Prompting Large Language Models with Raw HTML and JavaScript for Lightweight Phishing Website Detection

Lun-Ping Hung

National Taipei University of Nursing and Health Sciences, Taipei, Taiwan, R.O.C.

Shih-Yang Yang

Department of Media Arts, University of Kang Ning, Taiwan, R.O.C.

Kuan-Jung Chen *

National Taipei University of Nursing and Health Sciences, Taipei, Taiwan, R.O.C.

Syu-Bo Jhang

National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd., Da'an Dist., Taipei, Taiwan, R.O.C.

Abstract:

Phishing attacks have become increasingly sophisticated, exploiting dynamic web technologies and social engineering tactics to evade signature-based detectors. Traditional rule-driven systems struggle to generalize across diverse obfuscation techniques, leading to gaps in coverage. To address this challenge, we turned to large language models for their deep semantic reasoning capabilities. By directly using the raw HTML and JavaScript code as prompts, our framework can infer intent and detect subtle malicious behaviors, enabling an adaptable, dependency-light solution that can be rapidly deployed in constrained environments.

Our system retrieves web content using the requests library, capturing complete HTML and JavaScript for each URL. It then constructs structured prompts encapsulating key indicators—such as form behaviors, script execution flows, and embedded redirects—that are passed to a locally hosted large language model.

When evaluated on a balanced collection of phishing and legitimate URLs, the framework achieved a 79% detection accuracy. This level of performance highlights the model’s effectiveness for real-time threat triage and provides clear, explainable reasoning for each classification. Additionally, the modular architecture supports continuous improvement—such as fine-tuning on new phishing techniques—and easy integration into large-scale scanning pipelines, bridging the gap between research innovation and operational deployment.