

情報セキュリティとAI

DX・情報統括本部 情報基盤センター

システム情報学部

システム情報学研究科

内容

- 自己紹介
- 情報セキュリティインシデントの状況
- AIとの関連
 - 懸念点
 - AIエージェント
- AIの発展と情報セキュリティへの影響は？

自己紹介

- DX・情報統括本部 情報基盤センター
 - 情報基盤センター長・教授
 - ネットワーク基盤担当
 - CISO 補佐 (Chief Information Security Officer 補佐)
 - 神戸大学内における情報セキュリティ維持活動
 - 情報セキュリティインシデント対応
 - 情報セキュリティインシデントのための監視
 - 情報セキュリティポリシーの策定
- 大学院システム情報学研究科
 - 情報セキュリティ運用論講座担当
- システム情報学部
 - 授業担当

2026/10上旬のネットニュース記事

- 国内の不正アクセス、最多ペース 相次ぐ個人情報漏えい、対策急務
 - 国内企業・自治体からの漏洩が対象
 - AIを利用して脆弱性(欠陥)を探索？
- 個人ユーザにとっても．．．
 - 漏洩した個人情報に自分が含まれている場合も
 - 名前・住所などのCriticalな個人情報が含まれてない場合でも攻撃対象として記録される場合も
 - 詐欺メール(フィッシング)の品質(?)や手口の巧妙化がAIサポートで上がることが予想される．
- AIで武装（？）した攻撃者は脅威そのもの

AIの現状 (詳しい技術解説は行いません)

- LLM（大規模言語モデル）・生成AI の出現
 - 実現には莫大な計算機資源が必要
- 日進月歩（一ヶ月毎（？））に新しい機能がリリース
 - 頻繁に「XXがAIによってできるようになった」というアナウンス
- AIの発展の負の部分
 - 高度なLLM/生成AIの開発競争が勃発
 - 莫大な計算機資源が必要（メモリ・GPU・電力!)
 - AIのためのインフラ建設のための投資が増大→ 計算機・メモリの価格が高騰
 - AI実行の仕組み自体がサイバー攻撃の対象となる
 - 開発ベンダーのトップによる**AIの暴走や悪用に対する懸念の表明**
 - 社会におけるAIによって代替可能（であると考えられている）ジョブ消失に対する懸念

詳しくは**AI**に聞くことが可能

- 手元の端末で試してみてください。
 - Google 検索ならば AIモードで検索
 - Microsoft edge ならば copilot
 - ...

一般的に言われているAIの懸念点(一部)

- AIがサイバー攻撃に悪用され(てい)ないか？
 - OSやアプリケーションの脆弱性の発見がAIによって劇的に高速化
 - 「Linux」カーネルに三たび重大な脆弱性が見つかる--AIツールの登場で加速
 - Zero day 攻撃(脆弱性の発見とほぼ同時に攻撃が行われる)が増加する(と言われている)
- AIの応答に関するリスクは？
 - AIは正しい答えを返すことを保証しない(ハルシネーション)
 - 法的・権利リスク (著作権, 商標権)
 - 倫理・コンプライアンスリスク(応答に差別や偏見が含まれる)
- その他AI利用時における情報セキュリティ上のリスク(次スライド)

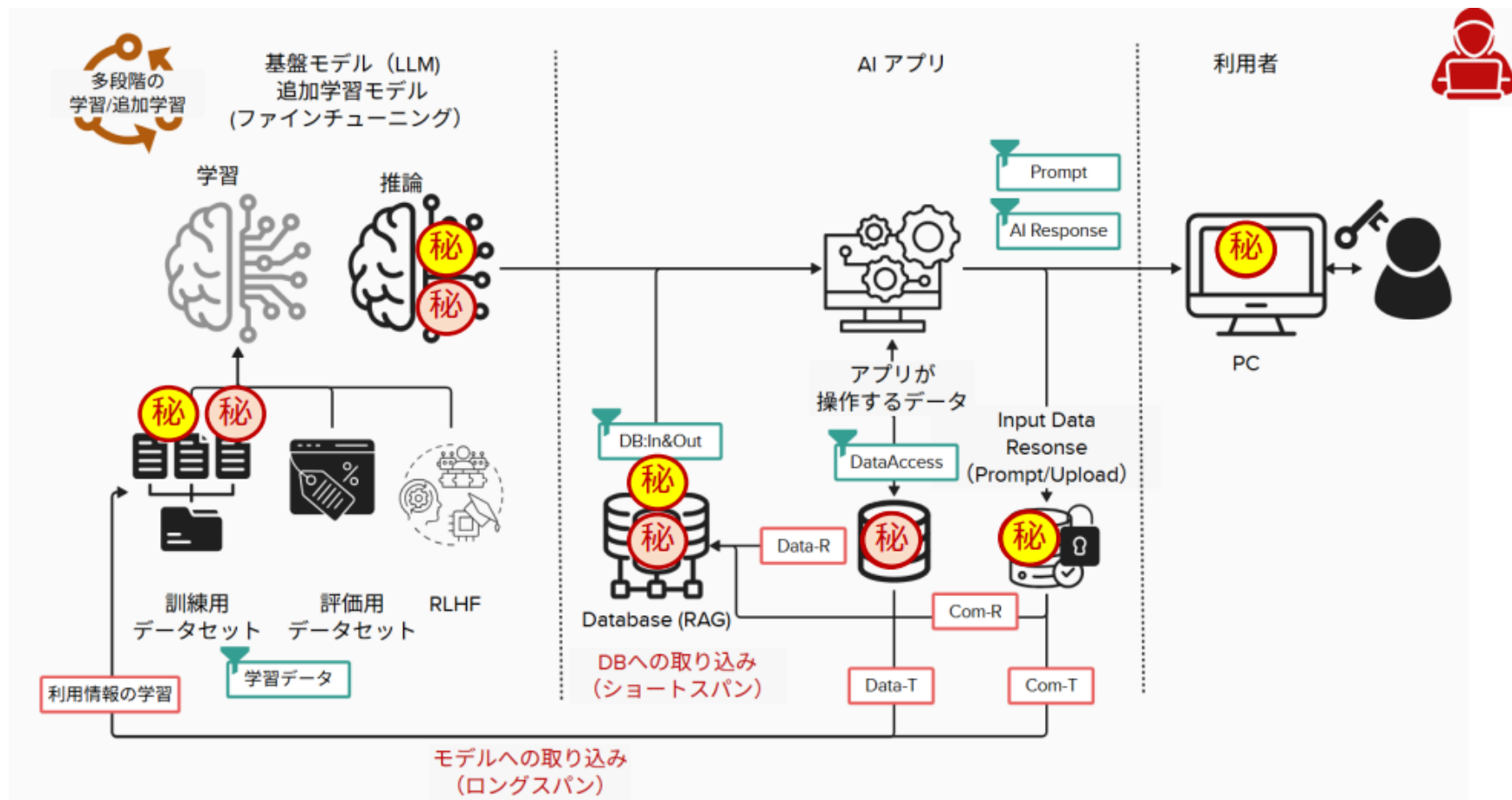
一般に言われている**AI**の情報セキュリティ上のリスク(一部)

- **AI**は、悪用を考えている人を大幅に強化する可能性大
 - 手口が巧妙化
 - 世の中に知られていない欠陥(脆弱性)が使われる
- **AI**を実行するための仕組み自体が攻撃対象となる可能性
 - 特に **AI**エージェント

(個人の)AI利用時の情報セキュリティ上の懸念点

- AIサービス利用時のデータ(ファイルアップロード)やデータ入力
 - 個人情報・機密情報が学習に利用されてしまう(情報漏えい)
 - 一旦学習されたデータは完全に取り消すのは難しい
- 組織等で比較的安全なAI環境が用意されているが個人的な環境を利用してしまう.
 - 組織などでは、入力データを学習に利用することが制限されたサービスが提供されてしまうことが多い.
 - 個人の環境を利用してしまうと入力した情報が学習に利用されてしまう場合も(情報漏えいにつながる.)
 - シャドーAIと呼ばれる.
- AIシステム・アプリケーションへのサイバー攻撃
 - AIシステム基盤へのサイバー攻撃により学習データが改ざんされる.
 - AI特有の攻撃リスク → プロンプトインジェクションによるAIの悪用
 - パソコン側のAIアプリ自体が改ざんされる or Fake のアプリを(騙されて)利用させられる
- AIエージェントの利用上のミス
 - AIエージェントへのプロンプトのミスで、重要なファイルを消去した
- AIサービスのアカウント BAN
 - AIベンダーの利用規約に反する利用・不正利用でアカウントが停止・強制削除される.
 - AIに入力しているデータ等が利用できなくなる

AIサービスにおけるデータ・ファイル利用時のリスク(1)



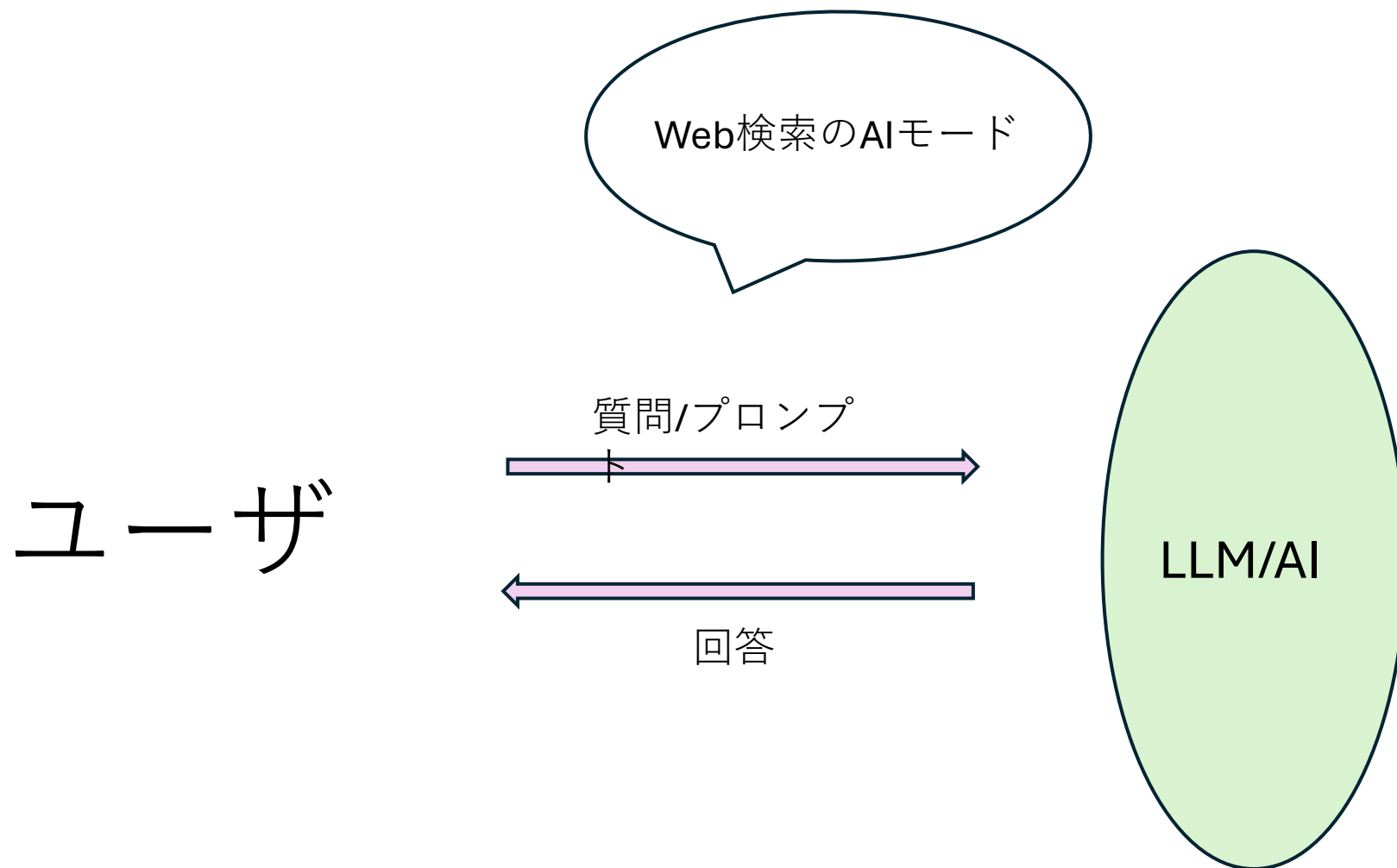
設定によっては利用者が入力した情報がモデルに学習される＝他のユーザに公開される可能性

https://www.jnsa.org/result/act_ciso/2026/data/2026-01.pdf から引用

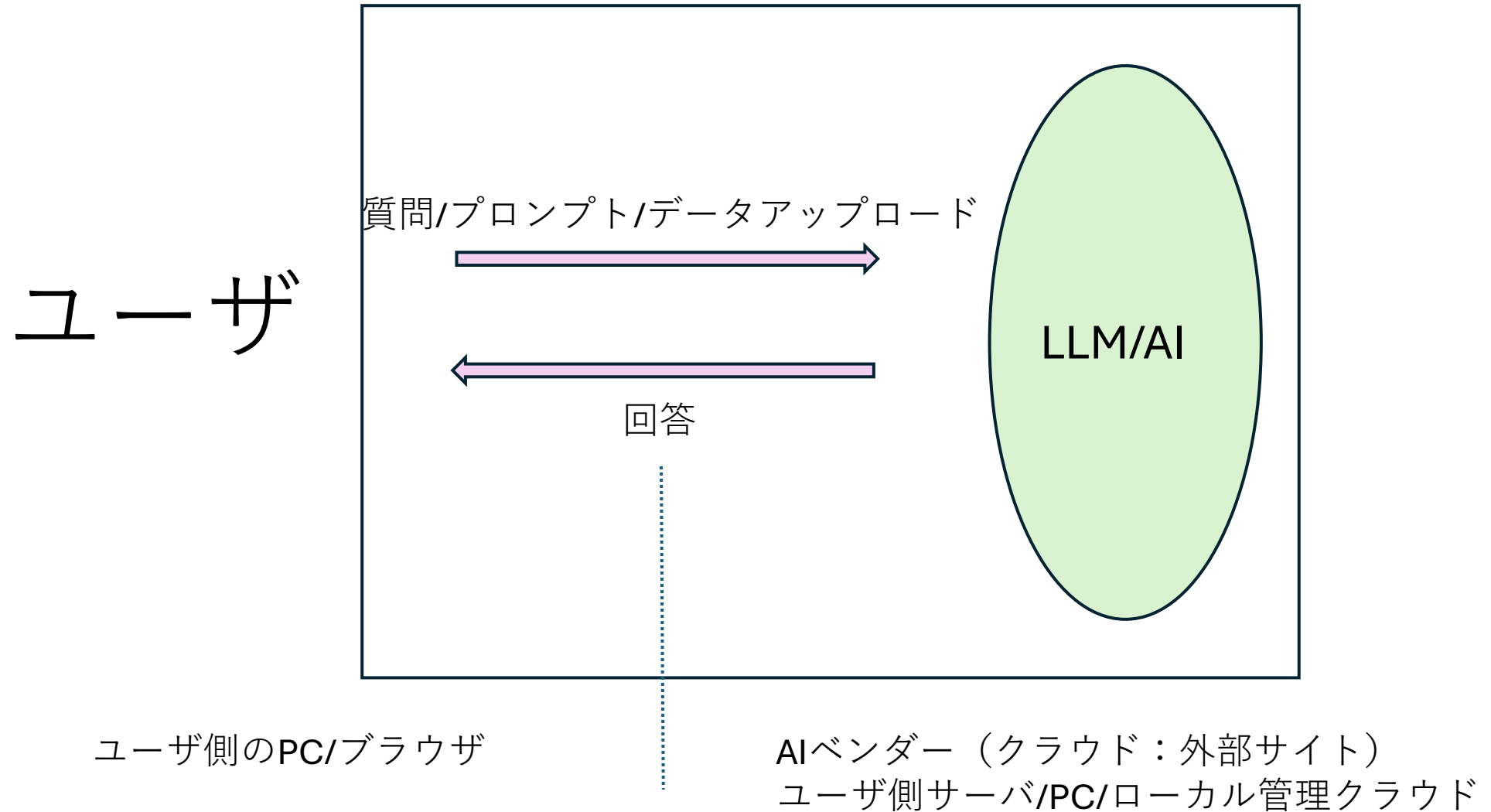
AIサービスにおけるデータ・ファイル利用時のリスク(2)

- 利用者がAIを操作するために入力する「プロンプト」も一定期間記憶される。
 - 入力したプロンプトが自由に消去可能化はシステム次第.
 - プロンプトに入力する情報も熟慮した方がよい.
- 設定をよく確認して利用する必要あり.

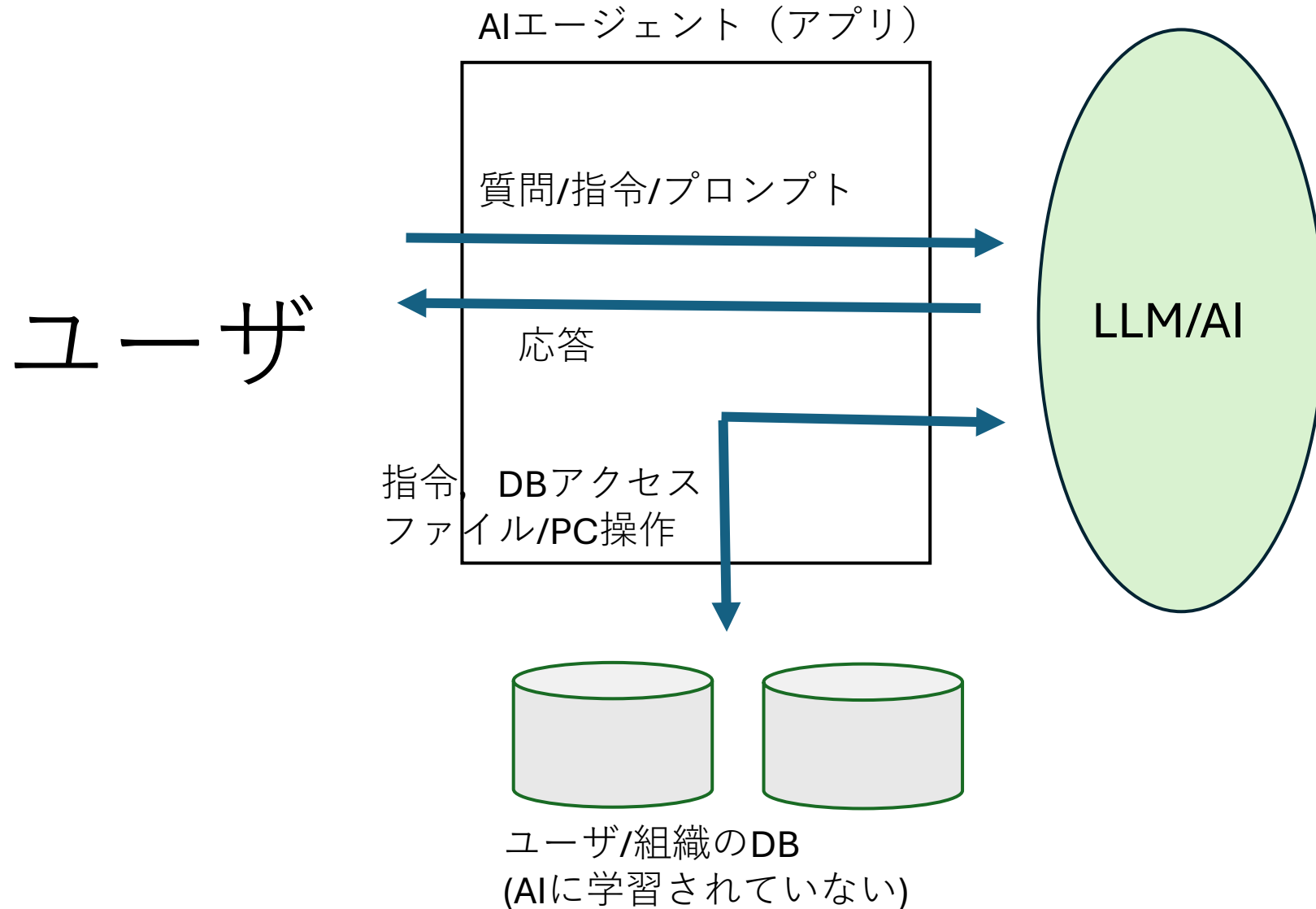
AIの仕組み1（回答のみ：シンプルなケース）



AIの仕組み2（回答のみ：シンプルなケース）



AIの仕組み3 (AIエージェント)



AIの仕組み・データがどこへ行く？

- データ処理がどこで実行されるかはケースバイケース

- ブラウザ(アプリ) による入力 → AI (クラウド/AIベンダー)

- ブラウザ(アプリ) による入力 → AI (ローカル/組織内)

- **個人情報・機微情報などをどこに入力してよいかは組織で規定されている。**

- AIエージェント

- ユーザによる指示に応じてLLMが複雑な処理を生成してPC内でAIエージェントアプリが実行

- データ分析・プログラムコード生成

- **操作ミスによる事故が発生し得る。**

AIシステム・アプリケーションへのサイバー攻撃

- AIによる便利な仕組み自体がサイバー攻撃の対象となる.
 - 学習データの操作
 - AIへの指令への介入(プロンプトインジェクション)
- 操作するユーザが騙される可能性
- 操作ミス・指令ミスも致命的な損害をもたらす可能性
 - サイバー攻撃だけではない.
 - 必要なデータの削除などが発生する可能性

AIエージェントの利用上のミスによる情報流出・消失

- AIエージェントが可能な処理が増大
 - PC内のファイルの読み書き
 - ファイル操作
 - データ転送
 - . . .
- AIエージェントにも事故防止の仕組み
 - ユーザが解除してしまえば仕組みは役に立たない

アカウントBANによる情報消失

- AIベンダーのアカウント
 - アカウントが停止
 - ルールが必ずしも公開されている訳ではない.
 - 規定に反する利用が一因？
 - 異議申し立てが却下されたらアップロードしているデータが失われる.
 - **重要なデータはローカルにもバックアップが必要**

再掲

(個人の)AI利用時の情報セキュリティ上の懸念点

- AIサービス利用時のデータ(ファイルアップロード)やデータ入力
 - **個人情報・機密情報が学習に**利用されてしまう(**情報漏えい**)
- 組織等で比較的安全なAI環境が用意されているが個人的な環境を利用してしまう.
 - 組織などでは、入力データを学習に利用することが制限されたサービスが提供されてしまうことが多い.
 - 個人の環境を利用してしまうと入力した情報が学習に利用されてしまう場合も (情報漏えいにつながる.)
 - **シャドーAI**と呼ばれる.
- AIシステム・アプリケーションへのサイバー攻撃
 - AIシステム基盤へのサイバー攻撃により学習データが改ざんされる.
 - AI特有の攻撃リスク → **プロンプトインジェクションによるAIの悪用**
 - **パソコン側のAIアプリ自体が改ざんされる or Fake のアプリを (騙されて) 利用させられる**
- AIエージェントの利用上のミス
 - AIエージェントへのプロンプトのミスで、重要なファイルを消去した
- AIサービスのアカウント BAN
 - AIベンダーの利用規約に反する利用・不正利用でアカウントが停止・強制削除される.
 - AIに入力しているデータ等が利用できなくなる

まとめ

- **AIの発展は急速**
 - 予測は難しい
- **AIの利用には情報の入力が必要**
 - どこにデータがアップロードされるのかの理解が必要
 - 特に**個人情報・機微情報・秘匿すべき情報**
- **AIの仕組み自体がサイバー攻撃の対象となる可能性大**
- **AIの仕組み自体がサイバー攻撃への対策していてもユーザ自身が騙されれば事故が発生する。**
 - 一般の詐欺事件と同様。

AIは便利・必要不可欠なツール

- 様々な職業でAI利用が必須になることが予想される.
- リスクがどこにあるかを理解しておく必要がある.
 - 報道やセキュリティ情報サイトを参考に
- AIの長期的な社会への影響は不明確
 - AIバブルの崩壊？
 - AIへの丸投げによる人間の能力の劣化？
 - AIが人間を支配する？
 - AIが人間を滅ぼす？

情報セキュリティの実務者の立場から一言

- 具体的なことは言えませんが. . .
 - なぜこんなことに（詐欺に）引っかかる??と思われる事例も
 - ICTに強い方でも油断は大敵！！
 - 最悪の事態を想定して置くべき