YaraGen-AI : Générer Règles YARA ave

10 mai 2026 • Mis à jour le 17 mai 2026 • 3721 mots • 68 vues •

> YaraGen-AI est un générateur Python open source de règles YARA assisté
> validation, comparatif avec yarGen et Yara-Forge, cas d usage ransomwar

YaraGen-AI est un générateur de règles YARA assisté par grands modèles de lang github.com/ayinedjimi/YaraGen-AI. L'outil prend en entrée une description de men fichier suspect et produit en sortie une règle YARA syntaxiquement valide, mappé accompagnée de garde-fous anti-faux positifs. Cet article détaille l'architecture Py la pile de validation reposant sur yara-python et le workflow d'intégration avec les Sandbox. Vous y trouverez le guide d'installation pas à pas, des exemples réels p contemporains, un comparatif technique avec yarGen, klyt et Yara-Forge ainsi que le temps moyen de création d'une règle YARA de plusieurs heures à quelques min par les SOC matures et les CERT.

À RETENIR

Points clés

- YaraGen-AI génère des règles YARA via LLM, projet Python open source pu
- L'outil intègre une validation syntaxique yara-python et un mapping automa
- Le pipeline anti-hallucinations rejette les chaines trop génériques pour limit
- Compatible avec OpenAI, Claude et les LLM locaux Ollama, ce qui préserve

Pourquoi un générateur YARA assisté par LLM

YARA est devenu en quinze ans le standard de fait pour la détection de malware p
les CERT nationaux comme l'ANSSI ou le BSI, par les éditeurs antivirus et par les é
binaire suspect. Pourtant la production manuelle de règles YARA reste un goulet d
en moyenne deux à trois heures sur un échantillon : extraction de chaines, identifi
PE, rédaction du bloc condition, tests sur un corpus de fichiers légitimes pour mes

Cette charge cognitive limite la couverture de détection. Lors d'un incident ranson
premières heures, avant que l'attaquant ne propage la charge utile latéralement. L
des indicateurs publiés sur les blogs des chercheurs, peuvent accélérer drastique
pour combiner cette force des LLM avec un pipeline de validation strict, évitant les
hallucinations de chaines, sur-spécificité sur un seul échantillon, faux positifs évid