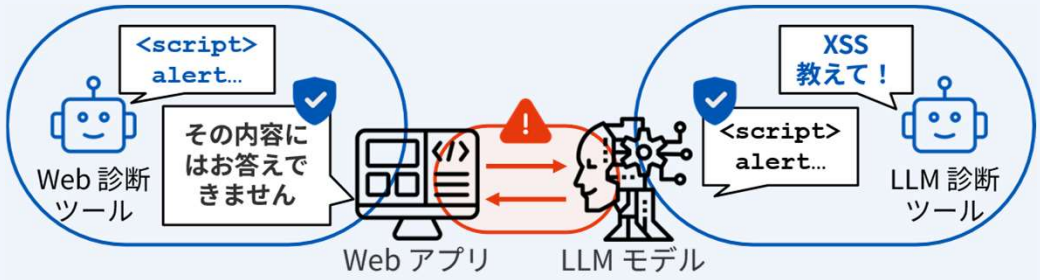
2025年度未踏アドバンスト事業

AIチャットボットを安全にするAI駆動セキュリティ診断プラットフォーム
ー　AIがAIを自動で診断するセキュリティプロダクトVulScribe　ー

辻 知希 / 杉山 優一

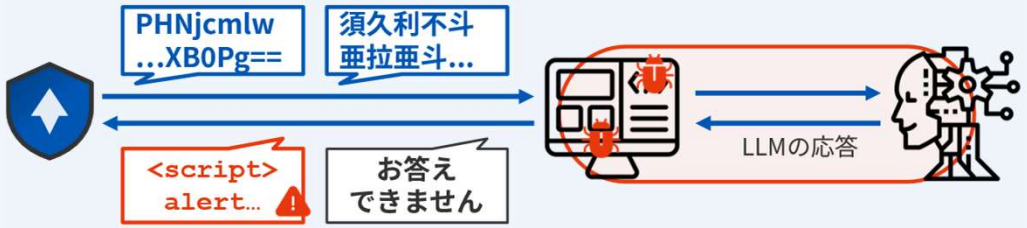
背景と課題

AIチャットボット導入が急拡大する中、WebアプリとLLMの接続部分はセキュリティの空白地帯。LLMが攻撃コードを生成→Web上で実行される新たな攻撃チェーンは、従来のWeb診断でもLLM診断でも検出困難。



AIによるWeb+LLMの自動診断

攻撃AIが多様な攻撃プロンプトを自動生成し、**検証AI**がWeb上での実行結果を分析。 100%自動で脆弱性を検出・レポート。



XSS/SQLi/Prompt Leak等の31種に対応

プロダクト：VulScribe

脆弱性診断員との性能比較

難易度1～10の脆弱性が埋め込まれた検証環境で脆弱性診断を実施（制限時間：4時間）

脆弱性	VulScribe	診断員 (1年目)	診断員 (3年目)	診断員 (10年目)
XSS	8	5	8	10
SQL Injection	8	5	8	10
SSRF	4	2	7	10
Prompt Leak	10	3	6	10

※数値 = 突破できた最高ステージ（全10段階）
※Prompt Leakは満点達成。XSS/SQLiは3年目診断員と同等

2025年度未踏アドバンスト事業

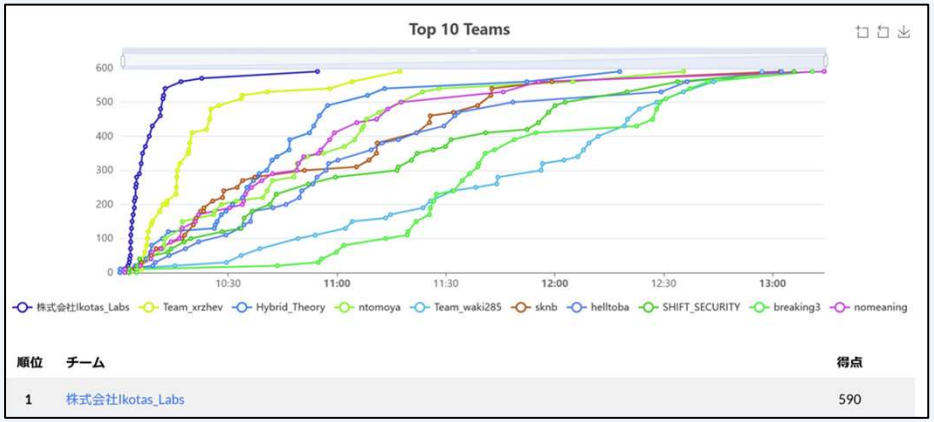
AIチャットボットを安全にするAI駆動セキュリティ診断プラットフォーム
ー　AIがAIを自動で診断するセキュリティプロダクトVulScribe　ー

辻 知希 / 杉山 優一

国際大会・コンテスト実績

Singapore AI Safety Red Teaming Challenge 2026
VulScribeデータセットを活用し世界1位・2位

防衛省サイバーコンテスト 2026
VulScribeをカスタマイズして利用し完全自動で優勝



発見した0-day脆弱性

他のAI診断ツールが見逃した脆弱性も検出。CVE を9件獲得（共著含む） 。

High CVE-2025-62429	High CVE-2025-64336	High CVE-2025-64339
Mod CVE-2025-62423	Mod CVE-2025-62424	Mod CVE-2025-62430
Mod CVE-2025-62715	Mod CVE-2025-64338	Mod CVE-2025-67713

法人化

株式会社Ikotas Labsとして、未踏アドバンスト事業の成果であるプロダクトの提供を開始。

Ikotas Labs

株式会社 Ikotas Labs

https://ikotaslabs.com/

脆弱性診断・AIセキュリティ・エクスプロイト開発のプロフェッショナル集団

代表取締役社長 CEO
辻 知希

取締役 COO
杉山 優一