生成AIとOSINTを利用したサイバーセキュリティ対策の提案

Proposal for Cybersecurity Measures using GenAI and OSINT

江 鴻浩・システム分科会・情報セキュリティ大学院大学

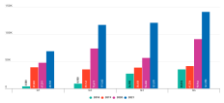
Abstract – This proposal suggests a method for cybersecurity using Generative AI and OSINT, aiming to validate its effectiveness. Specifically, by utilizing the generative models of Llama and GPT-4o, data related to cybersecurity will be collected and analyzed from social media platforms such as X (formerly Twitter) and Weibo. Linking with Link AI, the proposal aims to generate strategies to address cybersecurity issues. Furthermore, the validity of the proposed method will be confirmed through surveys, evaluating accuracy, effectiveness, and response efficiency through demonstration.

背景

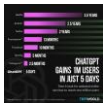
① ネットワーク・インフラに対するサイバー攻撃が頻発

2020年1月：マイクロソフトはデータ漏洩事件で、2.5億件以上の顧客記録が流出
2021年5月：コロニアル・パイプラインがランサムウェア攻撃を受け、米国東海岸へのガス供給が数日間完全に遮断

② サイバー攻撃も常に進化



③ 生成AIの活用の拡大



④ ソーシャルメディアの広範な利用
ユーザーも組織も潜在的な脅威を報告するためにXを利用

OSINTデータソース

今のインターネットにおける特別状況 →

X（旧Twitter）

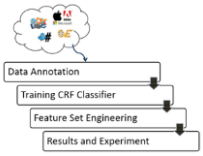
中国のWeibo（微博）

使用ツール・アルゴリズム

① **GPT-4o**：OpenAIによって作られた多言語対応かつマルチモーダルなGPT

② **Llama**：Meta が開発している大規模言語モデル

③ **Link AI**：Simple Futureが開発しているAI開発プラットフォーム、RAGに基づく知識データベースが提供



④ **SVCE**：セキュリティ脆弱性コンセプトエクストラクター、セキュリティブログ、CVEの説明、MicrosoftとAdobeのセキュリティ情報を使用して訓練したセキュリティ脆弱性の関連用語を抽出するNER

⑤ **ワードクラウド**：単語の出現頻度にあわせて文字の大きさを変えて視覚化

⑥ **リッカート尺度**：
提示された文に
回答者がどの程度合意できるかを回答

実証・評価方針

アンケート

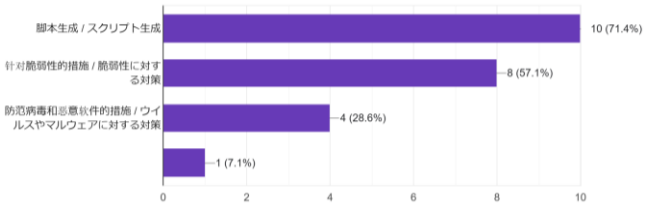
実証・評価結果

総参加人数：14人
・学生：12人
・個人開発者：1人
・IT業界関係者：1人

利用時間：
1週間以内：5人
1～2週間：3人
2～4週間：5人
4 週間以上：1人

利用頻度：
週に3～5回：6人
週に1～2回：8人

問題のカテゴリー別割合：
「スクリプト生成」：**71.4%**「脆弱性に対する対策」：**57.1%**「その他」：**7.1%**
「ウイルスやマルウェアに対する対策」：**28.6%**



評価（精度）：
・平均値：4.00
・中央値：4.00
・最頻値：4.00

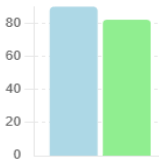
評価（総合満足度）：
・平均値：4.21
・中央値：4.00
・最頻値：4.00

評価（助けられた程度）：
・平均値：4.43
・中央値：4.50
・最頻値：5.00

評価（有効性）：
・平均値：3.36
・中央値：3.00
・最頻値：3.00

評価（対応効率）：
・平均値：4.07
・中央値：4.00
・最頻値：5.00

解決状況：
・総問題数：122件
・解決問題数：109件



① スクリプト生成
・合計：90件
・解決：82件
・解決率：91.11%

② 脆弱性に対する対策
・合計：17件
・解決：14件
・解決率：82.35%

③ ウイルスやマルウェアに対する対策
・合計：10件
・解決：8件
・解決率：80.00%

④ その他
・合計：5件
・解決：5件
・解決率：100.00%

① 対応事例：
・スクリプト生成　・ログ分析
・自動化テスト　・脆弱性修正

② 効率向上の表現：
・スクリプトの作成　・問題の修正
・セキュリティテスト

③ 不足・改善点：
・正確性　・応答速度
・安定性

達成した：
・高精度なサイバーセキュリティ問題自動分析
・効果的なサイバーセキュリティ対応策を自動生成
・サイバーセキュリティ問題の解決において、効率が大幅に向上

実証した：
・生成AIはサイバーセキュリティの問題において大きな助けとなる
・OSINTの活用、生成AIの問題解決能力を向上させるのに役立つ
・ソーシャルメディアをOSINTのデータソースとして活用、非常に有用

今後の課題

- 知識データベース内のコンテンツを最適化し、対策の自動生成速度を向上させる
- OSINTデータのフィルタリングを最適化し、データ品質を向上させ、自動生成される対策内容の品質を向上させる
- 生成AIの設定とプロンプトを最適化して、対策の自動生成の安定性を向上させる
- アンケートの回答数を増加させるため、多様な回答者からデータを収集して、分析結果の代表性と信頼性を高める