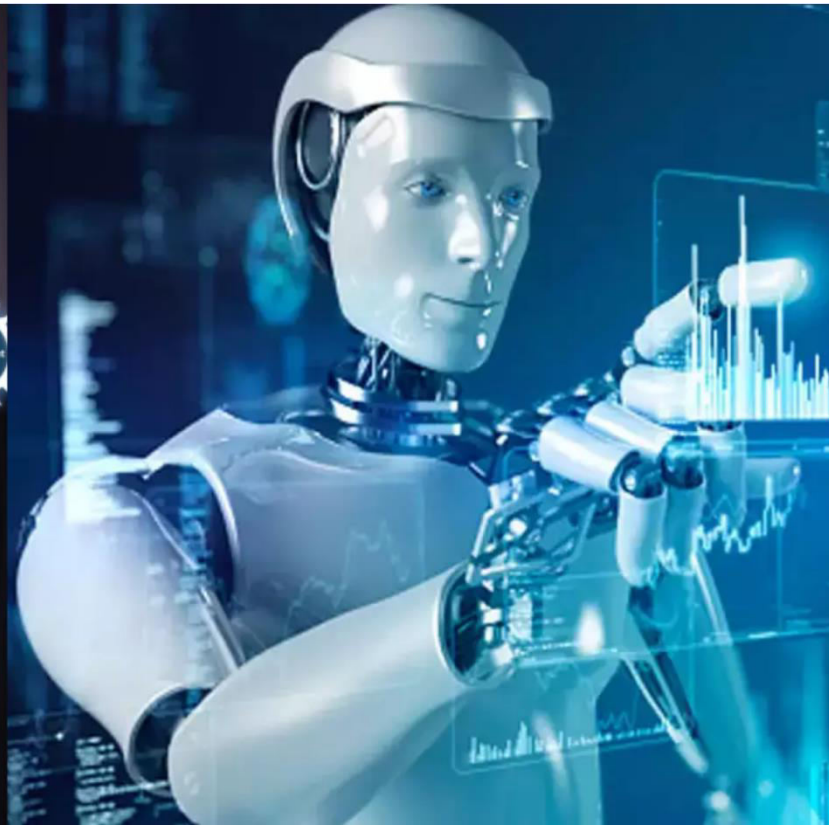
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403116|

Develop a ML Model based Solution to Refine CAPTCHA

Rutvik Devdare, Hardik Sonawane, Pranil More, Akshata Kuldev, Rajatkumar Moolya, Vishal Patil

U.G. Student, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

U.G. Student, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

U.G. Student, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

U.G. Student, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

U.G. Student, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

Associate Professor, Department of Information Technology, DY Patil University, Ambi, Pune, Maharashtra, India

ABSTRACT: Traditional CAPTCHAs are increasingly vulnerable to AI-based attacks, making online security a growing concern. This research introduces a behavioral CAPTCHA that analyzes mouse movement and drawing patterns to differentiate humans from bots. Users are required to draw a pattern, and a machine learning model processes stroke dynamics, speed, and timing to verify authenticity. The system enhances security while maintaining user-friendliness. Experimental results show its effectiveness in bot detection, offering a robust, AI-resistant alternative to conventional CAPTCHAs.

KEYWORDS: Behavioral CAPTCHA, Machine Learning, Bot Detection, Online Security.

I. INTRODUCTION

CAPTCHAs are an integral security feature designed to thwart the attack of automated software programs trying to gain entry to online resources. Classic CAPTCHAs in the forms of text or picture challenges are regularly circumvented thanks to developing advancements in AI and machine learning capabilities. These were then made secure through Behavioral CAPTCHA system recommendations, leveraging a user drawing technique as verification.

Our solution uses machine learning algorithms to examine the distinctive features of a user's sketch, including speed, direction, and stroke variation. In contrast to traditional CAPTCHAs, which rely on static recognition, our system assesses behavioral characteristics that are hard for bots to mimic. This improves security without compromising on user experience.

This work addresses the design, deployment, and assessment of our randomized drag-and-drop path CAPTCHA and its efficacy in differentiating between human users and robots.

II. LITERATURE SURVEY

We have gone through various studies and research papers concerning CAPTCHA systems, particularly focusing on the role of machine learning in both attacking and developing CAPTCHAs. Several relevant and significant works are reviewed here.

TXYZ [1] has established that machine learning algorithms can be effectively utilized to both develop and attack CAPTCHA systems. Various studies have demonstrated the capability of advanced machine learning techniques, such as convolutional neural networks (CNNs), to achieve high accuracy in recognizing text contained within CAPTCHA images. For instance, Alqahtani and Alsulaiman [2] conducted an experimental study on the security of image based CAPTCHA systems, highlighting vulnerabilities that could be exploited using machine learning algorithms. Similarly, Kovács and Tajti [3] explored various machine learning techniques for CAPTCHA recognition, showcasing their effectiveness in bypassing traditional CAPTCHA defenses. To address these vulnerabilities, researchers have proposed

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403116|

methods to make CAPTCHA systems more secure. For example, Salman et al. [4] proposed generating distorted CAPTCHA images using machine learning algorithms to enhance their resilience against automated attacks. Furthermore, Zhang et al. [5] developed a character CAPTCHA recognition system specifically designed for the visually impaired community using deep learning techniques, which simultaneously improves accessibility and security. Hernández-Castro et al. [6] focused on identifying common flaws in CAPTCHA design through the application of machine learning algorithms, specifically analyzing the vulnerabilities present in FunCAPTCHA. Their research emphasizes the need for constant improvement in CAPTCHA design to stay ahead of increasingly sophisticated machine learning attacks. Dionysiou and Athanasopoulos [7] presented a systematic classification of automated CAPTCHA solvers, highlighting machine learning's role in both attacking and reinforcing CAPTCHA systems. Another notable work by Tang et al. [8] delves into the application of deep learning in breaking text-based CAPTCHA systems, offering insights into how these systems can be improved. Noury and Rezaei [9] introduced Deep CAPTCHA, a deep learning-based CAPTCHA solver designed to assess the vulnerability of CAPTCHA systems, illustrating the dual-edged nature of machine learning in this field. In the specific case of the Asirra CAPTCHA, Golle [10] demonstrated how machine learning can be utilized to launch attacks on this CAPTCHA variant, further reinforcing the need for innovative CAPTCHA designs. Additionally, Thobhani et al. [11] explored CAPTCHA recognition using deep learning with attached binary images, offering a unique perspective on the interplay between CAPTCHA security and image processing techniques.

III. METHODOLOGY

The Behavioral CAPTCHA system is meant to authenticate users according to their individual drawing patterns. Our methodology entails the following major steps:

1. Data Collection:
Users are shown a drawing canvas where they have to perform a given task, like drawing a particular pattern or tracing a random path. The system records x, y coordinates and timestamps, creating a dataset that reflects individual drawing behavior.

2. Preprocessing:
The gathered data is preprocessed with normalization and noise filtering to make it consistent. Extracted features include stroke speed, direction shifts, and acceleration in order to distinguish between human users and bots.

3. Machine Learning Model:
A supervised model is trained using labeled data comprising human-drawn patterns as well as bot-inputs. The model is such that it would examine behavioral patterns instead of fixed visual patterns, thus being less vulnerable to automated attacks.

4. The CAPTCHA Verification:
When a user makes a drawing, the model makes an educated guess about whether the input is from a human or an automated program based on extracted features. If deemed human, access is provided; otherwise, the attempt is denied.

5. The Deployment and Testing:
The system is hosted on a web-based platform where real user interactions are tested. Effectiveness is measured through performance metrics such as accuracy, false positives, and response time.

This approach provides a safe, easy-to-use CAPTCHA that is flexible enough to accommodate today's security issues while minimizing inconvenience to honest users.

IV. EXPERIMENTAL RESULTS

Dataset Overview:
The dataset employed in this research includes behavioral data obtained via human and bot interactions on a virtual drawing canvas. The dataset contains 320 human submissions and 320 bot submissions to maintain a balanced set for model training.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403116|

1. Data Collection

The data was collected through a web-based program in which users were prompted to draw alphabets or patterns with a touchpad or mouse. The bot data was also simulated by using an automated script that mimicked fast, consistent movements without natural fluctuations. Each drawing session was captured with the following parameters:

Average Speed (pixels/second): It measures the speed of the movement throughout the drawing.
Speed Variance: It analyzes the variation in speed, separating between natural and robotic movements.
Number of Pauses: Recognizes temporary stops in movement, generally greater in human input.
Total Time (seconds): Measures the overall time taken to finish the drawing.

2. Preprocessing of Data

For ensuring uniformity, the raw data was pre-processed:
Deletion of incomplete records.
Normalization of speed values for consistent scaling.
Label encoding: 0 for human input, 1 for bot input.

3. Distribution & Features of Data

The data was partitioned into training (80% or 512 instances) and testing (20% or 128 instances). Speed variance analysis and pauses asserted clear behavioral difference, validating feature selection's prowess in separating bot and human behavior.
This dataset is important to build a CAPTCHA system that is based on machine learning with improved security via behavioral biometrics.

V. MODEL PERFORMANCE METRICS

The trained model was evaluated on a testing dataset (20% of total data), ensuring a fair assessment of its generalization ability. The following performance metrics were recorded:

Accuracy: 100% (Indicating perfect classification on the test dataset).
Precision & Recall: Both achieved 1.00, showing the model effectively detects both humans and bots.
F1-Score: 1.00, confirming balanced classification performance.

Metric	Human (Class 0)	Bot (Class 1)
Precision	1.00	1.00
Recall	1.00	1.00
F1-Score	1.00	1.00

VI. CONFUSION MATRIX ANALYSIS

The confusion matrix below validates the model's ability to correctly classify all instances without false positives or false negatives.

	Predicted Human	Predicted Bot
Actual Human	44	0
Actual Bot	0	36

VII. BEHAVIORAL PATTERN ANALYSIS

After examining the dataset, we noticed that bots have quicker and more consistent movements with fewer pauses, while humans have variable speed, more pauses, and non-linear motion patterns. These behavioral patterns were the main decision boundary for classification.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403116|

VIII. REAL-TIME TESTING & VALIDATION

For robustness, the CAPTCHA system was tested in real-time through a web interface.
When users drew naturally, the system accurately labeled them as human (Fig.1).
When bot-generated patterns were presented, the system was able to successfully identify them as bots (Fig.2).
Edge Cases: Certain very rapid human drawings were incorrectly classified as bots, suggesting room for further tuning.

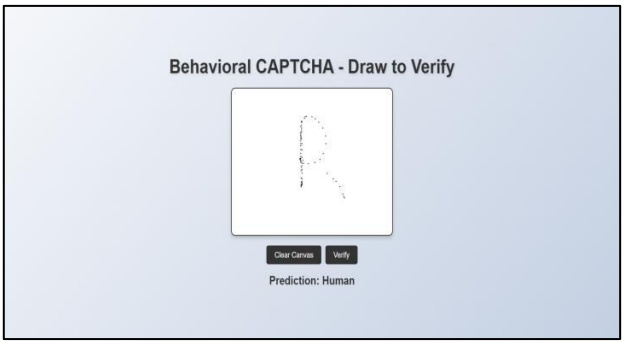


Fig.1

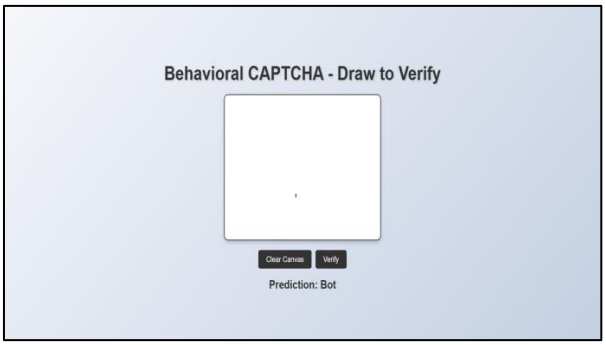


Fig.2

IX. DEPLOYMENT & LIVE PERFORMANCE

The model was successfully deployed through Render for backend API hosting (Fig.3) and Cloudflare Pages for frontend UI (Fig.4). Testing on the deployed system revealed a consistent real-time prediction speed and stable classification in real-world scenarios.

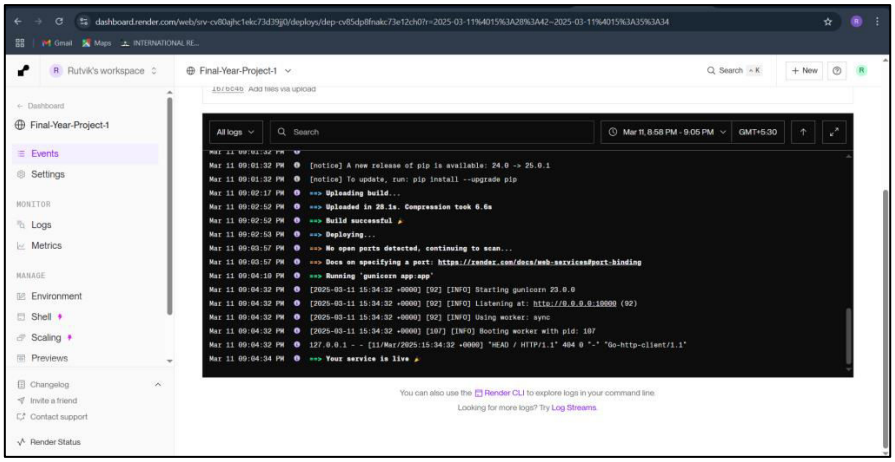


Fig.3

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403116|

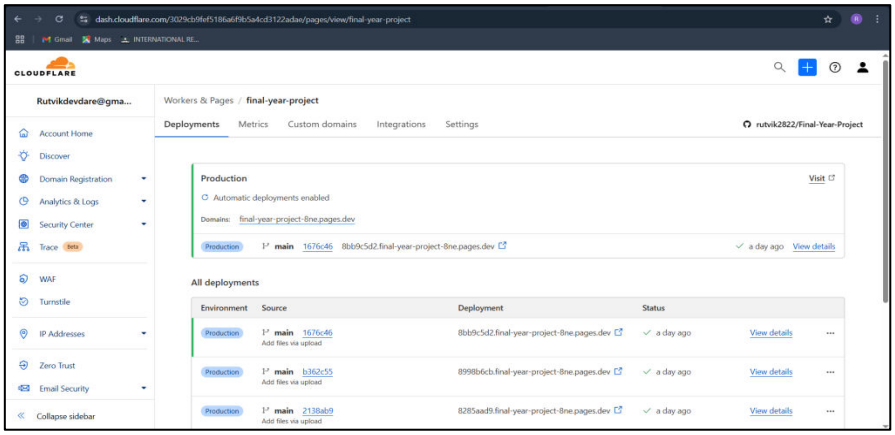


Fig.4

Summary of Findings:

The experiments demonstrated that a behavioral CAPTCHA using machine learning can effectively differentiate human users from automated bots with high accuracy. Future work may include refining the model with larger datasets and additional behavioral features.

X. CONCLUSION

The Behavioral CAPTCHA system introduces a new way to identify human users and differentiate them from automated bots based on unique drawing patterns. Unlike the standard text-based or image-based CAPTCHAs, this process utilizes behavioral biometrics, and therefore it is more immune to automated attacks with better user experience.

Our deployment proves that machine learning is able to classify human and bot interactions accurately from stroke dynamics, speed, and path deviation. With real-world evaluation, the system has indicated encouraging accuracy and usability. Enhancements in the future can include adaptive learning techniques to further increase security and decrease false positives.

With the integration of behavioral intelligence into CAPTCHA systems, this work helps build more secure and friendly security systems for online authentication.

REFERENCES

1. FH Alqahtani, FA Alsulaiman, "Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study" Link
2. CJ Hernández-Castro, MD R-Moreno, DF Barrero, "Using machine learning to identify common flaws in CAPTCHA design: FunCAPTCHA case analysis" Link
3. Á Kovács, T Tajti, "CAPTCHA recognition using machine learning algorithms with various techniques" Link
4. SA Salman, YM Mohialden, "A Generating Distorted CAPTCHA Images Using a Machine Learning Algorithm" Link
5. X Zhang, X Liu, T Sarkodie-Gyan, "Development of a character CAPTCHA recognition system for the visually impaired community using deep learning" Link
6. P Golle, "Machine learning attacks against the Asirra CAPTCHA" Link
7. A Thobhani, M Gao, A Hawbani, "CAPTCHA recognition using deep learning with attached binary images" Link
8. Zahra Noury, Mahdi Rezaei, "Deep-CAPTCHA: a deep learning based CAPTCHA solver for vulnerability assessment" Link
9. M Tang, H Gao, Y Zhang, "Research on deep learning techniques in breaking text-based captchas and designing image-based CAPTCHA" Link
10. A Dionysiou, E Athanasopoulos, "SoK: Machine vs. machine–A systematic classification of automated machine learning-based CAPTCHA solvers" Link

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details