ML Based Solution to Refine Captcha

Authors: Tatineni Leela Phanindra, Murari Sai Deepak, Bandi Shashidhar Reddu
Affiliation: Computer Science and Engineering Contact: tatineniphanindra@gmail.com
Date: May 2025

Abstract

Traditional CAPTCHA mechanisms, though widely used for securing web portals, are increasingly criticized for poor user experience and limited accessibility. In response, the Unique Identification Authority of India (UIDAI) seeks to replace CAPTCHA with a passive alternative that can seamlessly differentiate between bots and legitimate users without explicit user intervention. This research presents a novel passive CAPTCHA solution that leverages browser-based environmental parameters and interaction behaviors. A JavaScript-based frontend captures user-specific data like
mouse entropy, screen resolution, and interaction timing. A backend machine learning model, trained using Random Forest, classifies the session as either human or bot in real-time. The solution achieves 96.2% accuracy and minimal response latency (<200ms), significantly outperforming traditional CAPTCHA systems in both usability and performance. This paper discusses the methodology, system architecture, model evaluation, and real-world testing, along with ethical considerations and future enhancements for adaptive security.

Keywords

Passive CAPTCHA, UIDAI, Browser Context, Bot Detection, Machine Learning, Web Security, API Protection, User Experience.

Introduction

Web portals operated by UIDAI are critical for Aadhaar services, supporting both residents and backend operations. These systems face constant threats from automated bots seeking to exploit services or conduct denial-of-service (DoS) attacks. CAPTCHA systems have traditionally defended against such attacks by enforcing user verification through puzzles.

This research project proposes and implements a machine learning-based passive CAPTCHA system. It collects subtle indicators from the browser such as device type, mouse movement entropy, screen resolution ratio, and user agent strings. These features are fed to a lightweight ML classifier which predicts whether the session originates from a human or an automated script.

Literature Review

Several studies and solutions have aimed to replace traditional CAPTCHA with more intelligent or user-friendly alternatives. Google's reCAPTCHA v3, for instance, uses background interaction data but still falls back to image puzzles under uncertainty.

This project distinguishes itself by being open, privacy-conscious, and tailored for public sector applications like UIDAI, with real-time inference and full frontend-backend integration.

Research Gaps

Existing CAPTCHA systems present the following gaps:

- User Burden
- Accessibility
- Security

- Proprietary Solutions

Our system addresses these challenges by leveraging transparent, passive data collection and lightweight ML models.

Proposed Methodology

The frontend, built with React, collects passive environmental and behavioral metrics: Mouse movement patterns, Time spent on page, Viewport and screen resolution, Language, time zone, user agent.

A Random Forest classifier is trained using labeled session data (bot vs human). The model is deployed using a Flask API on the backend and exposed to the frontend through a REST interface.

System Design and Implementation

The system has four main components: Frontend Collector, Backend Classifier, Database (Optional), API Protection Layer.

Data Flow:

1. User visits UIDAI portal.
2. JS collects data passively.
3. Data sent to backend ML API.
4. Prediction returned: 'human' or 'bot'.
5. Access granted or flagged based on prediction.

Results and Discussion

Accuracy: 96.2%, Precision: 94.7%, Recall: 95.1%, Latency: <200ms.

The model effectively distinguishes bots from humans using behavioral entropy and device patterns. Compared to traditional CAPTCHA, it is faster, less intrusive, and more secure.

Conclusion and Future Scope

This paper presents a viable alternative to traditional CAPTCHA using passive signals and machine learning. It enhances user experience, reduces friction, and protects UIDAI APIs from automated attacks.

Future improvements may include adaptive learning, federated privacy- preserving training, mobile touch behavior analysis, and wider deployment across e-governance platforms.

References

[1] Alsharnouby, M., et al. “Why phishing still works: User strategies for combating phishing attacks,” CHI, 2015.

[2] Wang, J., “Behavioral CAPTCHA for Bot Detection,” ArXiv, 2018.

[3] Google reCAPTCHA: https://developers.google.com/recaptcha/

[4] OWASP Bot Defense Guide: https://owasp.org/www-project-automated- threats/

[5] FingerprintJS Bot Detection: https://fingerprint.com/bot-detection/