AI利活用と倫理

熊本保健科学大学「データサイエンス入門」

熊本保健科学大学　保健科学部　共通教育センター

水本　豪

AIってどんなもの？

※この講義では、「AIは人間の知能を人工的に実現したもの」と考える。

レベルで考えるAIの分類

レベル	説　明	例
レベル1 単純な制御	入力されると決められたルールに従って出力する → こういう場合はこのようにする	エアコンの湿度自動調整　など
レベル2 ルールに基づく	蓄積しておいたルールをもとに探索し、出力する 自ら学習することはない	メールのフィルター機能　など
レベル3 機械学習	学習したデータからルールやパターンを抽出する 注目するポイントは教えておく必要あり	検索エンジン　など
レベル4 深層学習	機械学習の手法のひとつである深層学習を用いることで、データについて注目するポイントを自動的に抽出可能になった	顔認証、自動翻訳、自動運転　など

次のレベル：人間と同等かそれ以上の思考を持つAI（これはまだ実現していない）

AIの歴史を振り返ってみよう

1950年代後半～1960年代（第１次AIブーム）　キーワード：推論と探索

単純で明確なルールを持つ問題（パズルや迷路など）に対し、場合分けや試行錯誤を経て解に到達。
人間の脳を再現したかのように見えたため、はじめはとても注目された。
しかし実世界の曖昧で複雑な問題に対しては解を出せなかった（膨大な時間をかければ可能かもしれない）
身の回りの問題が解けない所詮はおもちゃみたいなもの ⇒ トイ・プロブレム

→ ブームは去り、**1970年代　AI冬の時代**へ

1980年代（第２次AIブーム）　キーワード：知識の獲得

実社会の問題に対応できるAIとするため、専門家（エキスパート）による情報を大量に蓄積したシステム（エキスパートシステム）を構築。質問に対し、エキスパートシステムから答えを抽出して解を出した。対話システムや医療（診断）に活用。
汎用的なものとするためには、あらゆる知識を蓄積し、管理する必要があるが、現実的に困難。

→　ブームは去り、**再び冬の時代**

AIの歴史を振り返ってみよう

2000年代以降（第３次AIブーム）　キーワード：機械学習、深層学習

コンピュータや通信の進歩に伴い、ビッグデータを活用できるようになると、機械学習の技術により、膨大な情報からルールやパターンを見つけられるようになった。さらに、機械学習の技術自体も進歩し、ディープラーニング（深層学習）のような方法によりデータのどこに注目すればよいか（特徴量）を自動的に抽出できるようになった。**⇒ AIとビッグデータは相性抜群！**

現在のAIは何でもできるのか？

現在のAIはレベルで言えば「レベル４」、つまり人間と同等のところまではいっていない！

現在のAIは「特化型AI」＝特定の問題に特化したAI　⇒　何でもできるわけではない

※状況に応じて総合的な判断をするAI（考えるAI；汎用AI）

生成AIの登場

「与えられたデータを処理するAI」から「新しいコンテンツを生み出すAIへ」

生成AIが登場するまでのAI　⇒　与えられたデータに対して識別したり予測したりする
生成AI　⇒　既存の情報の再構成や新しいコンテンツの生成を行う

なぜそんなことができるようになったのか？

深層学習（ディープラーニング）によるAI性能の飛躍的な向上
GAN（敵対的生成ネットワーク, 2014）などの技術の蓄積
急に可能になったというよりは、この10年くらいの進歩が可能にした技術

主な生成AIサービス

言語を生成するAI	画像や動画を生成するAI	音を生成するAI
ChatGPT(Open AI) Gemini (Google)	StableDiffusion (Stability AI) Midjourney (Midjourney) Adobe Firefly (Adobe)	MusicGen (Meta)

AIのリスクとは？

AI戦略会議（2023/05/26）公開資料より

・機密情報の漏洩や個人情報の不適切な利用のリスク

・犯罪の巧妙化・容易化に繋がるリスク

・偽情報などが社会を不安定化・混乱させるリスク

・サイバー攻撃が巧妙化するリスク

・教育現場における生成AIの扱い

・著作権侵害のリスク

・AIによって失業者が増えるリスク

AI活用による負の事例

誤情報の生成

・福岡県の魅力を発信する目的で開設されたサイトの記事を生成AIで作成したことで、実在しない観光名所やご当地グルメが紹介されていた。
（記事がインターネット上の情報をもとにAI生成していること、情報の正確性を保証するものではないことが記載）。

・名義上の後援を行っていた福岡市と飯塚市は、後援申請の時点で生成AIによる記事作成を知らされていなかった。（後に後援は取り消し）

誤情報の例

福岡市「うみなかハピネスワールド」　・・・ 実在しない
福岡市「かしいかえんシルバニアガーデン」　・・・ 2021年末に閉園
古賀市「福津大自然公園」「鹿児島湾」「恋の浦海岸」　・・・ 古賀市内にはない

ディープフェイク技術を利用した詐欺

・香港の多国籍企業の財務担当者が、AIを利用して合成・複製されたディープフェイクによる同僚の姿が映ったビデオ通話に騙され、約38億円を送金。

AI活用による負の事例

実際の状況に対応できなかったり、情報が不足したことで誤った予測

・アメリカの不動産関連企業がAIを使って不動産を迅速に査定する事業を展開していたが、コロナ禍による状況の激変に対応できず、物件の価格を誤って予測し、数億ドルの損失を招いた。最終的に企業は事業から撤退。

・アメリカの企業が糖尿病網膜症の兆候を見つけるAIシステムを開発したが、医療現場の画像の品質に対応できず、適切な判断を下せないことが明らかになった。

・子ども家庭庁が、虐待が疑われる子どもの一時保護の必要性をAIに判定させるシステムの開発を進めてきたが、テスト段階で判定ミスが6割にも及び、実用化は困難と結論づけた。

差別的な判断

・アメリカの企業でAIによる人材採用システムを導入したところ、技術職において男性ばかりが採用され、女性を差別している可能性があるという欠陥が発覚。システムの運用を中止した。（過去の履歴書や採用結果のデータをもとに5点満点でランク付けするシステムであったが、過去に技術職に応募した人の大半が男性であったために、AIが「男性の採用が望ましい」と認識してしまったと思われる。）

AIとどのように向き合うか

AI倫理に関する指針やガイドライン、国際的な取り組み

年	事項
2016	アシロマAI原則（Future Life Institute）
2017	IEEE EAD ver. 2（アメリカ電子情報通信学会；IEEE） 人工知能学会　学会員が守るべき倫理指針（人工知能学会）
2019	AI原則（経済協力開発機構；OECD）　※2024年改訂 信頼できるAIのための倫理ガイドライン（EU） 人間中心のAI社会原則（内閣府）
2021	UNESCO AI倫理に関する勧告（国連教育科学文化機関；UNESCO）
2023	熊本保健科学大学における学生の生成AIの利用について
2024	AI Act（EU AI規制法）（EU）　※規制の内容により段階的に施行（最終的に2030年末にはすべて施行）
2024	初等中等教育段階における生成AIの利活用に関するガイドライン（文部科学省） AI事業者ガイドライン（総務省、経済産業省） 医療デジタルデータのAI研究開発等への利活用に係るガイドライン （保健医療分野におけるデジタルデータのAI研究開発等への利活用に係る倫理的・法的・社会的課題の抽出及び対応策の提言のための研究班）
2025	行政の進化と革新のための生成AIの調達・利活用に係るガイドライン（デジタル庁） 人工知能関連技術の研究開発及び活用の推進に関する法律（AI法案）[国会審議中]

多くの原則・ガイドラインの共通要素

- 人間中心であること（人類の成長と幸福への寄与）
- 人権、プライバシー、公平性の尊重
- システムの頑健性と安全性
- 透明性、説明可能性
- 説明責任

【参考】人間中心のAI社会原則（内閣府）

AIによる恩恵が受けられる社会（AI活用に対応した社会；AI-Readyな社会）を実現し、適切かつ積極的なAIの社会実装を推進するために留意すべき事項。

- 人間中心の原則
- 教育・リテラシーの原則
- プライバシー保護の原則
- セキュリティ確保の原則
- 公正競争確保の原則
- 公平性、説明責任及び透明性の原則
- イノベーションの原則

【参考】AI原則（OECD）

AIが信頼できるものであるために必要な事項。

- すべての人や社会の成長、持続可能な開発、幸福 Inclusive growth, sustainable development and well-being
- 法の支配、人権、民主主義的価値の尊重（公平性、プライバシーを含む）
  Respect for the rule of law, human rights and democratic values, including fairness and privacy
- 透明性及び説明可能性 Transparency and explainability　※説明可能性：AIの判断や結果が人々に理解できるものであること
- システムの頑健性、セキュリティ、安全性 Robustness, security, and safety
- 説明責任 Accountability　※説明責任：AIの影響や結果に対する責任を負うこと

【参考】信頼できるAIのための倫理ガイドライン（EU） Ethics guidelines for trustworthy AI

信頼できるAIの要件

- 人間が主体であること、人間が監視すること Human agency and oversight
- 技術的な頑健性と安全性 Technical robustness and safety
- プライバシーとデータガバナンス Privacy and data governance
- 透明性 Transparency
- 多様性、差別をしないこと、公平性 Diversity, non-discrimination and fairness
- 社会的幸福、環境的幸福 Societal environmental wellbeing
- 説明責任 Accountability

AI（特に生成AI）とどのように向き合うか

AIを正しく理解しよう

・AIが解として出すものには誤りを含む可能性や事実と異なる内容である可能性がある。
・AIがどのようなデータを学習しているのか、どのようなアルゴリズムで解を出しているのかなど「透明性」が担保されない可能性がある。
・AIを活用することで機密情報や個人情報の漏洩・不適切な使用、著作権侵害、偏見に基づく回答など、出力される内容について「信頼性」が担保されない可能性がある。
・AIが解として出したものにすべてを委ねるのではなく、わたしたち人間がその真偽や適切性をよく吟味し、自ら考えて判断を行う必要がある。

不適切な使用はしない

・生成AIによる生成物を自らの成果物として提出しない。
・自分が課された小テスト等の問題をAIに回答させない。

熊本保健科学大学における学生の生成AIの利用について (2023.09.26)

生成AIの利用について留意すること

１ 試験・レポート等での取扱いについて

試験やレポート等において、生成AIが作成した文章を自らの成果物として使用し提出することは不正行為に該当しますので控えてください。生成AIが作成した情報には他人の著作物の内容がそのまま含まれていることもあるため、意図せずに盗用や剽窃を行ってしまう可能性もあります。

２ 授業での利用について

授業における生成AIの利用については、この文書に示す留意点をよく理解したうえで当面は各教員の指示に従って行ってください。

３ 情報流出・情報漏洩のリスクについて

個人情報や未発表の研究成果などは、厳重な管理が必要です。こうした情報を生成AIに対する質問や指示として入力すれば、それが他の誰かの質問の回答に使用されるなど、情報の流出や漏洩に繋がる可能性があります。第三者に公開できない情報は絶対に入力しないでください。

４ 誤情報の可能性について

生成AIにより作成された情報には誤った内容が含まれていることもあります。そのため、生成AIが作成した情報を鵜呑みにせず、信頼できる情報で裏付けるよう、自ら確認することが必要です。

AI（特に生成AI）とどのように向き合うか

今まさに進行中の問題

・AIは適切に使用すればとても便利なツール。しかし、まだまだ不透明な部分も多いため、開発側は透明で説明可能なAIの開発を進めている。その結果、新たな技術や活用方法が日々生み出されており、半年先には今の「常識」が通用しない可能性も否定できない。

・AIが活用される時代に生きるわたしたちは、AI利活用と規制に関する情報にアンテナを向け、正しい知識を身につけて活用していくことが求められる。