「責任ある生成AIのガバナンス・デザインと産学の役割」

論点提起

大阪大学 社会技術共創研究センター 特任准教授
工藤郁子
2024年6月26日

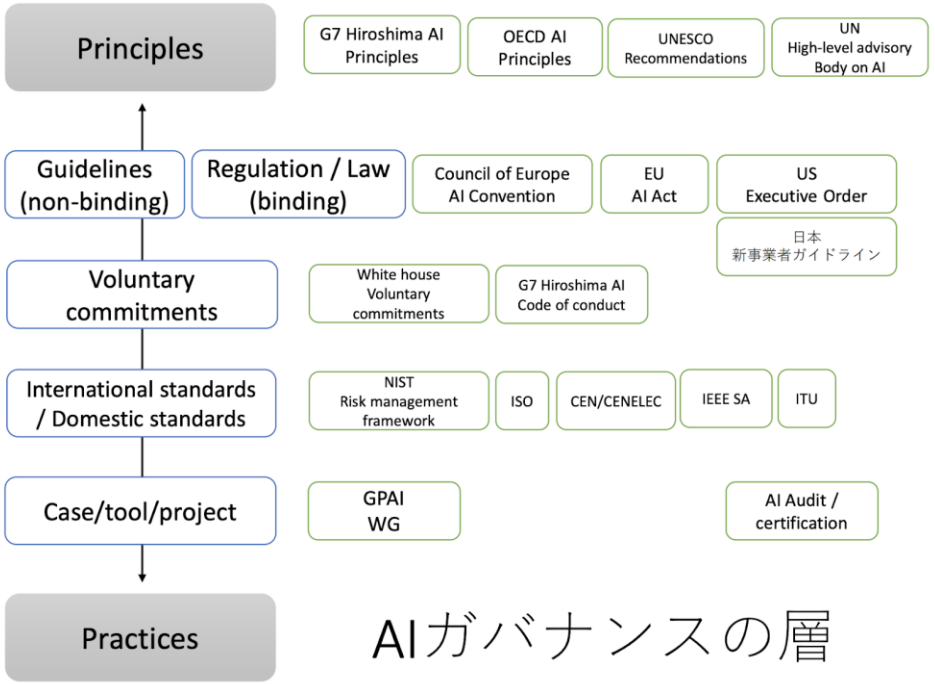
Osaka University
Research Center on
Ethical, Legal and
Social Issues

論点

- Why + What
- Who + How
- Where
- When

AIガバナンスの重層化

- 2010年代後半、AI倫理原則（principle）を議論/合意形成
  - What + Whyをめぐるもの
- 2020年代以降、AIに関する倫理・法の実践（practice）が増加
  - Who + Howをめぐるもの
  - とはいえ、What + Whyをめぐる議論が沈静化した訳ではない



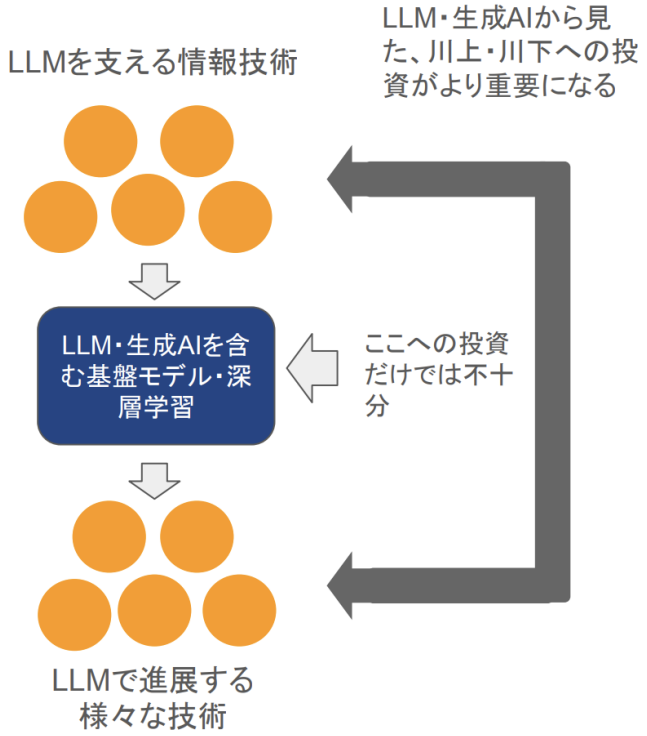
江間有沙「AIガバナンスの Principles and Practices」
（総務省「デジタル空間における情報流通の健全性確保の在り方に関する検討会」）
https://www.soumu.go.jp/main_content/000913994.pdf

各国のAIガバナンスの概況

EU	• AI法には「基本権（fundamental rights）」という言葉が100回以上登場 • 差別されない権利やプライバシー等の基本権を重視	（近代立憲主義は国家権力を制限し統制することで個人の自由と権利を守るとの発想に立つが）Big Tech が国家に匹敵する権力を持つとの現状認識のもと、国家権力だけでなくデジタル空間を対象とした統制も必要との考え	**デジタル立憲主義**
US	• 2023年1月、商務省NISTが「AIリスク管理フレームワーク」を公開したが、あくまでガイドライン • 2023年7月、AI企業7社、同年9月には新たに8社をホワイトハウスに招集し、自主規制推進で合意 • ただし、2023年10月、「AIの安心安全で信頼できる開発と利用に関する大統領令」では法的拘束力	市場と技術の力を信じ、企業の自主的な取組みを促進しつつ必要なら政府が規制をかける、との発想	**デジタル自由主義**
CN	• 2023年8月に施行された「生成AIサービス管理暫定弁法」では、生成AIサービスの提供・利用にあたって「社会主義の核心的価値観を堅持する」ことを義務付け	（個人の自由・権利より）国家・政権の安定を最も重視する思潮	**デジタル権威主義**

Cf. 見落とされがちなリスク？

- LLM/生成AIから見た、川上・川下の科学技術に関するガバナンスのあり方
  - 計算資源等：需要が増加傾向にあり、電力供給/冷却、ネットワーク帯域などがボトルネックに。また、日本では半導体産業が重視される中、その競争力を高められるようなAI開発ができるかも課題に。
  - 環境性能： 環境負荷の増大が懸念されており、EUのAI規則の目的規定でも環境保護が記載。
  - データガバナンス：必ずしも巨大なモデルを作れなくても競争力のある研究開発が可能と示唆されてきたことから、（学習以上に）学習するためのデータを整備することへの注力が予期。
  - AI4Science: 各分野の研究開発で、責任あるAIを利用する取組みや人材育成・能力開発が必要。



川原圭博「生成AIの登場と情報学の諸分野の役割」
（文部科学省「情報科学技術分野における戦略的重要研究開発領域に関する検討会」）
https://www.mext.go.jp/content/20240514-mxt_jyohoka01-000036161_05.pdf

論点

- Why + What
- Who + How
- Where
- When

マルチステークホルダー

- AIガバナンス分野では、従前よりマルチステークホルダープロセスを指向
  - 利害関係者たちが課題解決の議論や合意形成過程に参画すること
  - 政府関係者、ビジネスリーダー、研究者、市民社会代表などとの議論を重ねる場づくりが実践されてきた
- グローバルでの協働も進む

Cf. ダボス会議（2023年1月）

Cf. G7公式官民イベント「DXサミット」(2024年4月)

Cf. DXサミット「AIセッション」(2024年4月)

松本剛明 総務大臣

Mike Yeh, Microsoft

Mia Garlick, Meta

Michaela Browning, Google

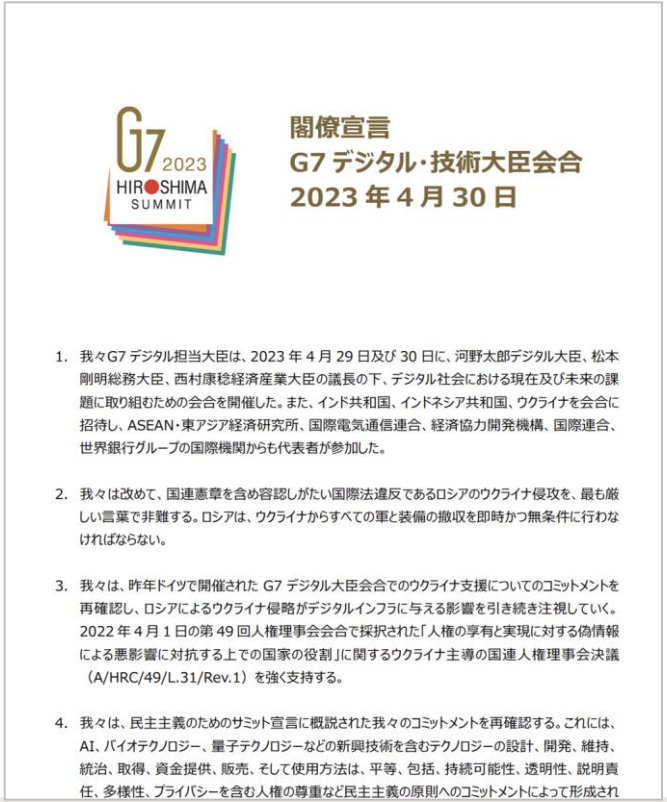
田中繁広, NEC

Ugo Pagallo, University of Turin

Cf.「G7デジタル・技術大臣会合閣僚宣言」(2024年4月)

> 35. 我々はまた、マルチステークホルダーコミュニティがG7閣僚と2023年4月28日の**G7デジタルトランスフォーメーションサミット**で、AI、デジタルインフラ、またタスクフォースが取りまとめた「ガバナンス原則」の成果に基づくアジャイルガバナンスの議論を含む、デジタルトランスフォーメーションに係る重要な課題について行った**討議とそれに基づく要請を承認**する。

「G7群馬高崎デジタル・技術大臣会合を開催しました」
(経済産業省サイトより抜粋)
https://www.meti.go.jp/press/2023/04/20230430001/20230430001.html



G7 2023 HIROSHIMA SUMMIT

閣僚宣言
G7 デジタル・技術大臣会合
2023 年 4 月 30 日

1. 我々G7 デジタル担当大臣は、2023 年 4 月 29 日及び 30 日に、河野太郎デジタル大臣、松本剛明総務大臣、西村康稔経済産業大臣の議長の下、デジタル社会における現在及び未来の課題に取り組むための会合を開催した。また、インド共和国、インドネシア共和国、ウクライナを会合に招待し、ASEAN・東アジア経済研究所、国際電気通信連合、経済協力開発機構、国際連合、世界銀行グループの国際機関からも代表者が参加した。

2. 我々は改めて、国連憲章を含め容認しがたい国際法違反であるロシアのウクライナ侵攻を、最も厳しい言葉で非難する。ロシアは、ウクライナからすべての軍と装備の撤収を即時かつ無条件に行わなければならない。

3. 我々は、昨年ドイツで開催された G7 デジタル大臣会合でのウクライナ支援についてのコミットメントを再確認し、ロシアによるウクライナ侵略がデジタルインフラに与える影響を引き続き注視していく。2022 年 4 月 1 日の第 49 回人権理事会会合で採択された「人権の享有と実現に対する偽情報による悪影響に対抗する上での国家の役割」に関するウクライナ主導の国連人権理事会決議（A/HRC/49/L.31/Rev.1）を強く支持する。

4. 我々は、民主主義のためのサミット宣言に概説された我々のコミットメントを再確認する。これには、AI、バイオテクノロジー、量子テクノロジーなどの新興技術を含むテクノロジーの設計、開発、維持、統治、取得、資金提供、販売、そして使用方法は、平等、包括、持続可能性、透明性、説明責任、多様性、プライバシーを含む人権の尊重など民主主義の原則へのコミットメントによって形成され

AIに関する次世代リーダーとの車座対話
（2023年5月）
https://www.kantei.go.jp/jp/101_kishida/actions/202305/09kurumaza.html

“

デジタルについて、G7首脳は、G7の価値に沿った生成系AIや没入型技術のガバナンスの必要性を確認するとともに、特に生成系AIについては、「**広島AIプロセス**」として担当閣僚のもとで速やかに議論させ、本年中に結果を報告させることとなりました。

G7広島サミット セッション1（ワーキング・ランチ）「分断と対立ではなく協調の国際社会へ／世界経済」概要(外務省サイトより抜粋)
https://www.mofa.go.jp/mofaj/ecm/ec/page1_001683.html

AIガバナンス拠点の政策的形成

- 従前、AIガバナンスに関する研究・実践の拠点は、大学・研究機関や民間企業・コンソーシアム・NGO等により運営されてきた
  - 米 ニューヨーク大学 AI Now Institute、米 スタンフォード大学 Institute for Human-Centered AI (HAI)、英 Alan Turing Institute、米 Future of Life Institute、米 Partnership on AI、英 Centre for the Governance of AI など多数。日本国内にも多くの拠点・コミュニティが存在
  - AIに特化していないものの、新規科学技術の ELSI（Ethical, Legal and Social Issues）等に関する組織・コミュニティでも研究・実践がされてきた
- 2023年11月に開催された「AI Safety Summit」にて、英国政府と米国政府が AI Safety Institute 設立を発表。2024年2月、日本政府も AI Safety Institute を設立
  - 政策的な拠点形成とプロジェクト推進が行われる傾向が見られる
  - とりわけ日本では、AIガバナンスに関わるエフォートが特定の研究者等に集中しがちであり、組織的な対応・体制ができていない点が課題だったところ、改善が期待される

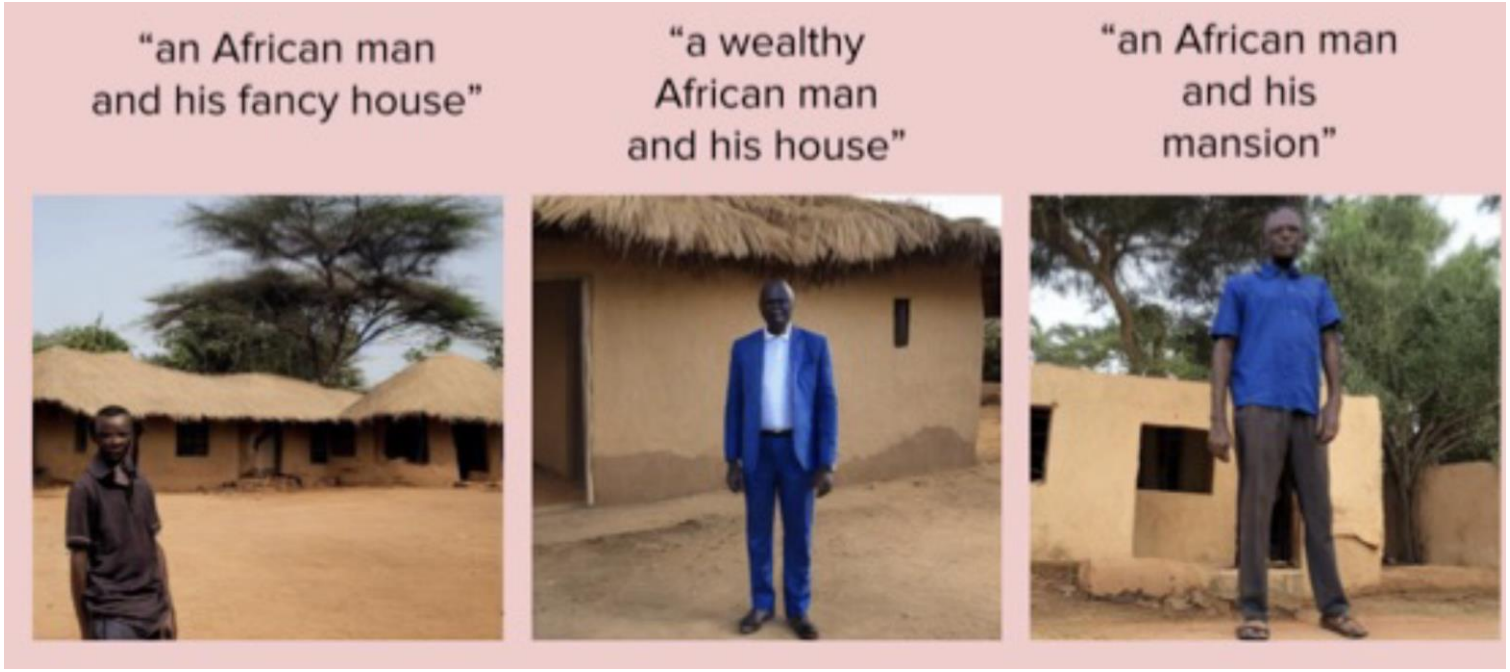
論点

- Why + What
- Who + How
- Where
- When

生成AIとフェアネス

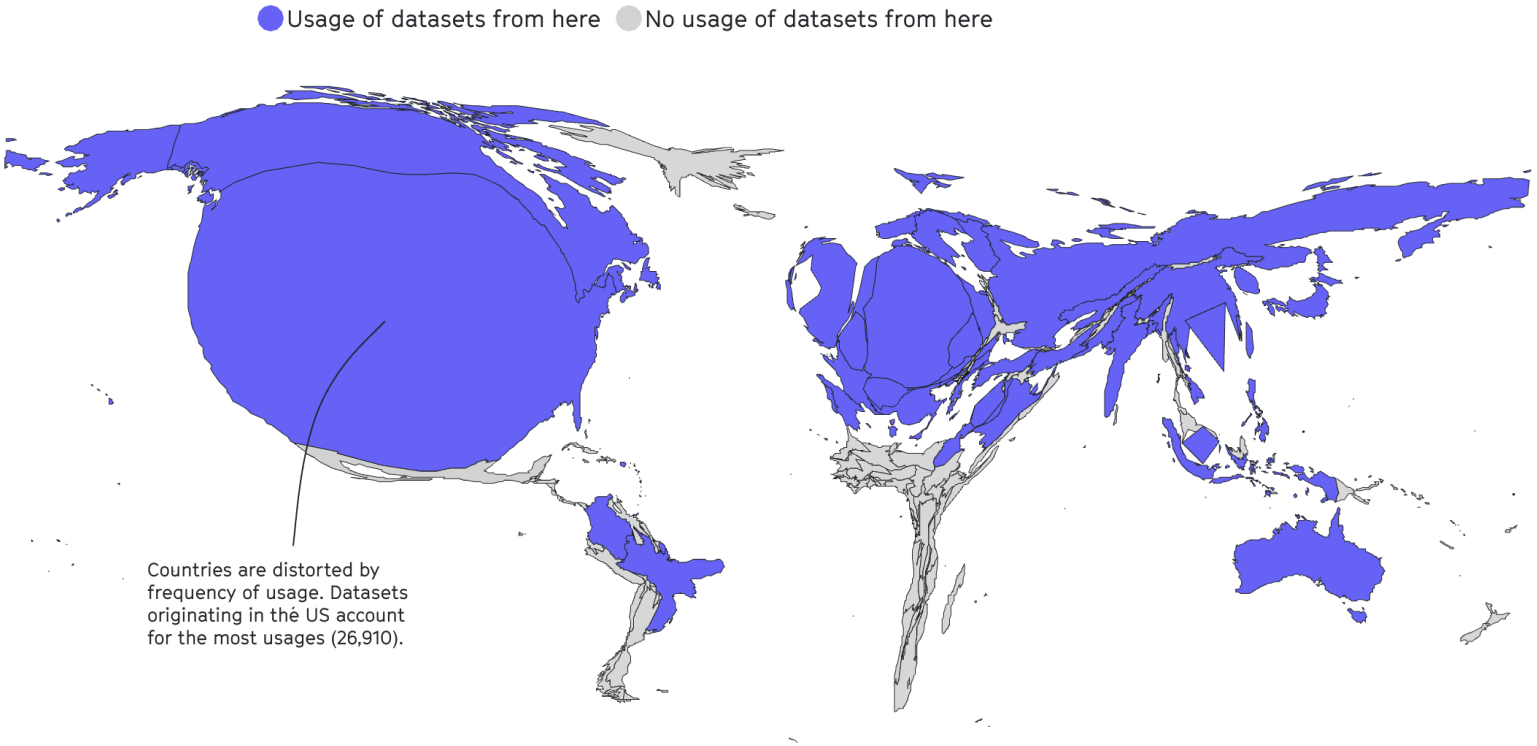
- 「生成AI以後」に重みが増した問題の一つとして、AIの開発・利用過程を通じた社会的不平等の再生産がある：
  - 画像生成AIによって生成される人物の描写にジェンダー、人種、社会経済的地位、性的指向を含む複数の種類のバイアスが反映され、建物や車などの背景要素の描写にも影響する [Bianchi+ 22]
  - DALL-E Miniのアウトプットは人種やジェンダーに関するバイアス等を含み、AI を介したフィードバックループ が観察される。生成AIのモデルに埋め込まれた社会的バイアスが、人間のユーザーによるバイアスのある意思決定を促し、その結果、システムと社会全体の両方でバイアスがさらに固定化される [Cheong + 23]
  - OpenAI社は、ChatGPTの学習データから有害なコンテンツを取り除くための作業をケニアのSama社に外注。トラウマを伴う作業だが、時給として支払われるのは1.32から2ドル程度(約175～265円)
  - インターネットからスクレイピングにより収集されたデータの中には著作権等で保護されたものが含まれている可能性があるものの、公正利用(fair use)に該当するとして、著作権者の許諾なく、また、使用料を支払わずにデータを利用できるとされているが、一部の著作権者からは作品の「収奪」や「搾取」ではないかといった批判や、経済的利益を還元する仕組みづくりを求める声もある

Bianchi, F, Kalluri, P, Durmus, E, Ladhak, F, Cheng, M, Nozza, D, Hashimoto, T, Jurafsky, D Zou, J, and Caliskan, A. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. arXiv preprint arXiv:2211.03759



Bianchi, F, Kalluri, P, Durmus, E, Ladhak, F, Cheng, M, Nozza, D, Hashimoto, T, Jurafsky, D Zou, J, and Caliskan, A. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. arXiv preprint arXiv:2211.03759

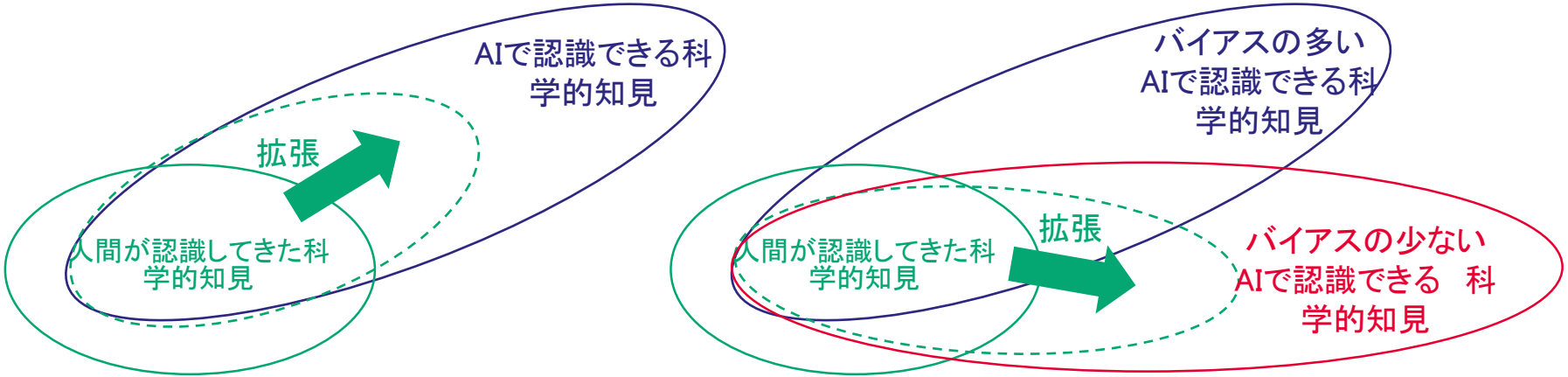
Frequency of dataset usage by country



https://2022.internethealthreport.org/facts/

Cf. バイアスの影響：AI4Science

- AIの科学への応用（AI for Science / AI4Science）をはじめ、AI利用が、今後の研究開発能力や新しい知見の探索を左右すると言われている
- バイアスが大きいと、新しい知見の方向性が歪んでしまうのではないか？



※上記は仮で作ったモデル図であり、あくまでイメージ。包含関係や大きさは今後検討を要する。

論点

- Why + What
- Who + How
- Where
- When

知的負債

- 「知的負債（intellectual debt）」は、ソフトウェア開発における「技術的負債」を拡張した概念
  - 例えば、透明性・解釈性・アラインメント等が低いままのAIが急速に普及すること、ネットワーク化したAI によるシステミック・リスクなどが想定されている
  - 広義には、生成AIによる偏見・差別の助長、プライバ シー侵害、情報の信頼性や民主主義的価値への悪影響などを含む
- （その定義上）普及度や時間経過により「負債の返済」＝対応コストが増大するため、早期の対応が期待される

Zittrain, Jonathan. “Intellectual Debt: With Great Power Comes Great Ignorance.”
The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives, 176–84. Cambridge University Press, 2022.

ありがとうございました

大阪大学
社会技術共創研究センター
ELSIセンター

https://elsi.osaka-u.ac.jp