AI基礎教育資料

情報処理推進機構
産業サイバーセキュリティセンター5期
「初めてのセキュリティ塾」プロジェクト

本書の目的と対象読者

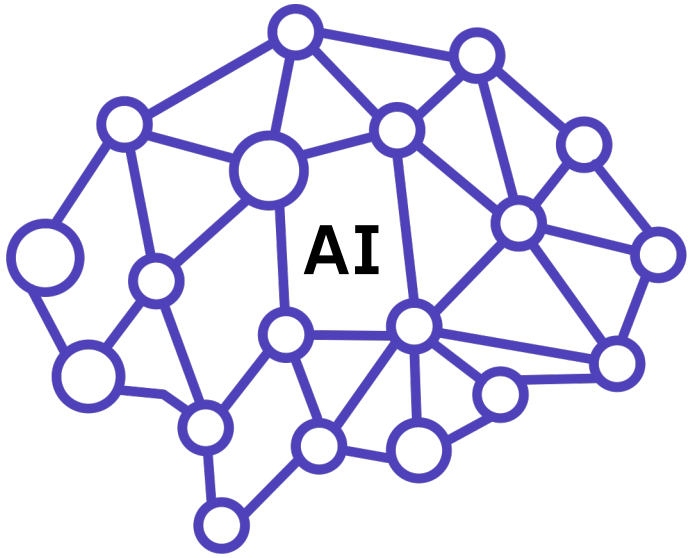
AIの特徴とリスクを理解する

目的 ：

- AIの簡単な仕組みを理解する
- AIを使用する際のリスクについて知見を得る

対象読者 ：

- AIが使われているサービスや製品を利用している方
- AIが使われているサービスや製品を利用する予定のある方
- 非技術者の方



第1章　AIとは

1.1 AIとは

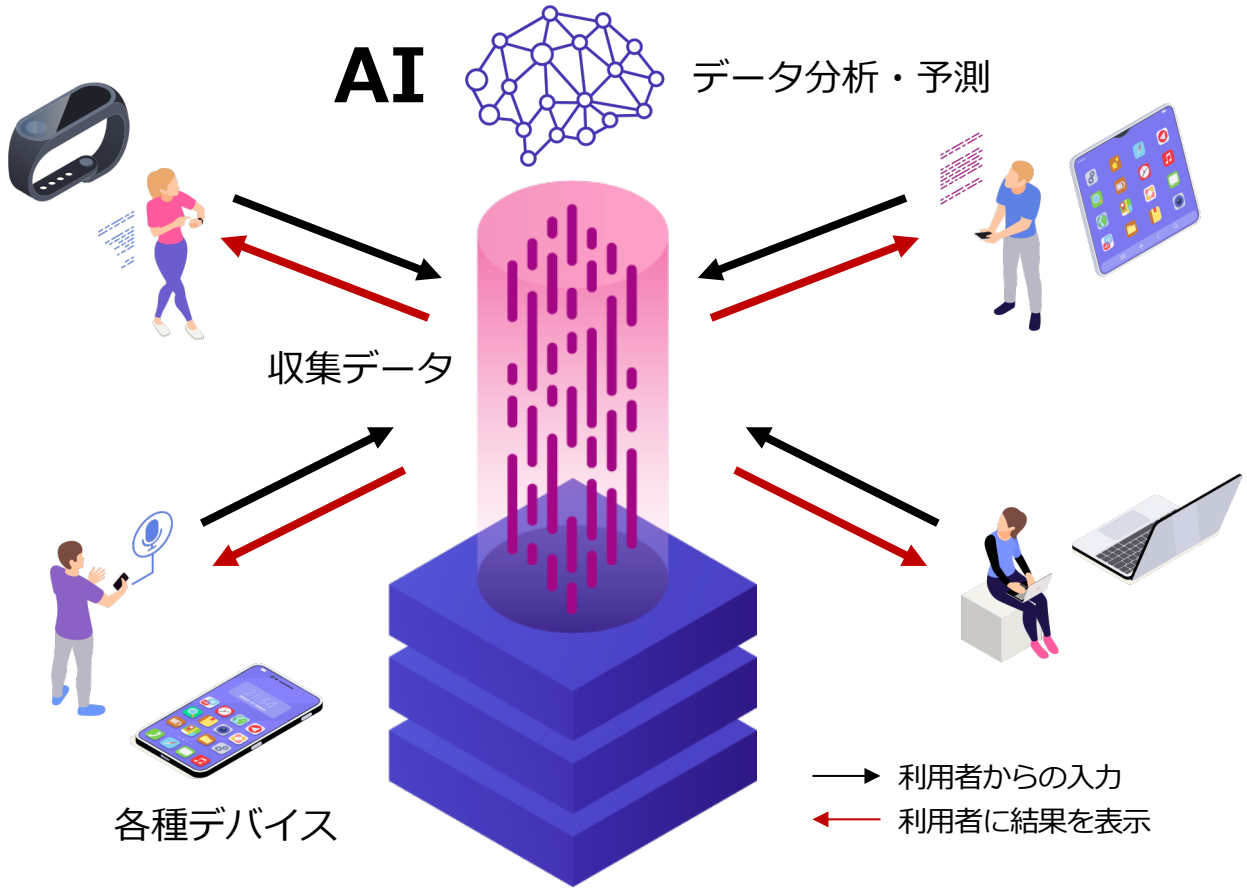
AIとはArtificial Intelligence（人工知能）の略称である

AIとは、一般に人間の知能や
その一部を模したアルゴリズムや
ソフトウェアを指します。

AIは、

① PCやスマートフォンなどの
各種デバイスに内蔵されている
センサーで収集したデータ

② 利用者からの入力などに
基づいて分析や予測を行い、
その結果をデバイスに表示します



1.2 AIの活用

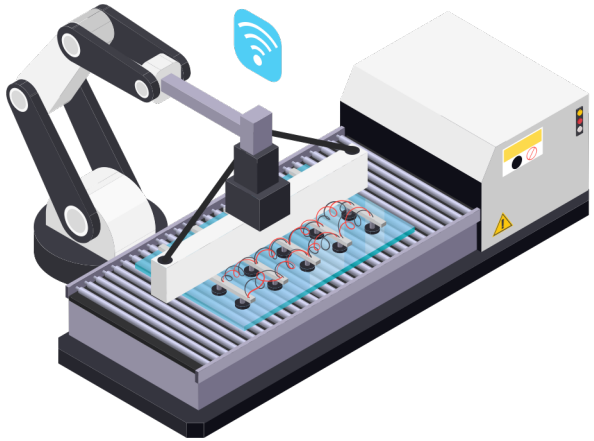
我々の身近な生活の中にもAIが活用されています

スマートフォンの音声認識

入力された音声をデータベースなどと照合し認識

自動運転技術

カメラ画像で車間距離や道路の白線を認識し動作を判断



製造品の品質確保

カメラ画像から不良品を検知

1.3 AIの得意 不得意

AIには得意なことと不得意なことがあります

AIが得意なこと

① **ルールやゴールが明確な作業**
将棋やチェスの戦術分析

② **大量のデータを蓄積し、分析する作業**
異常検知など

AIが不得意なこと

① **創造的な作業**
デザイン、研究など

② **言葉の意味を理解すること**
細かいニュアンス、文脈

③ **少ないデータで分析や予想すること**

第2章　AIのリスク

2.1 AIは万能ではない

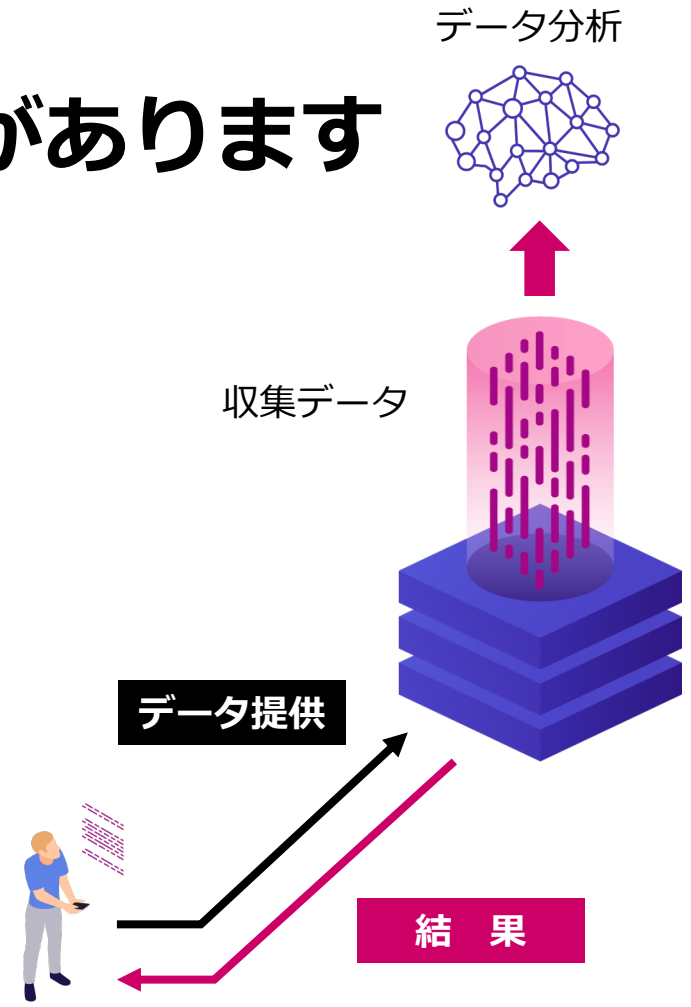
AIは便利ですが万能ではありません
その特性と制約を踏まえて利用する必要があります

① プログラムである

- AIはあくまでアルゴリズムやソフトウエアです
- あらかじめプログラムされた機能以上のことはできません

② 分析対象の十分かつ適切なデータが必要

- AIは入力されたデータに基づいてのみ学習を行います
- 裏を返すと、入力データから学習できない事柄について分析や予測することはできません



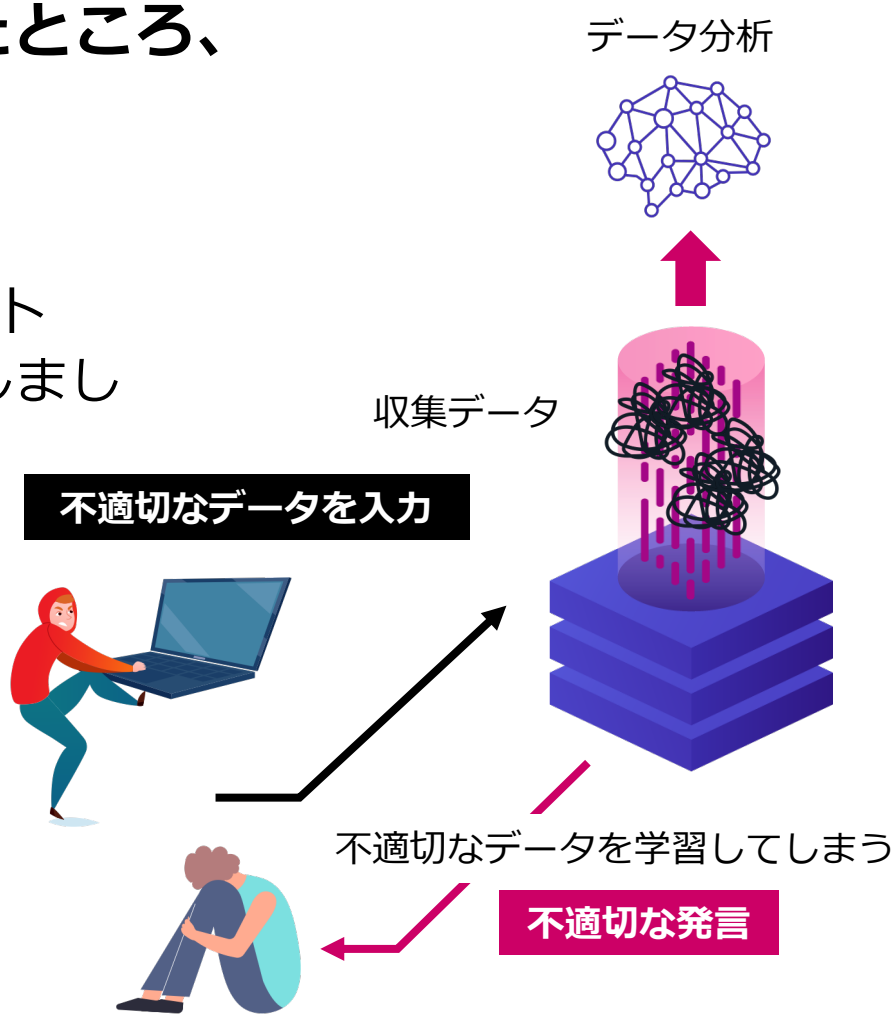
2.2 事例①：AIチャットボットによる差別的な発言

AIチャットボットの学習データをSNSから収集したところ、AIが不適切な発言を学習してしまい、差別的発言を繰り返すようになってしまった

2016年、米国IT企業がAIを用いたオンラインチャットボット（問いかけに対して返答するもの）を開発しSNS上に公開しました。

しかし、悪意のあるユーザーによって投稿された人種差別・性差別・暴力表現をAIが学習してしまい、不適切な発言を繰り返すようになりました。
その結果、公開から約16時間後にチャットボットは停止されました。

この事例によって、AIが不適切なデータを学習すると、不適切な出力を行ってしまうということが現実的なリスクであると分かりました。



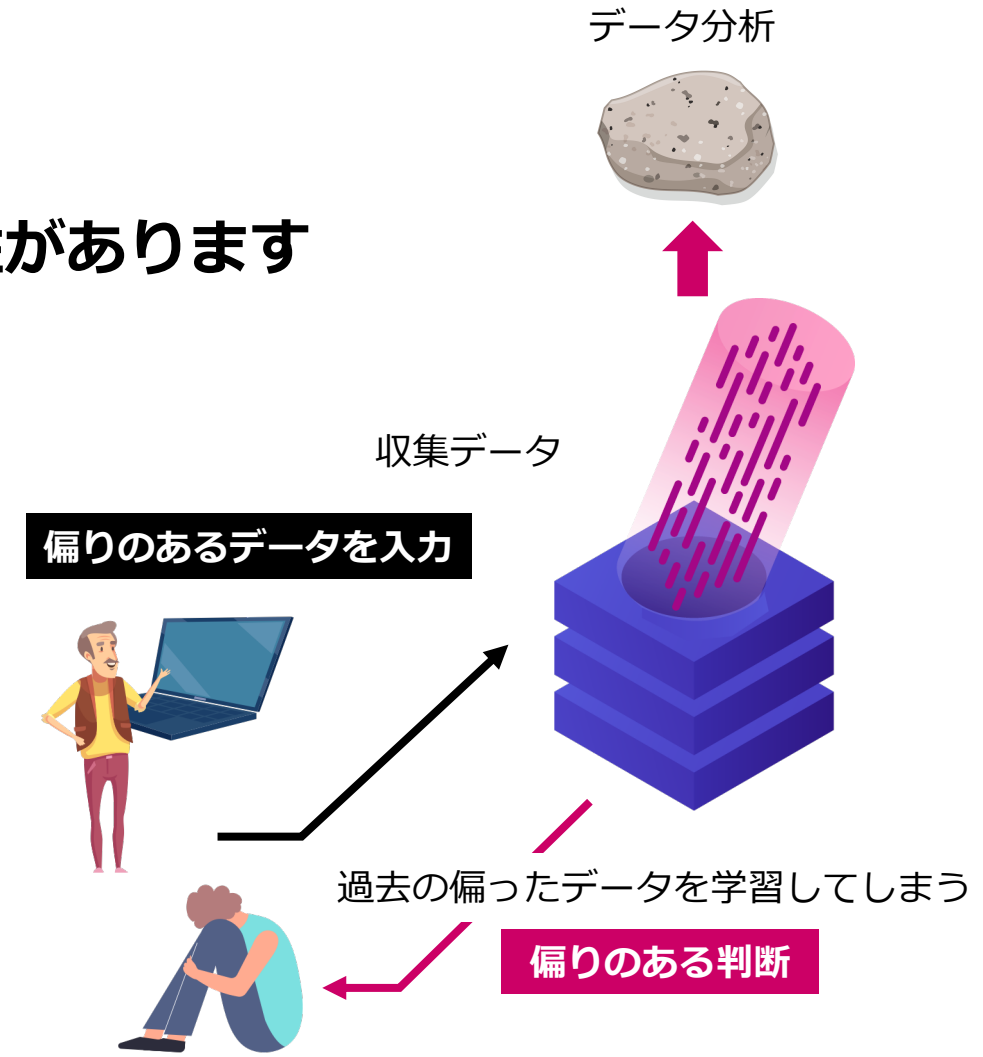
2.3 事例②：AIによる偏った採用の判断

AIが偏りのあるデータを学習したことにより、
偏った判断をしてしまう事例もあります。
AIによる偏った判断が倫理的問題に発展する可能性があります

大手IT企業が過去の人事データをAIに学習させて、
採用活動に利用したところ、
技術関係の職種において性別に偏りが生じました。

そもそも過去の人事データが偏っており、
そのデータを学習したAIは
偏った判断を行うようになりました。

この事例では、AIを盲信した採用活動が、
意図しない多様性の排除につながってしまいました。
企業のイメージダウンを招くリスクもあるため、
AIの利活用には慎重な判断が必要と言えます。



第3章　AI活用における注意点

3.1 気をつけるべきポイント

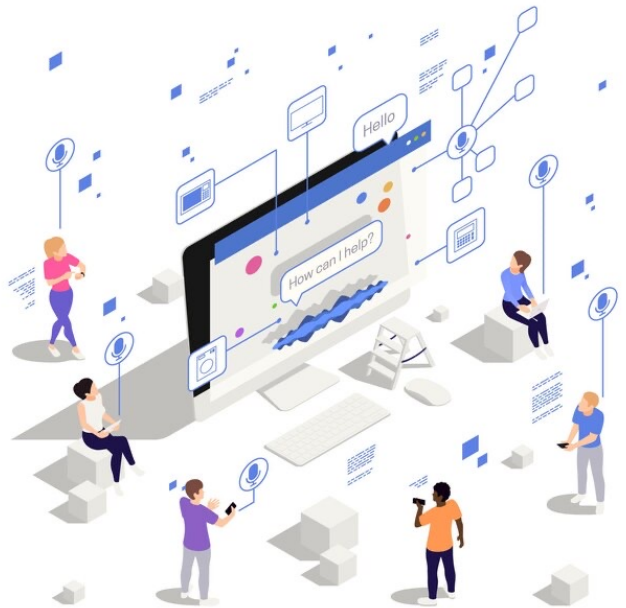
AIは、これからも発展・普及していく可能性がありますが過信は禁物です

① AIを過信しない

AIは完璧ではなく、まだまだ発展途上な技術です。
人間が意図しない動作をすることにも留意が必要です。

② AIの学習に使用するデータの重要性を理解する

学習データはAIの性能や判断に大きく影響します。
適切かつ十分な量のデータを収集することが重要です。



AIに頼りきるのではなく共存を目指しましょう。異常を感じたら周囲（*）に相談を！

＊企業で導入したAIシステムであればシステム担当者に相談しましょう。私的に利用しているAIサービスであればサービス窓口に相談しましょう

3.2 国際的な動き

こうした考え方は国際的にも注目を集めています

- 内閣府に設置された統合イノベーション戦略推進会議では「人間中心のAI　社会原則」を公開しており、AIへの過度な依存に対して警鐘を鳴らしています
- また、経済や社会福祉の向上を促進する国際機関OECD（Organisation　for　Economic Co-operation　and　Development：経済協力開発機構）においても、人間中心の価値観と公平性（原則1.2）として、AIの利用により人間中心の価値が侵害されるリスクについて述べています
- こうしたAIに依存し過ぎたり、AIを信用し過ぎて人間が振り回されたりすることなく、人間を中心としてとらえる考え方は国際的な潮流となっています

人間の尊厳が尊重される社会（Dignity）

我々は、AI を利活用して効率性や利便性を追求するあまり、人間が AI に過度に依存したり、人間の行動をコントロールすることに AI が利用される社会を構築するのではなく、人間が AI を道具として使いこなすことによって、人間の様々な能力をさらに発揮することを可能とし、より大きな創造性を発揮したり、やりがいのある仕事に従事したりすることで、物質的にも精神的にも豊かな生活を送ることができるような、人間の尊厳が尊重される社会を構築する必要がある。（引用：「人間中心のAI社会原則」）