

ABSTRAK

Perkembangan pesat model kecerdasan buatan (Artificial Intelligence/AI) multimodal menimbulkan pertanyaan kritis tentang efektivitas mekanisme verifikasi visual seperti reCAPTCHA v2 dalam membedakan manusia dari agen otomatis. Penelitian ini mengevaluasi kerentanan reCAPTCHA v2 berbasis gambar (grid 3×3) terhadap agen AI multimodal modern, khususnya varian GPT-4o, dengan membandingkan tingkat keberhasilan dan waktu penyelesaian antara pengguna manusia dan agen otomatis. Evaluasi ini dilakukan pada lingkungan uji terkontrol dengan empat skenario otomasi yang bervariasi berdasarkan antarmuka (web vs API) dan strategi eksekusi (sekuensial vs paralel), dengan masing-masing skenario diuji sebanyak 100 kali untuk menghasilkan metrik kuantitatif. Hasil menunjukkan bahwa model AI multimodal mampu menyelesaikan reCAPTCHA v2 dengan performa setara atau lebih baik dibanding manusia, di mana pengguna manusia mencapai tingkat keberhasilan 84% sementara agen otomatis berkisar 80-88%. Pendekatan berbasis API untuk agen otomatis dengan eksekusi paralel menunjukkan efisiensi optimal dengan tingkat keberhasilan 88% dan waktu total ~6,4 detik, secara signifikan mengungguli metode berbasis web yang memerlukan waktu ~20,8 detik. Temuan ini menegaskan bahwa reCAPTCHA v2 berbasis gambar rentan terhadap serangan agen AI otomatis. Penelitian ini memberikan kontribusi berupa benchmark empiris yang membandingkan performa manusia dan agen AI otomatis pada tantangan CAPTCHA aktual serta rekomendasi teknis untuk memperkuat mekanisme autentikasi.

Kata Kunci: reCAPTCHA v2, AI multimodal, GPT-4o, keamanan autentikasi, CAPTCHA breaking

ABSTRACT

The rapid advancement of multimodal artificial intelligence (AI) models raises critical questions about the effectiveness of visual verification mechanisms such as reCAPTCHA v2 in distinguishing humans from automated agents. This study evaluates the vulnerability of image-based reCAPTCHA v2 (3×3 grid) to modern multimodal AI agents, particularly GPT-4o variants, by comparing success rates and completion times between human users and automated agents. The evaluation was conducted in a controlled testing environment with four automation scenarios varying by interface (web vs. API) and execution strategy (sequential vs. parallel), with each scenario tested 100 times to generate quantitative metrics. Results demonstrate that multimodal AI models can solve reCAPTCHA v2 with performance equal to or exceeding that of humans, where human users achieved an 84% success rate while automated agents ranged from 80-88%. The API-based approach for automated agents with parallel execution demonstrated optimal efficiency with an 88% success rate and total time of ~6.4 seconds, significantly outperforming web-based methods that required ~20.8 seconds. These findings confirm that image-based reCAPTCHA v2 is vulnerable to attacks by automated AI agents. This research contributes empirical benchmarks comparing human and automated AI agent performance on live CAPTCHA challenges and provides technical recommendations for strengthening authentication mechanisms.

Keywords: *reCAPTCHA v2, multimodal AI, GPT-4o, authentication security, CAPTCHA breaking*