**Research Proposal Presentation Transcript**

|  |  |
| --- | --- |
|  | Welcome to the presentation of my research proposal titled, “Enhancing Fake News Detection on Twitter through Graph Neural Networks:  An Analysis of User Behavior, Network Structure, and Content Engagement.” |
|  | The agenda is aligned with the specifications outlined on Moodle and encompasses the topics listed there. |
|  | The information landscape has faced a significant decline, highlighted by the World Health Organization's 2020 declaration of a global "infodemic," a term brought to attention by Van Der Linden in 2022. The challenge of misinformation has reached such a height that, as Lewandowsky and colleagues pointed out in 2020, individuals may continue to support political figures even when they recognize the untruths in their statements. This creates a paradox, emphasizing technology's dual role—not just as a catalyst for spreading misinformation but also as a crucial part of the solution, a perspective earlier suggested by Lewandowsky et al. in 2017.  Fake news, which deliberately presents false information in the guise of legitimate news, has garnered significant attention from both the media and the research community. The slide before you showcases a recent instance of fake news circulated on Twitter. This is just one example among countless others, yet it's critical to note the extensive interaction such false information receives. This highlights the urgent need for research focused on the automatic detection of fake news, underscoring its importance in today's digital information age. |
|  | Although this research proposal stems from the literature review completed earlier in the course, it has been carefully developed to serve as a viable postgraduate computing project. The project outlined here is contemporary and challenging, designed to autonomously showcase the skills acquired, making it equally suitable as a final computing project. |
|  | To ensure the presentation flows cohesively, the results of the literature review are presented at this early stage. In this way, a thematic introduction can be conveyed directly.  The literature review adopted a methodology from the University of Zurich, uniquely using recent Literature Reviews in place of standard literature recommendations. This approach was supported by extensive searches across major databases, including IEEE Xplore Digital Library, ACM Digital Library, arXiv, ScienceDirect, Google Scholar, and Consensus. For a comprehensive account of the methods and findings, refer to the literature review submitted in week 7 of the course. |
|  | Numerous scholars, including Altay et al. (2023) and Ecker et al. (2022), have emphasized the highly interdisciplinary nature of fake news detection. A critical aspect of this research involves pinpointing the key factors that heighten individuals' vulnerability to fake news. People frequently assess the veracity of news through the lens of their political biases, often overvaluing news that aligns with their views and discounting that which does not. Furthermore, certain elements have been identified as increasing the likelihood of falling for fake news, such as biases, partisanship, identity, and the level of trust in media and political institutions. On the other hand, a propensity for analytical thinking has been shown to mitigate these vulnerabilities. Arin et al. (2023) highlight that individuals who are older, male, have higher incomes, and lean politically to the left tend to be more adept at identifying fake news. |
|  | The latest advancement in incorporating network details for fake news detection was made by Soga et al. (2024). Their approach involved the development of a Graph Neural Network (GNN) model that integrates the stance of users towards the source tweet, essentially capturing whether individuals agree or disagree with it. This determination of stance was innovatively calculated through direct tweet interactions, including comments and likes. |
|  | As shown, Graph Neural Networks (GNNs) have been central to the latest advancements in the field. These successes present an opportunity to be leveraged and expanded upon in a targeted manner. The research question is hence formulated as “What specific attributes of user behavior, network structure, and content engagement on Twitter enhance the efficacy of Graph Neural Network (GNN)-based models for detecting fake news?“  In simpler terms: The development of a GNN-based model is proposed, designed to intricately incorporate user interactions, network structure, and content engagement into the detection mechanism, offering a comprehensive approach to analysis. |
|  | The project's objective is to surpass the F1 score of the model introduced by Soga et al. (2024) by 5%, utilizing the same dataset, albeit extended. Additionally, it seeks to identify which features enhance the model's performance and which do not contribute significantly. To achieve this, comparisons will be made using randomized features to establish a clear understanding of the impact of various factors on model efficacy. |
|  | Utilizing theories from psychology and related disciplines, this project adopts a deductive approach to investigate fake news detection on Twitter. It employs a quantitative, exploratory research method aimed at identifying relevant socio-affective determinants. These determinants are derived from the data accessible on Twitter, offering a novel perspective on how psychological insights can enhance the accuracy of fake news identification. |
|  | To facilitate a direct comparison of outcomes, this project employs the benchmark datasets originally used by Soga et al. (2024). These datasets are rehydrated via the Twitter API—a process that will be elaborated upon subsequently—and enriched with previously quantified metrics. Model development is carried out using PyTorch Geometric and BERT, ensuring the use of cutting-edge technology in natural language processing and graph-based learning. The exploration of individual socio-affective determinants is conducted iteratively, with each determinant being progressively integrated into the model. This iterative process, inspired by principles of agile software development, enables a structured yet flexible approach within the exploratory nature of this research. |
|  | The ethical foundation of this project is grounded in the BCS Code of Conduct and the Menlo Report, with a particular emphasis on personal autonomy. Despite the public nature of tweets, research by Kaphle et al. (2022) reveals that Twitter users frequently lack awareness of the platform's terms and conditions. Moreover, the public availability of data does not negate the presence of personal information within it. In addition, the challenge of anonymization is notable; simply removing names does not effectively obscure an author's identity, a task that becomes increasingly difficult as more network-specific data is incorporated. This highlights the complexity of ethical considerations in handling publicly available, yet personally identifiable, information. |
|  | To safeguard users' rights, particularly the right to be forgotten, two prevalent methods are employed. The first method, data hydration, is depicted in the slide. This process involves retaining only Twitter user or post IDs in a dataset for distribution. Should a researcher wish to replicate the study, they must reacquire the data associated with these IDs via the API. This ensures that if a user has subsequently deleted their post or account, such data will be absent from the new researcher's dataset, aligning with the user's right to be forgotten.  However, this approach introduces a challenge to the 100% reproducibility of research projects. To counteract this, initiatives like the Social Media Archive have been developed. This project anonymizes data and secures it, granting access only to individuals who meet specific security criteria. The choice between these methods is determined by the project's advancement and the requisite level of data protection. |
|  | The project's objective is to enhance fake news detection by incorporating additional network-specific information. However, there's a potential ethical concern: this data could also enable the classification of users based on personal characteristics. Utilizing such information for classification purposes, akin to social scoring, poses an unacceptable risk under the AI Act due to the profound implications for privacy and individual rights.  Moreover, there are inherent risks associated with relying on Twitter (now referred to as X) data, which is often biased. Additionally, X's recent adjustments to its terms of use could further complicate research efforts, either by restricting data availability for scholarly purposes or imposing prohibitive costs for access. These challenges underscore the need for careful consideration of ethical standards and legal compliance in the project's design and implementation. |
|  | Given the project's exploratory and incremental nature, the specific artefacts to be generated remain undefined at this stage. Nevertheless, ethical considerations are paramount. To address these, the project will either provide a dataset comprised solely of Twitter IDs, enabling data rehydration, or utilize a service like SOMAR (Social Media Archive) to ensure data remains reproducible yet securely protected. This approach balances the need for research integrity with the imperative to uphold ethical standards, ensuring that personal privacy is respected and that data usage complies with legal and ethical guidelines. |
|  | The project follows a standard timeline, with a significant portion already dedicated to conducting the literature review and formulating the research proposal. Should the project commence as scheduled, the development of the model will adopt an interactive approach. Subsequent phases will allocate time specifically for finalizing the documentation. Considering the ECTS credits assigned to this computing project, an estimated maximum of 600 hours has been outlined. This estimation includes a substantial buffer to accommodate any unforeseen circumstances that may arise during the project's progression. |
|  | For a comprehensive understanding of the sources and insights that have shaped this presentation, the following references have been utilized extensively. |
|  | Thank you very much for your attention. |