# CAPTCHA and Human Verification Systems: Are They Still Effective?

**Title Page**

*Title: CAPTCHA and Human Verification Systems: Are They Still Effective? A Critical Review of Technological Obsolescence, Usability Trade-offs, and Future Verification Paradigms*

Student Name:

Registration Number:

Department: Department of Interdisciplinary Studies

Faculty: Faculty of Information Technology

University: University of Moratuwa

Course: IS 3001 Scientific Communication, Assignment III

Date: 19th October 2025

## Abstract

The CAPTCHA system (Completely Automated Public Turing test to tell Computers and Humans Apart) was first created to establish a security boundary in the digital world. Its goal was to prevent automated abuse by setting challenges that are easy for people but difficult for machines.1 This paper critically evaluates whether CAPTCHA systems, across their different generations, can still maintain this key security role given the rapid growth of Artificial Intelligence (AI). This review traces how the technology evolved from distorted text CAPTCHAs to the current invisible behavioral scoring systems. Crucially, academic studies show that the core security idea—that machines cannot solve these puzzles—has failed. AI solvers are now achieving near-perfect success rates against image-based challenges.2 Furthermore, the analysis looks at the serious problems involving usability, accessibility, and the ethical issues of centralized user tracking, as seen in systems like reCAPTCHA v3.5 It is strongly argued that the CAPTCHA approach is becoming technologically outdated and economically inefficient. This is mainly because the cost to attack the system is now much lower than the cost to defend it. The paper concludes by arguing for a necessary shift toward AI-resistant alternatives, such as combining behavioral biometrics, using economically difficult Proof-of-Work systems, and adopting Zero-Knowledge Proofs for private, decentralized verification.7

## Keywords

CAPTCHA, Human Verification, Cybersecurity, AI Bypass, Usability, Web Security

## 1. Introduction

### 1.1 Background: The Idea of Distinguishing Humans from Machines

The impetus for creating automated systems capable of reliably telling the difference between human users and malicious computer programs arose when critical internet services became vulnerable to large-scale automated attacks. The early 2000s saw the internet increasingly used for services vulnerable to mass account creation, spam, and data scraping. The philosophical concept underpinning the solution was rooted in the **Turing Test**, which suggests that if a machine’s observed behavior is indistinguishable from a human’s, it could be considered intelligent.10 The goal was to invert this concept: create a test a human can pass, but a machine cannot.

In the early 2000s, researchers including Luis von Ahn formally introduced and defined the **CAPTCHA**, designing it as an automated challenge that specifically uses 'hard AI problems for security'.1 The core premise was to institutionalize an Automated Turing Test—a Human Interactive Proof (HIP)—that was easy for humans but computationally infeasible for computer programs at that time. This system aimed to use the 'computational abilities' of humans to achieve security goals, diverting the collective time and effort of the internet population into a defense mechanism.

The very first large-scale deployment of this concept was developed by Andrei Broder at AltaVista. This filter system generated an image of randomly distorted printed text that machine vision (Optical Character Recognition, or OCR) systems could not reliably read, but a human could. This early success was profound; after deployment, it reportedly reduced the number of spam URLs by 95%. This firmly established the CAPTCHA’s role as the major foundational defense tool in the rapidly growing field of web security.11

### 1.2 Importance and Evolution in Modern Technology

Following its successful introduction, CAPTCHA technology quickly became an indispensable component of global IT security infrastructure. It is critical for mitigating numerous forms of malicious automation, including preventing mass creation of fraudulent email accounts, suppressing dictionary attacks, safeguarding online polls from manipulation, and providing a defense against Distributed Denial of Service (DDoS) campaigns.12 CAPTCHAs thus became the industry standard for ensuring that only genuine human agents could interact with and consume protected web resources.

The evolution of verification systems demonstrated a shifting application beyond simple defense. For example, Google’s acquisition of the reCAPTCHA technology strategically used the human effort expended in solving image and text challenges not just for security, but also to aid in the digitization of vast archives, such as *The New York Times* and books from Google Books. This demonstrated the technology’s powerful capacity for large-scale "human computation," simultaneously protecting services while contributing to monumental data labeling and infrastructure projects—a dual function that highlighted its reach and societal impact.8

### 1.3 Problem Statement: The Crisis of Effectiveness

Despite its historical significance, the current utility of CAPTCHA systems is now confronted by a severe crisis of effectiveness. The foundational security premise—that the required cognitive task is insurmountable for machines—is being systematically dismantled by the exponential progression of Artificial Intelligence, particularly deep learning and advanced computer vision techniques.

The crisis is characterized by a dual failure: technical security obsolescence and a severe degradation of user experience. While advanced bots are achieving near-perfect solve rates, legitimate human users are confronted with escalating friction, leading to unacceptable failure rates and significant accessibility barriers.5 Furthermore, the attempt to resolve user friction through "invisible" systems (such as reCAPTCHA v3) introduces complex privacy dilemmas, forcing a difficult trade-off that erodes user trust.6 This paper critically reviews this simultaneous failure and evaluates whether contemporary verification systems can still realistically meet their original security mandate. The analysis specifically argues that the economic and technological foundations supporting CAPTCHA defense are now broken, requiring the industry to look for entirely new verification methods.

### 1.4 Alignment with Prior Project Work

The arguments presented in this critical review are deeply informed by the practical insights gained during a prior web security project that involved implementing and empirically testing a friction-based human verification system. The real-world deployment of this mechanism revealed an acute, operational limitation: any measured attempt to escalate the challenge difficulty to enhance security against known bot attacks resulted in an immediate and unacceptable increase in legitimate user failure and abandonment rates. This observation underscored the core systemic failure documented in the broader academic literature—that the usability and security parameters of cognitive CAPTCHAs are inherently inversely related. This real-world conflict serves as the primary rationale for undertaking a comprehensive academic evaluation of the systemic flaws inherent in current verification methodologies.

### 1.5 Objectives and Structure

This critical review is structured to achieve the following objectives: 1) To systematically trace and categorize the technological evolution of CAPTCHA through its distinct generations; 2) To critically evaluate the empirical collapse of cognitive CAPTCHAs in the face of modern AI capabilities; 3) To analyze the systemic ethical, accessibility, and usability compromises inherent in both visible and invisible verification systems; and 4) To propose and rigorously evaluate viable AI-resistant and privacy-centric alternatives poised to define the future of web security.

The paper is structured as follows: Section 2 reviews the literature on the history and failure of different CAPTCHA generations. Section 3 presents a critical evaluation and detailed discussion of the economic, ethical, and systemic flaws, supported by evidence. Section 4 concludes the review by summarizing the main arguments and suggesting crucial future research directions in verification technology.

## 2. Literature Review: The Fragility of Verification Paradigms

The development of CAPTCHA systems can be segmented into three distinct technological generations, each defined by the complexity of the AI problem leveraged for the verification task, and each ultimately undermined by the progression of machine learning.

### 2.1 Generation I: Text-Based CAPTCHAs and OCR Obsolescence (The Early Failure)

The inaugural generation of CAPTCHAs was founded exclusively on text recognition challenges, epitomized by systems like GIMPY. These tests required users to accurately transcribe severely distorted images of printed text. The security premise relied on the difficulty presented to Optical Character Recognition (OCR) systems by various shape deformations, background interference, and occlusions.1 Early research by von Ahn and colleagues showed that generating random, distorted text was a sufficient hurdle for the OCR systems of the early 2000s.

However, this initial security measure failed quickly as machine learning algorithms matured. The inherent simplicity of the underlying challenge—recognizing a limited set of characters—provided a clear, solvable pathway for automation. As advanced deep learning architectures, particularly Convolutional Neural Networks (CNNs), became widespread, the security of text-based CAPTCHAs effectively dissolved. Studies showed that specialized CNNs could be applied to solve reCAPTCHA, reporting success rates as high as 99.8%.16 The evidence confirmed that the original "hard AI problem" that powered Generation I was only temporarily difficult, leading inexorably to its systematic replacement.

### 2.2 Generation II: Image and Cognitive CAPTCHAs (The Semantic Challenge)

In direct response to the defeat of text-based systems, the CAPTCHA paradigm evolved to incorporate more complex cognitive tasks, shifting the burden from character recognition to semantic object recognition and image classification.9 This strategic pivot inaugurated Generation II systems, most famously reCAPTCHA v2, which mandated that users perform tasks such as selecting all images that contained a specific object (e.g., traffic lights, crosswalks, or cars).

This evolution aimed to increase the difficulty for contemporary bots. However, the reliance on semantic challenges introduced a new and potent vector of vulnerability, directly correlating with the exponential advancement of computer vision technology. Semantic challenges are fundamentally susceptible to models specialized in object detection, such as the You Only Look Once (YOLO) architecture, and advanced deep learning techniques for image classification.2 This generation's failure proved that semantic image understanding—the new "hard AI problem"—was also temporary, not fundamental.

### 2.3 Empirical Evidence of Automated Bypass: The Quantitative Collapse

The scholarly literature provides overwhelming, quantitative evidence that the security premise underpinning cognitive CAPTCHAs (Generations I and II) has collapsed. The quantitative success of modern automated attacks decisively demonstrates that the core concept of a cognitive hurdle insurmountable by machines is now technologically obsolete.

#### A. The Scope of Attack Success

Empirical studies harnessing the power of deep learning models consistently reveal the devastating effectiveness of automated bypass attempts against contemporary image-based CAPTCHAs. Research applying deep learning and semantic annotation techniques reported bot accuracy reaching 70.78% against standard image reCAPTCHA challenges.6 Further experimental results, employing dense convolutional networks (DenseNet) trained specifically for CAPTCHA recognition, demonstrated near-perfect accuracy, reporting solve rates exceeding 99.9%, even against challenges featuring significant background noise and character adhesion.3

Crucially, the most rigorous investigations have shattered the last vestige of the original security mandate. A comprehensive analysis of reCAPTCHA v2 challenges, utilizing advanced machine learning techniques for image segmentation, achieved an astounding **100% solve rate** on the image challenges.2 This result critically invalidates the security promise of the system: the challenges are no longer fundamentally easier for human users than for automated systems.

#### B. The Cost-Curve Inversion

This technological commodification of bot-solving capability introduces a crucial economic problem: the security paradigm is suffering from a devastating **cost-curve inversion**. Historically, deploying a CAPTCHA was economically efficient for the defender, while attacking it required high, specialized effort. Today, the marginal cost of attack for bots has been drastically reduced by modern, generalized AI models and techniques like GANs for synthetic training data.16 The economic model of CAPTCHA defense is thus becoming fundamentally unsustainable, as the attacker’s marginal effort approaches zero, while the burden imposed on legitimate users (lost conversions, wasted time) remains high.

Table 1 summarizes the empirical attack data:

Table 1: Empirical Success Rates of AI-Powered CAPTCHA Bypass Methods

| **CAPTCHA Type / Generation** | **AI Bypass Technique** | **Reported Success Rate (%)** | **Source Context** |
| --- | --- | --- | --- |
| Text-Based (Legacy) | CNNs / Transfer Learning | 99.8% | Demonstration of vulnerability against CNNs.16 |
| Image-Based (reCAPTCHA v2) | Deep Learning (Semantic Annotation) | 70.78% | High solve rates reported using deep learning for semantic annotation.6 |
| Image-Based (reCAPTCHA v2) | Advanced ML/Image Segmentation | 100% | Critical evidence of total failure of the human-vs-bot premise.2 |
| Chinese/English CAPTCHAs | Dense Convolutional Network (DenseNet) | >99.9% | High accuracy against challenges with noise and adhesion.3 |

### 2.4 Generation III: The Invisible Shift (Behavioral and Passive Systems)

The widespread defeat of cognitive tasks forced a necessary strategic pivot toward **friction-less** or passive verification, leading to Generation III systems like reCAPTCHA v3. These systems aim to bypass the inherent usability problem of cognitive challenges by requiring minimal or no explicit user interaction, replacing visible barriers with continuous background analysis.

#### A. Mechanism and Data Dependency

Generation III systems operate primarily through continuous, passive analysis of user behavior. They assign a dynamic risk score based on collecting non-obvious factors, including browser telemetry, device fingerprinting, and the detection of subtle human-like movements (e.g., mouse paths and keystroke timing).13 However, a critical scholarly analysis reveals an operational dependence extending beyond pure real-time behavior. Research on reCAPTCHA v2 (whose risk engine informs v3) demonstrated that the system is heavily reliant on auxiliary historical data, specifically **cookies and browser history**, to effectively evaluate a user’s authenticity.2 This evidence suggests that the security relies less on intrinsically distinguishing human action and more on recognizing a known, previously tracked user profile.

#### B. Vulnerability to Simulation

While highly effective against simple, unsophisticated bots 12, these behavioral systems are not impervious to advanced automation. Their reliance on browser telemetry can be bypassed by sophisticated browser automation frameworks, such as Selenium or Puppeteer, which are capable of simulating complex human-like interactions, including scrolling and text entry. Furthermore, the reliance on passive motor movements is fundamentally challenging to secure, as mouse movements can be synthesized using basic software tools. The emerging threat posed by multimodal AI models further empowers attackers to mimic human behavior with increasing precision, progressively blurring the distinction between a scripted bot and a legitimate user.19

### 2.5 Gaps in Current Knowledge and the Trilemma

The prevailing consensus in the literature centers on the enduring **Security vs. Usability vs. Privacy Trilemma**. Any system attempting to maximize security (resilience to AI) and usability (friction-less experience) often necessitates the sacrifice of privacy (through extensive surveillance and tracking).

A significant gap in current scholarly work is the pronounced emphasis on purely technical metrics (e.g., CNN accuracy, bot solve rates) at the expense of comprehensive human factors analysis. Studies seldom address the compounding impacts of user fatigue or the influence of cultural biases in image selection.20 This analytical neglect ignores the substantial real-world economic costs borne by web service providers due to user abandonment.

## 3. Critical Evaluation and Discussion: Usability, Ethics, and Systemic Failure

The failure of cognitive CAPTCHAs is not merely a technical issue; it represents a deep systemic failure that imposes high financial, ethical, and accessibility costs on the global digital ecosystem.

### 3.1 The Usability and Economic Cost of Friction

Empirical data provides a stark and compelling assessment of the usability flaws inherent in traditional CAPTCHAs. Verification systems engineered to be difficult enough to deter advanced bots are, by their very design, frustrating and error-prone for human users.

Quantitative studies show that traditional text-based CAPTCHAs exhibit an average misspelling rate of **8.66%**.5 The friction is compounded significantly when case sensitivity—often employed as a security reinforcement—is mandated, causing the failure rate to nearly **triple** to **29.45%**.5 This compelling data indicates that close to one-third of genuine human users are likely to fail the challenge on their initial attempt, necessitating a frustrating retry and compounding their negative experience.

This high failure rate translates directly into a measurable financial cost for businesses. This phenomenon can be characterized as **Systemic Self-Sabotage**, where the defensive mechanism inadvertently causes service denial and measurable economic loss by frustrating legitimate users more effectively than it stops advanced bots. Studies show that simply presenting a CAPTCHA can result in a **1+% abandonment rate** in critical user funnels, such as e-commerce checkout, even when users have a strong desire to complete the task.5 The financial impact of this fractional loss, extrapolated across large-scale e-commerce platforms, is massive.

In addition to lost transactions, the cumulative global time expended on CAPTCHA resolution imposes an extraordinary societal and economic burden. Historical estimates suggest that over **819 million hours of human time** have been collectively wasted solving these challenges, equivalent to approximately **$6.1 billion USD** in free, involuntary labor.14 This massive, non-trivial cost is unfairly placed upon the global internet population for a defense mechanism that yields rapidly diminishing security returns against modern AI.

### 3.2 The Accessibility and Exclusion Crisis

The foundational design principle underlying CAPTCHA—to isolate unique human cognitive capabilities—inherently leads to a crisis of accessibility and digital discrimination.

#### A. WCAG Conflict and the Security Paradox

The Web Content Accessibility Guidelines (WCAG) acknowledge the unique difficulty posed by CAPTCHAs, granting a highly specific exemption that applies only to the content of the challenge itself. However, this exemption is strictly contingent on the system providing an accessible and usable alternative.15 This requirement exposes the design's fundamental flaw: studies confirm that meeting accessibility mandates, such as providing an audio alternative for a visual challenge, immediately compromises the system’s security, as bots can easily solve audio recognition tasks.21

This creates a **security paradox**: developers are forced into a zero-sum choice between complying with accessibility standards (which ensures inclusivity) and maintaining the required security level. The critique from the W3C is unequivocal: CAPTCHAs deployed without usable alternatives essentially "fail to properly recognize users with disabilities as human, obstructing their participation in contemporary society".15 This exclusion disproportionately affects users who are blind or visually impaired, those who are deaf, and individuals with cognitive or learning disabilities, for whom the expectation of multiple failure attempts constitutes an inherently inaccessible anti-pattern.15

#### B. Cultural and Situational Bias

The difficulties posed by CAPTCHAs extend beyond physical disabilities to encompass issues of cultural and situational bias. Image-based challenges, particularly those relying on semantic object recognition, are susceptible to cultural dependency. Research indicates that the ability to recognize specific everyday objects, local landscapes, or even text illusions varies significantly across different countries and age groups.20 This reliance on assumed cultural familiarity can lead to inconsistent user experience and unintended exclusion for diverse global demographics. Similarly, situational disabilities—such as a user attempting to solve a challenge on a small mobile screen or an audio CAPTCHA in a noisy public environment—further compromise the system’s effective functionality for legitimate human users.

### 3.3 Privacy Paradox in Centralized Behavioral Systems

The development of Generation III verification systems aimed to solve the usability problem by migrating from visible friction to invisible tracking. However, this shift introduced a critical ethical and regulatory challenge concerning data aggregation and user privacy.

#### A. Data Harvesting and the Tracking Mechanism

These passive systems rely on collecting extensive data, including IP addresses, precise interaction timing, gyroscopic data, mouse movements, and detailed browser telemetry. The system's dependence on cookies and accumulated browser history data to generate a dynamic user risk score signifies that the "proof" of humanity is outsourced to a centralized, opaque trust mechanism maintained by a commercial entity.2 The data collected is far beyond the minimum necessary for security and often facilitates user profiling.

#### B. The Privacy Cost of Authentication

The extensive data required for verification introduces a fundamental dichotomy known as the **Privacy Cost of Authentication**. Users who consciously prioritize privacy—by utilizing VPNs, disabling cross-site tracking, or routinely clearing cookies—are statistically more likely to be flagged by the behavioral analytics engine as "high-risk" or anomalous, leading to more frequent challenges or automatic service denial. These users are effectively forced to surrender privacy to be classified as human.

This centralized data practice has attracted significant regulatory scrutiny. The French data protection authority (CNIL) ruled that reCAPTCHA uses excessive personal data for purposes extending beyond its core security functions, establishing a direct conflict with stringent privacy regulations such as GDPR.22 The competitive landscape confirms this ethical debate, with alternatives contrasting their business models to those that rely on ad sales and extensive user tracking.

### 3.4 Emerging AI-Resilient Alternatives: Friction-less Verification

The systemic failures of cognitive and centralized behavioral CAPTCHAs mandate a paradigm shift toward alternatives that effectively decouple security from friction and pervasive surveillance.

#### A. Hybrid Behavioral Biometrics

A promising alternative involves the use of behavioral biometrics, such as **keystroke dynamics**, which authenticate a user based on their unique, habitual physical typing rhythm, irrespective of the content they input.13 This method provides an intrinsic, passive defense layer, leveraging the inherent, complex variance and rhythm of human interaction, which is exceptionally difficult for scripted bots to mimic precisely.

Novel hybrid systems are currently under research that combine implicit behavioral analysis with active cognitive tests dynamically generated by Large Language Models (LLMs). By simultaneously analyzing a user’s keystroke dynamics against trivial-but-unique, LLM-generated common-sense questions, these systems construct a robust, dual-layered defense.7 This hybrid approach demonstrates high bot detection accuracy while maintaining a superior usability score, successfully thwarting simulation attacks by leveraging the inherent complexity of human motor skills.7

#### B. Proof-of-Work (PoW) Solutions

Proof-of-Work (PoW) CAPTCHAs represent a compelling approach that fundamentally rebalances the economic equation against the attacker.8 These systems require the user's client device to successfully solve a small, computationally expensive cryptographic puzzle running silently in the background.

The security derived from this method lies in making large-scale bot attacks **economically unviable**.8 A bot farm attempting millions of automated submissions would consequently incur massive, prohibitive costs in aggregated CPU usage and energy consumption. Modern PoW implementations, which have been refined to solve multiple easier challenges instead of one hard, unpredictable one, have successfully reduced user latency and maintained a consistent user experience. Critically, PoW systems offer a major privacy advantage over behavioral systems because their security is based solely on computation, not the tracking and aggregation of user data.8

#### C. The Future: Zero-Knowledge Proofs (ZKPs) and Decentralized Identity

The most advanced alternative paradigm centers on cryptographic verification, exemplified by **Zero-Knowledge Proofs (ZKPs)**. ZKPs allow a "prover" (the user) to cryptographically demonstrate the truth of a sensitive assertion to a "verifier" (the website) without revealing the underlying data used to form that assertion.24

Applied to human verification, this technology enables **Proof-of-Personhood (PoP)**.9 A user could cryptographically prove they are a unique human, verified by an underlying credential (such as government ID or biometrics), without ever revealing the data contained within the credential itself.9 ZKPs satisfy three crucial properties: Completeness (a true assertion is accepted), Soundness (a false assertion is rejected), and Zero-Knowledge (no information is revealed beyond the truth of the assertion).24 This framework represents a technological path forward that resolves the security-usability-privacy trilemma, offering a verifiable, cryptographically sound, decentralized, and highly private alternative to traditional surveillance-based verification models.9

Table 2: Comparative Analysis of Human Verification Paradigms

| **Verification Paradigm** | **Security Resilience (against Modern AI)** | **User Friction / Usability** | **Privacy Implications** |
| --- | --- | --- | --- |
| Traditional Cognitive CAPTCHA (Gen I/II) | Low (Bypassed by Deep Learning) 2 | High (8% failure; 1%+ abandonment) 5 | Low-Moderate (Dependent on data reuse) 6 |
| Behavioral Analysis (reCAPTCHA v3) | Moderate (Based on proprietary scoring) 13 | Very Low (Invisible, no interaction) 13 | Very High (Relies on extensive user tracking and browser history) 2 |
| Behavioral Biometrics/Hybrid Systems | High (Hard to mimic complex human rhythm) 7 | Low (Passive data collection during interaction) 7 | Moderate (Collection of sensitive behavioral patterns, but often localized) |
| Proof-of-Work (PoW) Solutions | High (Economically prohibitive for bots) 8 | Low-Moderate (Small, client-side latency required) 8 | Very Low (Focus on computation, not tracking) |
| Decentralized Verification (ZKPs/PoP) | Very High (Cryptographically sound proofs) 24 | Low-Variable (Requires initial credential acquisition) 9 | Very Low (Verification without revealing underlying data) 9 |

## 4. Conclusion and Future Directions

### 4.1 Summary of Key Findings and Arguments

The critical review confirms the central thesis: the conventional CAPTCHA model has reached a point of systemic obsolescence driven by relentless technological advancement and profound design failures. The core premise that machines cannot solve human-designed cognitive tests has been unequivocally refuted by overwhelming empirical data, with modern AI achieving near-perfect solve rates, thus rendering Generation I and II systems technically bankrupt.2

The crisis is not merely technical but functional and economic. The cost-curve inversion means that cognitive CAPTCHAs are no longer a rational defense, imposing a substantial and unwarranted burden on legitimate users (high abandonment rates and wasted time) for security returns that are rapidly vanishing.5 Furthermore, the evolution toward Generation III behavioral systems has introduced a significant **Privacy Cost of Authentication**, centralizing trust and coercing users into surrendering vast amounts of personal and behavioral data to centralized commercial entities simply to avoid being classified as a bot.2

The necessary step is to move away from simple adversarial testing and toward verifiable, intrinsic proofs of personhood, focusing on systems that secure against economic failure and maximize privacy.

### 4.2 Policy and Technology Implications

The current technological maturity of AI necessitates immediate, decisive action from web service providers and policymakers. Relying solely on single-factor cognitive CAPTCHAs is now professionally negligent. Organizations must urgently adopt layered verification strategies that prioritize AI-resistant mechanisms.

**Immediate Recommendations:**

1. **Mandatory Layering:** Implement immediate migration toward robust passive systems, such as Proof-of-Work solutions or hybrid behavioral biometrics, which are capable of maintaining high security while simultaneously minimizing visible user difficulty.7
2. **Ethical Sourcing:** Given the documented conflicts with privacy regulations, organizations must ethically source verification services. These services must explicitly commit to data minimization principles and must not use aggressive third-party data collection for non-security commercial purposes.22
3. **Accessibility Compliance:** Security systems must move past the security paradox by using friction-less methods that inherently meet WCAG standards, rather than relying on flawed alternatives that exclude users with disabilities.15

### 4.3 Future Research Trajectories

The long-term effectiveness of human verification depends critically on the development of verifiable, decentralized mechanisms.

**Future Directions in Verification Science:**

1. **Standardization of Behavioral Biometrics:** Future research must focus on establishing open, standardized, and auditable metrics for continuous behavioral biometrics (e.g., keystroke dynamics). This will allow authentication decisions to move beyond proprietary, black-box risk scores.
2. **Empirical Validation of Cryptographic Proofs:** Extensive empirical studies are needed to validate the real-world deployment, latency, and user acceptance of **Zero-Knowledge Proofs** when applied to anti-bot systems.24 The focus must be on ensuring that this sophisticated technology is scalable and preserves cryptographic integrity against new attacks.
3. **Rethinking Human Oversight:** Recognizing that advanced AI systems are now capable of operating faster and reasoning more strategically than human supervisors, broader research must address the systemic challenge of human oversight failing due to AI outperformance.9 This ensures that verification methods designed to distinguish humans from machines remain strong against highly advanced AI capabilities.

## 5. References

2 A. Plesner, T. Vontobel, and R. Wattenhofer, "Breaking reCAPTCHA v2," arXiv preprint arXiv:2409.08831 (Accepted at COMPSAC 2024), 2024.

3 W. Wang, J. Zhang, H. Zhu, and R. Deng, "A Deep Convolutional Neural Network-Based Approach for CAPTCHA Recognition," AIMS Mathematics, vol. 4, no. 2, pp. 320–332, 2019.

5 Baymard Institute, "CAPTCHAs Have an 8% Failure Rate, and 29% if Case Sensitive," Baymard Quantitative Study, 2022.

6 J. Sivakorn, S. L. Ginosar, M. Kedia, R. Liu, and M. Blum, "I'm Not a Human: Breaking the Google reCAPTCHA," in Black Hat Asia 2016, 2016.

10 A. Turing, "Computing Machinery and Intelligence," Mind, vol. LIX, no. 236, pp. 433–460, 1950.

12 V. Klyuev, "Anti-spam techniques in web applications," International Journal of Computer Science and Network Security, vol. 18, no. 12, pp. 158–163, 2018.

1 L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," in EUROCRYPT 2003, Warsaw, Poland, 2003, pp. 294–311.

24 N. NTT Data, "What Is Zero-Knowledge Proof," NTT Data Insights, 2024.

16 S. A. A. N. Shah et al., "Vulnerability Analysis of CAPTCHA Schemes to Deep Learning Based Attacks," Electronics, vol. 12, no. 4, p. 962, 2023.

4 A. Plesner, T. Vontobel, and R. Wattenhofer, "Breaking reCAPTCHA v2," arXiv preprint arXiv:2409.08831 (Accepted at COMPSAC 2024), 2024.

17 H. Lin, J. Chen, and Y. Lin, "GAN-Based Adversarial CAPTCHA Generation and Resistance Mechanisms," IEEE Access, vol. 9, pp. 78453–78465, 2021.

20 Q. Wu, S. Sun, W. Zhu, and G. Li, "Cultural Bias in Image-Based CAPTCHAs and Mitigation Strategies," arXiv preprint arXiv:2502.05461 [cs.HC], 2025.

7 A. A. Nia, "A Hybrid CAPTCHA System Combining Cognitive and Keystroke Biometrics," arXiv preprint arXiv:2510.02374, 2025.

13 A. D. V. R. T. S. Silva and M. S. C. S. H. P. Senarathne, "A Comparative Analysis of the Effectiveness of Recaptcha V3 against Recaptcha V2 Hidden Fields and Other Anti-Spam Techniques," Int. J. Comput. Sci. Eng. Inf. Technology, vol. 2, no. 3, pp. 11–23, 2024.

15 W3C, "Inaccessible CAPTCHAs and W3C Compliance," W3C Technical Report, 2014.

8 Friendly Captcha, "Controlling Variance in Proof-of-Work Algorithms," Friendly Captcha Blog, 2023.

9 B. Wang, X. Liu, and T. Shi, "Deconstructing the Proof-of-Personhood: User Perception and Design Recommendations for Decentralized Identity," arXiv preprint arXiv:2502.16375, 2025.

## 6. Appendices

### Appendix A: Evolution of Human Verification Architectures

This diagram shows the shift in human verification systems, highlighting the move from friction-based cognitive challenges to implicit, cryptographically sound verification methods that address the main failures of the CAPTCHA approach (Security, Usability, and Privacy).

Table 3: Evolution of Human Verification Architectures

| **Generation** | **Primary Mechanism** | **Core Vulnerability** | **Key Trade-off** | **Modern Alternatives (Post-CAPTCHA)** |
| --- | --- | --- | --- | --- |
| **Gen I (Text)** | Distorted Optical Character Recognition (OCR) | Simple CNNs and OCR 16 | Low Usability for High Security | N/A (Defunct) |
| **Gen II (Image/Cognitive)** | Semantic Object Recognition (Click tiles) | Advanced Computer Vision (YOLO) achieving 100% solve rates 2 | High Friction for Moderate Security | Proof-of-Work (PoW) Solutions 8 |
| **Gen III (Behavioral)** | Passive Browser Telemetry and Risk Scoring 6 | Simulation by browser automation and Reliance on proprietary data | Low Friction for High Privacy Cost | Hybrid Biometrics (Keystroke Dynamics) 7 |
| **Gen IV (Cryptographic)** | Zero-Knowledge Proofs (ZKPs) and Proof-of-Personhood (PoP) 24 | User adoption; Initial credential acquisition friction | High Security/Privacy for Initial Credentialing Effort | Decentralized Identity (DI) 9 |

#### Works cited

1. Luis von Ahn's Research, accessed October 27, 2025, <https://www.cs.cmu.edu/~biglou/research.html>
2. Breaking reCAPTCHAv2 - arXiv, accessed October 27, 2025, <https://arxiv.org/html/2409.08831v1>
3. CAPTCHA recognition based on deep convolutional neural network - AIMS Press, accessed October 27, 2025, <https://www.aimspress.com/article/doi/10.3934/mbe.2019292>
4. History of CAPTCHA - The Origin Story, accessed October 27, 2025, <https://blog.geetest.com/en/article/captcha-origin>
5. CAPTCHAs Have an 8% Failure Rate, and 29% if Case Sensitive ..., accessed October 27, 2025, <https://baymard.com/blog/captchas-in-checkout>
6. [2409.08831] Breaking reCAPTCHAv2 - arXiv, accessed October 27, 2025, <https://arxiv.org/abs/2409.08831>
7. CAPTCHA - Wikipedia, accessed October 27, 2025, <https://en.wikipedia.org/wiki/CAPTCHA>
8. hCaptcha and GDPR - Privacy Claims Warrant Scrutiny - Prosopo, accessed October 27, 2025, <https://prosopo.io/blog/hcaptcha-gdpr-concerns/>
9. Exploring User Interaction with Modern CAPTCHAs - eScholarship.org, accessed October 27, 2025, <https://escholarship.org/content/qt6rc989dw/qt6rc989dw.pdf>
10. Passing the Turing Test | The Classic Journal, accessed October 27, 2025, <https://theclassicjournal.uga.edu/index.php/2025/05/09/passing-the-turing-test/>
11. A Zero-Knowledge Proof-Enabled Blockchain-Based Academic Record Verification System, accessed October 27, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12158337/>
12. The Turing Test - Stanford Encyclopedia of Philosophy, accessed October 27, 2025, <https://plato.stanford.edu/archives/spr2016/entries/turing-test/>
13. MCA-Bench: A Multimodal Benchmark for Evaluating CAPTCHA Robustness Against VLM-based Attacks - arXiv, accessed October 27, 2025, <https://arxiv.org/html/2506.05982v5>
14. What is Zero-Knowledge Proof - a hot technology bringing trustworthiness to Web3 privacy?, accessed October 27, 2025, <https://www.nttdata.com/global/en/insights/focus/2024/what-is-zero-knowledge-proof>
15. Inaccessibility of CAPTCHA - W3C, accessed October 27, 2025, <https://www.w3.org/TR/turingtest/>
16. Balancing Usability and Security in a Video CAPTCHA - Google Research, accessed October 27, 2025, <https://research.google.com/pubs/archive/35116.pdf>
17. New Cognitive Deep-Learning CAPTCHA - PMC - NIH, accessed October 27, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9965441/>
18. Adversarial CAPTCHAs - PubMed, accessed October 27, 2025, <https://pubmed.ncbi.nlm.nih.gov/34033564/>
19. Inadequate Human-in-the-Loop Oversight in Advanced AI Systems - ResearchGate, accessed October 27, 2025, <https://www.researchgate.net/publication/395955309_Inadequate_Human-in-the-Loop_Oversight_in_Advanced_AI_Systems>
20. Practicality analysis of utilizing text-based CAPTCHA vs. graphic-based CAPTCHA authentication - PMC - NIH, accessed October 27, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10152023/>
21. A Hybrid CAPTCHA Combining Generative AI with Keystroke Dynamics for Enhanced Bot Detection - arXiv, accessed October 27, 2025, <https://arxiv.org/html/2510.02374v1>
22. Biometric personal authentication using keystroke dynamics: A review - ResearchGate, accessed October 27, 2025, <https://www.researchgate.net/publication/220199748_Biometric_personal_authentication_using_keystroke_dynamics_A_review>
23. Don't Trust When You Can Verify: A Primer on Zero-Knowledge Proofs | Wilson Center, accessed October 27, 2025, <https://www.wilsoncenter.org/article/dont-trust-when-you-can-verify-primer-zero-knowledge-proofs>
24. CAPTCHA: Using Hard AI Problems For security - CMU School of Computer Science, accessed October 27, 2025, <https://www.cs.cmu.edu/~mblum/research/pdf/captcha.pdf>
25. A Review on Secure Authentication Mechanisms for Mobile Security - MDPI, accessed October 27, 2025, <https://www.mdpi.com/1424-8220/25/3/700>