ONLINE RISKY BEHAVIOUR (ORB) DETECTORS

Enhancing Online Safety in Technical Communication

December 3rd , 2024

Group Members/Authors:

Tauha Imran (22i1239)

Minahil Ali (22i0849)

Table Of Contents

Abstract

This project explores the balance between user privacy, safety, and transparency in Online Risky Behavior (ORB) Detectors. Technical studies and expert interviews highlighted the importance of AI technologies for real-time detection, the challenge of balancing privacy with safety, and the need for transparent monitoring practices. Addressing ethical and legal concerns, such as AI biases and data protection laws, is crucial.ORB Detectors can currently provide online safety by preventing harmful activities such as cyberbullying and harassment. Future improvements should focus on contextual analysis, user feedback, and multi-language support to ensure effective and compliant ORB Detectors.

Introduction

Online Risky Behavior Detectors are an emerging technology designed to identify and prevent harmful activities online. It focuses on enhancing online safety by detecting behaviors such as cyberbullying, harassment, and other forms of digital misconduct. This technology can significantly impact technical communication by ensuring safer online interactions.It offers the potential to protect brand reputation, reduce legal risks, improve multicultural communication and enhance user trust in online services. Nowadays Online communication is an important part of our personal and professional lives, hence the need to detect malicious behavior online can not be compromised. However communication challenges may arise in implementing this technology. These challenges include ensuring user privacy, addressing ethical concerns, and maintaining transparency about detection methods. Previous research related to this involves assessment tools for risky behaviors [1] and Identifying risky Internet use along with associating negative online experiences with specific online behaviors [2].

This research project investigates the understanding, significance, and implications of Online Risky Behavior (ORB) Detectors. By exploring expert insights on how ORB Detectors work, their role in enhancing online safety, and the potential for mitigating risky behaviors of both the practical and theoretical aspects of the technology. The research also examines the impact of these detectors on user trust in online platforms and considers personal willingness to use such technology. Key ethical and legal challenges, including issues of transparency, AI biases, and privacy concerns, are addressed, alongside suggestions for enhancing the effectiveness and ethical use of ORB Detectors.

Investigating the role of technical, research, and business writing skills in the context of AI-driven landscapes is crucial for developing clear, ethical, and effective communication strategies. These skills ensure that complex AI technologies are accessible and transparent to diverse audiences, fostering trust and compliance with legal standards. Additionally, they play a pivotal role in translating technical insights into actionable business strategies, driving innovation and competitive advantage.

Methodology

The project uses two major research methodologies. Expert interviews and thematic analysis.Expert interviews were taken of professionals developing these ORB detectors and deploying these technologies for testing. For thematic analysis , multiple articles and studies were used to better understand the nature and current progress of these technologies.
Each of these methodologies are discussed in detail below.

Expert Interviews:

A Cyber Security expert developing ORB technology was interviewed in great detail. We will address them as Z for the remainder of the report (they're identity is to remain anonymous). The interviewee, Z , was approached by our team formally and provided their consent for the interview. The interview consisted of a total ten questions that were addressed over the span of 30-40 minutes.

The result of the interview with Z emphasizes the use of AI technologies, especially Natural Language Processing (NLP), Social Media Forensics and Machine Learning (ML), for detecting harmful online behaviors like cyberbullying and harassment. Z, along with other experts also stress the importance of real-time detection to prevent escalation. Balancing privacy and safety is a key concern, with discussions on ethical challenges and the need for compliance with data protection laws such as GDPR (General Data Protection Regulation). Building user trust through transparency about how detectors function and data usage is highlighted. Concerns about AI biases and the potential invasion of privacy are noted, with a need for safeguards. Suggestions for future improvements include contextual analysis, user feedback mechanisms, and multi-language support to enhance detection accuracy and fairness.

Looking at the results of the Expert interviews , we can measure the frequency of mentions of important aspects , as shown in the figure below.

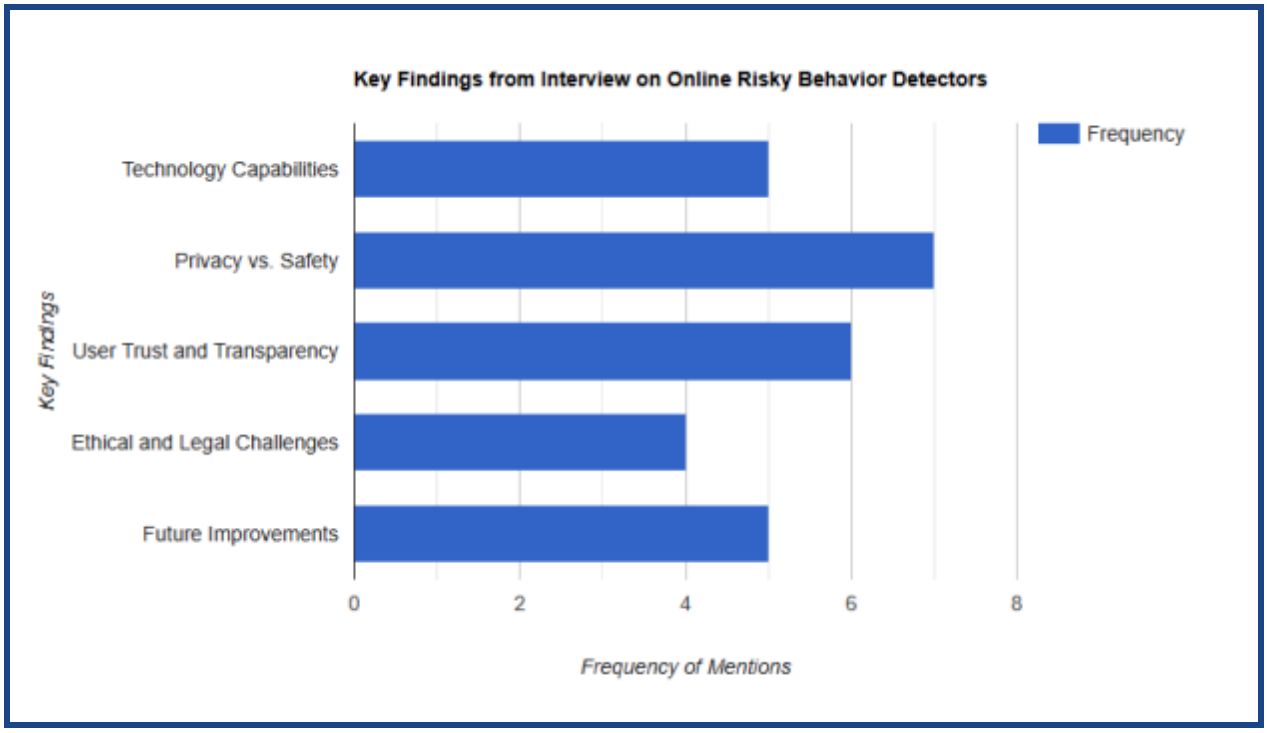

Figure 1: Histogram of frequency of mentions of Key Findings from Expert Interview with Z.

Looking at the Expert interviews , our interviewee and research partner, Z, emphasized on the legal issues caused by AI-bias.

"Legal issues, especially around data protection and the misidentification of risky behaviors, will need to be addressed before this technology can be widely deployed."

Z , CyberSecurity Expert

Z, further emphasized ,

"If users do not feel confident that their interactions are being handled fairly and securely, they will reject the technology."

Z , CyberSecurity Expert

Overall, expert interviews gave us valid insights into the technical and social aspects of these ORB detectors

Thematic Analysis:

The thematic analysis of online risky behavior detectors reveals a complex interplay between technological advancements in AI, ethical challenges related to privacy and bias, and the critical need for transparency to build user trust.

The results of one study show that both youth and secondary schools' students are likely to engage in risky behaviours online and mostly it is related to sharing their personal information online or texting.[3] Articles online

also mention how "*It's easy for busy people to put off dealing with online privacy and device security.*"[4] which in turn promotes and favours people putting off risky behaviours. This aspect can prove very useful for ORB detectors becoming popular amongst the public due to the provided hassle free security and safety.

In technical aspects ,a recent research conducted in 2024 into the ORB technology's development shows quantitative results show that the *proposed model achieved good accuracy of about 86.86%,* which outperforms recent works.[5]

As we continue to go deeper into the details, the findings shed light on the intricate balance between leveraging AI for online safety and addressing the ethical, legal, and privacy concerns that arise.

Findings

The study of Online Risky Behavior Detectors reveals several key takeaways crucial for enhancing online safety. These detectors have the potential to significantly improve online environments by identifying and preventing harmful activities such as cyberbullying and harassment. Real-time detection capabilities are vital for user protection, enabling immediate intervention to prevent escalation.

However, one of the most significant challenges identified is balancing user privacy with the need for safety. Transparency in monitoring practices is essential to build trust among users, who may otherwise feel uneasy about the extent of surveillance. Ethical and legal implications also arise, particularly concerning user privacy and AI biases. Compliance with regulations like GDPR (General Data Protection Regulation) is necessary to avoid legal issues and ensure that data handling practices are ethical and transparent.

User trust is another critical component. Trust can be enhanced by clearly communicating how online behaviors are monitored and giving users control over their privacy settings. One expert highlighted the importance of user confidence, stating that without a sense of fair and secure handling of their interactions, users are likely to reject the technology. Thus, providing a safer and more positive online experience can help businesses foster a sense of trust and confidence among their users. We can see a distribution of the most mentioned difficulties in implementing these technologies.

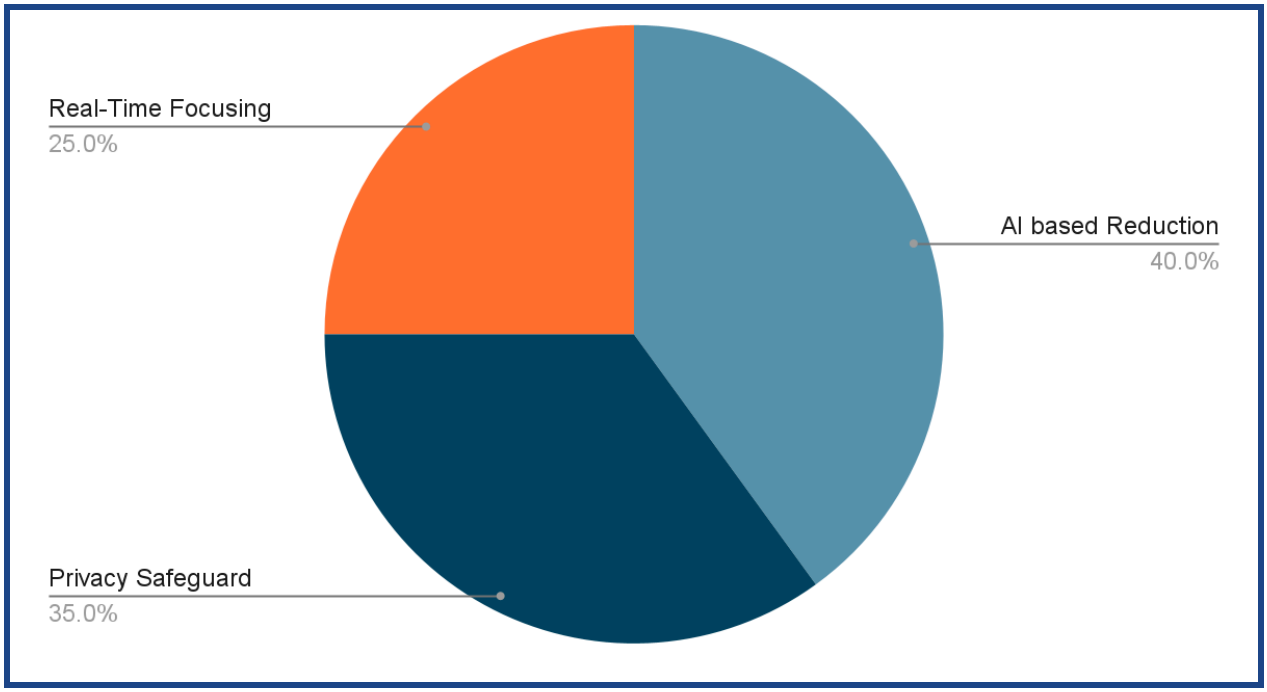

Figure 2 : Pie Chart Showcasing division of mentions of difficulties in ORB Technologies.

To improve accuracy, the technology should focus on understanding the context of online interactions. This approach can help reduce false positives and negatives by recognizing the nuances of language, humor, and cultural differences. Future enhancements suggested include incorporating user feedback mechanisms, contextual analysis, and support for diverse languages.

"Bias in AI is a real risk, and if these tools misidentify risky behaviors because of flawed algorithms, it could damage reputations and even breach laws."

- Z , CyberSecurity Expert

These improvements will not only enhance the effectiveness of the technology but also address ethical concerns.

Discussion

The key findings of this study underline the crucial role of Online Risky Behavior (ORB) Detectors in enhancing online safety by identifying harmful behaviors such as cyberbullying and harassment.The emphasis on real-time detection highlights its importance in preventing the escalation of these activities, making it a vital feature for user protection. Balancing privacy and safety remains a significant challenge, as transparency in monitoring practices is essential to build trust among users. Ethical and legal implications, particularly regarding user privacy and AI biases, demand strict compliance with regulations like GDPR (General Data Protection Regulation) to avoid legal issues and ensure ethical data handling practices.

Furthermore, user trust emerged as a critical factor for the adoption of these technologies. Providing clear information about monitoring practices and giving users control over their privacy settings can significantly enhance trust. The findings also suggest that improving the accuracy of an ORB Detector requires focus on contextual analysis, which helps reduce false positives and negatives by recognizing the nuances of language, humor, and cultural differences. Future enhancements such as incorporating user feedback mechanisms, contextual analysis, and multi-language support are crucial for improving the technology's effectiveness while addressing ethical concerns.

Below is a suggested pie chart of enhancements based on the key findings

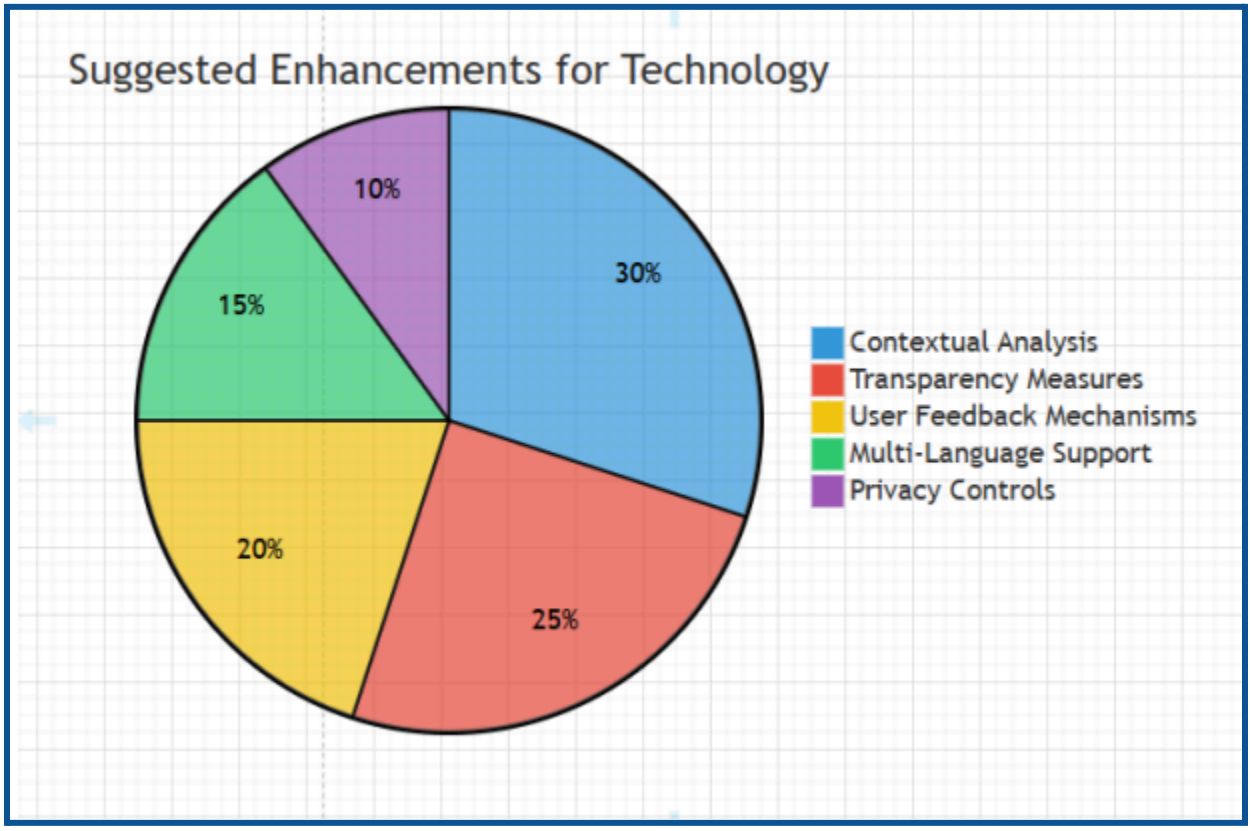

Figure 3: Pie Chart showing the distribution of the quantitative distribution of suggested enhancements for ORB technologies derived from both research methods.

To build on these findings, it is recommended to conduct quantitative analysis to validate the qualitative insights obtained. Developing ethical AI guidelines will ensure that the ORB Detectors operate within an ethical framework, addressing concerns related to biases and privacy. Prototyping ORB detectors can provide practical insights and identify areas for improvement. Finally, collaborating with policymakers is essential to ensure that the development and deployment of ORB detectors are aligned with legal standards and societal expectations. This collaborative approach will help create a safer and more trustworthy online environment.

Conclusion

This research investigates the development and implementation of Online Risky Behavior (ORB) Detectors, emphasizing the balance between user privacy, safety, and transparency. Through expert interviews, it uncovers ethical, technical, and legal challenges, highlighting the importance of AI technologies for real-time detection, the necessity for transparent practices, and addressing biases and data protection concerns.

Key takeaways include the critical role of AI in enhancing online safety, the delicate balance between monitoring for harmful behaviors and protecting user privacy, and the need for transparent communication to build user trust. Addressing ethical and legal issues, incorporating user feedback, and ensuring multi-language support are essential for effective deployment.

Clear communication skills are paramount in AI-driven environments to ensure complex technologies are understandable, transparent, and trustworthy. These skills help translate technical insights into actionable strategies, fostering innovation and maintaining compliance with legal standards.

Appendix

AI (Artificial Intelligence):

The simulation of human intelligence in machines that are programmed to think and learn. AI enables machines to perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation.

ML (Machine Learning):

A subset of artificial intelligence that involves the development of algorithms that allow computers to learn from and make decisions based on data. Machine learning models improve their performance over time without being explicitly programmed for specific tasks.

ORB (Online Risky Behavior Detectors):

Technologies designed to identify and mitigate harmful online behaviors such as cyberbullying, harassment, and aggressive language. These detectors use AI and machine learning techniques to analyze and respond to potential risks in real time.

Social Media Forensics:

The process of analyzing social media platforms to gather and examine digital evidence. Social media forensics involves the collection, preservation, and interpretation of data from social networks to investigate and resolve legal, criminal, or ethical issues.

GDPR (General Data Protection Regulation):

A comprehensive data protection law implemented in the European Union in May 2018. GDPR aims to protect individuals' personal data and privacy by giving them more control over their information. The regulation mandates clear consent for data collection, data subject rights, transparency, data protection measures, and breach notification requirements.

References

[1]

Castellano S, Platania GA, Varrasi S, Pirrone C, Di Nuovo S. Assessment tools for risky behaviors: Psychology and health. Health Psychol Res. 2020 Oct 1;8(2):9235. doi: 10.4081/hpr.2020.9235. PMID: 33123648; PMCID: PMC7588849.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7588849/

[2]

Gainsbury, S. M., Browne, M., & Rockloff, M. (2019). Identifying risky Internet use: Associating negative online experience with specific online behaviors. New Media & Society, 21(6), 1232-1252. Retrieved from https://doi.org/10.1177/1461444818815442

[3]

Paluckaitė, U., & Žardeckaitė-Matulaitienė, K. (2016). Students' engagement in risky online behaviour: The comparison of youth and secondary schools. ResearchGate. Retrieved from https://www.researchgate.net/publication/305747462_Students'_Engagement_in_Risky_Online_Behaviour_The_Comparison_of_Youth_and_Secondary_Schools'

[4]

Johnson, A. (2021, January 23). 7 risky behaviors you should stop right now. Norton. Retrieved from https://us.norton.com/blog/privacy/risky-online-behaviors

[5]

Salim, L. M., & Celik, Y. (2024). Detection of Dangerous Human Behavior by Using Optical Flow and Hybrid Deep Learning. Electronics, 13(11), 2116. https://doi.org/10.3390/electronics13112116