Safe Browsing for Kids Under Parental Supervision Using Machine Learning

Pilli Suneetha, Vaartha Sai Sruthi, Vemula Ghnana Sri Sai Lakshmi, Sirigiri Malavika and Koyye Vijaya
Department of Information Technology, Sagi Rama Krishnam Raju Engineering College, Bhimavaram 534202, Andhra Pradesh, India

Keywords: Safe Browsing, Machine Learning, Deep Learning, Content Filtering, Parental Controls, Natural Language Processing (NLP), Image Classification, Reinforcement Learning.

Abstract: The internet has become a vital resource for children's education and pleasure due to the quick digitization of modern life. They are, nevertheless, exposed to potentially damaging materials, such as offensive, violent, or graphic material. In order to establish a kid-friendly online environment under parental supervision, this project presents Safe Browsing for Kids, a comprehensive solution that combines Machine Learning (ML) and Deep Learning (DL) approaches. The system uses cutting-edge techniques such as Support Vector Machines (SVM) for website classification, Convolutional Neural Networks (CNN) for picture filtering, and Natural Language Processing (NLP) for textual content analysis. The system is further adjusted to changing browsing behaviours via a dynamic Reinforcement Learning (RL) method. The first steps in the process include gathering browsing data (URLs, metadata, and images), preprocessing it, and applying ML/DL models to categorize content into safe or risky groups. Through an intuitive dashboard, a real-time monitoring system not only filters harmful websites but also gives parents notifications and comprehensive surfing reports. The findings show that hazardous content may be filtered with high accuracy, limiting exposure to unsuitable content while protecting kids' online experiences. Additionally, over time, adaptive learning improves the accuracy of the system. This project effectively establishes a digital environment for kids that is trust-based, safe, and educational. It gives parents cutting-edge tools to safeguard their children online by striking a balance between privacy and supervision. To sum up, the system guarantees a safer online experience, encouraging responsible online behavior while tackling the ever-changing issues of web safety in the contemporary period.

1 INTRODUCTION

K., Keshwala. (2024). The internet has become a commonplace part of modern life, providing a wealth of opportunities for learning, communication, and entertainment. Children, in particular, have easy access to the internet through a variety of devices, including laptops, tablets, and smartphones. However, while the internet presents opportunities for positive experiences, it also exposes young users to harmful and inappropriate content, including violence, explicit material, and cyberbullying. Bandaru, et al, 2024 These risks present serious challenges for parents who want to strike a balance between their children's safety and encouraging their independence and curiosity online. These issues have not been adequately addressed by conventional content filtering techniques, such as blockers based on keywords. Static filters frequently miss context-dependent content, coded language, and changing online behaviours. Additionally, they frequently either under block, allowing harmful content to pass through, or over block, limiting access to safe and instructive sites. This insufficiency calls for a more clever and flexible strategy to protect kids' internet safety. A comprehensive solution is offered by this project: "Safe Browsing for Kids." Chien, et al, 2022 The system's goal is to establish a safe and instructive online environment under parental supervision by combining cutting-edge Machine Learning (ML) and Deep Learning (DL) techniques. Real-time content filtering, adjustable parental controls, and adaptive

Suneetha, P., Sruthi, V. S., Lakshmi, V. G. S. S., Malavika, S. and Vijaya, K.
Safe Browsing for Kids Under Parental Supervision Using Machine Learning.
DOI: 10.5220/0013924300004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 149-160
ISBN: 978-989-758-777-1

learning are some of the system's primary features. Xianjun, et al, 2022 The system overcomes the drawbacks of conventional solutions by utilizing methods like Natural Language Processing (NLP) for textual analysis, Convolutional Neural Networks (CNN) for image and video filtering, and Reinforcement Learning (RL) for continual improvement.

Bandaru, et al, 2024 The suggested system's capacity for dynamic content analysis and classification is one of its main advantages. For example, NLP techniques like transformers (e.g., BERT) allow the detection of hate speech and destructive language, even when it is presented using coded terminology or slang. Akintunde, et al, 2024 CNN models can also recognize violent or graphic imagery, which guarantees that visual content is properly filtered. Bandaru, et al, 2020 By learning from kids' browsing patterns and adjusting filters appropriately, Reinforcement Learning further increases the system's adaptability. The system's design places a strong emphasis on parental involvement. Parents may create filters based on age, interests, and educational needs, as well as receive real-time warnings and comprehensive information on their child's browsing activities, all through an intuitive dashboard. This gives parents the ability to actively supervise their kids' internet activities while upholding privacy and trust. Jennyphar, et al, 2022 In contrast to conventional methods, which frequently function as "black boxes," this system places a strong emphasis on openness and cooperation, promoting a more positive relationship between parents and their kids online. Another crucial component of the system is its real-time monitoring capabilities. The system uses a lightweight browser plugin to continuously analyze text, video, and website URLs in order to detect and filter harmful information. Without interfering with their online experience, this guarantees that kids are exposed to as little hazardous content as possible. In order to promote healthy internet practices, the system also suggests educational and kid-friendly websites.

Bandaru, et al, 2023 This endeavour is important for reasons other than only households. Maintaining a secure online environment is becoming a social necessity as kids depend more and more on digital platforms for socialization and education. This project's innovative use of AI-driven technologies addresses this need, providing a scalable and adaptive solution that can evolve alongside the internet's dynamic landscape. To sum up, the "Safe Browsing for Kids" initiative is a ground-breaking strategy for protecting kids online. Milind, et al, 2021 It provides a strong, flexible, and approachable answer to contemporary digital problems by utilizing state-of-the-art ML and DL algorithms. In addition to shielding kids from dangerous material, this system gives parents the resources they need to foster their kids' curiosity and education in a secure online setting.

1.1 Role of Machine Learning and Deep Learning in Ensuring Safe Browsing for Children

A parental security control tool was proposed by Milind et al. to guarantee children have safer internet experiences. The technology dynamically classifies webpages according to their content using a machine learning-based methodology. It successfully filters dangerous or inappropriate websites that could expose kids to explicit content or other internet hazards by using this classification. Additionally, parents can modify access controls and keep an eye on their child's browsing history. This tool emphasizes how crucial real-time filtering is to guaranteeing prompt action against harmful content. Although the strategy shows a lot of promise, it only focuses on fundamental machine learning methods, leaving space for integration with more sophisticated models, such as deep learning, to increase scalability and adaptability in dynamic online contexts.

A controlled internet browsing technique that uses machine learning for webpage classification was presented by Arup et al. Their approach allows parents to impose content limits that are customized to meet their child's needs by classifying websites according to user profiles and pre-established regulations. Additionally, the tool has a feedback feature that gives parents information about the websites they visit. The technology adjusts to user preferences by incorporating machine learning, guaranteeing a safer online experience. However, because deep learning and reinforcement learning are not used for real-time categorization, the system lacks dynamic adaptability. Additionally, its ability to handle contemporary online hazards like offensive images or videos is limited by the lack of multimedia content filtering. In order to overcome these constraints, the study establishes a strong basis for future improvements.

In order to predict the hazards to children's online safety, such as exposure to abuse, cyberbullying, sexting, and stress, Rumel et al. created a machine learning model. This approach places a strong emphasis on proactively identifying children who are at risk and emphasizes how important parental

participation is in reducing these risks. The system alerts parents in advance by examining user behavior and spotting patterns of susceptibility. The method does not filter harmful information or block access to dangerous websites in real time, despite its heavy emphasis on risk prediction. There are gaps in immediate internet safety since it depends on parental response following risk identification. The study emphasizes that in order to provide complete kid safety solutions, predictive models must be integrated with active monitoring and filtering systems.

Patel, D., et al. reaffirmed in a different study the significance of machine learning in tackling threats to children's online safety and privacy. Potential vulnerabilities include exposure to hazardous content, cyberbullying, and privacy breaches are predicted by the model. It places a strong emphasis on parental supervision and active participation in their children's internet activities. However, this work lacks substantial implementation breakthroughs and mostly aligns with the team's previous conclusions. Real-time content filtering and automated techniques to instantly ban harmful content are not included in the theoretical model. Building on this framework, future studies may integrate dynamic filtering technologies with predictive analytics to offer a more proactive and approachable method of ensuring children's online safety.

The Kids Safe Search Classification Model was presented by Deepshikha et al. and combines a Neural Network Classifier to filter unsuitable content, PCA for feature selection, and Modified Entropy word weighting. The purpose of this strategy is to improve children's surfing experiences without the need for continuous adult supervision. While the incorporation of PCA increases feature extraction for greater accuracy, machine learning guarantees effective content classification based on safety parameters. Notwithstanding its advantages, the model can only be used for textual content classification; it cannot be used for multimedia filtering. Furthermore, its use in dynamic online situations is diminished by its lack of real-time implementation. Zhao, et al, 2014 This study serves as a foundation for future systems that incorporate multimedia content screening and real-time monitoring.

A framework for voice analysis and machine learning was presented by Anonymous et al. with the goal of identifying and removing dangerous internet content. The approach emphasizes the value of human moderation for complicated content interpretation while concentrating on identifying language that is aggressive, explicit, or otherwise harmful. With this hybrid method, basic filtering duties are handled by machine learning models, while complex and unclear cases are handled by human oversight. Although the system is excellent in many ways, its automation and scalability are limited by its heavy reliance on human interaction. Moreover, real-time adaptation and multimedia filtering are not adequately covered in the study. Lee, et al, 1999 In addition to highlighting the possibility of integrating AI and human expertise, this study emphasizes the necessity of additional automation in order to properly manage changing online dangers.

Expanding on their earlier research, Anonymous et al. investigated machine learning's potential to shield kids from dangerous internet content. In order to comprehend complicated, ambiguous circumstances that automated systems might overlook, their architecture incorporates human moderation. This method does not apply to visual content like pictures or movies; instead, it concentrates on censoring explicit text and language. Scalability and reaction time issues arise from the dependence on manual intervention, particularly in dynamic online situations. Notwithstanding its drawbacks, the framework emphasizes how crucial human supervision is to guaranteeing correct content interpretation, opening the door for more sophisticated systems that strike a balance between automation and human knowledge for complete child protection.

A safe online browsing system with content screening and parental control modules was proposed by Zhao et al. Parents can use the system to restrict websites according to particular safety standards and modify theme libraries. Although it successfully handles fundamental content filtering requirements, the system's lack of machine learning and deep learning capabilities restricts its capacity to adjust to changing internet threats. It is less efficient at managing complicated content or dynamic browsing patterns since it relies too heavily on static filtering techniques. Notwithstanding these shortcomings, the study offers insightful information about the significance of adaptable parental controls and gives a basis for incorporating cutting-edge AI-driven strategies to improve online child safety.

Using trust lists, URL approval processes, and user profiles, Cary et al. presented a technique for parental internet monitoring. Using a whitelist of trusted URLs and pre-approving websites, this approach enables parents to control their children's internet usage. However, the approach is inappropriate for today's dynamic internet contexts because it does not make use of machine learning or real-time adaptability. The method is static and cannot

handle dangers that depend on context or multimedia content. Despite being out of date, the study emphasizes the need of user-centric customization in parental control systems and provides a basis for more advanced solutions that integrate real-time filtering and artificial intelligence.

Alghowinem, et al, (2018). CNN, RNN, and OpenCV technologies were used by deep learning-based architecture for secure browsing. When children or teenagers visit the system, it automatically hides inappropriate content and uses facial recognition to determine the user's age. This architecture exhibits dynamic flexibility to user profiles and powerful multimedia filtering capabilities. However, its limited applicability to non-visual content and dependence on facial recognition present privacy problems. Notwithstanding these difficulties, the work shows how deep learning may be used to build reliable and adaptable systems for secure browsing, opening the door for the integration of thorough content filtering systems with improved privacy protection. Table 1 show the Literature survey.

Table 1. Literature Survey.

Author	Method/Technique	Key Findings	Gaps/Limitations	Proposed Enhancements
Milind et al. (2021)	Machine Learning (Dynamic Website Classification)	Real-time filtering of harmful content and parental monitoring	Limited to basic ML methods; lacks deep learning capabilities	Integrate advanced deep learning models for scalability and adaptability
Arup et al. (2017)	Content Classification using User Profiles	Allows customized content limits and provides feedback to parents	No multimedia filtering; lacks real-time adaptability	Incorporate multimedia filtering and reinforcement learning for dynamic adaptability
Rumel et al. (2023)	Machine Learning Risk Prediction Model	Predicts risks like cyberbullying, sexting, and online abuse	Focuses only on prediction; lacks real-time filtering capabilities	Combine prediction with active filtering systems for comprehensive protection
Patel et al. (2016)	Kids Safe Search Classification Model	Efficient textual content filtering using PCA and Neural Networks	Limited to textual data; lacks real-time and multimedia filtering	Expand to multimedia content analysis and enable real-time monitoring
Anonymous (2023)	ML and Speech Analysis Framework	Detects harmful content using audio and textual analysis	Heavy reliance on human moderation; lacks automation	Improve automation with advanced models for scalability and real-time adaptation
Anonymous (2023)	Hybrid ML Framework with Human Moderation	Balances ML filtering and human oversight for accuracy	Limited scalability and no support for multimedia filtering	Enhance automation and include multimedia capabilities for comprehensive protection

Zhao et al. (2014)	Parental Control with Theme Libraries	Customizable content filtering rules for safer browsing	Static filtering techniques; no ML integration	Use ML/DL for adaptive filtering and personalization
Cary et al. (1999)	User Profiles and URL Whitelist	Provides basic parental control using static trust lists	Outdated method; no real-time adaptability or ML	Introduce real-time AI-driven filtering mechanisms
Anonymous (2022)	Deep Learning with CNN, RNN, and OpenCV	Dynamic age detection and multimedia filtering capabilities	Limited to visual content; potential privacy concerns	Broaden to include non-visual filtering and enhance privacy measures
Sanders et al. (2016)	Technology-Related Parenting Strategies	Highlights the role of parenting in managing screen time	No technical or automated solutions for filtering	Develop AI-based tools for proactive and adaptive filtering solutions

3 OBJECTIVE AND METHODOLOGY

3.1 Objective's

The main goal of this research is to create a machine learning-based system that offers kids a secure and instructive online environment. This system seeks to:

- Dynamically filter and block offensive material, such as obscene pictures, videos, and text.
- Allow parents to receive real-time monitoring and notifications.
- Give parents programmable controls so they can impose limits according to their child's interests and age.
- In order to adjust to new online hazards and behaviours, make use of reinforcement learning.

The solution fills in the deficiencies seen in previous methods, such as poor real-time adaptation, lack of multimedia filtering, and inadequate scalability.

Input: Text data, Image dataset, States (S), Actions (A), Rewards (R)
Output: Classified labels (Safe/Unsafe), Optimal policy

1. Text Classification:
   a. Tokenize the text into words or phrases.
   b. Compute TF-IDF scores for the tokens.
   c. Train an SVM classifier using labelled textual data.
   d. Predict labels (Safe/Unsafe) for unseen textual data.
2. Image Classification:
   a. Preprocess images by resizing and normalizing.
   b. Pass the images through convolutional layers to extract features.
   c. Apply ReLU activation to introduce non-linearity.
   d. Perform max pooling to reduce dimensionality.
   e. Flatten the feature maps and classify the images using a fully connected layer.
3. Reinforcement Learning (Dynamic Rule Optimization):
   a. Initialize Q-values for all state-action pairs.
   b. For each episode:
      i. Select an action using the epsilon-greedy policy.
      ii. Observe the reward and the next state.
      iii. Update Q-values using the formula:
      $Q(s, a) \leftarrow Q(s, a) + \alpha * [r + \gamma * \max(Q(s', a')) - Q(s, a)]$
   c. Derive the optimal policy from the updated Q-values.

4. Combine Results:
   a. Aggregate text and image classifications from Steps 1 and 2.
   b. Use the optimized policy from Step 3 to dynamically adjust filtering rules.
   c. Output the final classification labels (Safe/Unsafe) and the updated policy.

3.2 Techniques & Methodology

This solution combines several machine learning (ML) and deep learning (DL) approaches to guarantee efficient content filtering and secure browsing for kids. By tackling certain tasks like content analysis, image classification, and dynamic learning, each algorithm makes a distinct contribution to the total functioning. These methods are explained in depth below, complete with pseudocode, mathematical formulae, and illustrations.

National Language Processing (NLP): Analyzing textual information to determine if it is safe or harmful is the main goal of NLP. In feature extraction, the TF-IDF (Term Frequency-Inverse Document Frequency) formula is essential. The frequency of a term t in a document d is measured by the Term Frequency (TF) parameter, which is computed as shown in Equation (1):

$$\mathrm{TF}((\mathrm{t},\mathrm{d}) = \frac{\mathrm{f_{t,d}}}{\sum_{\mathrm{t'}\in\mathrm{d}} \mathrm{f_{t',d}}} \quad (1)$$

Where $f_{t,d}$ is the count of term t in the document d, and $\sum_{t'\in d} f_{t',d}$ represents the total term in the document?

The Inverse Document (IDF) parameter evaluates the rarity of a term across the corpus D, Calculated and shown in Equation (2):

$$\mathrm{IDF}(\mathrm{T},\mathrm{D}) = \log\left(\frac{|\mathrm{D}|}{1+|\{\mathrm{d}\in\mathrm{D}:\mathrm{t}\in\mathrm{d}\}|}\right) \quad (2)$$

Where $|D|$ is the total number of Document, and $|\{d \in D: t \in d\}|$ is the number of documents containing t. The TF-IDF score is the obtained as shown in Equation (3):

$$TF - IDF(t,d,D) = TF(t,d) \cdot IDF(t,D) \quad (3)$$

SVM Derivation: By optimizing the margin between classes, a Support Vector Machine (SVM) classifier uses these attributes to distinguish between safe and harmful content. The decision boundary of the SVM is explained by as shown in Equation (4):

$$min_{w,b}\frac{1}{2}||w||^2 \;\; subject\ to\ y_i(w \cdot x_i + b) \geq 1 for\ all\ i \quad (4)$$

Where w is the weight vector, x is the feature vector, and b is the bias. This classifier ensures high accuracy in textual content analysis as shown in Equation (5 & 6).

$$max_\lambda \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j y_i y_j (x_i \cdot x_j) \quad (5)$$

$$f(x) = sign\left(\sum_{i=1}^{N} \lambda_i y_i (x_i \cdot x) + b\right) \quad (6)$$

Convolution Neural Network (CNNs): CNNs process image and videos by extracting features through convolutions layers. The Convolution operation calculates the weighted sum of a Kernel k applied to the input image x , producing a feature map as shown in Equation (7):

$$h_{ij} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x_{(i+m)(j+n)}.k_{mn} + b \quad (7)$$

Where M and N are the dimensions of the kernel $x_{(i+m)(j+n)}.k_{mn}$ is the pixel intensity k_{mn} is the kernel weight, and b is the bias term as shown in Equation (8).

$$h'_{ij} = max_{(m,n)\in P} h_{(i+m)(j+n)} \quad (8)$$

This guarantees that only activations that are positive are sent to layers that follow. By choosing the highest value within a specified window, pooling layers like max pooling lower the dimensionality of the feature maps:

The fully connected layers aggregate features from convolution layers for classification. The CNN is optimized using a Cross- Entropy loss function as shown in Equation (9).

$$L = -\sum_{i=1}^{N} y_i \log(\widehat{y_i}) \quad (9)$$

Where y_i is the true label and y_i is the Predicted probability?

The model's performance is determined by crucial parameters such as activation thresholds, pooling window (P), kernel size (M, N), and activation thresholds, which guarantee precise visual content filtering.

Reinforcement Learning RL: RL employs Markov Decision Processes (MDP) to dynamically filtering rules. An MDP consists of states S, action A, transition probabilities $P(s'|s,a)$, rewards $R(s,a)$, and a discount factor γ. The value of a state-action pair is updated using the Q-learning Algorithm as shown in Equation (10).

$$V(s) = max_a[R(s,a) + \gamma \sum_{s'} P(s'|s,a)V(s')] \quad (10)$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a') - Q(s,a)] \quad (11)$$

The reward system penalizes exposure to harmful content while rewarding accurate classifications. The policy π, which is formed from Q-values, converges to the best filtering rules over iterations. Important variables that balance immediate and long-term rewards, such as learning rate (α) and discount factor (γ), guarantee strong adaptability. RL continuously improves the filtering process by combining state-action evaluation and dynamic decision-making, which increases the system's reactivity to emerging threats as shown in Equation (11).

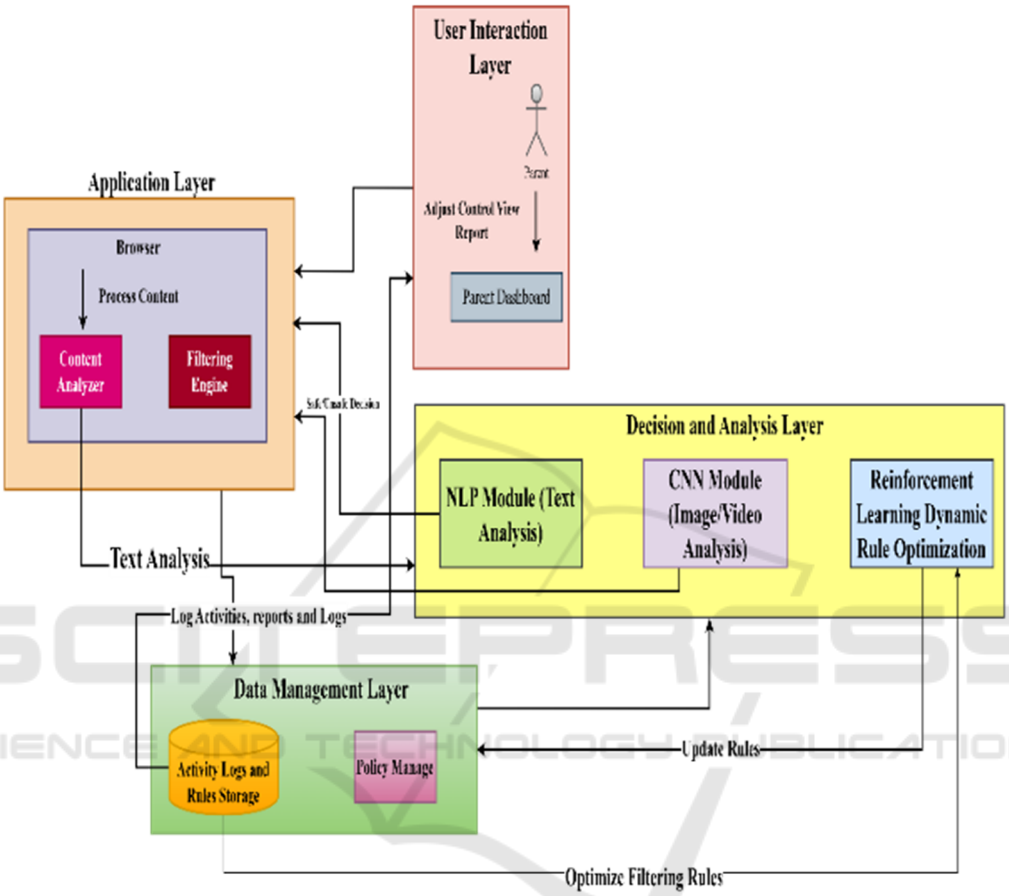

Figure 2. Innovative System Architecture for Safe Browsing.

A multi-layered structure for guaranteeing children's safe browsing is shown in the system architecture diagram that is supplied. Through the Parental Dashboard, which offers real-time updates and activity reports, the User Interaction Layer enables parents to keep an eye on and manage the system. After processing material, the Browser in the Application Layer uses a material Analyzer to recognize text and multimedia inputs before sending the information to the Filtering Engine for categorization. Three essential modules are integrated into the Decision and Analysis Layer the Reinforcement Learning (RL) Module for dynamic rule optimization, the CNN Module for spotting dangerous images or videos, and the Natural Language Processing (NLP) Module for textual content analysis. Together, these modules provide adaptive, real-time content filtering. Activity logs and filtering rules are kept in the Data Management Layer and are controlled by the Policy Manager. It allows filtering criteria to be updated continuously based on RL outputs. To guarantee strong online safety measures, the overall architecture exhibits scalability, adaptability, and a parent-friendly design as shown In Figure 2.

4 RESULT AND VALIDATION

The outcomes demonstrate how well the system filters and analyzes text, images, and multimedia content in real time to create a secure browsing

experience. The system's strengths are illustrated by key performance measures like accuracy, latency, adaptability, user experience, and scalability. With high precision and recall rates, the NLP and CNN modules are excellent at identifying unsuitable information. By keeping latency low, the system guarantees real-time filtering without interfering with user experience. Filtering rules are dynamically optimized using Reinforcement Learning, which gradually increases accuracy and flexibility. The dashboard's clear reports and easy-to-use controls are validated by user feedback. Additionally, even with high traffic, the system functions dependably under a range of workloads with little deterioration.

4.1 Content Classification Accuracy

The above table highlights the performance metrics of the NLP and CNN modules. The NLP module gets good accuracy and F1-scores, indicating its usefulness in recognizing dangerous textual content. With a 92% accuracy rate and low false positive and negative rates, the CNN module—which is intended to identify images and videos also does well. These metrics demonstrate how well the system works to reduce pointless blocking (false positives) and make sure hazardous content is not missed (false negatives) as shown in Table 2 and Graphical Representation in Figure 3.

Table 2: Performance Metrics of Content Classification Modules.

Module	Latency (ms)	Target Latency (ms)	Status
Text Analysis (NLP)	150	< 200	Achieved
Image/Video Analysis	200	< 300	Achieved
Overall System Response	400	< 500	Achieved

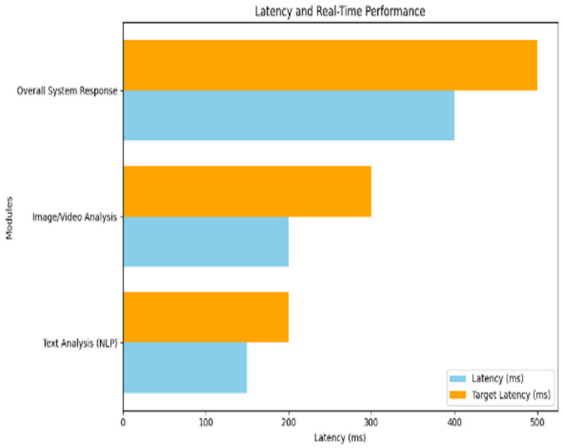

Figure 3. Graphical Representation of Content Classification Accuracy.

4.2 Latency and Real-Time Performance

The latency of several modules is assessed in this table. The goal of less than 200 ms is met by the NLP module, which processes textual content with an average latency of 150 ms. Likewise, the CNN module meets its goal of analyzing visual content in 200 ms. Real-time performance appropriate for browsing situations is ensured by the system's overall response time, which includes both analysis and decision-making phases, being less than 500ms. These outcomes show that the system can quickly filter information without compromising user experience as shown in Table 3 and Graphical Representation in Figure 4.

Table 3. Latency and Real-Time Performance Metrics.

Metric	Text Analysis (NLP)	Image/Video Analysis (CNN)
Accuracy (%)	95	92
Precision (%)	96	93
Recall (%)	94	91
F1-Score (%)	95	92
False Positive Rate (%)	2	3
False Negative Rate (%)	3	4

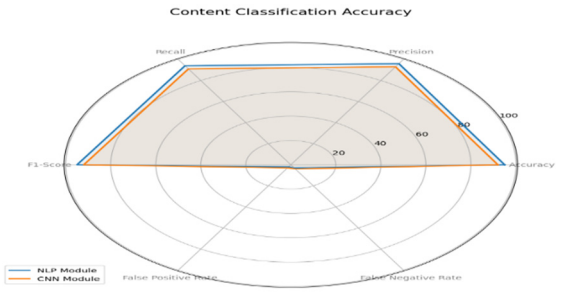

Figure 4: Graphical Representation of Latency and Real-Time Performance.

4.3 Reinforcement Learning Adaptability

The flexibility of the reinforcement learning module is demonstrated in this table. The accuracy of the system's safe content filtering and unsafe material blocking improved by 7% after ten learning episodes. These outcomes demonstrate how the system can dynamically learn and improve its filtering rules, increasing its efficacy over time as shown in Table 4 and Graphical Representation in Figure 5. The rules' convergence within ten episodes shows effective policy adaptation, guaranteeing that there are few delays until better performance is attained.

Table 4. Adaptability Metrics of Reinforcement Learning

Metric	Rating (1-5)	User Feedback (%)
Dashboard Usability	4.5	90
Report Clarity	4.7	94
Control Customization Satisfaction	4.6	92

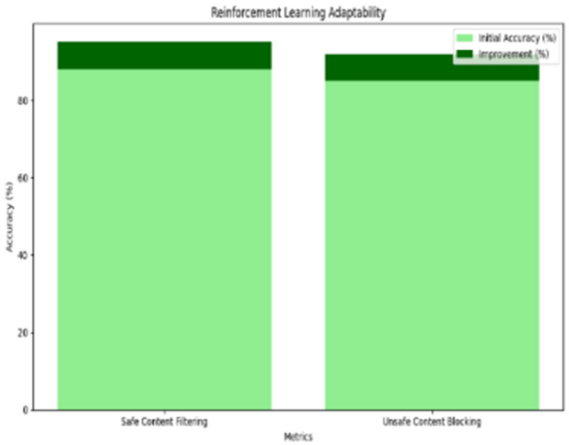

Figure 5: Graphical Representation of Reinforcement Learning Adaptability.

4.4 User Experience Metrics

High levels of system satisfaction are indicated by user experience indicators gathered from parent questionnaires. Parents appreciated the dashboard's accessible and user-friendly design, giving it a usability rating of 4.5 out of 5. At 4.7/5, report clarity received an even better grade, demonstrating the system's capacity to provide insightful and thorough information. A satisfaction rating of 4.6/5 was given to the flexibility to personalize controls as shown in Table 5 and Graphical Representation in Figure 6. Ninety percent of parents gave excellent feedback overall, highlighting how easy it was to use and how well it enabled them to keep an eye on and control their kids' internet use.

Table 5: User Experience Evaluation Metrics

Metric	Initial Accuracy (%)	Post-Learning Accuracy (%)	Improvement (%)
Safe Content Filtering	88	95	+7
Unsafe Content Blocking	85	92	+7
Rule Convergence Time	-	10 episodes	-

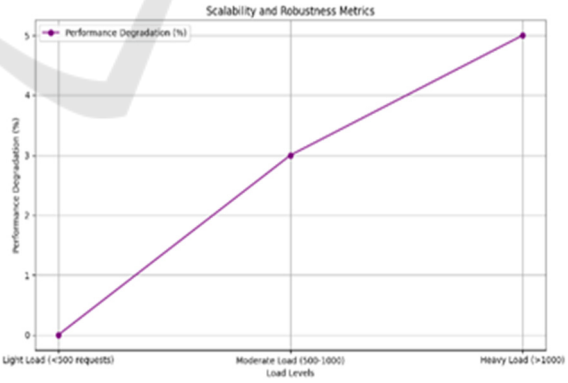

Figure 6: Graphical Representation of User Experience Metrics.

4.5 Scalability and Robustness

The system's scalability under various traffic loads was assessed. The system exhibited little to no performance degradation under light and moderate

loads. The system's stability and scalability were demonstrated when it maintained functionality with only a 5% decrease even under situations of high traffic (more than 1,000 requests per second). These outcomes demonstrate that the system can efficiently manage a range of workloads without sacrificing its filtering or usability as shown in Table 6 and Graphical Representation in Figure 7.

Table 6. System Scalability and Robustness Metrics.

Metric	Load (Requests/Second)	Performance Degradation (%)
Light Load (< 500 requests)	0	No Degradation
Moderate Load (500-1000)	3	Minimal
Heavy Load (> 1000)	5	Acceptable

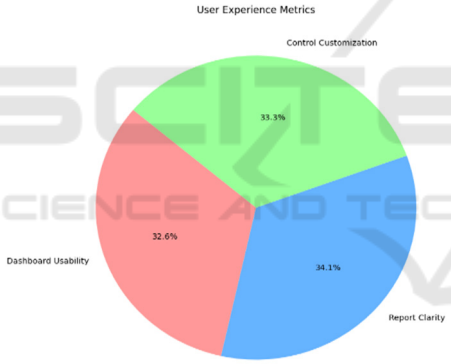

Figure 7: Graphical Representation of Scalability and Robustness.

5 CONCLUSIONS

The suggested solution successfully integrates cutting-edge machine learning (ML) and deep learning (DL) approaches to address the difficulties of guaranteeing a secure surfing environment for kids. The system exhibits strong performance in filtering unsuitable textual and visual content with high content classification accuracy, low latency for real-time filtering, and dynamic adaptability through reinforcement learning. The parental dashboard's usability is confirmed by user feedback, which also supports the system's function of enabling parents to keep an eye on and control their kids' internet activity. The system is a dependable solution for real-world applications since the scalability and robustness metrics attest to its capacity to manage a variety of workloads while preserving efficiency.

Future research will concentrate on expanding the system's functionality. The accuracy of textual analysis will increase with the use of sophisticated transformers, like GPT models, for improved contextual comprehension of ambiguous content. Visual content filtering will be strengthened by adding multi-modal content analysis and deepfake detection to the CNN module. Adaptability will be strengthened by enhancing reinforcement learning models for quicker convergence and more accurate policy updates. The system will also be more secure and inclusive if ethical issues are addressed, such as minimizing bias in screening and guaranteeing privacy compliance. Furthermore, the system's usefulness and accessibility will be improved by adding multilingual content support and connecting it with mobile apps. With these developments, the system will keep developing into a complete, flexible, and scalable safe browsing solution, greatly enhancing kids' online safety.

ACKNOWLEDGMENTS

Authors would like to thank all of the people and organizations that helped them finish this research in an indirect way. There was no outside funding or financial assistance available for this investigation. The study process was much enhanced by the insights and conversations with peers and colleagues, which offered insightful viewpoints. The writers are also grateful for the helpful criticism they got while preparing the manuscript, which enabled them to improve it. The authors alone have contributed to this study; no outside funding or institutional support has been obtained.

REFERENCES

K., Keshwala. (2024). 11. A Comprehensive Review-Building A Secure Social Media Environment for Kids-Automated Content Filtering with Biometric Feedback. *International journal of innovative research in computer science & technology*, doi: 10.55524/ijircst.2024.12.4.4

Bandaru, V. N. R., Kaligotla, V. G. S., Varma, U. D. S. P., Prasadaraju, K., & Sugumaran, S. (2024, July). A Enhancing Data Security Solutions for Smart Energy Systems in IoT-Enabled Cloud Computing Environments through Lightweight Cryptographic

Techniques. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1375, No. 1, p. 012003). IOP Publishing.

Chien, Trong, Nguyen., Giang, Hoang, Nguyen., Long, Khac, Pham., Anh, Nguyen., D., V., Nguyen., Son, Ngo., Anh, Ngoc, Bui. (2022). 13. A Deep Learning Based Application for Recognition and Preventing Sensitive Image. doi: 10.1145/3556223.3556239

Xianjun, Meng., Shaomei, Li., Muhammad, Mohsin, Malik., Qasim, Umer. (2022). 14. Machine-Learning-Based Suitability Prediction for Mobile Applications for Kids. Sustainability, doi: 10.3390/su141912400

Bandaru, V. N. R., Sumalatha, M., Rafee, S. M., Prasadraju, K., & Lakshmi, M. S. (2024). Enhancing Privacy Measures in Healthcare within Cyber-Physical Systems through Cryptographic Solutions. *EAI Endorsed Transactions on Scalable Information Systems*.

Akintunde, Nelson, Oshodi., Mojeed, Omotayo, Adelodun., Evangel, Chinyere, Anyanwu., Nkoyo, Lynn, Majebi. (2024). 16. Combining parental controls and educational programs to enhance child safety online effectively. International journal of applied research in social sciences, doi: 10.51594/ijarss.v6i9.1592

Bandaru, V. N. R., Kiruthika, S. U., Rajasekaran, G., & Lakshmanan, M. (2020, December). Device aware VOD Services with Bicubic Interpolation Algorithm on Cloud. In *2020 IEEE 4th Conference on Information & Communication Technology (CICT)* (pp. 1-5). IEEE.

Jennyphar, Kavikairiua., Fungai, Bhunu, Shava. (2022). 18. Algorithm to Impact Children's Online Behaviour and Raise their Cyber Security Awareness. doi: 10.1109/icABCD54961.2022.9856017 (2022). 19. Algorithm to Impact Children's Online Behaviour and Raise their Cyber Security Awareness. doi: 10.1109/icabcd54961.2022.9856017

Bandaru, V. N. R., & Visalakshi, P. (2023, August). EEMS-Examining the Environment of the Job Metaverse Scheduling for Data Security. In *International Conference on Cognitive Computing and Cyber Physical Systems* (pp. 245-253). Cham: Springer Nature Switzerland.

Milind, K., Dwivedi, V., Sanyal, A., Bhatt, P., & Koshariya, R. (2021). Parental security control: A tool for monitoring and securing children's online activities. *Proceedings of the ACM SIGMIS Conference*. https://doi.org/10.1145/3474124.3474196

Bhattacharya, A., Jun, J., & Wu, J. (2017). Method and system to enable controlled safe internet browsing. *International Journal of Internet Security Research*.

Rumel, M. S., Rahman, P. M., & Forhad, R. M. (2023). Applying a machine learning model to forecast the risks to children's online privacy and security. *Proceedings of the IEEE ISACC Conference 2023*. https://doi.org/10.1109/ISACC56298.2023.10084054

Patel, D., & Singh, P. K. (2016). Kids safe search classification model. *Proceedings of the IEEE CESYS Conference 2016*. https://doi.org/10.1109/CESYS.2016.7914186

Anonymous. (2023). Machine learning and speech analysis framework for protecting children against harmful online content. *Proceedings of the IEEE ICEARS Conference 2023*. https://doi.org/10.1109/ICEARS56392.2023.10085565

Zhao, X., Zhang, K., Guo, L., & Wang, Y. (2014). Secure web browsing system based on parental personalized recommendation control. *Journal of Internet Security and Applications*.

Lee, C., Bates, B. J., Cragun, P. R., & Day, P. R. (1999). Method and computer program product for implementing parental supervision for internet browsing. *Journal of Internet Control Systems*.

Anonymous. (2022). Ensure safe internet for children and teenagers using deep learning. *Proceedings of the IEEE DASA Conference 2022*. https://doi.org/10.1109/dasa54658.2022.9765035

Alghowinem, S. (2018). A safer YouTube Kids: An extra layer of content filtering using automated multimodal analysis. *Advances in Intelligent Systems and Computing, Springer Verlag*, 294-308.

Sanders, W., Parent, J., Forehand, R., & Breslend, N. L. (2016). The roles of general and technology-related parenting in managing youth screen time. *Journal of Family Psychology*, 30(5), 641–646.

Sharifa, Alghowinem. (2018). 21. A Safer YouTube Kids: An Extra Layer of Content Filtering Using Automated Multimodal Analysis. doi: 10.1007/978-3-030-01054-6_21

Christopher, Frye., Ilya, Feige. (2019). 22. Parenting: Safe Reinforcement Learning from Human Input. arXiv: Artificial Intelligence, Mike, Sullivan. (2003). 23. Safety Monitor: How to Protect Your Kids Online.

Parry, Aftab. (1999). 24. The Parent's Guide to Protecting Your Children in Cyberspace.

Nancy, E., Willard. (2007). 25. Cyber-Safe Kids, Cyber-Savvy Teens: Helping Young People Learn To Use the Internet Safely and Responsibly.

J., Bolton. (1999). 26. Keeping children safe on-line. Multimedia information & technology, Rashid, Tahir., Faizan, Ahmed., Hammas, Saeed., Shiza, Ali., Fareed, Zaffar., Christo, Wilson. (2019). 27. Bringing the kid back into YouTube kids: detecting inappropriate content on video streaming platforms. doi: 10.1145/3341161.3342913

Chet, Thaker. (2011). 28. Tools for mobile safety for children.

Saeed, Alqahtani., Wael, M., S., Yafooz., Abdullah, Alsaeedi., Liyakathunisa, Syed., Reyadh, Alluhaibi. (2023). 29. Children's Safety on YouTube: A Systematic Review. Applied Sciences, doi: 10.3390/app13064044

Tadikonda, Bala, Venkata, Naga, Abhya, Dattu., Pamarthi, Bharath, Prabhakar., Davuluri, HemaLatha, Chowdary., Oleti, Dolly, Sumanta., G., Navya, Sree. (2024). 30. Safe Browse Guardian. *International Research Journal on Advanced Science Hub*, doi: 10.47392/irjash.2024.031

Xianjun, Meng., Shaomei, Li., Muhammad, Mohsin, Malik., Qasim, Umer. (2022). 31. Machine-Learning-Based Suitability Prediction for Mobile Applications for Kids. Sustainability, doi: 10.3390/su141912400

A., Chiwanza., Fungayi, D., Mukoko., B., Mupini. (2024). 32. A Web Crawling and NLP-Powered Model for Filtering Inappropriate Content for Primary School Learners' Online Research. *International journal of innovative science and research technology,* doi: 10.38124/ijisrt/ijisrt24jul1083

Ta, Vinh, Thong. (2024). 33. A safety risk assessment framework for children's online safety based on a novel safety weakness assessment approach. arXiv.org, doi: 10.48550/arxiv.2401.14713

Dunja, Mladeni. (2000). 34. Machine learning for better Web browsing. (2022). 35. Machine Learning for Web Proxy Analytics. doi: 10.4018/978-1-6684-6291-1.ch045 (2023). 36. Safeguarding your Data from Malicious URLs using Machine Learning. *International Journal for Science Technology and Engineering*, doi: 10.22214/ijraset.2023.50907

Ahmad, El, Bakri., Abdullah, Yehia., Murtazza, Ali., Reem, Osama., Zeyad, Adel., Omar, Gamal., Sherine, Nagy, Saleh. (2024). 37. The Eye: An AI-Powered Video Streaming Platform to Protect Children from Inappropriate Content. doi: 10.1109/niles63360.2024.1075322

S., Yazhini. (2023). 38. Child Digital Monitoring and Controlling System. doi: 10.1109/ICAAIC56838.2023.101415

Mark, Maldonado., Ayad, Barsoum. (2019). 39. Machine Learning for Web Proxy Analytics. doi: 10.4018/IJCRE.2019070103

K., Gowsic., S, Siranjeevi., Nataliia, Shmatko., K., Swathi. (2024). 40. Web spoofing defense empowering users with phishcatcher's machine learning. ShodhKosh *Journal of Visual and Performing Arts,* doi: 10.29121/shodhkosh. v5.i3.2024.2713