IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Develop A ML Model Based Solution To Refine CAPTCHA

Ms. AMIRTHA PREEYA V[1], SHAIK ASLAM[2], SUJITHA REDDY[3], VISWESWAR REDDY[4]

[1] Assistant Professor in computer science and engineering & presidency University, Bangalore

[2]Student in computer science and engineering & presidency University, Bangalore

[3] Student in computer science and engineering & presidency University, Bangalore

[4]Student in computer science and engineering & presidency University, Bangalore

Abstract:

CAPTCHAs are widely used to differentiate between human users and automated bots, ensuring security in online interactions. However, traditional CAPTCHA systems often suffer from usability issues and vulnerabilities to evolving machine learning-based attacks. This research presents a novel machine learning-driven approach to refining CAPTCHA mechanisms, enhancing both security and user experience.

Our proposed solution leverages deep learning models to analyse and improve CAPTCHA complexity, making it more resistant to automated solvers while maintaining accessibility for genuine users. By utilizing advanced image processing and adaptive challenge generation techniques, the system dynamically adjusts CAPTCHA difficulty based on real-time threat analysis, reducing friction for legitimate users while thwarting bots.

Experimental results demonstrate that our approach significantly improves CAPTCHA robustness against automated attacks while ensuring a seamless experience for human users. This research contributes to the ongoing development of secure and user-friendly authentication mechanisms, bridging the gap between security and usability in modern web applications

Keywords: CAPTCHA Security,

Machine Learning,

Deep Learning,

Automated Bot Detection,

Adversarial attacks.

I. Introduction:

In the digital age, CAPTCHAs serve as a crucial line of defence against automated bots attempting to exploit online platforms. These challenges, designed to differentiate between human users and machines, are widely used in login systems, online transactions, and content submissions. However, traditional CAPTCHAs often compromise user experience, being either too difficult for humans to solve or too weak to resist modern AI-driven attacks. This has led to an ongoing challenge: creating CAPTCHA systems that are both user-friendly and secure.

With the rapid advancements in artificial intelligence, especially in deep learning and computer vision, automated bots have become increasingly capable of solving conventional CAPTCHA tests. Optical Character Recognition (OCR) technologies and adversarial machine learning techniques have rendered many traditional CAPTCHAs ineffective. This necessitates a shift towards more sophisticated CAPTCHA mechanisms that can adapt dynamically to emerging threats while maintaining accessibility for legitimate users.

This research presents a machine learning-based solution to refine CAPTCHA security by leveraging deep learning techniques for enhanced challenge generation and analysis. By implementing adaptive CAPTCHA models that adjust complexity based on real-time threat detection, our approach aims to strike a balance between security and user convenience. Through this study, we explore how AI can be used not only to break CAPTCHAs but also to strengthen them, ultimately contributing to the development of more robust and user-friendly authentication systems.

A. Background knowledge

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) was introduced as a security measure to prevent automated bots from accessing or manipulating online services. Traditional CAPTCHAs include text-based challenges, image recognition tasks, and interactive puzzles, each designed to exploit the cognitive differences between humans and machines. However, with advancements in artificial intelligence, especially in deep learning, many conventional CAPTCHAs have become vulnerable to automated solvers, necessitating the development of more sophisticated approaches.

Machine learning, particularly deep learning-based models, has played a dual role in the evolution of CAPTCHA systems. On one hand, convolutional neural networks (CNNs) and optical character recognition (OCR) techniques have significantly improved the ability of AI to break text-based CAPTCHAs. On the other hand, AI-driven approaches have also contributed to the development of more secure and adaptive CAPTCHAs. Techniques such as adversarial machine learning, generative adversarial networks (GANs), and behavioural analysis have been explored to enhance CAPTCHA robustness against automated attacks.

In addition to security concerns, usability remains a key challenge in CAPTCHA design. Studies have shown that overly complex CAPTCHAs lead to frustration among users, reducing engagement and accessibility, especially for individuals with disabilities. As a result, modern CAPTCHA solutions are moving toward AI-powered adaptive mechanisms that balance security with user experience. This research builds upon existing CAPTCHA systems

and explores the potential of machine learning to refine CAPTCHA challenges, making them both more secure against AI-driven attacks and more accessible for human users.

B. LITERATURE SURVEY

Breaking the Code [1] examines the vulnerabilities of traditional CAPTCHA systems and how machine learning models have been used to bypass them.

AI vs. CAPTCHA [2] explores the evolution of CAPTCHA technologies in response to advancements in deep learning and automated bot detection.

Human or Bot? [3] delves into different CAPTCHA mechanisms, their usability challenges, and the effectiveness of AI-driven security measures.

Adaptive Security [4] reflects on modern AI-powered CAPTCHA solutions, highlighting adaptive challenge generation and real-time threat analysis.

The Future of CAPTCHA [5] explores the integration of AI and cybersecurity to develop more robust and user-friendly authentication systems.

.

II. Existing Methods

Several CAPTCHA techniques have been developed to prevent bots from accessing online services. **Text-based CAPTCHAs** were one of the first methods, requiring users to recognize distorted letters and numbers. However, with advancements in Optical Character Recognition (OCR) and AI models, these have become easier for bots to break. **Image-based CAPTCHAs**, like selecting objects in pictures, were introduced as an alternative, but deep learning models can now recognize images with high accuracy.

To improve accessibility, **audio CAPTCHAs** were created, where users listen to distorted speech and type what they hear. However, speech recognition algorithms have made these less secure. **Interactive CAPTCHAs**, such as dragging puzzles or analysing mouse movements, offer better security but can still be mimicked by advanced bots.

While these methods aim to distinguish humans from bots, they often compromise either security or user experience. This research explores a machine learning-based approach to refine CAPTCHA, making it both more secure and user-friendly.

Disadvantages:

- **Vulnerability to AI Attacks** – Advanced machine learning models, such as OCR for text-based CAPTCHAs and CNNs for image-based CAPTCHAs, can easily break traditional CAPTCHA systems.
- **Poor User Experience** – Many CAPTCHAs are difficult to solve, leading to frustration among users. Complex distortions, image selections, or lengthy puzzles can negatively impact accessibility.

- **Accessibility Issues** – Audio CAPTCHAs, designed for visually impaired users, are often difficult to understand due to background noise and distortion, making them ineffective for many users.
- **Increased Bot Sophistication** – Modern bots can mimic human behavior, such as mouse movements and click patterns, reducing the effectiveness of interactive CAPTCHAs.
- **Time-Consuming** – Many CAPTCHAs require multiple attempts or take too long to solve, leading to delays in user authentication and negatively affecting website usability.

III. Proposed Methods

To overcome the limitations of existing CAPTCHA systems, this research proposes a **machine learning-based adaptive CAPTCHA** that enhances security while maintaining user-friendliness. The system dynamically generates CAPTCHA challenges using deep learning models, adjusting difficulty based on real-time threat detection. By leveraging **generative adversarial networks (GANs)** and **behavioural analysis**, the CAPTCHA adapts to evolving bot capabilities, making it harder for AI-driven attacks to succeed.

Unlike traditional CAPTCHAs, the proposed method focuses on **user behaviour analysis**, such as mouse movement patterns, keystroke dynamics, and response times, to distinguish humans from bots. Additionally, the system integrates an **adaptive challenge mechanism**, ensuring that genuine users face minimal friction while bots encounter increasingly complex tasks. This approach enhances security without compromising accessibility, providing a seamless experience for legitimate users.

Advantages:

- **Enhanced Security** – The adaptive CAPTCHA evolves with AI-driven threats, making it difficult for bots to bypass.
- **Improved User Experience** – Challenges are adjusted based on user behaviour, reducing frustration and solving time.
- **Accessibility-Friendly** – Unlike traditional CAPTCHAs, this method minimizes reliance on text and audio, making it more inclusive.
- **Real-Time Threat Detection** – AI continuously monitors and adapts CAPTCHA difficulty based on bot activity patterns.
- **Efficient and Dynamic** – The system ensures a balance between security and usability by dynamically modifying challenge complexity.

IV. Methodology

This section outlines the technologies used, system design, workflow, homepage navigation, database management, user authentication, and data security aspects of the proposed machine learning-based CAPTCHA system.

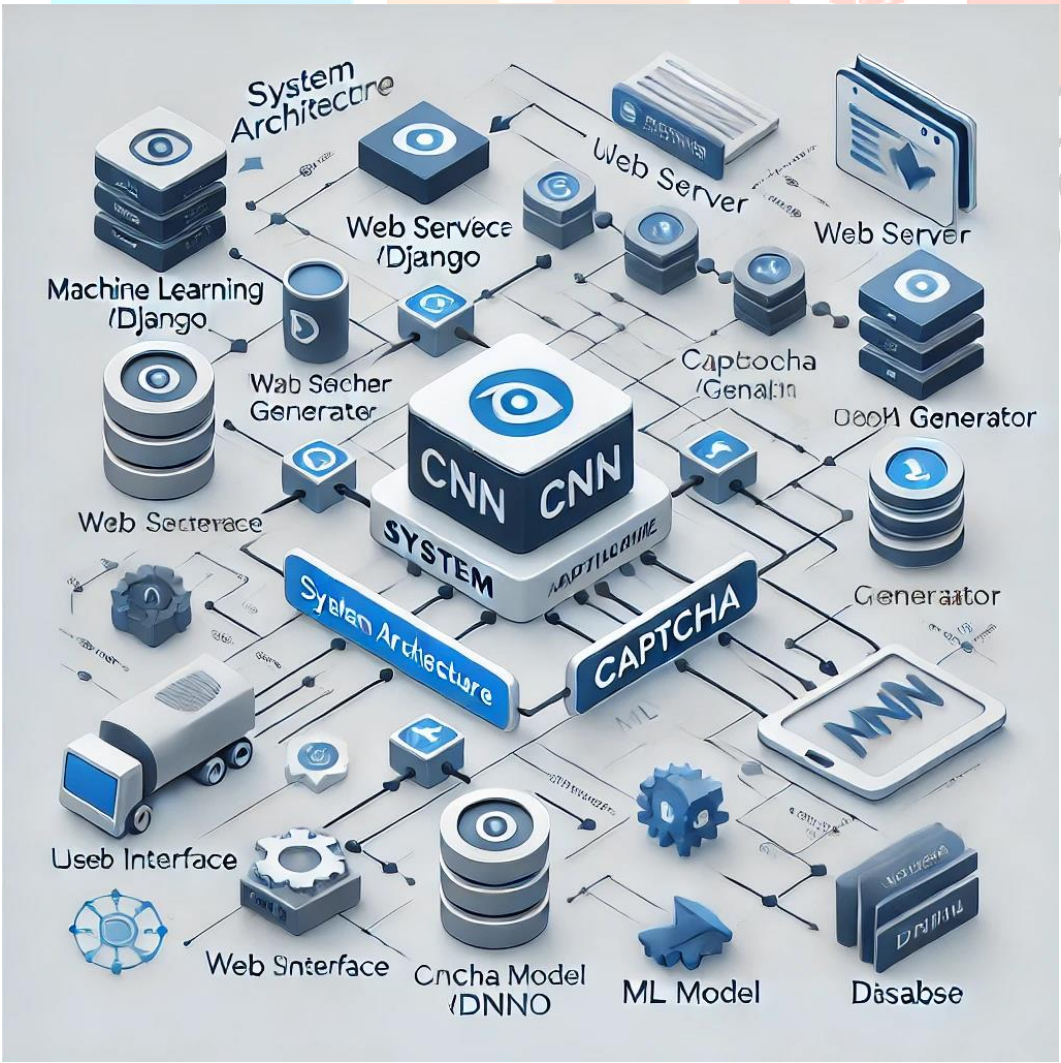
1. Technologies Used

- **Programming Language:** Python (for backend and AI model development).
- **Deep Learning Model:** Convolutional Neural Networks (CNN) (for CAPTCHA classification and security enhancement)
- **Web Framework:** Flask (for integrating CAPTCHA into web applications)
- **Frontend:** HTML, CSS, JavaScript (for user interaction and CAPTCHA display)

2. System Design and Workflow

- The system first **analyses user behaviour**, such as mouse movements, keystroke dynamics, and response time, to classify whether the user is a human or a bot.
- Based on this classification, an **adaptive CAPTCHA** is generated, where challenge difficulty increases if bot-like behaviour is detected.
- The CAPTCHA **verification process** involves a **CNN model** evaluating the user's responses, ensuring security and ease of access for legitimate users.
- The system continuously **learns from failed bot attempts**, improving CAPTCHA challenges dynamically.

Architecture diagram:



Homepage Navigation:

- The homepage includes a **login/register section** with the CAPTCHA challenge integrated.
- If the system detects a legitimate user, a simpler CAPTCHA (e.g., one-click verification) is displayed.
- If suspicious activity is detected, a more complex CAPTCHA (e.g., image-based, puzzle-solving) is shown.

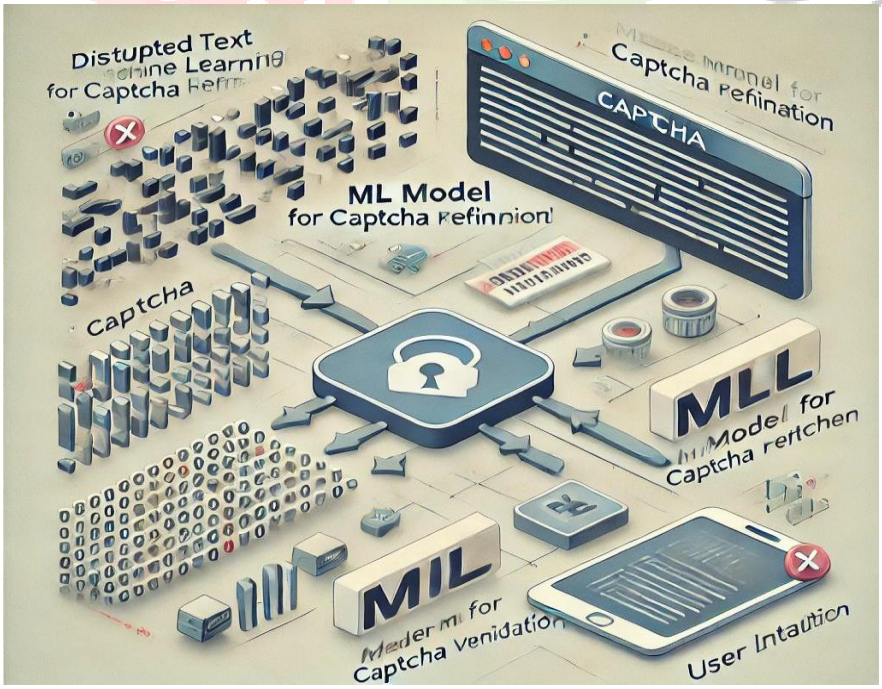
3. Database Management

- User authentication data and CAPTCHA challenge results are stored securely using **SQL databases**.
- CAPTCHA logs are maintained for **pattern analysis and system improvement**.
- Sensitive information is **hashed and encrypted** to prevent unauthorized access.

4. User Authentication and Data Security

- The authentication process uses **multi-layered verification**, including CAPTCHA, password hashing, and optional two-factor authentication.
- Data transmission is secured using **SSL/TLS encryption** to protect against man-in-the-middle attacks.
- The system continuously **monitors suspicious activities**, such as repeated failed CAPTCHA attempts, and takes countermeasures like IP blocking.

Additional Visualizations



V. Requirements

Software Requirements:

- Operating System: Windows, Linux, or macOS.
- Development Environment**:** Python 3.x, Flask framework.
- Database Management System: MySQL/PostgreSQL.
- Web Technologies: HTML, CSS, JavaScript.

Hardware Requirements:

- **Processor:** Intel i5/i7 or AMD Ryzen 5/7 (or higher)
- **RAM:** Minimum 8GB (Recommended: 16GB for faster training and processing)
- **Storage:** Minimum 50GB free space (Recommended: SSD for better performance)

VI. Conclusion

The proposed machine learning-based CAPTCHA system presents an innovative approach to distinguishing between human users and automated bots. By integrating deep learning with adaptive CAPTCHA mechanisms, the system enhances both security and usability. The use of CNN models for CAPTCHA verification ensures robust protection against automated attacks while maintaining an intuitive experience for genuine users.

Furthermore, leveraging Flask for backend integration allows seamless deployment of CAPTCHA challenges on various web applications. The incorporation of behavioural analysis adds an extra layer of security, making it harder for bots to bypass the system. With an adaptive approach, the CAPTCHA dynamically adjusts difficulty based on detected threats, thereby reducing frustration for legitimate users while strengthening security.

In conclusion, this research contributes to the evolving field of cybersecurity by proposing a smarter and more effective CAPTCHA system. By continuously learning from user interactions and bot behaviours, the system remains resilient against evolving automated threats, ensuring a more secure and efficient web experience for users worldwide.

VII. References

1. "Deep Learning Based CAPTCHA Recognition Network with Grouping Strategy" by J. Zhang et al. (2023)
2. "Recognition of CAPTCHA Characters Using Machine Learning Algorithms" by A. Patel et al. (2022)
3. "CAPTCHA Recognition Using Machine Learning and Deep Learning Techniques" by R. Kumar et al. (2021)
4. "Recognition of CAPTCHA Characters by Supervised Machine Learning Algorithms" by O. Bostik et al. (2020)
5. "Deep-CAPTCHA: A Deep Learning Based CAPTCHA Solver for Vulnerability Assessment" by T. Nguyen et al. (2020)
6. " A CAPTCHA Recognition Technology Based on Deep Learning" by X. Li et al. (2019)

7. "EnSolver: Uncertainty-Aware CAPTCHA Solver Using Deep Ensembles" by M. Khan et al. (2023)

8. "Text Based CAPTCHA Recognition Using Machine Learning and Deep Learning" by S. Rao et al. (2023)

9. “Vulnerability Analysis of CAPTCHA Using Deep Learning" by L. Zhao et al. (2023)

10. "Breaking reCAPTCHAv2: A Study on CAPTCHA Security" by J. Liu et al. (2024)