Smart Technologies Academic Press

Jordanian Journal of Informatics and Computing
https://www.jjic.thestap.com/

AI-Powered Cyberbullying Detection and Prevention for Safer Digital Spaces

Soumiya Marichelvan[1], Aitizaz Ali[1*], Uzair Ali[2]

[1] School of Technology, Asia Pacific University of Technology and Innovation (APU), Malaysia
[2] Department of Computer Science, AWKUM, KPK, Pakistan

ARTICLE INFO

Article History

Received: 27-04-2026
Revised: 29-05-2026
Accepted: 02-09-2026

Vol.2027, No.1

DOI:

Corresponding author.
Email:
Aitizaz.ali@apu.edu.my

Orcid:
https://orcid.org/0000-0002-4853-5093

ABSTRACT

This research aims to address the growing issue of cyberbullying, particularly among youth, by developing an AI-powered system for real-time detection and prevention on digital platforms. The research focuses on understanding the various forms of cyberbullying and their detrimental effects on mental health. Using the Agile system development methodology, the research will follow an iterative process involving design, development, and testing phases to create a reliable detection tool. The system is intended to identify harmful online behaviour and alert parents or guardians promptly. Preliminary results suggest that the tool will feature an intuitive user interface and effective algorithms capable of detecting abusive interactions. This research is significant, as it proposes a technological solution to a pressing social issue, aiming to enhance the safety and well-being of young people online. Additionally, the research aligns with SDG 16 Peace, Justice, and Strong Institutions by contributing to the protection of children from online harm and promoting safer, more just digital environments.

Keywords: Cyberbullying, AI, Detection, Prevention, NLP, Agile

How to cite the article

1. Introduction

In today's world, the internet is a big part of everyday life, especially for children and teenagers who use social media and messaging apps to stay connected. But along with the good things about being online, there's a serious issue called cyberbullying. Cyberbullying is when people use digital platforms to hurt or harass others, and it can have a big emotional and psychological impact, especially on young people. In Malaysia, cyberbullying is becoming a bigger problem, with over 3,000 complaints reported in 2023 alone, showing how urgently we need better ways to prevent it (Rashidi, 2024) (Nortajuddin, 2020).

Although organizations like the Malaysian Communications and Multimedia Commission (MCMC) have tried to fight this problem through rules and awareness campaigns, these efforts haven't been enough to stop the growing number of cases (Rashidi, 2024). Schools, social media platforms, and tech companies are also trying to find solutions, but most of what they do is reactive, meaning they only respond after the bullying happens. This is where technology can really make a difference in how we handle cyberbullying.

This research is about creating AI-powered software that can spot and stop cyberbullying in real-time. By using Natural Language Processing (NLP) and machine learning, the software will detect harmful messages before they are sent, block them, and let parents or guardians know if needed. This proactive approach is important because the traditional way of dealing with cyberbullying is to wait for it to be reported and then react, which often happens too late (Kuppusamy, 2023) (Nortajuddin, 2020). The research required for this research will focus on:

• Investigating NLP techniques for detecting harmful language patterns.
• Exploring machine learning models that can adapt to evolving online behaviors, such as sarcasm or indirect threats (Kuppusamy, 2023).
• Ensuring the software is user-friendly and can be easily adopted by parents to monitor their children's online interactions.

This research aims to create a safer digital environment for Malaysian youth by addressing the limitations of current systems and offering a real-time solution that protects children and teenagers from the harmful effects of cyberbullying. This study aims to:
1. To detect harmful and bullying content in text messages using AI and Natural Language Processing (NLP).
2. To block users from sending messages that contain cyberbullying content.
3. To notify parents or guardians when harmful messages are detected.
4. To provide a user-friendly interface for monitoring and preventing harmful communication.

1.2 Research Background

1.2.1 Escalating Cyberbullying

In the digital age, the rise of social media and online communication platforms has transformed how people interact, especially among young people. While these platforms offer numerous benefits, they also expose users, particularly children and adolescents, to harmful behaviors such as cyberbullying. Cyberbullying is defined as the use of digital technology to harass, intimidate, or harm others. Unlike traditional bullying, cyberbullying can occur anonymously, at any time, and can reach a wider audience, making it especially insidious.
In Malaysia, as well as globally, the incidence of cyberbullying has surged, with numerous reports highlighting its severe psychological, emotional, and social consequences for victims. According to a 2023 report, over 3,000 cases of cyberbullying were recorded in Malaysia alone, signaling a growing public concern and the urgent need for more effective preventive measures.

Despite various anti-cyberbullying initiatives, existing solutions, such as reporting mechanisms and educational programs, are often insufficient in preventing or detecting cyberbullying in real-time. Current tools for identifying harmful online behaviors lack the ability to automatically detect and address incidents across multiple digital platforms. Moreover, these solutions often fail to provide timely intervention or support for victims, leaving them vulnerable to ongoing harassment.

1.2.2 Limitations of Traditional Detection Methods

Existing cyberbullying detection techniques, which primarily rely on traditional machine learning and keyword-based methods, struggle with understanding the complex, evolving language of online interactions. This challenge includes recognizing nuanced forms of cyberbullying, such as sarcasm, implicit threats, and contextual language, which often evade

simple keyword detection systems (Walli et al., 2024). Additionally, traditional models lack the capacity to process the large volume of data generated on platforms like Instagram, WhatsApp, and Twitter, which further limits their effectiveness.

1.2.3 Limitations of Deep Learning Models in Real-World Data

Recent research demonstrates that advanced deep learning techniques, like Bidirectional Encoder Representations from Transformers (BERT), have shown promise in improving cyberbullying detection by leveraging contextual understanding of language (Fadlallah et al., 2024). BERT and similar models excel at analyzing complex text structures, enabling them to detect bullying language that may be indirect or non-explicit. However, while these deep learning models have achieved notable success in experimental environments, they still face limitations when applied to real-world social media data. Studies have shown that such models can suffer from accuracy drops due to factors like diverse linguistic styles, cultural differences, and rapidly changing online slang, which require continuous model training and updates (Mukta et al., 2023).

1.2.4 Anonymity and Accountability Issues

Moreover, the anonymous nature of online interactions allows perpetrators to evade consequences, which exacerbates the cyberbullying problem by reducing accountability (H. D et al., 2023). Even with advances in machine learning, current systems frequently fail to address this anonymity factor effectively, limiting the success of detection tools in providing comprehensive solutions.

1.2.5 Lack of Real-Time Prevention and Contextual Depth

In sum, while traditional cyberbullying detection tools have made strides in identifying harmful content, they often lack the contextual depth and real-time preventive capabilities needed to provide safe digital spaces for young people. Through advancements in deep learning, particularly using BERT and other NLP techniques, this research intends to create a robust system tailored to the dynamic and culturally nuanced environment of Malaysian social media interactions (Fadlallah et al., 2024; Walli et al., 2024; Mukta et al., 2023).

This research seeks to address these gaps by developing an AI-powered cyberbullying detection and prevention system. The system aims to automatically detect cyberbullying behaviors across digital platforms, providing real-time alerts to parents, educators, or administrators, and offering interventions that could reduce the impact of cyberbullying. By utilizing advanced machine learning algorithms, this system will improve the accuracy and efficiency of detecting harmful online behaviors, addressing the need for timely, automated solutions to a growing societal issue.

2. Literature Review

2.1 Cyberbullying

Cyberbullying is defined as the use of electronic communication to harass, intimidate, or harm others, often repeatedly and with malicious intent. It takes place across digital platforms such as social media, messaging apps, gaming forums, and emails. Unlike traditional bullying, cyberbullying transcends physical boundaries, making it a pervasive issue that victims often cannot escape. The virtual nature of cyberbullying enables perpetrators to remain anonymous, which not only reduces accountability but also emboldens harmful behavior (Milosevic et al., 2022). The unique features of digital environments, such as permanence and viral spread, further intensify the impact on victims, amplifying psychological harm and social ostracism (Ambareen & Sundaram, 2023).

The types of cyberbullying are varied and constantly evolve. One prominent form is harassment, characterized by the persistent sending of offensive messages intended to distress the recipient. Similarly, flaming involves hostile and confrontational interactions, often in public forums, where individuals exchange insults or aggressive language. Exclusion, another form, deliberately isolates individuals from online groups or communities, thereby exacerbating feelings of loneliness and rejection. Impersonation, where a perpetrator assumes someone else's identity to post false or harmful content, is particularly damaging, as it tarnishes the victim's reputation while concealing the bully's identity. Outing and trickery involve exposing sensitive personal information or deceiving someone into sharing private details, which are then publicly disclosed to humiliate the victim (Balakrishnan & Kaity, 2023).

More extreme forms of cyberbullying, such as doxing and cyberstalking, illustrate the ways in which online harassment can spill over into the physical world. Doxing involves publishing an individual's personal information, such as their address or phone number, to enable further harassment or threats. Cyberstalking, on the other hand, involves persistent monitoring of a victim's online activities, often accompanied by threatening messages or actions, creating a climate of fear and uncertainty (Kumar et al., 2024). These behaviors not only cause significant psychological distress but also pose tangible safety risks.

Emerging trends in cyberbullying reflect the adaptability of perpetrators to new technologies and online behaviors. For instance, memes and GIFs, often used as tools for humor, are increasingly weaponized to mock and ridicule individuals. Similarly, the proliferation of artificial intelligence tools has enabled the creation of deep-fake images and videos, which are manipulated to humiliate victims in sophisticated ways. These advancements complicate the detection and prevention of cyberbullying, highlighting the importance of adaptive, context-sensitive solutions (Ambareen & Sundaram, 2023). The diverse forms of cyberbullying underscore its complexity and the challenges associated with addressing it. As online communication evolves, so too do the methods used by perpetrators, necessitating continuous updates to detection and prevention mechanisms. Addressing cyberbullying requires a nuanced understanding of its types and the contexts in which it occurs, forming the foundation for developing effective intervention strategies.

2.2 Factors Influencing Cyberbullying

Technological advancements, especially the rise of social media and smartphone usage, have significantly contributed to the spread of cyberbullying, making it a pervasive issue in contemporary society. According to Shahzad et al. (2024), these technologies facilitate the anonymity and reach of online harassment, enabling bullies to target victims in ways that were previously not possible. The ease with which individuals can interact online, often without immediate consequences, creates an environment where harmful behaviors can thrive.

Another contributing factor to the prevalence of cyberbullying is the lack of awareness regarding digital etiquette and the consequences of online behavior. Many individuals, especially adolescents, may not fully understand the impact of their online actions. Shahzad et al. (2024) emphasize the need for increased awareness about the ethical use of technology, suggesting that fostering a better understanding of digital communication norms could help reduce cyberbullying incidents.

2.3 Cyberbullying in Malaysia

Cyberbullying has become an increasingly pressing issue in Malaysia, particularly among youths and young adults who are highly active on social media platforms. Malaysia is the second most cyberbullied country in Asia after India, according to a survey by Ipsos. According to the study, which involved 20,793 respondents from all over the world, 23 per cent or one in four Malaysian parents said their children had been the victims of cyberbullying. The prevalence of cyberbullying is alarming, with studies indicating significant psychological and social repercussions for victims. In Malaysia, a study by Samsudin et al. (2023) found that cyberbullying victimization is prevalent among young adults, particularly in urban areas such as Selangor. The study reported associations between cyberbullying victimization and family dysfunction, unhealthy behaviors, and heightened levels of psychological distress, highlighting the multifaceted nature of this issue.

Cyberbullying in Malaysia disproportionately affects young individuals, often occurring on social media platforms that facilitate easy and anonymous interactions. Samsudin et al. (2023) revealed that young adults with strained family relationships and inadequate social support are more likely to experience cyberbullying, exacerbating their risk of mental health issues such as anxiety, depression, and stress. These findings underscore the need to address underlying social and familial factors contributing to the problem. Similarly, Mui et al. (2023) explored the psychological consequences of cyberbullying among Malaysian youth, demonstrating that victimization significantly increases the likelihood of suicidal ideation, mediated by heightened psychological distress. These outcomes highlight the severe mental health risks posed by cyberbullying and the urgent need for intervention strategies.

The relationship between psychological distress and cyberbullying behaviors in Malaysia is well-documented. Ayub et al. (2023) examined the psychological determinants of cyberbullying behaviors among youth in Selangor, finding that higher levels of psychological distress, such as anxiety and depression, are associated with both victimization and perpetration. These findings suggest that psychological well-being is not only a consequence of cyberbullying but also a contributing factor, as distressed individuals are more likely to engage in or become targets of such behavior. Addressing mental health issues through counseling and support systems is crucial for breaking this cycle.

Efforts to address cyberbullying in Malaysia often rely on reactive measures, such as social media reporting mechanisms. However, these approaches are limited in effectiveness, as they primarily address incidents after the harm has occurred. Samsudin et al. (2023) emphasized the importance of preventive measures that focus on strengthening family relationships and promoting healthier behaviors among youths. In addition, Mui et al. (2023) highlighted the potential for integrating mental health support services within schools and communities to mitigate the psychological impact of cyberbullying.

Educational initiatives aimed at promoting digital literacy and responsible online behavior have shown promise in reducing the prevalence of cyberbullying in Malaysia. Ayub et al. (2023) suggested that fostering empathy and self-awareness among youth through school-based programs can reduce both victimization and perpetration rates. These initiatives should be complemented by broader community engagement efforts that encourage open dialogue about cyberbullying and its consequences, particularly within families and peer groups.

The prevalence of cyberbullying in Malaysia highlights the urgent need for a multifaceted approach to prevention and intervention. By addressing the psychological, social, and familial dimensions of the problem, stakeholders can develop comprehensive strategies to mitigate its impact. Incorporating mental health support, educational programs, and community-based interventions will be essential in creating a safer digital environment for Malaysian youths. Figure 1 represents the Malaysia's Cyberbullying Statistics between 2022 and 2024.

Cyber-bullying statistic (2022-2024)

Year	2022	2023	2024 (as of July)	Total
Cyber-bully Case Recorded	4,128	3,737	1,618	**9,483**

Statistic above refers to the number of complaints received by MCMC regarding cyber-bullying. The matters falls under the jurisdiction of the Royal Malaysia Police.

Figure 1. Malaysia's Cyberbullying Statistics (Dzul, 2024).

2.4 Current Measures to Combat Cyberbullying

Efforts to combat cyberbullying have adopted a multifaceted approach, combining educational initiatives, technological tools, and community involvement. The success of these measures depends on the collaboration between schools, families, mental health professionals, and policymakers. Despite significant progress, the evolving nature of digital platforms and cyberbullying behaviors continues to challenge the effectiveness of these strategies. Educational interventions remain at the forefront of efforts to address cyberbullying, emphasizing prevention through awareness and curriculum integration. Awareness programs implemented in schools aim to educate students about the nature of cyberbullying, its psychological and social consequences, and ways to seek help. A study by Tozzo et al. (2022) highlighted the importance of integrating such programs into school activities to foster a culture of empathy and responsible digital behavior. Similarly, Nayak et al. (2023) suggested that embedding cyberbullying topics into school curricula provides students with the tools to identify, prevent, and respond to incidents effectively.

The use of technology has advanced significantly as a means to address cyberbullying. Digital tools, such as reporting systems and monitoring software, empower victims and facilitate timely intervention. For example, mobile applications designed for anonymous reporting enable students to report incidents without fear of retaliation (Tozzo et al., 2022). Furthermore, monitoring software used by parents and schools helps track online activity, alerting stakeholders to potentially harmful interactions before they escalate (Vismara et al., 2022). However, these tools often face limitations, such as balancing privacy concerns with the need for surveillance and adapting to the complexities of detecting nuanced forms of cyberbullying.

Community-based approaches emphasize the importance of a supportive environment to combat cyberbullying. Collaborative programs involving schools, families, and community organizations have shown promise in addressing the issue holistically. Akbar et al. (2020) emphasized the role of such partnerships in fostering a sense of accountability and shared responsibility among stakeholders. Additionally, peer support systems have emerged as an effective intervention,

encouraging students to support their peers and report cyberbullying incidents. Chakraborty et al. (2021) noted that peer-led initiatives can empower students to act as bystanders and advocates, fostering a safer digital space within their communities.

Despite these efforts, challenges persist in implementing and sustaining these measures. Educational programs often struggle with consistency and funding, while technological tools require constant updates to keep pace with evolving digital behaviors. Moreover, cultural and contextual differences affect the adoption and effectiveness of these strategies, necessitating localized solutions tailored to specific communities (Nayak et al., 2023). Continuous research and adaptation are essential to address these gaps and enhance the overall effectiveness of cyberbullying interventions.

Current measures to combat cyberbullying leverage education, technology, and community engagement to create a comprehensive framework for prevention and intervention. However, the dynamic nature of online interactions and the psychological complexities of cyberbullying require ongoing efforts to refine these strategies. By fostering collaboration among stakeholders and integrating innovative approaches, it is possible to build a safer and more supportive digital environment for all users.

3. NLP and AI Techniques for Cyberbullying Detection

Natural Language Processing (NLP) has become a crucial tool in detecting and mitigating cyberbullying, leveraging advanced computational techniques to analyze and understand human language in digital environments. As cyberbullying is often manifested through text-based communication, NLP enables the automatic identification of harmful language patterns and behaviors across various platforms. The application of NLP in cyberbullying detection encompasses a range of methodologies, from traditional machine learning models to advanced deep learning architectures, each contributing to improving the accuracy and efficiency of detection systems.

Supervised machine learning techniques have been widely used for cyberbullying detection, where models are trained on labeled datasets to classify text as bullying or non-bullying. Commonly employed models include Support Vector Machines (SVM), Naïve Bayes, and Decision Trees, all of which have demonstrated significant effectiveness in identifying instances of cyberbullying across social media platforms (Toktarova et al., 2024). These models typically rely on feature extraction methods such as word frequencies and syntactic patterns to detect malicious content. However, the accuracy of these models is often limited by the complexity and nuance of human language, which may not be fully captured by traditional approaches.

Hybrid approaches have emerged as a promising solution to overcome these limitations. By combining NLP techniques with SVM classifiers and tools like Term Frequency-Inverse Document Frequency (TF-IDF) and Linguistic Inquiry and Word Count (LIWC), researchers have achieved impressive detection accuracies. A study by Sathya and Fernandez (2024) demonstrated that hybrid models could detect cyberbullying with up to 93.15% accuracy on platforms such as Twitter. These hybrid systems integrate the strengths of feature engineering and machine learning classifiers, significantly enhancing the precision of detection systems.

Recent advances in NLP have seen the application of Transformer models like BERT and DistilBERT, which have dramatically improved the performance of cyberbullying detection systems. Transformer models excel in capturing contextual nuances and understanding the relationships between words in sentences, making them particularly effective for tasks that require understanding the meaning behind text, such as cyberbullying detection. According to Tapaopong et al. (2024), these models have outperformed traditional techniques, achieving accuracies above 94% in classifying the severity of cyberbullying. BERT's ability to model complex contextual relationships in text allows it to identify more subtle forms of harassment, including sarcastic remarks and covert threats, which may be overlooked by simpler models. Another critical NLP technique used in cyberbullying detection is sentiment analysis. By analyzing the emotional tone of messages, sentiment analysis helps to identify toxic or harmful content that may not always be explicitly offensive but still carries negative emotional weight. Sentiment analysis, often used alongside other NLP tools, aids in detecting messages that express aggression, anger, or frustration-common indicators of cyberbullying (V, 2023). These techniques are instrumental in identifying harmful content even when direct insults or explicit threats are not present.

Despite the progress made, several challenges remain in the application of NLP for cyberbullying detection. One major issue is the dynamic nature of social media, where language evolves rapidly, and new slang or expressions emerge frequently. Platforms like Twitter, for example, exhibit continuous changes in how users communicate, making it difficult

for detection systems to keep pace with the ever-changing linguistic patterns (Rishi et al., 2024). This poses a challenge for models that require constant retraining to remain effective.

Additionally, the anonymity of online interactions adds another layer of complexity to the problem. Many perpetrators of cyberbullying use pseudonyms or anonymous accounts to evade detection, making it harder to track and mitigate harmful behaviors. As such, future research must focus on developing more robust models that can adapt to these evolving digital environments and identify new forms of cyberbullying effectively. Researchers must also explore cross-platform detection to ensure that bullying behaviors are identified consistently, regardless of the medium. In conclusion, while NLP has made significant strides in improving cyberbullying detection, the complexity of human language and the anonymity of online interactions continue to present challenges. The future of NLP in cyberbullying detection lies in creating more adaptable, context-aware models that can navigate the rapidly evolving landscape of online communication. Figure 2 shows the overview of cyberbullying classification pipeline using NLP and ML.

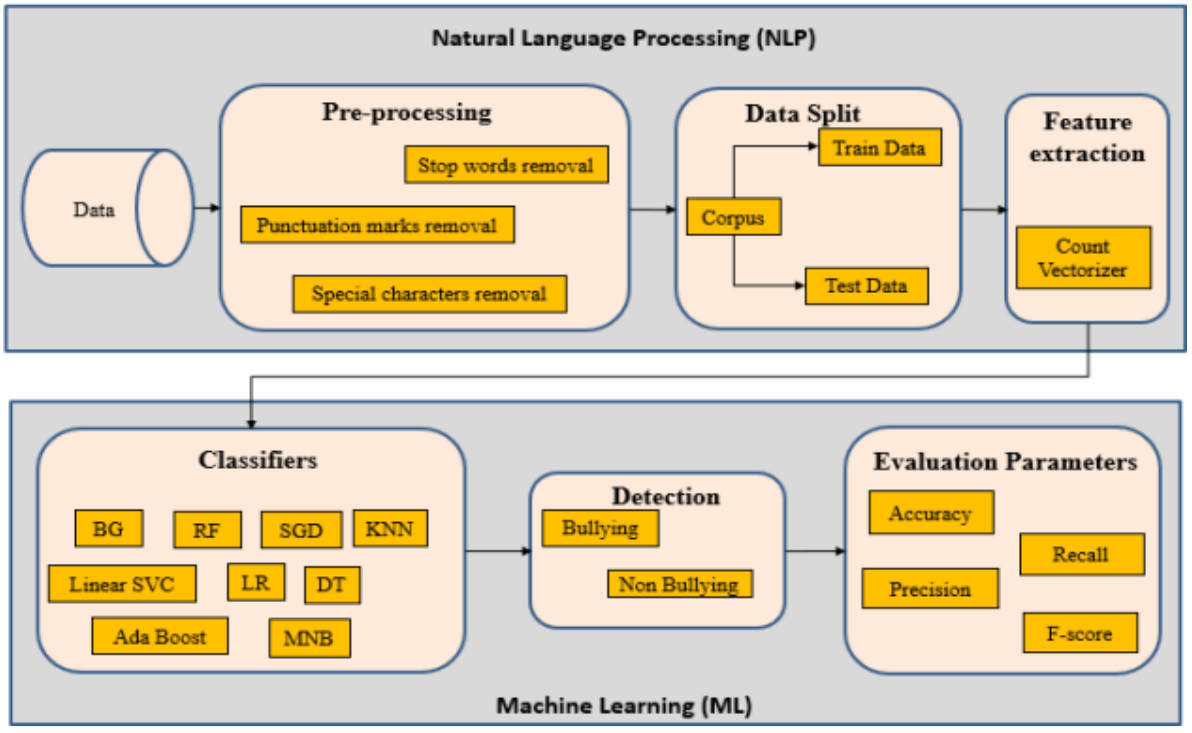

Figure 2. Overview of cyberbullying classification pipeline using NLP and ML (Sahana V et al., 2023)

3.1 Sentiment Analysis in Cyberbully Detection

Sentiment analysis is a critical component in the detection of cyberbullying, as it allows systems to assess the emotional tone of text and identify potentially harmful content. By leveraging Natural Language Processing (NLP) techniques, sentiment analysis extracts sentiments from unstructured data, a necessity in the anonymous and unregulated environment of social media platforms. The integration of sentiment analysis into cyberbullying detection systems enhances their ability to distinguish between benign and malicious interactions, thereby improving both the accuracy and efficacy of these systems. Below are key insights from the literature regarding sentiment analysis in the context of cyberbullying detection.

One widely used method for sentiment analysis is the lexicon-based approach, which involves the application of predefined lists of words associated with specific sentiments. For example, the SentiStrength method categorizes sentiments into positive, negative, or neutral and has shown high accuracy in detecting cyberbullying content on platforms like Instagram, with an accuracy rate of 93.85% (Yunior et al., 2024). This method offers a robust way to identify the emotional intensity of text, which is crucial for detecting both explicit and implicit forms of bullying.

Another effective technique is the Adaptive External Dictionaries (AED) approach. This involves creating dynamic and adaptable dictionaries that list words with positive or negative sentiments and refining machine learning models to improve their sensitivity to the evolving nature of language on social media. The AED method has been particularly effective in detecting cyberbullying on platforms like Twitter, as it allows for continuous updates and adjustments to the lexicon based on real-time trends (Hamzeh Jehad et al., 2023). Figure 3 represents the sentiment analysis technique.

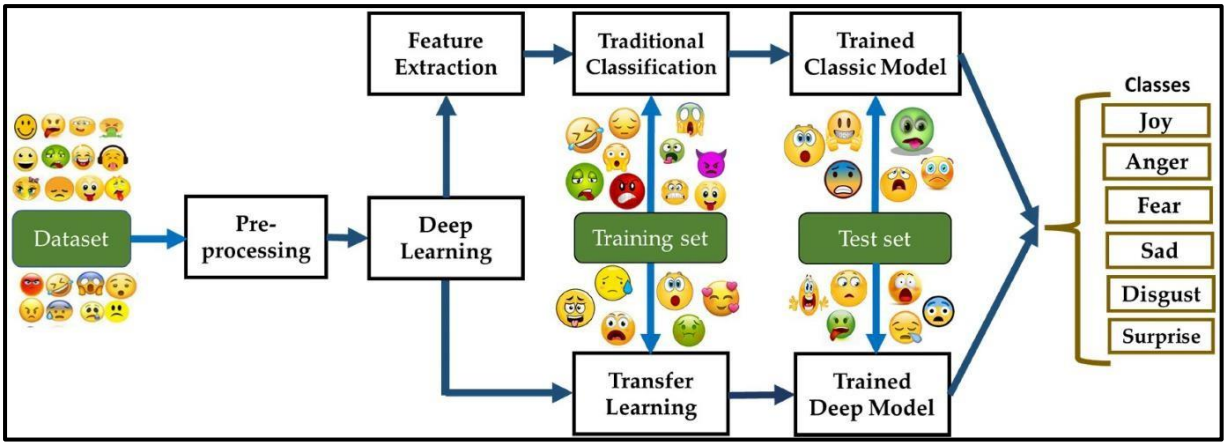

Figure 3. Sentiment Analysis Technique (Lanz, 2024)

3.2 Real-Time Detection and Prevention Systems

Real-time detection and blocking systems play a vital role in a variety of domains, including cybersecurity, financial risk management, fire safety, and network security. These systems rely heavily on technologies such as machine learning (ML), artificial intelligence (AI), and the Internet of Things (IoT) to provide immediate responses to detected threats or anomalies. The integration of these advanced technologies enhances the systems' ability to identify and mitigate risks in real-time, improving both detection accuracy and system responsiveness. Below, key insights from the literature on real-time detection systems across multiple domains are discussed, highlighting their relevance to cyberbullying detection.

3.2.1 Real-Time Cyberbullying Detection and Prevention

In the context of cyberbullying detection, real-time systems are essential for promptly identifying harmful interactions on social media platforms. Machine learning algorithms are often employed to continuously monitor and analyze user interactions, automatically flagging potentially abusive or harmful content. This approach allows for swift interventions, such as warnings or content removal, minimizing the emotional distress caused by cyberbullying. By integrating AI-based systems with existing social media platforms, real-time detection systems can reduce the spread of harmful content and provide a safer online environment (Gunavathie et al., 2023).

Similar to other real-time detection systems, cyberbullying prevention systems must handle large volumes of data generated by online interactions, which presents challenges in terms of scalability and false positives. As seen in other domains such as intrusion detection systems (IDPS) for next-generation networks like 5G, these systems must operate efficiently under high data rates and complex user behavior patterns (Gunavathie et al., 2023). In the case of cyberbullying, reducing false positives is critical, as overly sensitive systems may flag non-harmful content, potentially violating freedom of expression.

3.2.2 Technological Integration in Real-Time Detection Systems

In cybersecurity, real-time detection systems like intrusion detection and prevention systems (IDPS) leverage machine learning and AI to identify unauthorized access or attacks in network traffic (Mönch & Roth, 2023). Similarly, real-time systems for detecting cyberbullying use machine learning to analyze text, detecting toxic language, aggression, or threatening behavior as it happens. The ability to integrate real-time blocking where harmful messages or interactions are blocked immediately adds an additional layer of protection, ensuring that harmful content does not spread further.

In financial systems, real-time financial monitoring systems, which use AI and machine learning to detect fraud and assess credit risk, have proven effective in improving response times and reducing risks (Abikoye et al., 2024). Similarly, by utilizing AI-powered detection in cyberbullying, harmful online interactions can be flagged and dealt with before they escalate. These models often improve over time, learning from previous data to better detect subtle forms of abuse that might not have been flagged in earlier iterations

4. Research Methodology

System development methodologies are frameworks that guide the entire process of creating, designing, implementing, and testing research. These methodologies play a crucial role in ensuring that research goals are achieved efficiently while addressing constraints such as time, resources, and user expectations. By providing structured processes, methodologies help in breaking down complex research into manageable phases, improving collaboration, reducing risks, and ensuring the delivery of high-quality outcomes.

For this research, which involves developing an AI-powered cyberbullying detection and prevention system, the choice of methodology is particularly significant. The nature of the research demands a flexible and iterative approach to accommodate constant refinements. Features such as real-time harmful message detection, parental notifications, and user-friendly interfaces require regular feedback from stakeholders like parents, educators, and young users. Moreover, integrating machine learning models, like Natural Language Processing (NLP), further emphasizes the need for adaptability to improve accuracy and performance through testing and evaluation.

4.1 Agile Methodology

Agile methodology is the ideal choice for the development of an AI-powered cyberbullying detection system due to its flexibility, focus on continuous improvement, and its ability to foster ongoing collaboration with stakeholders throughout the research lifecycle. These core features of Agile make it particularly suited for dynamic research that require frequent iteration and adaptation to meet the needs of users and respond to emerging challenges. The iterative nature of Agile ensures that the system is constantly evolving and improving. In the case of a cyberbullying detection system, this is particularly important as online behaviours and language patterns continuously evolve. Agile allows for frequent updates and refinements to the system's algorithms, which is critical when working with AI, ensuring that the system can adapt to new trends in harmful content and improve detection accuracy over time. This iterative development process also helps mitigate the risk of major issues arising later in the research, as feedback is integrated at each stage, leading to early identification and resolution of problems.

Moreover, stakeholder involvement is central to Agile's success, and for a system designed to protect young users from harmful content, constant feedback from parents, educators, and the users themselves is invaluable. Agile facilitates direct and regular communication with stakeholders, ensuring that the system aligns with their expectations and remains user-centric throughout development. By allowing for real-time feedback during each sprint, Agile ensures that the system evolves in a way that reflects actual user needs, enhancing the likelihood of creating a tool that is both effective and user-friendly.

The flexibility inherent in Agile is also crucial in an AI research. As machine learning models and algorithms improve, the system needs to be adaptable to new insights and data. Agile's adaptability ensures that the development team can pivot or adjust the system's features as required, based on real-world data and stakeholder feedback. This adaptability is essential for keeping the system relevant and responsive to both the technological landscape and the shifting needs of users. In addition to these advantages, Agile promotes faster delivery of working software. Instead of waiting until the end of a long development cycle, Agile delivers incremental functional components, allowing users and stakeholders to access a working version of the system earlier and regularly. This ensures that by the end of the research, the product is already aligned with user needs and business goals, as it has been continuously tested and refined in real-world conditions. In conclusion, Agile methodology's focus on collaboration, flexibility, rapid delivery, and continuous feedback makes it the best choice for developing a complex, evolving system like AI-powered cyberbullying detection. By embracing Agile, the research will benefit from an approach that ensures the system remains effective, adaptable, and closely aligned with user expectations, helping to create a solution that not only meets current needs but can also evolve to address new challenges in the future. Figure 4 represents the Agile methodology.

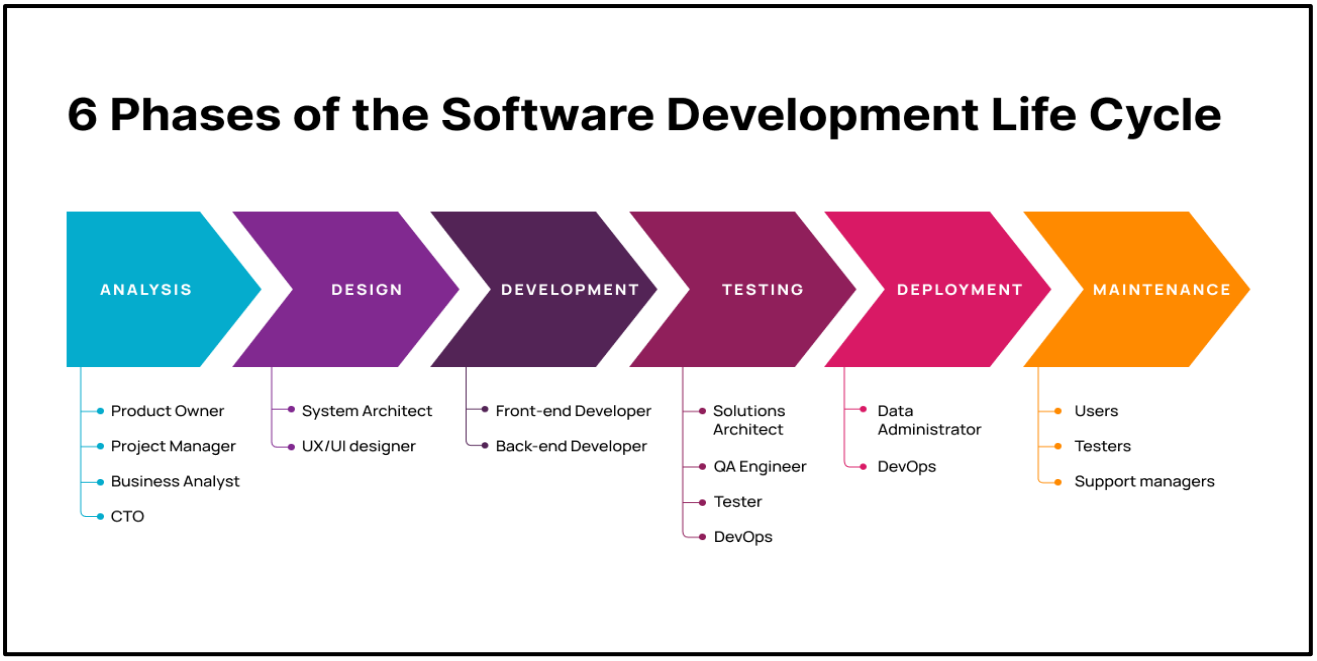

Figure 4. Agile Methodology.

4.2 System Requirement Analysis

The proposed system's technical requirements for building a prototype are designed to balance functionality with simplicity. While the final implementation of the system will necessitate more complex infrastructure and software components, the prototype aims to demonstrate the core functionalities of cyberbullying detection and real-time intervention. The following outlines the necessary hardware and software components needed for the prototype's development, ensuring that the system is both efficient and feasible to construct within the scope of early-stage testing. Table 1 summarizes the proposed system's technical requirements.

Table 1. The proposed system's technical requirements.

Hardware Requirement		
Component	**Requirement**	**Justification**
Local Development Server	Personal computer	A personal machine sufficient for real-time text analysis and alert generation.
User Devices	Smart Phone, Tablet	Smartphone or tablet devices are needed for user interactions and monitoring.
Network Infrastructure	Reliable broadband connection.	A fast internet connection is necessary for real-time communication and updates.

Software Requirements		
Component	**Requirement**	**Justification**
Operating System	Cross-platform support for mobile (Android, iOS) and web-based platforms.	The system needs to work on both mobile platforms and web interfaces.
Programming Language	Python 3.x, with NLP and machine learning libraries (spaCy, Hugging Face Transformers, TensorFlow/PyTorch).	Python is ideal for implementing NLP, machine learning, and backend logic.

4.3 Data Collection

To ensure that the AI-powered cyberbullying detection system is effective and aligns with the needs of its users, it is essential to gather relevant insights from individuals who interact with online platforms regularly. In this study, data will be collected using two main techniques: surveys and interviews. These methods have been selected to provide both a broad overview of general trends and in-depth perspectives from key stakeholders. Surveys will capture quantitative data from a larger sample, enabling the identification of common patterns and preferences among users, while interviews will allow for a more nuanced exploration of the experiences and concerns of educators and parents. Together, these data-gathering techniques will inform the design and development of the proposed system, ensuring that it addresses real-world challenges effectively and provides a user-friendly solution for all involved parties. The bellow sections will describe in detail about each chosen method.

4.3.1 Surveys

Surveys are a commonly used method for collecting data from a large group of participants in a structured and systematic manner. This technique allows researchers to gather quantitative data that can be analyzed for patterns and trends, as well as qualitative insights through open-ended questions. In this research, surveys will be used to collect data from a minimum of 30 participants, including parents, educators, and online users. The primary aim is to understand the experiences and perceptions of individuals regarding cyberbullying, the effectiveness of current prevention methods, and their preferences for an AI-powered detection and prevention system.
Survey Structure and Content

The survey will be divided into several sections to ensure a comprehensive understanding of the issue. The first section will collect demographic information, such as age, gender, and role, to find whether the parent, educator, or student, which will help contextualize the responses. The second section will focus on participants' online behaviours, specifically the frequency of social media and messaging app usage, which will provide context for their experiences with online interactions. The third section will address experiences with cyberbullying, including whether the participants have personally been victims or witnesses of cyberbullying and the impact it had on them or others. The fourth section will assess awareness and satisfaction with existing tools and systems designed to prevent cyberbullying, allowing the researcher to gauge the effectiveness of current solutions. Lastly, the survey will explore participants' preferences for the proposed AI-powered cyberbullying detection system, including desired features such as real-time alerts, blocking harmful messages, and providing notifications to guardians.

The survey will consist exclusively of closed-ended questions, which will provide structured responses that can be easily quantified and analyzed. The questions will be designed to assess various aspects, such as participants' online behavior, their awareness and experiences with cyberbullying, and their views on existing solutions. For example, participants will be asked to rate the perceived safety of online platforms, the frequency of encountering harmful content, and their satisfaction with current tools aimed at preventing cyberbullying. Additionally, the survey will explore preferences for

key features of the proposed system, such as real-time detection of harmful messages and the ability to block them before they are sent.

4.3.2 Interviews

Interviews are a valuable qualitative research method used to gain in-depth insights from key individuals who have direct experience with the subject matter. In this research, interviews will be conducted with educators to gather detailed perspectives on the challenges they face in addressing cyberbullying, their thoughts on existing detection and prevention methods, and their opinions on the proposed AI-powered detection system. Educators play a critical role in identifying and managing instances of cyberbullying in school environments, making their input essential for the development of an effective system that can be practically implemented in educational settings.

5. Analysis and Results

5.1 Survey Results

This section presents a detailed analysis of the survey responses gathered for the research titled "AI-Powered Cyberbullying Detection and Prevention for Safer Digital Spaces." The purpose of this analysis is to extract insights from participant feedback to understand the prevalence, impact, and user perception of cyberbullying, as well as the effectiveness of existing preventive measures. These insights will inform the development of an AI-powered system aimed at addressing the limitations of current approaches to cyberbullying prevention. The survey gathered responses from individuals between the ages of 18-45, focusing on their experiences and perceptions of online safety, cyberbullying, and the use of preventive tools. Key areas explored include the frequency of social media use, exposure to harmful content, emotional impact, and the perceived effectiveness of existing reporting mechanisms. By analysing these responses, this research aims to identify critical gaps in current solutions and justify the need for an AI-based proactive system that enhances online safety and reduces the emotional and psychological toll of cyberbullying. The insights derived from the survey will be used to refine the objectives of the research, ensuring that the proposed system meets the needs of the target users effectively. Furthermore, this analysis will guide the development of specific features that align with the concerns and expectations expressed by participants, ultimately contributing to a safer digital environment for all users.

When asked about the platforms they use frequently, participants indicated a preference for social media in Figure 5 (e.g., Instagram, Facebook, TikTok) and messaging apps (e.g., WhatsApp, Discord, Teams). Social media was mentioned by 85.3% of respondents, while messaging apps were used by 94.1%, and online gaming platforms were mentioned by 20.6%. These results suggest that the proposed system should be compatible with a range of popular platforms where users are most active. By focusing on integrating with social media and messaging apps, the system can effectively monitor and detect harmful interactions where they are most likely to occur. Additionally, the lower use of online gaming platforms indicates that, while still important, this may be a secondary focus for initial development efforts.

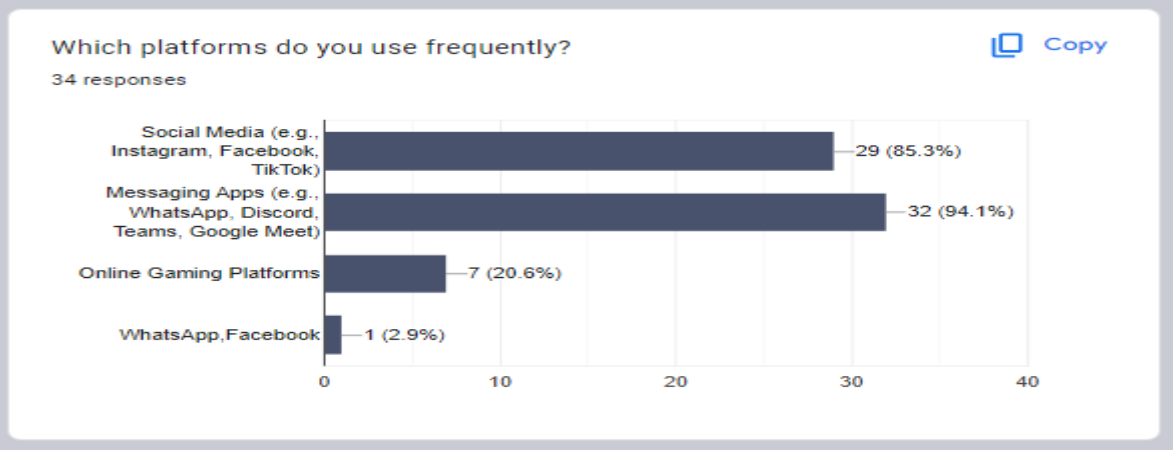

Figure 5. Use frequently results.

In Figure 6, Participants were also asked to rate how safe they feel while interacting online, on a scale of 1 to 5. The majority of respondents (64.7%) rated their feelings of safety as moderate (3 out of 5), with smaller percentages feeling either very safe or unsafe. This data suggests that there is a general sense of caution or concern when interacting online, which underlines the need for effective prevention tools that can help users feel more secure. The proposed system should aim to enhance users' sense of safety by providing clear notifications and proactive interventions. By empowering users with the knowledge that harmful content can be detected and prevented before it reaches them, the system can contribute to a more positive online environment.

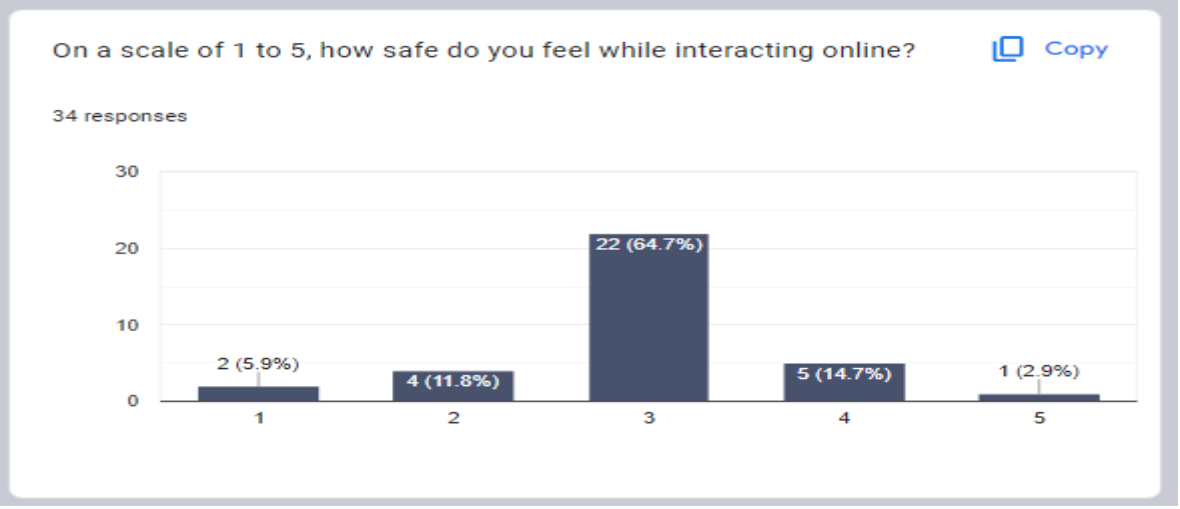

Figure 6. Interacting online results.

The survey also explored in Figure 7 the impact of cyberbullying on victims. Respondents who indicated that they or someone they know had been a victim of cyberbullying were asked to describe the effects experienced. The results showed that the most common impacts were mental and emotional, with 47.1% of respondents mentioning stress, anxiety, depression, or other mental health challenges. Additionally, 41.2% of respondents highlighted emotional impacts such as loss of confidence and fear, while 32.4% reported physical effects like sleep disturbances and fatigue. These findings underscore the significant toll that cyberbullying takes on individuals, affecting not only their mental health but also their emotional well-being and physical state. The proposed AI-powered system must address these issues by intervening before cyberbullying escalates, thereby preventing the negative consequences associated with prolonged exposure to harmful behaviour. The system should also include resources to help victims recover from these impacts, such as educational materials on mental health and emotional support, as well as connections to professional counselling services where necessary.

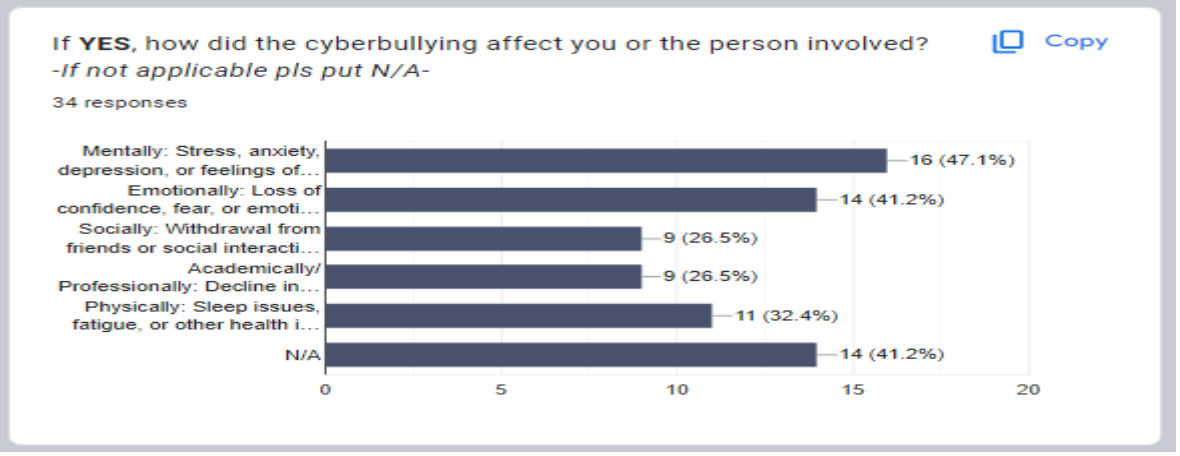

Figure 7. Cyberbullying affect results.

The survey responses in Figure 8 reveal that the most commonly witnessed or experienced forms of text-based cyberbullying are hate speech (67.6%), insulting messages (64.7%), and mockery (64.7%). These are followed by rumour spreading (52.9%), gossip sharing (50%), and threats (47.1%). Less frequently reported forms include sexual harassment (44.1%) and persistent messaging or spamming (20.6%). These results indicate that while overtly aggressive behaviours like hate speech and insults are predominant, subtler forms of bullying such as gossip and persistent messaging also significantly affect users. This data highlights the need for the proposed cyberbullying detection system to handle a diverse range of harmful behaviours, from explicit threats to more nuanced forms like sarcasm or rumour spreading. Incorporating advanced natural language processing (NLP) models capable of detecting both direct and indirect bullying tactics will ensure the system effectively addresses the wide spectrum of cyberbullying behaviours identified by the respondents.

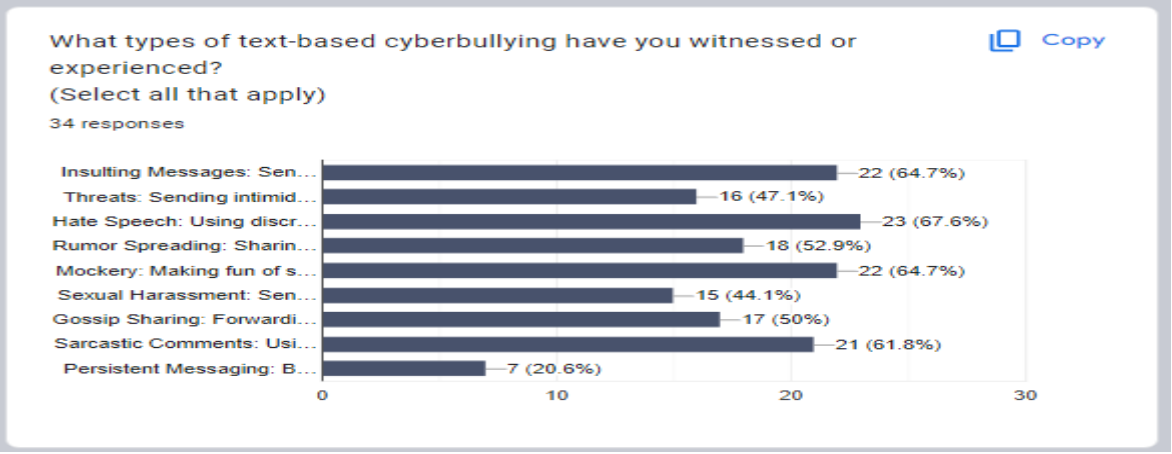

Figure 8. Types of text-based Cyberbullying results.

The survey responses reveal in Figure 9 that 58.8% of participants are not aware of any tools or systems designed to prevent cyberbullying, while 41.2% are aware. This indicates a significant lack of awareness among users about existing solutions, which may contribute to underreporting and the persistence of cyberbullying incidents. These findings emphasize the importance of raising awareness about available resources and developing systems that are not only effective but also well-promoted and accessible. For the proposed cyberbullying detection and prevention system, an emphasis on outreach, education, and visibility will be crucial to ensuring widespread adoption and usage. Additionally, embedding user-friendly features and educational resources directly within the system can help bridge the awareness gap and empower users to take proactive measures against cyberbullying.

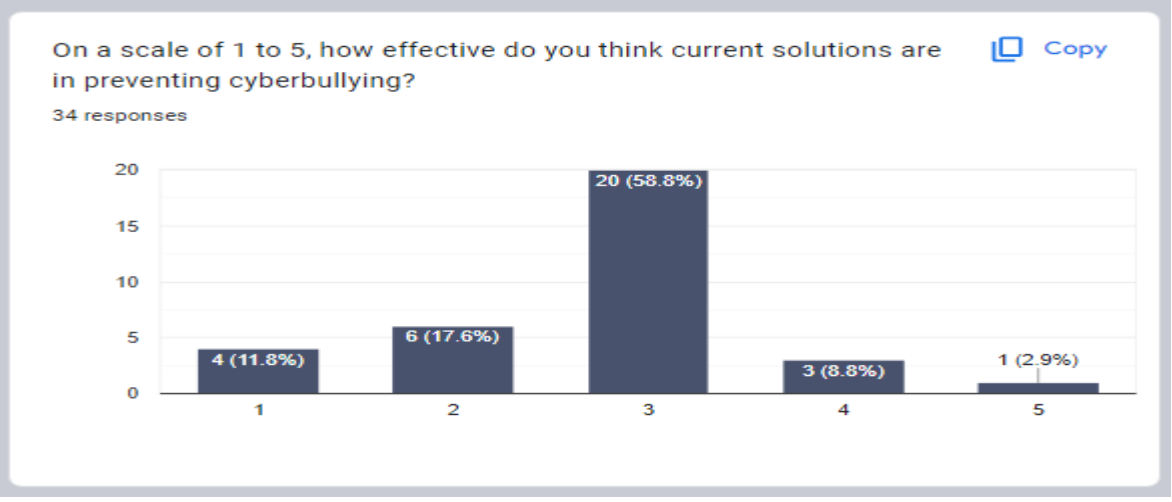

Figure 9. Preventing Cyberbullying results.

The survey responses in Figure 10 reveal that users place a high priority on features that enhance the effectiveness and accessibility of a cyberbullying detection system. Real-time detection of harmful messages was rated as very important by the majority, reflecting the need for immediate identification of harmful content to prevent its impact. Similarly, the ability to block harmful messages before they are sent was widely emphasized, highlighting a preference for proactive measures that stop cyberbullying at its source. Other key features include notifications to parents, guardians, or educators, which underscores the importance of involving external stakeholders to address incidents promptly. Additionally, reporting functionality for serious incidents is considered crucial for accountability and ensuring that severe cases are addressed effectively. Lastly, a user-friendly interface was universally deemed critical, indicating the need for an intuitive design that can be easily adopted by diverse users. These insights suggest that the proposed system must focus on preventive, collaborative, and accessible features to effectively meet user needs.

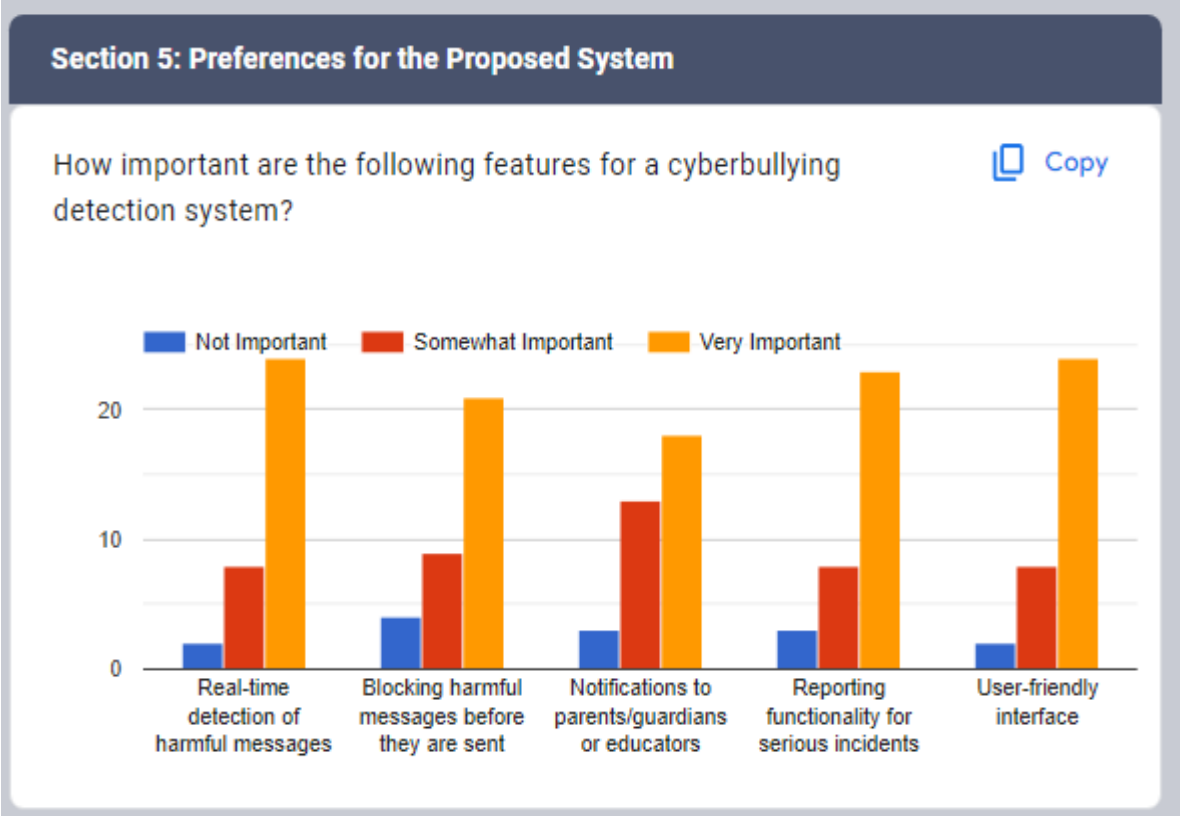

Figure 10. A high priority on features that enhance the effectiveness and accessibility of a cyberbullying detection system.

The survey in Figure 11 responses highlight significant privacy concerns among participants who expressed apprehensions about data security. The most commonly cited issue, at 67.6%, is unauthorized access to sensitive information, reflecting fears of data breaches or malicious misuse. This is closely followed by data misuse (64.7%), where users are concerned about their personal information being repurposed for unintended activities such as marketing or surveillance. Additionally, 55.9% of respondents expressed concerns about the lack of transparency in data handling, underscoring the need for clear and open communication regarding how user data is collected, processed, and stored.
These findings point to the importance of incorporating robust privacy measures in the proposed system. Strategies such as implementing strong encryption protocols, limiting data access through strict controls, and providing clear, user-friendly privacy policies can address these concerns. Moreover, giving users greater control over their data, such as options to opt in, review, or delete their information, can enhance trust and ensure compliance with ethical and legal standards. These measures will be critical to building a system that users feel confident adopting.

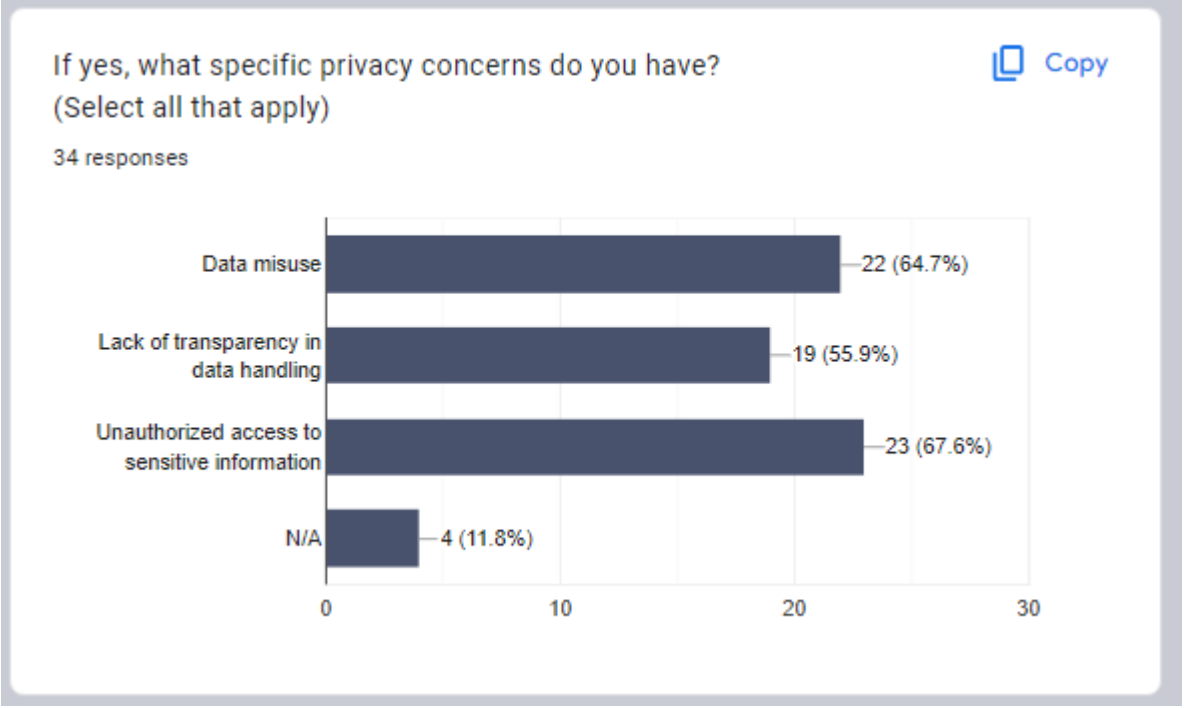

Figure 11. Significant privacy concerns among participants.

6. Conclusion

The first part of the research focused on laying the groundwork for the development of an AI-powered system to detect and prevent cyberbullying. The objective was to thoroughly investigate the issue of cyberbullying, explore existing solutions, and establish a solid foundation for the design and implementation of the proposed detection system. A significant achievement of this phase was the successful completion of the literature review, which provided a comprehensive understanding of the problem of cyberbullying, its psychological effects, and current prevention and detection methods. The research identified the various forms of cyberbullying, such as harassment, impersonation, and exclusion, and how these behaviours manifest across different digital platforms, including social media, messaging apps, and gaming forums. By exploring these aspects in-depth, the review laid the groundwork for the development of an AI-based solution that can address the complex and evolving nature of cyberbullying.

Additionally, this phase saw the selection of an appropriate system development methodology. The decision to adopt an iterative and flexible methodology, such as Agile, was driven by the need for continuous feedback and refinement, especially given the dynamic nature of both cyberbullying tactics and AI technology. This methodology aligns with the objectives of the research, ensuring that the system design can be improved and adjusted throughout its development based on testing and user input.

The research on AI algorithms used for cyberbullying detection was another key accomplishment. Through the investigation of existing AI-powered tools and models, insights were gained into the most effective techniques for detecting harmful online behaviour. These findings helped inform the design of the detection system, specifically the features that would need to be integrated, such as real-time message monitoring, context understanding, and classification of harmful content.

Moreover, the identification of key features for the AI system, such as real-time alerts, a user-friendly interface, and integration with existing social media platforms, was a significant achievement in this phase. These features were defined based on the needs of the target audience such as parents, educators, and system administrators and have laid the

foundation for the next phase of the research, which will focus on the actual system design and development. In conclusion, the first part of the research has achieved its goal of thoroughly investigating the problem of cyberbullying and preparing the necessary groundwork for the development of an AI-powered detection system. The research, methodology selection, and feature identification have collectively set the stage for the successful implementation of the research in subsequent phases.

6.1 Gaps and Areas for Further Exploration

While significant progress has been made in the first part of the research, there are several gaps in both the research and design areas that need to be addressed in the next phase of the research. These gaps present opportunities for further exploration and refinement, which will contribute to the overall success of the system.

One of the main gaps identified during the research phase is the limited analysis of different AI algorithms for detecting cyberbullying across diverse digital platforms. While initial research pointed to the use of natural language processing (NLP) and machine learning models for content classification, further investigation is needed to compare the effectiveness of various algorithms in accurately identifying harmful behaviors in different contexts, such as social media, messaging apps, and online gaming forums. Exploring additional models, such as deep learning techniques and reinforcement learning, could help improve the accuracy and robustness of the detection system.

Another important gap is the ethical considerations surrounding the use of AI in monitoring and detecting cyberbullying. Although the research has touched on some ethical issues, such as the importance of data privacy and user consent, these aspects require deeper exploration. The implementation of AI-powered monitoring systems raises questions about user surveillance, data protection, and the potential for algorithmic bias. Further research is needed to ensure that the system is ethically sound and that safeguards are in place to protect the privacy and rights of users. Additionally, investigating how the system can be made transparent and accountable to users will be an important area of focus.
In terms of system design and usability, while key features have been identified, there is still much to be done to ensure that the AI detection system is intuitive and accessible to non-technical users. For instance, the design of the user interface (UI) and user experience (UX) must be refined to ensure ease of use, especially for parents, educators, and administrators who may not have technical expertise. Further user testing and feedback will be crucial to improve the UI/UX, ensuring that the system is both functional and user-friendly.

Additionally, scalability and real-time performance are critical areas that need further investigation. As the research moves into the development phase, it will be important to ensure that the system can handle large volumes of data from various social media platforms and process it in real-time. Research into optimizing the system for performance, speed, and resource efficiency will be essential for the successful implementation of the research.
Finally, there is a need for a more comprehensive evaluation and testing of the system. Although the literature review has identified key factors to consider in the detection of cyberbullying, the actual performance of the AI system needs to be rigorously tested in real-world scenarios. This will include validating the system’s ability to accurately detect cyberbullying, assessing its effectiveness in different cultural contexts, and evaluating the overall impact of the system on reducing cyberbullying.

In summary, while the first part of the research has provided a solid foundation, there are several key gaps that need to be addressed. These include further research into AI algorithms, deeper exploration of ethical concerns, refinement of the system's design and usability, optimization for real-time performance, and comprehensive testing. Addressing these gaps in the next phase will be critical to the successful completion of the research and the development of an effective cyberbullying detection system.

Acknowledgements
NA.

Funding
No funding.

Contributions
All authors; Conceptualization, All authors; Investigation; All authors; Writing (Original Draft), All authors; Writing (Review and Editing) Supervision, All authors; Project Administration.

Data Availability Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Use of Generative Artificial Intelligence
The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript.

Competing interests
All authors declare no competing interests.

References

[1] Băroiu, A.-C., & Trăușan-Matu, Ș. (2022). Automatic sarcasm detection: Systematic literature review. *Information*, *13*(8), 399. https://doi.org/10.3390/info13080399
[2] Thesing, T., Feldmann, C., & Burchardt, M. (2021). Agile versus waterfall project management: decision model for selecting the appropriate approach to a project. *Procedia computer science*, *181*, 746-756.
[3] Balakrishnan, V. (2015). Cyberbullying among young adults in Malaysia: The roles of gender, age and Internet frequency. *Computers in Human Behavior*, *46*, 149–157. https://doi.org/10.1016/j.chb.2015.01.021
[4] Balakrishnan, V., & Fernandez, T. (2018). Self-esteem, empathy and their impacts on cyberbullying among young adults. *Telematics and Informatics*, *35*(7), 2028–2037. https://doi.org/10.1016/j.tele.2018.07.006
[5] Bull Stop Inc. (2024, October 27). *BullStop App*. Google Play. https://play.google.com/store/apps/details?id=com.bullstop.bullstopapp
[6] Coram, M., & Bohner, S. (2005). The impact of agile methods on software research management. In *Proceedings of the 13th IEEE International Conference on Engineering of Computer-Based Systems* (pp. 363–370). IEEE. https://doi.org/10.1109/ECBS.2005.68
[7] Gamal, D., Alfonse, M., Jiménez-Zafra, S. M., & Aref, M. (2023). Intelligent multilingual cyber-hate detection in online social networks: Taxonomy, approaches, datasets, and open challenges. *Big Data and Cognitive Computing*, *7*(2), 58. https://doi.org/10.3390/bdcc7020058
[8] Balakrishnan, V. (2015). Cyberbullying among young adults in Malaysia: The roles of gender, age and Internet frequency. *Computers in human behavior*, *46*, 149-157.
[9] Aljohani, E. J., Yafooz, W. M. S., & Alsaeedi, A. (2023). Cyberbullying detection approaches: A review. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1310–1316). IEEE. https://doi.org/10.1109/ICIRCA57980.2023.10220688
[10] Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). Sarcasm identification in textual data: Systematic review, research challenges and open directions. *Artificial Intelligence Review*, *53*(6), 4215–4258. https://doi.org/10.1007/s10462-019-09791-8
[11] Samsudin, E. Z., Yaacob, S. S., Wee, C. X., Mat, N., Azzani, M., Jamil, A. T., Muzaini, K., Ibrahim, K., Suddin, L. S., Selamat, M. I., Saman, A., Abdullah, N. N., Ismail, N., Yasin, S. M., Azhar, Z. I., Ismail, Z., Isa, M. R., & Mohamad, M. (2023). Prevalence of cyberbullying victimisation and its association with family dysfunction, health behaviour and psychological distress among young adults in urban Selangor, Malaysia: A cross-sectional study. *BMJ Open*, *13*(11), e072801. https://doi.org/10.1136/bmjopen-2023-072801
[12] Chbib, F., Sujud, R., Khatoun, R., & Fahs, W. (2024). Cyberbullying detection using Bidirectional Encoder Representations from Transformers (BERT). In *2024 Mediterranean Conference on Telecommunications and Networking (MeditCom)* (pp. 257–262). IEEE. https://doi.org/10.1109/MEDITCOM61057.2024.10621093
[13] Gunavathie, M. A., Sneha, P. D., Yuvarani, K., & Swathyshree, P. (2023). *An exploration of real-time intrusion detection and prevention systems for next generation networks* (pp. 1–5). IEEE. https://doi.org/10.1109/WCONF58270.2023.10235147
[14] Balakrishnan, V., Khan, S., & Arabnia, H. R. (2020). Improving cyberbullying detection using Twitter users' psychological features and machine learning. *Computers & Security*, *90*, 101710.
[15] Berger, H., & Beynon-Davies, P. (2009). The utility of rapid application development in large-scale, complex projects. *Information Systems Journal*, *19*(6), 549-570.
[16] Lee, M. H. L., Kaur, M., Shaker, V., Yee, A., Sham, R., & Siau, C. S. (2023). Cyberbullying, social media addiction and associations with depression, anxiety, and stress among medical students in Malaysia. *International journal of environmental research and public health*, *20*(4), 3136.
[17] Ambareen, K., & Sundaram, S. M. (2023). A survey of cyberbullying detection and performance: Its impact in social media using artificial intelligence. *SN Computer Science*, *4*(6). https://doi.org/10.1007/s42979-023-02301-2

[18] Kulkarni, S., & Biswas, A. (2020). Sarcasm detection methods in deep learning: Literature review. In *Lecture Notes in Networks and Systems* (pp. 507–512). Springer. https://doi.org/10.1007/978-981-15-0630-7_50
[19] Kumar, Y., Huang, K., Perez, A., Yang, G., Li, J. J., Morreale, P., Kruger, D., & Jiang, R. (2024). Bias and cyberbullying detection and data generation using transformer artificial intelligence models and top large language models. *Electronics*, *13*(17), 3431. https://doi.org/10.3390/electronics13173431
[20] Misra, R., & Arora, P. (2023). Sarcasm detection using news headlines dataset. *AI Open*, *4*, 13-18.
[21] Dingsøyr, T., Nerur, S., Balijepally, V., & Moe, N. B. (2012). A decade of agile methodologies: Towards explaining agile software development. *Journal of systems and software*, *85*(6), 1213-1221.
[22] Milosevic, T., Royen, K. V., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: New directions in the midst of complexity. *International Journal of Bullying Prevention*, *4*(1), 1–5. https://doi.org/10.1007/s42380-022-00117-x
[23] Mishra, A., Sinha, S., & George, C. P. (2024). Shielding against online harm: A survey on text analysis to prevent cyberbullying. *Engineering Applications of Artificial Intelligence*, *133*, 108241. https://doi.org/10.1016/j.engappai.2024.108241
[24] Mönch, S., & Roth, H. (2023). *Real-time APT detection technologies: A literature review* (pp. 136–141). IEEE. https://doi.org/10.1109/CSR57506.2023.10224983
[25] Mui, D., Anwar, A., & Vranješ, I. (2023). Cyberbullying victimization and suicide ideation: The mediating role of psychological distress among Malaysian youth. *Computers in Human Behavior*, *108000*. https://doi.org/10.1016/j.chb.2023.108000