AIと倫理

広島大学 AI・データイノベーション教育研究センター
村上 祐子

目標

AIの利活用について倫理を考慮することが必要不可欠であることを説明できる。

この授業で紹介すること

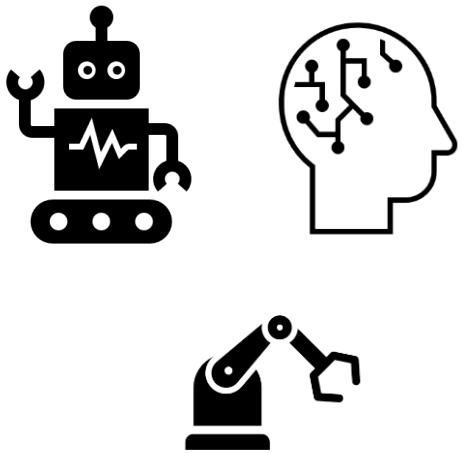
- AIが社会に与える関心や不安の高まりから、国際的にAIの利活用の方針が策定
- 日本でのAIに関する技術の活用の倫理指針

キーワード

AI倫理、AI脅威論、アシロマAI原則、人間中心のAI社会原則

こんなことはありませんか

AI(Artificial Intelligence: 人工知能)は人間にとって脅威になる？



v.s.

倫理とは？

社会規範

「社会において守らなければならないとされているルール(規範)」

道徳

- 人々の内心の規範
- 個人の考え方に起因

「ご飯の時にTVを見るのは行儀が悪い」

法律

- 社会規範の一つ
- 違反した場合、国家による制裁がある
- 国家が違反者に対して制裁を受けることを強制できる

AIは脅威なのか？

21世紀に入ってからAIが人間の存在を脅かすという不安が急激に増加

- The Singularity Is Near「シンギュラリティは近い」

  Ray Kurzweil (2005)

  - 機械の知能は人間の知能をすべて合わせたものより強力になる
  - シンギュラリティとは、機械の知能と人間が拮抗する時点
  - 2045年に到来する？

- *The future of employment: How susceptible are jobs to computerisation?*

  Carl Benedikt Frey and Michael A. Osborne (2016)
  https://doi.org/10.1016/j.techfore.2016.08.019

  発表は2013年

  - 米国の2010年の雇用統計を元に、人々が行っている業務について総雇用の約47%がコンピュータにとって代わられるリスクが高いと推定

Asilomar AI Principles（アシロマAI原則）

- 人工知能の可能性に対する広い社会からの関心の高まり
- 開発に関与する人々には、AI をより良いものにする責任と機会がある

2017年8月2日
非営利団体Future of Life Instituteは人工知能の「研究課題」、「倫理と価値観」、「長期的な課題」の３領域を含む計23項目のガイドラインを公表

https://futureoflife.org/open-letter/ai-principles/
https://futureoflife.org/open-letter/ai-principles-japanese/

アシロマAI原則：倫理と価値

6) 安全性：人工知能システムは、運用寿命を通じて安全かつロバストであるべきで、適用可能かつ現実的な範囲で検証されるべきである。

7) 障害の透明性：人工知能システムが何らかの被害を生じさせた場合に、その理由を確認できるべきである。

8) 司法の透明性：司法の場においては、意思決定における自律システムのいかなる関与についても、権限を持つ人間によって監査を可能としうる十分な説明を提供すべきである。

9) 責任：高度な人工知能システムの設計者および構築者は、その利用、悪用、結果がもたらす道徳的影響に責任を負いかつ、そうした影響の形成に関わるステークホルダーである。

10) 価値観の調和：高度な自律的人工知能システムは、その目的と振る舞いが確実に人間の価値観と調和するよう設計されるべきである。

11) 人間の価値観：人工知能システムは、人間の尊厳、権利、自由、そして文化的多様性に適合するように設計され、運用されるべきである。

12) 個人のプライバシー： 人々は、人工知能システムが個人のデータ分析し利用して生み出したデータに対し、自らアクセスし、管理し、制御する権利を持つべきである。

13) 自由とプライバシー：個人のデータに対する人工知能の適用を通じて、個人が本来持つまたは持つはずの自由を不合理に侵害してはならない。

14) 利益の共有：人工知能技術は、できる限り多くの人々に利益をもたらし、また力を与えるべきである。

15) 繁栄の共有：人工知能によって作り出される経済的繁栄は、広く共有され、人類すべての利益となるべきである。

16) 人間による制御：人間が実現しようとする目的の達成を人工知能システムに任せようとする場合、その方法と、それ以前に判断を委ねるか否かについての判断を人間が行うべきである。

17) 非破壊：高度な人工知能システムがもたらす制御の力は、既存の健全な社会の基盤となっている社会的および市民的プロセスを尊重した形での改善に資するべきであり、既存のプロセスを覆すものであってはならない。

18) 人工知能軍拡競争：自律型致死兵器の軍拡競争は避けるべきである。

例題

AIを医療分野に取り入れた実用例として、大腸内視鏡画像をAIで解析し、医師の診断を補助するというものがあります。一例を紹介します。

1. 医師が患者の大腸内視鏡画像を撮影します。
2. AIを用いた解析ソフトウェアを用いて、がんなどの異常な状況を検出します。
   - あらかじめソフトに集積されている多くの内視鏡画像のデータから学習されています。
3. 異常を発見した場合、その写真に対して警告を発します。
   - ポリープ、がんのステージなどの診断はしません。

一連の操作において、アシロマ原則の中でどこが留意されているか考えてみましょう。

解説

―例として、「16) 人間による制御」が考慮されている。

人間が実現しようとする目的の達成を人工知能システムに任せようとする場合、その方法と、それ以前に判断を委ねるか否かについての判断を人間が行うべきである。

- 人間が実現しようとする目的→臓器の異常を画像診断したい。
- 判断を委ねるか否かについての判断を人間が行うべき
  →異常が疑われる画像の検知はAIが行うが、異常かどうか診断するのは人間。

より多くの人の診断が可能になり、かつ、早期段階で異常を発見することができることは「 14) 利益の共有」も考慮されている。

AI倫理の意味の変遷

AIが脅威にならないようにどのように使うかという考え方

高い能力を持つAI(シンギュラリティの実現)によりAIが人間と敵対する脅威になる状況を避けたい

- シンギュラリティの実現時期は想定以上に遠い
- 「AIが人間と敵対する脅威になる」可能性は低い

AIの利活用に伴う社会的問題を解決するための考え方

- 2015～2017年：研究機関、専門機関が中心となりAI原則を策定
- 2018年以降：企業からのAI原則の発表が増加

https://aiindex.stanford.edu/ai-index-report-2021/

日本におけるAI倫理原則の策定

日付	作成者	名称
2017年2月28日	人工知能学会	人工知能学会倫理指針
2017年7月28日	総務省AIネットワーク社会推進会議	国際的な議論のためのAI開発ガイドライン案
2019年3月29日	統合イノベーション戦略推進会議	人間中心のAI社会原則
2019年8月9日	総務省AIネットワーク社会推進会議	AI利活用ガイドライン

人間中心のAI社会原則 https://www8.cao.go.jp/cstp/aigensoku.pdf

AIは情報システムに組み込まれる「高度に複雑な情報システム一般」として活用されることを想定し、AIの研究開発や社会実装において考慮すべき問題を列挙。

1. 人間中心の原則
2. 教育・リテラシーの原則
3. プライバシー確保の原則
4. セキュリティ確保の原則
5. 公正競争確保の原則
6. 公平性、説明責任及び透明性の原則
7. イノベーションの原則

問題

1. 既存のシステムでAIを取り入れたらよいと思うシステムを考えてみましょう。あるいは、既に取り入れられているAIに関連するシステムについて調べてみましょう。
2. 1で取り上げたシステムについて、７つの観点を考慮しているか評価してみましょう。
   - システムの構成がどのようにそれぞれの原則を守っているのか具体的に列挙しましょう。