IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DIGITAL SHIELD: AN AI-BASED PLATFORM FOR ONLINE SAFETY AND HARASSMENT REPORTING

[1] Vangari Kiran Kumar,[2] Artham Prashanthi,
[1,2] Assistant Professor,
[1,2] Department of CSE(AI&ML),
[1,2] CMR Technical Campus, Hyderabad, Telangana, India.

Abstract: One big issue today? Individuals who are abused and threatened online. Currently, aid tools are either disjointed, difficult to use, or sluggish. Enter Digital Shield an artificial intelligence based smart website that helps in fighting back. It is expected to educate, allow the users to report safely, as well as provide live chat where necessary. The more people raise their about their safety, the more they will be secure - this tool will help people speak easier. It directs more desirable decisions over the internet instead of disregarding injuries. Subsequent updates would be able to follow trends, speak various languages, even do it on phones.

Index Terms- Digital Safety , Online Harassment , AI Chatbot , Firebase, Cybersecurity, Web Application

I. INTRODUCTION

The rapid changes in internet technology nowadays are speedy. The use of social networks is becoming increasingly prevalent. Due to that, the manner in which individuals speak and exchange information has also evolved. Nevertheless, despite such advantages, such sites, such as Internet harassment, accompany him just as much. Cyberbullying is coming very often. Stalking occurs using screens more than in the past. Online abuse infiltrates at the low level. In the recent times, these ails are visible everywhere. They're seen as real dangers. It affects mental well-being negatively. Online safety does not seem as guaranteed to the user.

People tend to remain quiet even where a tool exists to indicate or prevent harmful posts in the internet. The fear of being exposed retards them. Anonymity feels uncertain. Most of them just do not understand the way the system works. Existing practices are diffuse and sluggish and one step behind. They do not prevent some harm before it occurs but wait to see what happens. Help during urgent moments?Rare. Live assistance is missing. Emotional care in crisis? Almost gone.

Recently, the literature indicates an increase in the use of smart digital solutions capable of managing education, avoiding and incident handling - all in a single system. Not a mere theory, artificial intelligence, as well as natural language processing, is highly promising in terms of easier interactions between individuals and machines, particularly in live scenarios, such as cyber threats. Nonetheless, the application of AI capabilities within the daily safety systems has not actually exploded yet - still in its infancy.

Here a new direction is developed. Digital Shield is an artificial intelligence-driven web tool that has been developed to enhance privacy on the internet. People can report on harassment without identifying themselves instead of remaining silent. The support is immediate with the help of an intelligent chatbot that works on the basis of intelligent algorithms. Notices about safety are placed at critical situations in

order to prevent damage before it occurs. It is supported by modern web frameworks where it can develop. The information is secured with the help of cloud protection systems built with privacy in mind.

II. RELATED WORK

Online abuse and how to be safe on the internet has grabbed a lot of attention as people have been spending more of their time on the internet. Researchers have explored the possibility of identifying offensive messages by tools such as machine learning and language detection. Instead of general methods, old systems rely more frequently based on predetermined rules combined with algorithmic models of Naive Bayes, Support Vector Machines or logic that continually builds out trees. Even if these methods work fairly well, they require tons of examples labeled by hand while struggling to adjust themselves to new forms of harassment when these emerge. Their rigidity makes them slow to catch shifts in the way in which people express hostility over the internet.

Recently research has shifted to application of deep learning and natural language processing to detect any toxic content on the internet.

Although the understanding of context is better in RNNs and transformers, they simply end up with detecting but do not assist the user immediately. Such systems perform poorly when they are put to the test in live web sites. Their cumbersome design makes everything slow down as speed is the most needed thing. Some of the existing methods of disseminating information about online safety are based on pre-stated guidelines or propositional online question and answer pages. Other users are connected to social networks by tools that allow the use of flags to identify problems. Even so, a number of these attempts are like pieces of a puzzle, there is no obvious organization that ties them together. Problems of reporting either require practical effort, and are too lengthy, and do not conceal the identity of the reporter of the issue. Once one has made a report, no one tends to offer help, be it advice or comfort, at the time when it is most needed.

Today, chatbots that can be used in various fields such as assisting customers, medical services, and so on are driven by artificial intelligence as they provide prompt responses. There are studies that examined the existence of bots that have provided assistance in the emotional health context and those that have examined how the devices can impart the online security behavior. The findings indicate that automated helpers do well in assisting individuals in the difficult subjects. Nonetheless, very few articles investigate their contribution to reporting abuse or disseminating information on how to be safe over the internet.

A major disparity with the existing tools? This is the new online arrangement whereby the reporting abuse, locked-down security, instant AI chat help, and learn-about-how-to-keep-safe lessons are merged in a single roof. It does not create its own data set, preferring to rely on ready-made AI-based language technology, updating it more quickly and expanding its scale. Tugging at fragments of what has already been put in place, the entire enterprise can be finished in one plan that is bigger than it was previously.

III. PROPOSED METHODOLOGY:

One tap provides the answers quickly with a clever software that learns whilst performing. In real time, responses are made instead of waiting to get an update. The flow of information is directly moved into the safe storage facility, hidden within the cloud database of Google. There are no scripts to follow when responding to - every action changes every time. Performance remains stable even where situation changes are fast. The security is strict with narrow access points and regular inspections. As it happens, changes occur behind the scenes without dragging things down.

The Digital Shield system is not another tool but it provides individuals that encounter damage on the Internet with an opportunity to take action without fear. One would be able to report by remaining anonymous - identity remains unknown. Information is stored safely in cloud storage with a high degree of protection in place keeping records intact. Reports are processed in a manner that there are no items that are lost or altered. In addition to the provision of details, every user is guided according to his or her circumstance. That personalized care would assist one to be more aware of risks at an earlier stage and make provision before it is too late. Safety tips do not look out as some general pieces of advice but are constructed as reactions on actual actions. The resultant effect is a space in which trust is built gradually, quietly, by steadily designing.

IV. METHODOLOGY

One line of work Digital Shield takes is a new beginning - creating security via smart web tools aimed at the victims of cyberbullying. Developed bit by bit in the cloud, it connects live interaction with behind-

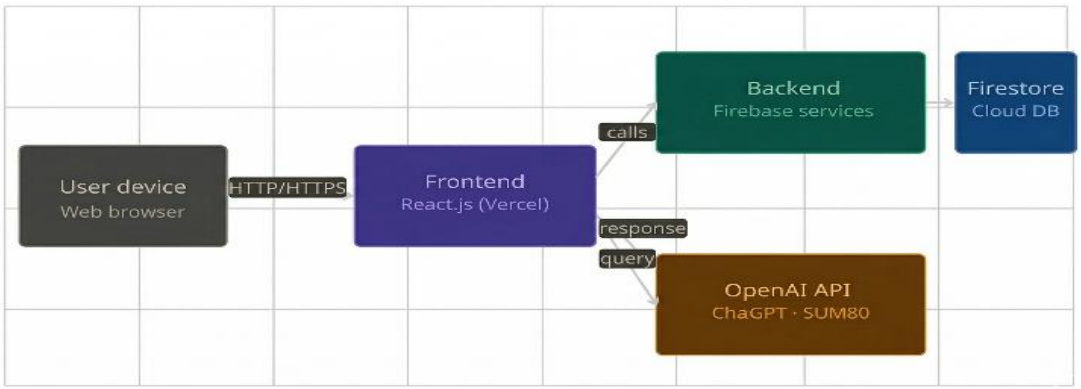
the-scene management and support and it learns on its own rather than based on predetermined information.

V. IMPLEMENTATION

The Digital Shield installation is based on existing web technologies, which contributes to the development, efficient protection, and effective implementation. The idea of development is divided into three sections - interface development, server functionality, due to the smart technology connections. In the case of the visible side, React.js creates the structure and Vite accelerated work and Tailwind CSS colored the appearance. The points of interaction are the possibility of sending reports, speaking with bots, seeing advice about how to keep safe on the Internet. Vercel Cloud pages are easy to load as they use automatic updates and are encrypted by default, which has led to high loading speeds.

The Digital Shield installation is based on existing web technologies, which contributes to the development, efficient protection, and effective implementation. The idea of development is divided into three sections - interface development, server functionality, due to the smart technology connections. In the case of the visible side, React.js creates the structure and Vite accelerated work and Tailwind CSS colored the appearance. The points of interaction are the possibility of sending reports, speaking with bots, seeing advice about how to keep safe on the Internet. Vercel Cloud pages are easy to load as they use automatic updates and are encrypted by default, which has led to high loading speeds.

The backend is built using Firebase, which consists of logging in and storing data in a cloud solution. User preports including associated information are stored in Firestore, a versatile NoSQL database, which is updated in real-time. Only authorized members can access it, and this is achieved by tight Firebase security controls. A smart chatbot is an AI chatbot developed with the OpenAI technology - ChatGPT on model SUM80. When one writes a question, the chatbot looks it up in real-time as it also provides helpful answers concerning staying online and avoiding harassment. The fact that the artificial intelligence is already trained by the already existing information will mean that there is no necessity to acquire new data or re-train the system with repetitively acquired information - this will assist in making the system more efficient without any additional measures.



Project Architecture of DIGITAL SHIELD (Protecting Women Against Online Harassment & Stalking**)**

1. AI Chatbot (SUM80) Based on ChatGPT

The conversation occurs between the user and the SUM80. After typing a question, a person is connected via a secure API connection directly to ChatGPT. It is the system that laps into the asked query, and based on the sense and background, it picks at the question before responding in a manner that fits. Nothing is withheld, no human kind has to aid. Even though older chatbots rely on predetermined rules or examples they have learned, this one ignores them altogether - there is no local instruction, no strict scripts here. Since it is powered by the existing capabilities of ChatGPT, which is made up of the large volumes of written text, the chatbot within SUM80 is capable of responding to a wide range of questions regarding how to stay safe on the internet. The questions about cyber threats, how to dodge abuse, or the perception of digital risk are answered in a straightforward manner courtesy of the same. The tool is always available when needed and it provides credible assistance based on question and answer conversations. Being a background, everything remains manageable despite the increased demand - no additional inconvenience, consistent level of delivery. Digital Shield relies on SUM80 to have a sense of protection that is personal, fast responsive, intelligent, but not flaunt-profanity.

2. Frontend Implementation

The HTTPS establishes a secure connection between the front end and the back end. It is run on the Vercel Cloud, which implies a quick loading time, a consistent uptime, since it does the deployment automatically. The advances are seamless and go round without any hiccups. Simple modules will take users through activities avoiding complicated processes and baffling layouts. Checks on the client-side prevent errors prior to their outward progression.

3. Backend Implementation

Harassment reports and additional information - including dates and names of the people were involved - are stored in Firebase Firestore, a NoSQL cloud store. Since it synchronizes data in real time, everyone will be able to access the correct pieces when they are needed without having to write much code on the backend. In that manner, changing user requirements do not slow down the things. No need to work with databases manually or processes of offline records. New supplies are received and this is stored in safe places. As individuals grow, the arrangement automatically increases to accommodate them.

4. Database Implementation

Firestore receives user reports in structured files. Each of them includes such information as the written content, the date of its arrival, as well as automatically added background tags. Different types of info are welcome as they are accommodated with the help of documents rather than a predetermined table. This will give it much ease in dealing with unpredictable submissions. It does not await batch updates, as well as, use of old snapshots. There is real-time updates and this has enabled expansion as more individuals come. When there is an increase in load, performance remains constant.

VI. RESULTS

The analysis of the results reveals that the set-up is reasonably effective in reporting abuse in a safe and confidential way. Firebase Firestore seamlessly transferred user entries to a storage and did not hiccup or delay. The updates remained real time even with high traffic, and this was made possible by the cloud infrastructure. Chatbot responses were realized in real-time whenever it was tested. One of the questions was the one that inquired about safety online and how not to be harassed, and answers were fast to reply, which suits the occasion well. There were positively responding responses that were circumstantial and scriptless. Assistance was not taken out of hoarded cases. What was found was not new data, but knowledge was spontaneously produced according to a design.

Fig. 1 Home Page

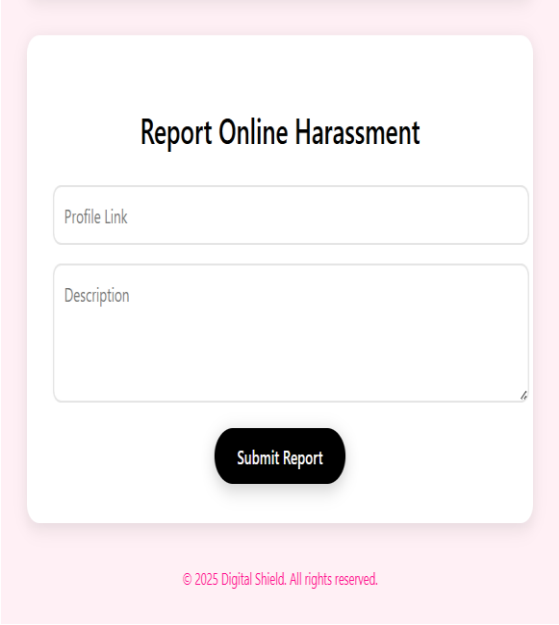


Fig. 2 Online Harassment Reporting Page

Digital Shield creates the space immediately where one feels safe in the platforms of online worlds. What stands out first? Certain rules as to what to do - also to leave out - when fighting digital threats. Anonymous assistance is offered in the form of a specially constructed form that contains only those victims of online abuse. The information remains secured since all the met cases remain secured on remote servers in encrypted format. Whenever queries arise, an active assistant called SUM80 intervenes, which is an intelligent technology that is aware of cyber
Problems.

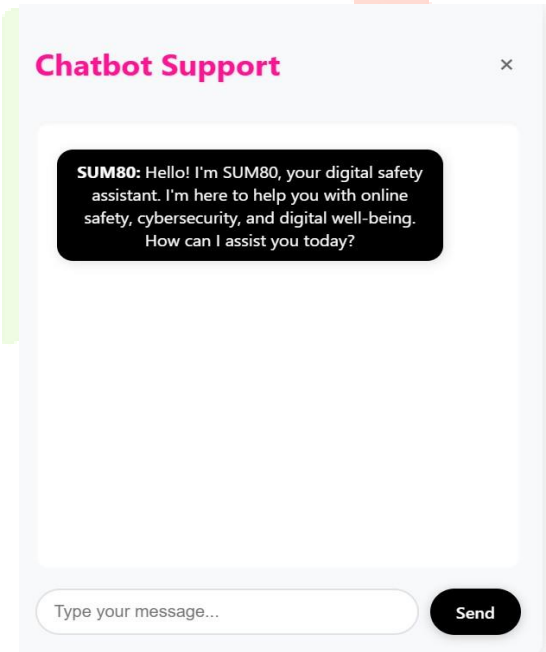


Fig:3. AI Chatbot Support Interface

VII. CONCLUSIONS

An artificial intelligence-driven Web tool called Digital Shield has appeared - it is a web browser extension. It does not only allow people to react to harm but assists them to educate about risks as they browse. It is easier to report abuse as there are no muddled steps. Privacy is maintained as there is nothing that is saved without authorization. A ChatGP based behind curtains is the quick chat helper in the form of a smart chat helper named SUM80. No loads of data required, no huge learning models to be trained either. It is developed on advanced web technologies, and it curves around any devices such as phones,

tablets, laptops. The information is defended in the background by cloud systems. Being tested, all pieces worked: messages were locked, responses were sensible and timing was satisfactory.

Safety will be an experience not an after-thought. And overall, it can be concluded that Digital Shield is a practical method of raising awareness of how to be safe over the internet and enabling people to confront the dangers of digital use with confidence. The platform has been designed to expand and can expand in the future by developing such facilities as mobile access, support of multiple languages, as well as assistance in the event of cyber attacks. It is designed to support new features without compromising what has been working reasonably well in the various applications.

REFERENCES

1. A. Kumar and S. Verma, "A Study on Online Harassment and Cyber Safety in Social Media Platforms," International Journal of Computer Applications, vol. 175, no. 20, pp. 15–20, 2020.
2. R. Smith and J. Brown, "Artificial Intelligence-Based Chatbots for User Support Systems," Journal of Information Technology and Systems, vol. 12, no. 3, pp. 45–52, 2019.
3. M. Patel, "Cloud-Based Web Application Architecture Using Firebase," International Journal of Cloud Computing and Services, vol. 8, no. 2, pp. 60–66, 2021.
4. OpenAI, "ChatGPT: Large Language Models for Conversational AI," OpenAI Documentation, 2023.
5. Google Firebase, "Firebase Firestore: Scalable NoSQL Cloud Database," Google Developers Documentation, 2023.
6. N. Johnson and K. Lee, "Improving Digital Safety Through AI-Driven Solutions," Proceedings of the International Conference on Artificial Intelligence and Applications, pp. 112–118, 2022.
7. S. Rao, "Privacy and Security Challenges in Cloud-Based Web Systems," International Journal of Cybersecurity, vol. 6, no. 1, pp. 25–31, 2020.
8. World Wide Web Consortium (W3C), "Web Security and Privacy Guidelines," W3C Recommendation, 2022.
9. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 4–37 (2000).
10. Mengel, K., Kirkby, E.A.: Principles of Plant Nutrition. Kluwer Academic Publishers, Dordrecht (2001).