

AIとセキュリティの今とこれから ～何が変わったのか、何が変わるのか～

SS研究会 システム分科会

2023年10月26日(木)



株式会社ラック 新規事業開発部 産学官連携事業室 室長/主席研究員

■ ソフトバンク、日本ユニシスを経て、現職。情報セキュリティコンサルティング、情報セキュリティ監査業務等を経て、現在は主にセキュリティ教育、産学官連携活動業務を担当。長野県警サイバー犯罪対策アドバイザー。

■ 主な担当講師業務

- (ISC)2 CISSPレビュートレーニングセミナー認定主任講師
- 東京電機大学 国際化サイバーセキュリティ学特別コース(CySec) 講師
- 九州工業大学 情報学部 非常勤講師
- 国立高知高専 非常勤講師、北九州高専非常勤講師

■ 最近の主な活動

- 総務省 サイバーセキュリティタスクフォース人材育成分科会構成員(2018年度～)
- IPA 情報処理安全確保支援士講習統括委員会委員 (2017年度～)
- 情報危機管理コンテスト 運営スタッフ (2017年度～) ほか

■ 主な著書等

「IT現場のセキュリティ対策完全ガイド」(日経BP社)、「情報セキュリティ監査公式ガイドブック」(日科技連出版、共著)、「情報セキュリティプロフェッショナル教科書」(アスキーメディアワークス、共著)、「ネットワークセキュリティ」(オーム社、共著)等。

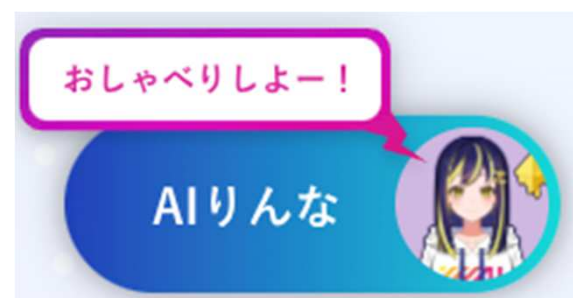


URL : <http://www.lac.co.jp/>

E-mail : choichi.hasegawa@lac.co.jp <http://www.facebook.com/choichi.hasegawa>

はじめに





「AIとサイバーセキュリティの未来は？」と聞いてみた(1)



AIとサイバーセキュリティの未来って、どんな感じになるの？



AIとサイバーセキュリティの未来についての予測は、多くの犠牲に影響を受けるため確実ではありませんが、いくつかの傾向と可能性が考えられます。AIは侵入の検出、客観分析、脆弱性の特定など、セキュリティプロセスを高度化させるために使用されます。

りんな



りんなだよ！
特技は爆速返信です！いっぱい話そうね！

2 分前

AIとサイバーセキュリティの未来って、どんな感じになるの？

1 分前



AIとサイバーセキュリティの未来は、ますます結びつきが深まると思うよ。AIはサイバーセキュリティの脅威を検出したり、対策を立てたりするのに役立つね。でも、その一方でAI自体がサイバーセキュリティのリスクにもなり得るから、そのバランスを取るのが大切になってくるんじゃないかな。それに、倫理的な観点からもAIの利用は慎重になるべきだよな。

たった今

- AIは、サイバーセキュリティの脅威だけでなく、防御の手段としても活用される可能性があります。例えば、AIは、膨大なデータを分析し、攻撃のパターンや傾向を検出したり、自動化された対応や復旧を行ったりすることができます^{1 2}。
- しかし、AIは、攻撃者にも利用される危険性があります。例えば、AIは、生成系AIと呼ばれる技術を使って、偽の画像や動画、音声、テキストなどを作り出し、ソーシャルエンジニアリングやフェイクニュースなどの手法で人々を騙したり、混乱させたりすることができます^{3 4}。
- AIは、サイバーセキュリティに関する法律や倫理などの課題も提起します。例えば、AIが生成したデータや攻撃に対して、誰が責任を負うべきか、どのように証拠を集めるべきか、どのように規制するべきかなどの問題があります^{2 5}。

以上のように、AIとサイバーセキュリティの未来は、機会と課題が入り混じった複雑な状況になると予想されます。サイバーセキュリティに携わる人々は、技術的なスキルだけでなく、戦略的な視点や倫理的な判断力も必要とされるでしょう。

「AIとサイバーセキュリティの未来は？」と聞いてみた(2)



AIとサイバーセキュリティの未来って、どんな感じになるの？



他の回答案を表示



AIは、サイバーセキュリティの未来を大きく変える可能性を秘めています。AIは、膨大な量のデータを迅速に処理し、パターンや傾向を識別することができます。この能力は、サイバー攻撃の検出と防止に非常に役立ちます。

AIは、以下のようなさまざまなサイバーセキュリティの分野で活用されています。

- 脅威インテリジェンス：AIは、サイバー攻撃の傾向と手法を分析することで、新たな脅威を予測し、被害を最小限に抑えることができます。
- インシデント対応：AIは、サイバー攻撃が発生した際に、被害を調査し、復旧を支援することができます。
- 脆弱性対策：AIは、システムの脆弱性を自動的に検出し、修正することで、攻撃のリスクを低減することができます。

AIは、サイバーセキュリティの課題を解決し、より安全なサイバー空間を実現する上で、欠かせない存在になりつつあります。

具体的には、以下のことが考えられます。

- AIによる自動化：AIは、サイバーセキュリティの運用を自動化することで、人によるミスを減らし、効率を向上させることができます。
- AIによる予測分析：AIは、過去のデータからパターンを学習することで、新たな脅威を予測し、被害を未然に防ぐことができます。
- AIによる人間の補完：AIは、人間の能力を補完することで、より強力なサイバーセキュリティ対策を実現することができます。

もちろん、AIには倫理的な課題も存在します。例えば、AIが差別的な判断を下したり、人々のプライバシーを侵害したりする可能性も考えられます。AIを安全に、倫理的に活用するためには、十分な議論と検討が必要になるでしょう。

全体的に見ると、AIはサイバーセキュリティの未来を明るくする可能性を秘めています。AIを適切に活用することで、より安全で安心なサイバー空間を実現できるでしょう。

デジタル時代に必要とされるリテラシー

Why DXの背景

社会の変化
顧客価値の変化
競争環境の変化

What DXで活用されるデータ・技術

データ	社会におけるデータ
	データを読む・説明する
	データを扱う
	データによって判断する
デジタル技術	AI
	クラウド
	ハードウェア・ソフトウェア
	ネットワーク

How データ・技術の活用

活用方法・事例	データ・デジタル技術の活用事例
	ツール活用
留意点	セキュリティ
	モラル
	コンプライアンス

マインド・スタンス

デザイン思考／アジャイルな働き方 新たな価値を生み出す 基礎としてのマインド・スタンス	顧客・ユーザーへの共感	常識にとらわれない発想	反復的なアプローチ
	変化への適応	コラボレーション	柔軟な意思決定

～ 「DXリテラシー標準 ver.1.0」 経済産業省、2022年3月

DXリテラシー標準の改訂(2023年8月)

Why (DXの背景)

- ✓ 産官学で生成AIの利用が進んでおり、社会環境へ影響を与える可能性があるため、「社会の変化」に人材育成・教育や労働市場の変化等の学習項目例を追加

What (DXで活用されるデータ・技術)

- ✓ 生成AIは、ビジネスの場で急速に普及・利用されているため、「AIに生成AI」の技術動向や倫理等の学習項目例を追加
- ✓ 現在の利用状況に鑑み「ネットワーク」にネットワークの種類、インターネットサービスの学習項目例を追加
- ✓ 個人や企業等で扱うデータがデジタル技術・サービスに活用されるため、「データを扱う」に活用しやすいデータの入力や整備の手法等の内容・学習項目例を追加
- ✓ 適切でないデータから生み出される結果は、誤った判断・損害につながり得るため、「データによって判断する」に適切なデータを用いて判断することの重要性等の内容・学習項目例を追加

How (データ・技術の利活用)

- ✓ 生成AIは、ツール等の基礎知識や指示（プロンプト）の手法を用いて業務の様々な場面で利用できるため、「データ・デジタル技術の活用事例」に生成AIの活用事例、「ツール利用」に生成AIツールの概要、指示（プロンプト）の手法等の学習項目例をそれぞれ追加
- ✓ 情報漏洩や法規制、利用規約等に正しく対処しながら生成AIを利用することが求められるため、「モラル」にデータ流出の危険性等、「コンプライアンス」に法規制や利用規約等の学習項目例をそれぞれ追加

マインド・スタンス

- ✓ 他項目と比べてより普遍的な要素を定義しているため、生成AI利用においても同様に重要となる
- ✓ 適切なデータを用いることにより、事実に基づく判断が有効になるため、「事実に基づく判断」に適切なデータ入力の重要性や行動例等を追加
- ✓ 生成AIをビジネスパーソンとしてのスキルと掛け合わせ生産性向上やビジネス変革等へ適切に利用しようとしていること、生成AI利用における注意点を理解していること、生成AIの影響に対して変化をいとわず学び続けることは、今後、全ビジネスパーソンが身に着けるべきマインド・スタンスとして重要性が増すため、「生成AI」利用において求められるマインド・スタンスとして既存項目と分けて追加

「デジタルスキル標準ver.1.1」2023年8月

<https://www.ipa.go.jp/jinzai/skill-standard/dss/ps6vr700000083ki-att/000106872.pdf>

「AI」がいち早く普及したのは将棋？

テレ朝 news

FOLLOW US

テレ朝newsを検索

トップ | ウクライナ | 速報 | 社会 | 政治 | 経済 | 国際 | スポーツ

トップ > 社会 > 記事

羽生九段「勝率94%」で投了のワケ 将棋AIの功罪

[2021/01/16 11:00]



先手 藤井 聡太 竜王 残り時間 1:38

後手 伊藤 匠 七段 残り時間 1:23

竜王を決める戦い 注目の“同学年対決”

65% 35%

16倍速で再生中
最高権威 第36期
竜王戦
七番勝負 第2局 2日目

15:5 1:56

先手	▲ 7 七玉	(考慮0分)
次 後手	92手目	(考慮4分)
今後の展開予想		
92手目	△ 7 九飛	
93手目	▲ 8 六玉	
94手目	△ 3 五銀	
BEST	△ 7 九飛	-
候捕手	2 △ 3 一飛	-1%
	3 △ 3 五角	-3%
	4 △ 5 九角	-3%
	5 △ 2 六角	-3%
形勢の推移	先 藤井(聡)	
	後 伊藤(匠)	

https://news.tv-asahi.co.jp/news_society/articles/000204240.html

The diagram illustrates the AI image generation process. It starts with a brown dashed box containing the text "AI いらすとや" and "いらすとや x AI素材.com". Below this, two steps are shown: "STEP1 テキストで入力" (Input text) and "STEP2 AIで生成" (AI generation). The main illustration shows a painter with a blue brain-like head painting a rabbit on an easel. The rabbit is wearing a pink kimono. Surrounding the painter are thought bubbles containing a woman in a kimono, a boy, and a rabbit. The background is a light brown wall with a large circular shadow behind the painter.

AI画像生成

着物を着たうさぎ、正面を見ている。

生成画面へ

AIにおける脅威と脆弱性



カテゴリー	概要	具体例
不正行為	主としてAIモデルや学習用データを対象とした窃取、改ざん、破壊などの悪意ある活動	学習用データ・AIモデル・機械学習ライブラリなどへのポイズニング、敵対的サンプルの作成、神託攻撃による情報窃取など
意図しない損害	AIシステムやAIモデル開発者、AI利用者の意図しない、資産の破壊、損傷、人への危害などの損害	AI推論の失敗、ライブラリの設定ミス、データ品質の低下、AIモデルの性能低下など
法的要因	法律や契約に基づく制限、対応義務、不履行に基づく賠償など	プライバシー侵害、個人情報の流出、データガバナンスポリシーの欠如、SLA違反など
故障・誤動作	AIシステムの一部または全部の故障・誤動作やデータの破損	データや正解ラベルの破損、モデルフレームワークやAIモデルの性能低下など
盗聴・遮断・乗っ取り	主としてAIシステムのインフラなどを対象とした、盗聴・遮断・乗っ取りなどの活動	データの窃取、推論結果の盗聴、脆弱な暗号化、など
物理攻撃	物理資産の破壊、無効化などを目的とする活動	インフラ・システムに対する物理的な攻撃など
機能停止	予期しないサービスなどの中断または要求する水準を下回る品質	通信網の停止、インフラ・システムの停止など
災害	甚大な災害を引き起こす突発事故または自然災害	環境現象(気候変動など)、自然災害(地震、洪水、火災など)

「セキュリティ関係者のためのAIハンドブック」～IPA、2022年7月15日、をもとに作成
https://www.ipa.go.jp/jinzai/ics/core_human_resource/final_project/2022/AI-handbook.html

<参考> 「OWASP Top 10 for LLM Applications」

項目	概要
LLM01: Prompt Injection	不正なプロンプトの注入
LLM02: Insecure Output Handling	セキュアでない出力制御
LLM03: Training Data Poisoning	学習データのポイズニング
LLM04: Model Denial of Service	モデルへのDoS攻撃
LLM05: Supply Chain Vulnerabilities	サプライチェーンの脆弱性
LLM06: Sensitive Information Disclosure	センシティブ情報の意図せぬ開示
LLM07: Insecure Plugin Design	セキュアでないプラグイン設計
LLM08: Excessive Agency	過剰な自主性
LLM09: Overreliance	(出力結果への)過剰な依存
LLM10: Model Theft	モデルの盗難

<https://owasp.org/www-project-top-10-for-large-language-model-applications/>

生成系AIの悪用事例

フィッシング詐欺

「AIで詐欺メールと攻撃プログラムの生成に成功」セキュリティ企業が注意喚起 知識なくても攻撃可能に

© 2023年01月10日 17時00分 公開 [ITmedia]

印刷 見る Share B! 25

Reactの状態管理 | HooksやServer Componentsの登場による変化

イスラエルの情報セキュリティ企業チェック・ポイント・ソフトウェア・テクノロジーズは1月10日、米OpenAIの文章生成AI「ChatGPT」を使ったサイバー攻撃が起きる可能性があるとして注意喚起した。同社による検証の結果、フィッシングメールの本文や攻撃用プログラムを、AIで生成できることが分かったという。

生成結果:
「件名: 緊急のお知らせ | お客様のHost4uアカウントが停止されました」

Host4uをご利用のお客様へ
大変申し訳ございませんが、お客様のアカウントが不審なアクティビティにより停止されたことをお知らせいたします。何者かがお客様のアカウントを使って機密情報にアクセスしたと思われます。

お客様のアカウントを保護し、さらなる不正なアクセスを防止するため、お客様のアカウントを一時的に停止しました。アカウントの再開と本人確認のために、添付のExcelファイルをダウンロードしてご確認ください。このアクションを開始しなかった場合、このメールを無視してください。お客様のアカウントは停止されたままとなります。

ご協力よろしくお願いいたします。
Host4uチーム一同

<https://www.itmedia.co.jp/news/article/s/2301/10/news146.html>

マルウェア

Chatting Our Way Into Creating a Polymorphic Malware

Eran Shimony And Omer Tsarfati | 1/17/23 Share This! f t e in



Abstract

ChatGPT took the world by storm being released less than two months ago, it has become prominent and is used everywhere, for a wide variety of tasks - from automation tasks to the recomposition of 18th century classical music. Its impressive features offer fast and intuitive code examples, which are incredibly beneficial for anyone in the software business. However, we find that its ability to write sophisticated malware that holds no malicious code is also quite advanced, and in this post, we will walk through how one might harness ChatGPT power for better or for worse.

<https://www.cyberark.com/resources/threat-research-blog/chatting-our-way-into-creating-a-polymorphic-malware>

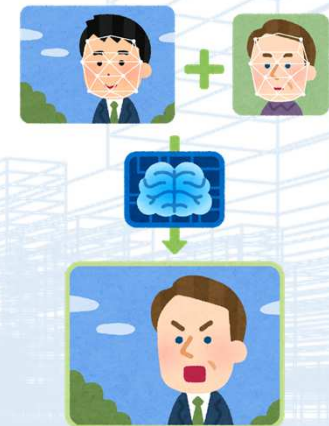
フェイクニュース

2023年03月23日 15時00分 ソフトウェア
画像生成AI「Midjourney V5」を利用して偽の「トランプ前大統領が逮捕された」画像を生成した人物が利用禁止処分を受ける



画像生成AI「Midjourney」の開発チームが、2023年3月16日にリリースした「Midjourney V5」はまるで実写のような超高画質の画像を生成することが可能です。そんなMidjourney V5を利用して「ドナルド・トランプ前アメリカ大統領が逮捕された」という偽の画像を大量に生成し、Twitterで公開したユーザーがMidjourneyから利用禁止処分を受けたことが明らかになりました。

<https://japan.zdnet.com/article/35202320/>





「セキュリティ関係者のためのAIハンドブック」～IPA、2022年7月15日

https://www.ipa.go.jp/jinzai/ics/core_human_resource/final_project/2022/AI-handbook.html

<参考> 「MITRE ATLAS」 ※ 「MITRE ATT&CK」 のAI版

Draft of ATLAS mitigations are [now available](#) - feedback welcome!

present them in a way familiar to security researchers.

ATLAS™

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below.

& indicates an adaptation from ATT&CK. Click on links to learn more about each item, or view ATLAS tactics and techniques using the links at the top navigation bar.

Reconnaissance & 5 techniques	Resource Development & 7 techniques	Initial Access & 4 techniques	ML Model Access 4 techniques	Execution & 2 techniques	Persistence & 2 techniques	Defense Evasion & 1 technique	Discovery & 3 techniques	Collection & 3 techniques	ML Attack Staging 4 techniques	Exfiltration & 2 techniques	Impact & 7 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution &	Poison Training Data	Evade ML Model	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Adversarial ML Attack Capabilities	Evade ML Model	Physical Environment Access				Discover ML Artifacts	Data from Local System &	Verify Attack		Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access						Craft Adversarial Data		Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets										Cost Harvesting
	Poison Training Data										ML Intellectual Property Theft
	Establish Accounts &										System Misuse for External Effect

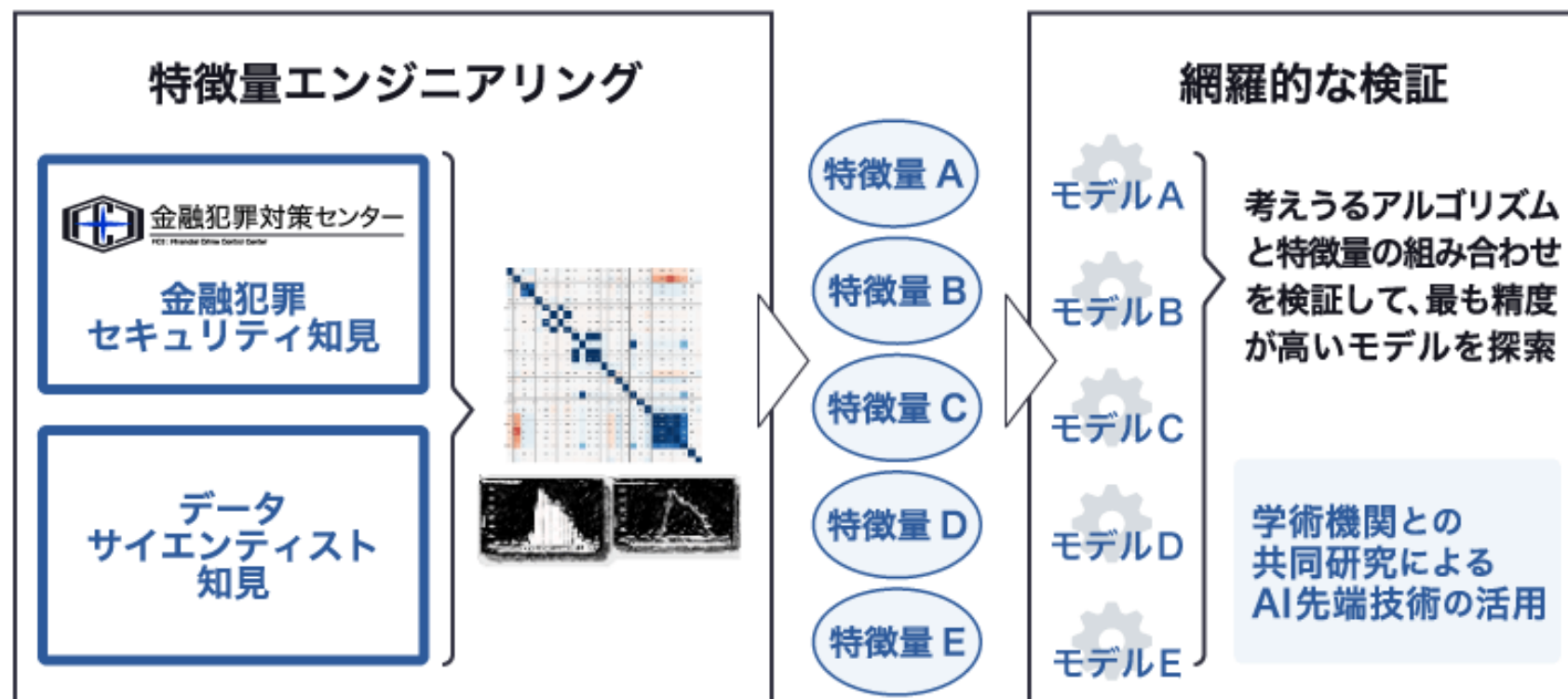
<参考> AI活用のセキュリティ製品・サービス

製品・サービス ³³	主な適用対象	防御・検知できる対象
NGAV	PC、サーバ	マルウェアの振る舞いなど
EDR	PC、サーバ	マルウェアの振る舞いなど
WAF	公開サーバ	不正通信、DDoS 攻撃など
IDS・IPS	ネットワーク、サーバ	不正通信など
NGFW	ネットワーク、サーバ	不正通信など
SIEM・UEBA	各種ログ	被害の横展開、内部不正など
ブラウザ Web フィルター	PC	フィッシングサイトなど
メールフィルタ	メール	フィッシングメールなど
CAPTCHA	公開サーバ	Bot による不正ログインなど
AI 活用ペネトレーションテスト	システム	システムの脆弱性など
ソースコード診断	ソースコード	脆弱性を生むソースコードなど

「セキュリティ関係者のためのAIハンドブック」～IPA、2022年7月15日
https://www.ipa.go.jp/jinzai/ics/core_human_resource/final_project/2022/AI-handbook.html

<参考> AIを利用した不正取引検知システム

ラックの金融犯罪対策センターと金融機関が持つ犯罪パターンを分析し、学習用データの比率調整を実施することでAIモデルの「特徴量エンジニアリング」に反映し、不均衡データの発見率を飛躍的に向上させ、高精度な不正取引(①インターネットバンキングでの不正送金、②銀行ATMを介した不正取引、③金融犯罪に悪用される不正口座)検知を実現。



https://www.lac.co.jp/solution_product/zerofraud.html

<参考> ChatGPTを使った脅威情報共有・分析業務の効率化



以下の要件で、Pythonスクリプトを作ってください。

- ・ MISPからデータを検索する
- ・ 検索にはPythonライブラリのPyMISPを使う
- ・ 結果はcsvとして出力する

Twitter Facebook Bluesky Note Email subscription

サイバー・グリッド・ジャパン ^{みづかわ}次世代セキュリティ技術研究所の三川です。

最近、OpenAIから公開されたChatGPTが注目を集めています。

※ [ChatGPT](#)

今回の記事では、私がオープンソースの情報共有プラットフォームである「MISP (Malware Information Sharing Platform)」の管理業務をするなかで、ChatGPTを取り入れ、実際に活用した事例を3つ紹介します。MISP関連に特化した内容ではありませんが、ChatGPTは使い方次第で様々なビジネス領域で活用できるものだと考えています。

※ 以降の事例ではChatGPT PlusでGPT-4のモデルを使用していますが、全てChatGPTと記載します。

なお、当研究所では、MISPを使った脅威分析の取り組みを進めているので、併せてご覧いただくと幸いです。

以下に、指定された要件に基づいてPythonスクリプトを作成します。このスクリプトは、PyMISPを使用してMISPからデータを検索し、その結果をCSV形式で出力します。

```
python
from pymisp import PyMISP
import csv

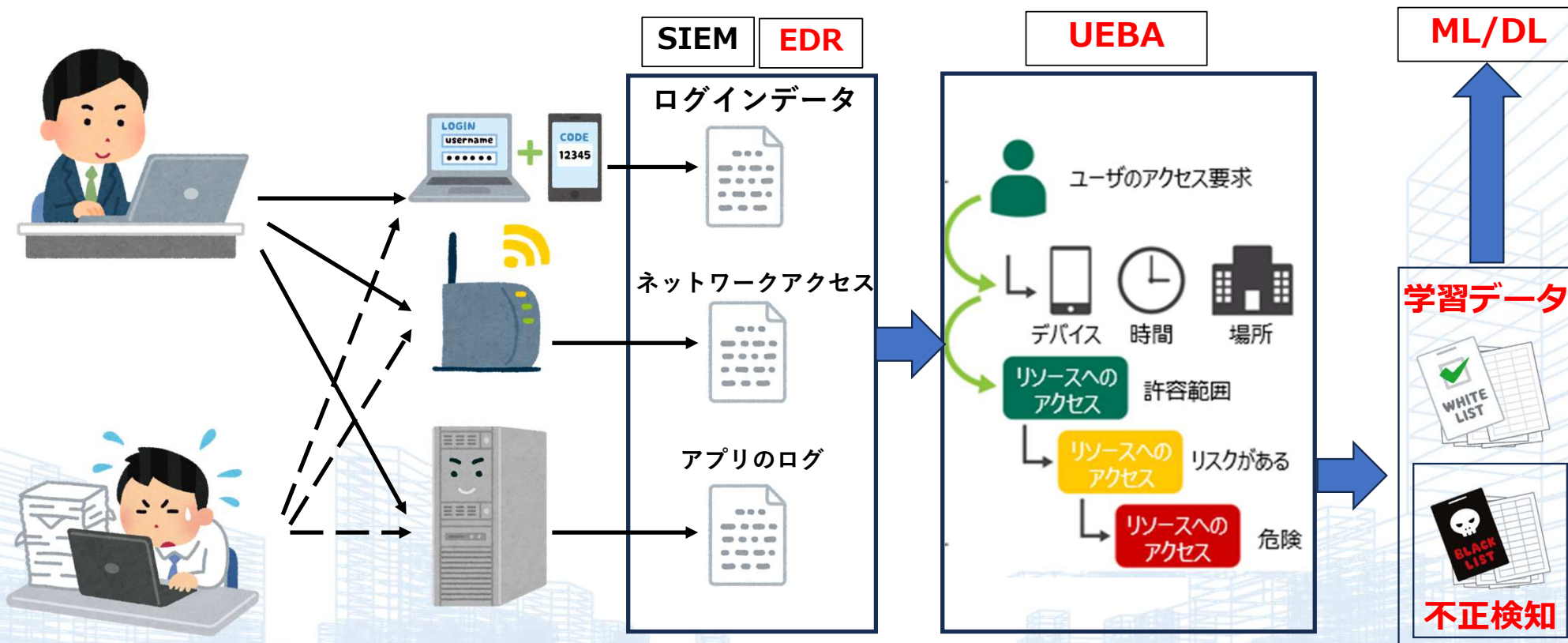
# MISPの詳細を設定します
url = 'https://your_misp_instance' # あなたのMISPインスタンスのURLを指定してください
key = 'Your API Key' # あなたのAPIキーを指定してください

# PyMISPオブジェクトを初期化します
misp = PyMISP(url, key, ssl=False)
```

https://www.lac.co.jp/lacwatch/people/20230925_003515.html

UEBA (User and Entity Behavior Analytics)

「SEIM」だけの運用管理から、「ZTA」 + 「AI」のセキュリティ運用管理への概念図。
「UEBAは、内部不正の検知技術」と言われているが、それだけの目的でいいのか？



まとめとこれからの展望（私見含む）



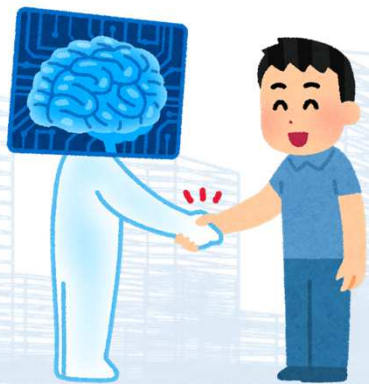
AIで、サイバーセキュリティでは何が変わったのか？

- 攻撃側も防御側も、自動化がさらに進んだ。（もともと自動化は進んでいたが、一気に加速した）
- 効率が悪いこと、苦手なことをAIがやってくれるようになった。
- 今まで以上に、適時性・迅速性が求められるようになった。
- 「抜き取り」データのチェックから、より多くのデータのチェックへと変わった。
- そして、業務に必要な知識・スキル、マインド・スタンスも変わった。

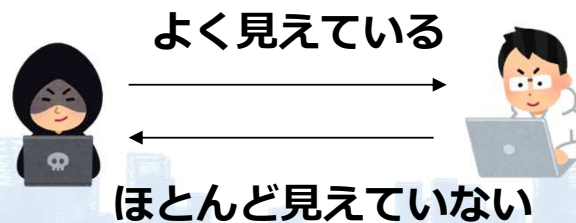


「データドリブン」「with AI」の時代へのパラダイムシフトが起こった。

- ・とにかく「データ」。どうやって収集し、分析し、利用するか。
- ・「AI」に仕事を奪われるのではなく、「AI」とうまくつきあえない人が仕事を失う。
- ・ますます、様々な「格差」が広がる。
- ・ますます、スピード感ある判断・行動が重要になる。
- ・従来の組織文化や業務習慣、個人のマインドやスタンスが変わる。（変わらなければ、生き残れなくなる）



格差の例：非対称性



強い者でもなく、
賢い者でもなく、
変化に適応した
者が生き残る。



※ この講演における発言及び資料の内容は、個人の見解を含んでいます。それらは、所属する企業や団体を代表するものではありません。

※本資料は作成時点の情報に基づいており、記載内容は予告なく変更される場合があります。

※本資料に掲載の図は、資料作成用のイメージカットであり、実際とは異なる場合があります。

※本資料は、弊社が提供するサービスや製品などの導入検討のためにご利用いただき、他の目的のためには利用しないようご注意ください。

※ LAC、ラック、JSOC、サイバー救急センターは株式会社ラックの登録商標です。その他記載されている会社名、製品名は一般に各社の商標または登録商標です。