ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 14, Issue 2, pp: (253-256), Month: April – June 2026, Available at: www.researchpublish.com

Machine Learning-Based Detection of Corporate Domain Impersonation and Brand Abuse: An Agentic AI Approach

[1]Muhammad Zaki Ahmad, [2]Asif Ali Khan

Affiliation: Corporate Content and Channels Department, Saudi Aramco

City, Country: Dhahran, Saudi Arabia

DOI: https://doi.org/10.5281/zenodo.20714361

Published Date: 16-June-2026

Abstract: **The proliferation of digital commerce has led to an unprecedented rise in corporate domain impersonation and brand abuse, where malicious actors leverage typo squatting, homograph attacks, and sophisticated social engineering to deceive consumers. Traditional detection mechanisms, often reactive and rule-based, struggle to keep pace with the dynamic nature of these threats. This paper proposes a novel Agentic AI-driven framework for the autonomous detection and mitigation of brand-related cyber threats. By employing a multi-agent system (MAS) where specialized agents handle domain scouting, visual brand analysis, and semantic content evaluation, the proposed system achieves high-fidelity detection with minimal human intervention. Our framework integrates advanced machine learning models, including Siamese Neural Networks for logo similarity and Transformer-based models for lexical analysis of URLs. Experimental results demonstrate that the agentic approach significantly reduces the time-to-detection and improves the accuracy of identifying sophisticated impersonation attempts compared to conventional centralized ML models. This research contributes a scalable, autonomous architecture for digital risk protection, offering a proactive defense mechanism for corporate brand integrity.**

Keywords: **Agentic AI, Machine Learning, Domain Impersonation, Brand Protection, Cybersecurity, Multi-Agent Systems, Phishing Detection.**

I. INTRODUCTION

Corporate domain impersonation and brand abuse represent significant threats to the global digital economy, resulting in billions of dollars in annual losses and severe erosion of consumer trust [1]. Malicious actors frequently register domains that closely resemble legitimate corporate entities a practice known as typo squatting or "cybersquatting" to host phishing sites, distribute malware, or sell counterfeit goods [2]. As these attacks become more sophisticated, often utilizing AI to generate realistic content, traditional security measures such as static blacklists and manual monitoring have become increasingly inadequate [6] [7].

The emergence of Agentic AI autonomous systems capable of reasoning, planning, and executing complex tasks presents a transformative opportunity for cybersecurity. Unlike traditional machine learning (ML) models that provide static predictions, Agentic AI systems can proactively navigate the internet, interact with suspicious environments, and adapt their strategies based on real-time observations [3] [5]. This paper introduces an "Agentic Brand Guardian" framework, which decomposes the complex task of brand protection into a coordinated effort among specialized autonomous agents.

The primary contributions of this research are:

1. The design of a multi-agent architecture specifically tailored for autonomous brand protection.
2. The integration of multi-modal machine learning models (visual, lexical, and behavioral) within an agentic workflow.
3. An evaluation of the system's performance in detecting "zero-day" impersonation domains that have not yet been blacklisted.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 14, Issue 2, pp: (253-256), Month: April – June 2026, Available at: www.researchpublish.com

II. PROPOSED AGENTIC FRAMEWORK

The proposed framework, "Agentic Brand Guardian," operates on a decentralized architecture comprising four primary agent types, coordinated by a central "Orchestrator Agent."

A. Multi-Agent System Architecture

The system utilizes a Multi-Agent System (MAS) approach to handle the diverse data streams involved in brand abuse. Table I outlines the specific roles and responsibilities of each agent within the framework.

TABLE I: AGENT ROLES AND RESPONSIBILITIES

Agent Type	Primary Responsibility	Key ML/AI Techniques
Scouting Agent	Monitors domain registries, Certificate Transparency (CT) logs, and social media for suspicious activity.	NLP-based string similarity, Pattern matching, Anomaly detection.
Visual Agent	Analyzes website screenshots for logo and UI abuse, including subtle visual discrepancies.	Siamese Neural Networks for logo comparison [4], Object Detection (e.g., YOLOv8) for UI element analysis.
Semantic Agent	Evaluates site content, metadata, and behavioral cues for malicious intent.	BERT-based Sentiment and Intent Analysis, Lexical analysis using Transformer models [10], Heuristic-based phishing indicators.
Risk Scoring Agent	Aggregates findings from all agents, calculates a comprehensive threat score, and prioritizes alerts.	Random Forest Classifier, Bayesian Inference, Explainable AI (XAI) for threat interpretation.
Mitigation Agent	Automates takedown requests, incident reporting, and communication with hosting providers and registrars.	LLM-based automated communication, pre-defined legal templates.

B. Machine Learning Methodology

The core detection logic resides within the Visual and Semantic agents. The **Visual Agent** employs a Siamese Neural Network to compare logos found on suspicious domains with a repository of authorized corporate assets. This allows for the detection of "fuzzy" impersonations where logos are slightly altered to evade simple hash-based detection [4]. This approach has shown robust performance in one-shot logo recognition tasks [9].

The **Semantic Agent** utilizes a fine-tuned Transformer model (e.g., RoBERTa) to analyze the textual content of the page, URL structure, and metadata. It looks for high-pressure language, credential harvesting forms, inconsistent brand messaging, and other phishing indicators [8] [11]. By combining these multi-modal signals, the **Risk Scoring Agent** can make high-confidence determinations, leveraging advanced ML techniques for threat detection [12]

C. Agentic Workflow and Reasoning

The Orchestrator Agent manages the workflow using a "Chain-of-Thought" (CoT) reasoning process. When the Scouting Agent identifies a suspicious domain (e.g., saudi-aramco.com), the Orchestrator assigns the Visual Agent to capture and analyze screenshots, while the Semantic Agent scrapes the HTML content. If both agents report high similarity to the target brand (Visa) but with malicious intent, the Risk Scoring Agent escalates the threat, and the Mitigation Agent initiates a takedown request to the hosting provider. This multi-agent approach allows for continuous optimization and contextual personalization of detection strategies [13].

III. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the effectiveness of the Agentic Brand Guardian, we conducted a series of experiments using a dataset of 5,000 newly registered domains and 1,000 known phishing domains targeting Fortune 500 companies. The dataset included various types of impersonation, such as typo squatting, homograph attacks, and visual brand abuse.

A. Detection Accuracy

The agentic framework was compared against a baseline centralized Random Forest model trained on URL features alone. As shown in Table II, the multi-agent approach achieved superior precision and recall by incorporating visual and semantic context, aligning with recent findings in advanced phishing detection [14].

TABLE II: PERFORMANCE COMPARISON

Model	Precision	Recall	F1-Score	Avg. Detection Time
Baseline (URL-only ML)	0.82	0.75	0.78	< 1 second
Agentic Brand Guardian	**0.96**	**0.93**	**0.94**	**45 seconds**

B. Discussion

While the agentic approach incurs a higher computational cost and longer processing time per domain (averaging 45 seconds due to browser rendering and multi-modal analysis), the significant increase in F1-score justifies the trade-off. The system successfully identified 92% of "homograph" attacks (e.g., using Cyrillic characters) which the baseline model missed.

Furthermore, the autonomous nature of the Mitigation Agent reduced the mean time to respond (MTTR) from hours to minutes, a critical factor in mitigating brand damage. The ability of Agentic AI to produce high-fidelity, individually tailored deception at mass-market scale also necessitates equally sophisticated detection mechanisms [15].

IV. CONCLUSION

This paper has presented a comprehensive Agentic AI framework for the detection of corporate domain impersonation and brand abuse. By leveraging a multi-agent architecture, we have demonstrated that autonomous systems can effectively mimic the reasoning of human security analysts at scale. The integration of specialized visual and semantic agents allows for the detection of sophisticated attacks that bypass traditional ML models. This proactive approach offers a robust defense against evolving cyber threats to brand integrity.

Future work will focus on enhancing the "adversarial resilience" of the agents, as malicious actors may begin to use AI to specifically deceive agentic detection systems. Additionally, the expansion of the framework to monitor decentralized web (Web3) and metaverse environments remains a critical frontier for brand protection. Further research will also explore the integration of real-time threat intelligence feeds and adaptive learning mechanisms to continuously improve the system's performance against novel attack vectors.

REFERENCES

[1] FBI Internet Crime Complaint Center (IC3), "2023 Internet Crime Report," Federal Bureau of Investigation, Mar. 2024.

[2] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013.

[3] J. S. Park et al., "Generative Agents: Interactive Simulacra of Human Behavior," arXiv preprint arXiv:2304.03442, 2023.

[4] A. Bozkir and M. Aydos, "Logo-based Phishing Website Detection via Siamese Networks," 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 2020.

[5] Tracer AI, "Understanding Agentic AI and Its Transformative Impact on Brand Protection," Tracer Blog, Sep. 2024. [Online]. Available: https://www.tracer.ai/tracer-blog/

[6] Sardine AI, "AI Fraud Vectors: 7 Agentic Attacks now Live in 2026," Sardine Blog, Feb. 2026.

[7] Microsoft Security, "Cyber Signals Issue 9: AI-powered deception," Microsoft, Apr. 2025.

[8] F. A. Omojowo, "Comparative evaluation of lexicon-based and transformer-based models for phishing URL detection," *SpringerLink*, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s10791-026-09967-1

[9] S. S. Lakshmi, M. Govindarajan, and A. Sreenivasulu, "Malware visual resemblance analysis with minimum losses using Siamese neural networks," *Theoretical Computer Science*, vol. 964, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304397522004406

[10] "SwiftURL: A Lightweight Transformer-Based Model for Malicious URL Detection," *MDPI*, 2026. [Online]. Available: https://www.mdpi.com/2076-3417/16/7/3366

[11] L. Chen and L. Meng, "Metadata driven malicious URL detection using RoBERTa large and multi source network threat intelligence," *Scientific Reports*, vol. 16, no. 1, 2026. [Online]. Available: https://www.nature.com/articles/s41598-025-34790-x

[12] J. Hwang and M. Min, "Survey of AI-Based Threat Detection for Illicit Web Ecosystems: Models, Modalities, and Emerging Trends," *Computer Modeling in Engineering & Sciences*, vol. 127, no. 3, 2026. [Online]. Available: https://search.proquest.com/openview/07941be07df13ed0771745bd36c77712/1?pq-origsite=gscholar&cbl=2048798

[13] "Phishing 2.0: exploring the capabilities and risks of agentic AI in social engineering," *Frontiers in Computer Science*, 2026. [Online]. Available: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2026.1795045/full [14]. R. Rawat et al., "An entropy-guided hybrid framework for real-time phishing detection in digital communication systems," *Scientific Reports*, vol. 16, no. 1, 2026. [Online]. Available: https://www.nature.com/articles/s41598-026-46430-z

[14] "The End of Trust: How Agentic AI Breaks Security Assumptions," *arXiv*, 2026. [Online]. Available: https://arxiv.org/html/2605.16436v1