Chapter 10
Adversarial AI-Driven Cyber Threat Intelligence: A Human-Centric Framework for Detecting and Mitigating AI-Augmented Social Engineering Attacks

Soumi Ghosh
https://orcid.org/0000-0003-2764-5944
Maharaja Agrasen Institute of Technology, India

Ritik Raj
Maharaja Agrasen Institute of Technology, India

Amita Goel
Maharaja Agrasen Institute of Technology, India

ABSTRACT

Artificial intelligence and cybersecurity convergence has fundamentally transformed the threat environment, in particular, the social engineering attacks. In this paper, a comprehensive holistic user-centered approach to detection and reduction of adversarial AI-assisted social engineering attacks is proposed with assistance of adversarial AI-determined menace insight. The proposed architecture takes into consideration the concepts of machine learning, behavioral analytics, and human-computer interaction to create a defense multilayer. The experiments prove that the rate of AI-generated phishing attacks and identification of deepfake-generated campaigns are high, respectively, 94.3 and 99.0, respectively. The viability of the framework is proven by the experiments that are founded on the real-life data and

DOI: 10.4018/979-8-3373-4898-8.ch010

simulated AI-based attacks that are serious challenges to the framework. It is an original work of literature that bridges many gaps in the current threat intelligence models through the application of human-centered principles of design, adversarial artificial intelligence detection systems, and training systems that adapt to new developments. Interdisciplinary character of the framework brings along with the ethical considerations and policy orientation and human behavioral analysis besides the technical innovations and thus it is a holistic approach to the next-generation cybersecurity issues.

1. INTRODUCTION

1.1 Background and Historical Context

The age of digital transformation has resulted in the unmatched modification in the environment of cyber threats, and artificial intelligence emerged as a protective barrier and instrument to attack. The evolution of AI in cybersecurity shows that it was first used as a rule-based detection system in the early 2000s, then in the 2010s as a machine learning algorithm, and in the modern world as adversarial AI (Anderson et al., 2024). This has been especially notable in the field of social engineering attacks whose conventional motive of human psychology and control has been adapted to the use of sophisticated AI technologies to develop attack vectors of high sophistication.

Adversarial AI has become a strategic cybersecurity issue, which can be traced to several major developments. Initial AI uses in cybersecurity were mainly based on the recognition of patterns and the detection of abnormalities (Bhattacharya & Kumar, 2023). Nevertheless, with the emergence of generative models or especially Generative Adversarial Networks (GANs) and Large Language Models (LLMs) the threat landscape has fundamentally changed. Such technologies also allow attackers to create convincing deepfakes, scale-based personalized phishing messages, and profile social media in a high level of sophistication never before (Chen et al., 2023).

1.2 The Evolution of AI-Enhanced Threats

The current threat intelligence systems continue to be mainly technicalistic with primary focus on network-based detection techniques at the expense of the human component of social engineering attacks. This shortcoming becomes especially acute when it comes to discussing AI-enhanced attacks that are specifically aimed at human cognitive biases and social interactions (Deb et al., 2024). The fact that generative AI models are now able to generate human-style text, create realistic

Broad Perspective of Smart Home Technology in 2024
Joseph M. Schulzand Jack S. Scilla (2024). *International Journal of Smart Technologies (pp. 1-27).*
www.irma-international.org/article/broad-perspective-of-smart-home-technology-in-2024/350186