# LLM-Guided Adversarial Executable Malware Generation

Md Mahmuduzzaman Kamol[1], Saeyeon Hong[2], Hyejin Woo[3], Se Eun Oh[2], Mohammad Saidur Rahman[1]

[1]Dept. of Computer Science, University of Texas at El Paso, USA
[2]Dept. of Computer Science and Engineering, Ewha Womans University, Korea
[3]Dept. of AI Cyber Security, Korea University, Korea

## Abstract

Malware is considered to be one of the persistence and evolving attacks in the cybersecurity domain. Recent advances in large language models (LLMs), especially their capability to generate, refactor, and obfuscate code, have lowered the technical barrier for attackers to craft and adapt malicious binaries. While LLMs have helped defenders with automated analysis and faster detection, they also enable adversaries to synthesize malware variants capable of evading existing defense mechanism. Prior work on adversarial malware has mainly explored perturbations in the feature space, overlooking their executability at the binary level. Yet, feature-space adversarial attacks become a real threat only when the modified features can be concretely manifested as functional executables. In order to address this research gap, We propose an LLM-guided feature-to-executable pipeline that (i) uses SHAP to produce target adversarial feature vectors and (ii) prompts LLMs to synthesize overlay and section modifications that preserve PE executability. On 10,000 VirusShare samples against the EMBER detector, our SHAP attacks achieve 87.2% Attack Success Rate (ASR) (vs. 71.4% for MAB-malware). All 100 tested binaries executed in sandbox environment. However, when we re-extract features from the synthesized binaries and re-evaluate the detector, the ASR falls to 55.1% due to the feature-to-binary reconstruction gap.

**Keywords:** LLM; malware generation; adversarial learning; cybersecurity