end free data access', *Nature* **614**(7949), 602–603.
- Mahyoob, M., Algaraady, J. & Alrahaili, M. (2021), 'Linguistic-based detection of fake news in social media', *International Journal of English Linguistics* **11**(1), 99–109.
- Mayring, P. (2015), Qualitative content analysis: Theoretical background and procedures', in A. Bikner-Ahsbals, C. Knipping & N. Presmeg, eds, 'Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods', Springer Netherlands, Dordrecht, pp. 365–380.
- Naeem, S., Bhatti, R. & Khan, A. (2021), 'An exploration of how fake news is taking over social media and putting public health at risk', *Health Information & Libraries Journal* **38**(2), 143–149.
- Paasch-Colberg, S., Strippel, C., Trebbe, J. & Emmer, M. (2021), 'From insult to hate speech: Mapping offensive language in german user comments on immigration', *Media and Communication* **9**(1), 171–180.
- Pennycook, G. & Rand, D. (2019), 'Lazy, not biased: Susceptibility to partisan news is better explained by lack of reasoning than by motivated reasoning', *Cognition* **188**, 39–50.
- Sample, C., Jensen, M., Scott, K., McAlaney, J., Fitchpatrick, S., Brockinton, A., Ormrod, D. & Ormrod, A. (2020), 'Interdisciplinary lessons learned while researching fake news', *Frontiers in Psychology* **11**, 537612.
- Shahi, G., Jaiswal, A. & Mandl, T. (2024), Fakeclaim: A multiple platform-driven dataset for identification of fake news on 2023 israel-hamas war, in 'European Conference on Information Retrieval', Springer Nature Switzerland, Cham, p. 66–74.
- Shu, K. (2017), 'Fake news detection on social media: A data mining perspective', *ACM SIGKDD Explorations Newsletter* **19**(1), 22–36.
- Soral, W., Liu, J. & Bilewicz, M. (2020), 'Media of contempt: Social media consumption predicts normative acceptance of anti-muslim hate speech and islamoprejudice', *International Journal of Conflict and Violence* **14**, 1–13.
- Tandoc Jr, E. C., Lim, Z. W. & Ling, R. (2018), 'Defining "fake news" a typology of scholarly definitions', *Digital Journalism* **6**(2), 137–153.
- Union, E. (2024), 'Germany – eu country profile'.
URL: https://european-union.europa.eu/principles-countries-history/eu-countries/germany_en
- Vasist, P., Chatterjee, D. & Krishnan, S. (2024), 'The polarizing impact of political disinformation and hate speech: a cross-country configural narrative', *Information Systems Frontiers* **26**(2), 663–688.
- Verma, P. (2021), 'Welfare: word embedding over linguistic features for fake news detection', *IEEE Transactions on Computational Social Systems* **8**(4), 881–893.
- Wardle, C. & Derakhshan, H. (2017), Information disorder: Toward an interdisciplinary framework for research and policymaking, Technical report, Council of Europe Strasbourg.

- Weinhardt, C., Fegert, J., Hinz, O. & Aalst, W. (2024), 'Digital democracy: A wake-up call: How is research can contribute to strengthening the resilience of modern democracies', *Business & Information Systems Engineering* **66**, 127–134.
- Yu, S. & Lo, D. (2020), Disinformation detection using passive aggressive algorithms', *in* 'Proceedings of the 2020 ACM Southeast Conference', pp. 324–325.