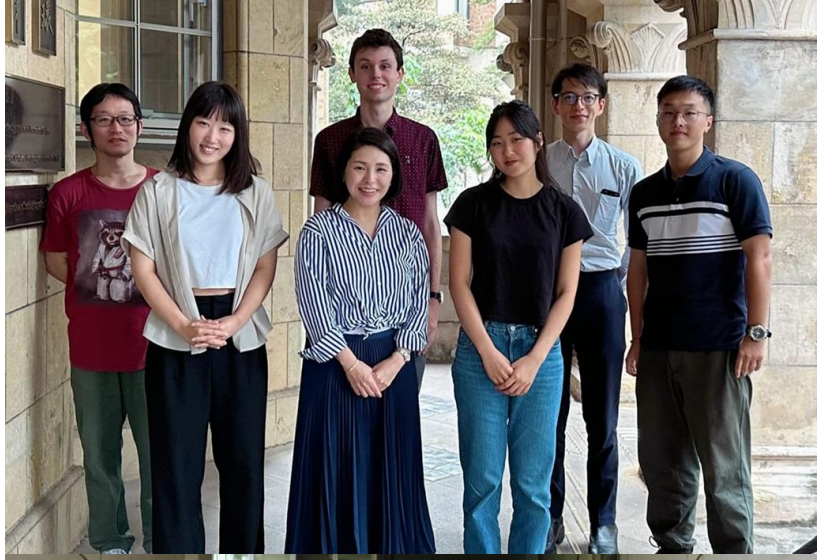
生成AI時代に必要とされる力

AIと倫理、科学技術社会論の観点から

横山広美（よこやまひろみ）
東京大学　国際高等研究所カブリ数物連携宇宙研究機構
学際情報学府

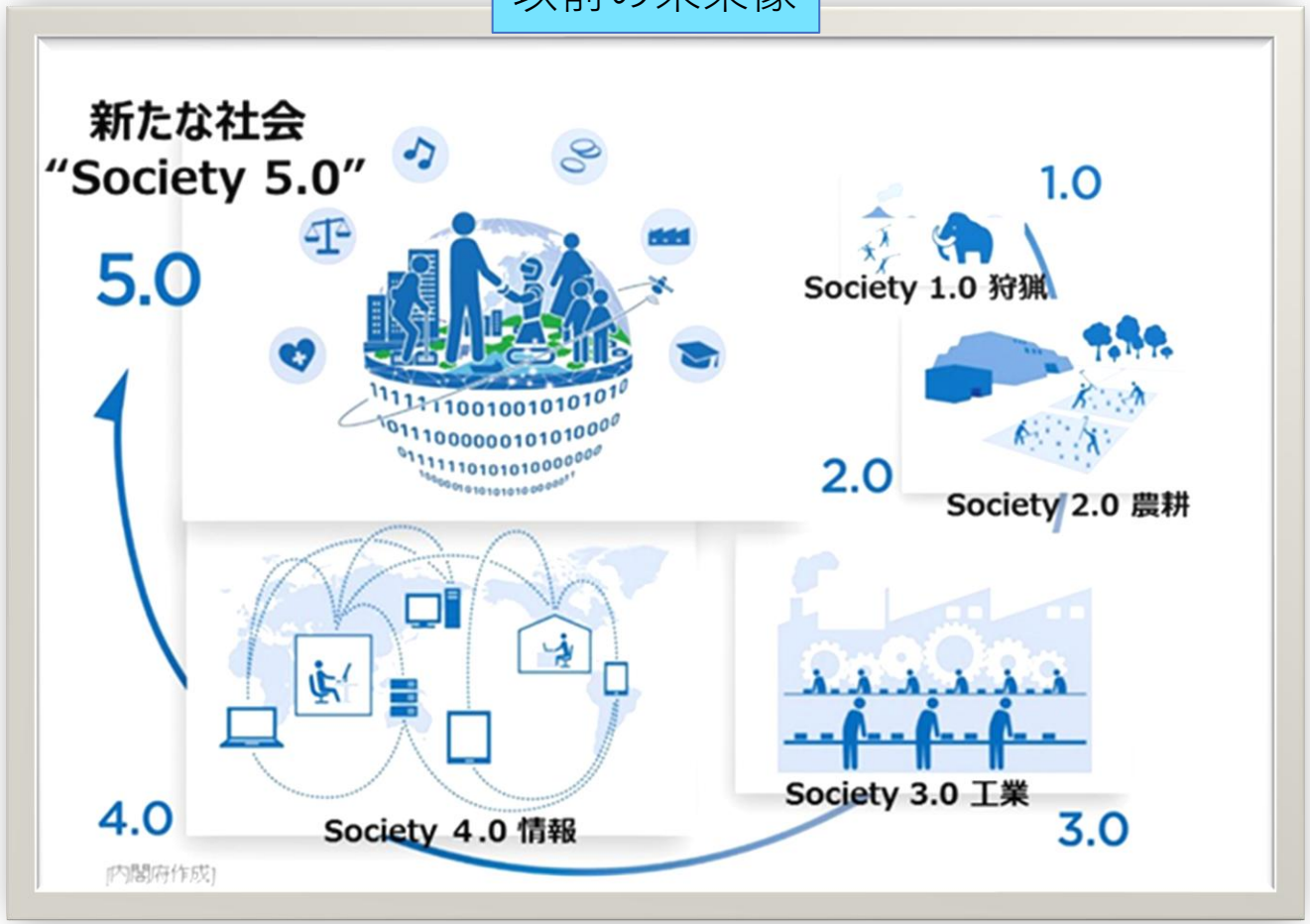


理系出身の文系研究者：科学技術社会論が専門

学際性、国際性が特徴

世界を一変させた「生成AI」

以前の未来像



- 2022年以前のディープラーニングまでのAIと、それ以後の生成AIの影響は桁違い
- Society 5.0という平和的な未来像はすでに変わってしまった
- 現在は、**生成AI時代の未来像**を共有できていない状況
  - **生活、仕事、教育、研究**、すべてが生成AIで激変している
- **人間は、AIを十分にコントロールできるのだろうか？**

どんな研究をしてきたか
生成AI前
AI倫理の程度を測る「ものさし」を開発

[AI] Ikkatai, Y., Hartwig, T., Takanashi, N. Yokoyama, H.M.(2022). Segmentation of ethics, legal, and social issues (ELSI) related to AI in Japan, the United States, and Germany. AI Ethics. https://doi.org/10.1007/s43681-022-00207-y

ELSI セグメント論文:前の論文で抽出された３つの項目を使って日米独の意見分布を調べたところ、４つのグループに分かれることがわかりました。

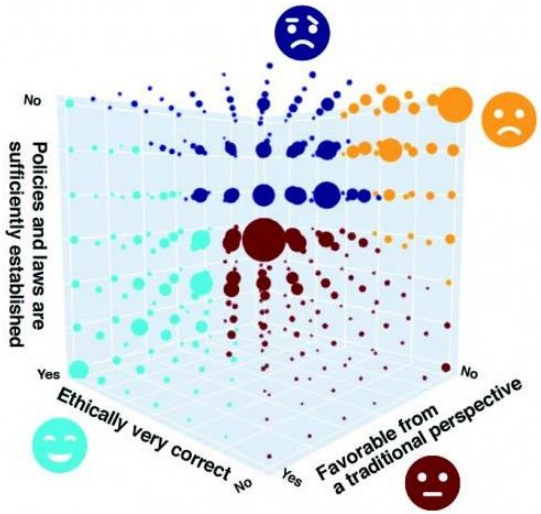
[AI] Hartwig, T., Ikkatai, Y., Takanashi, N. & Yokoyama, H.M. (2022). 'Artificial intelligence ELSI score for science and technology: a comparison between Japan and the US'. *AI & Soc*. https://doi.org/10.1007/s00146-021-01323-9

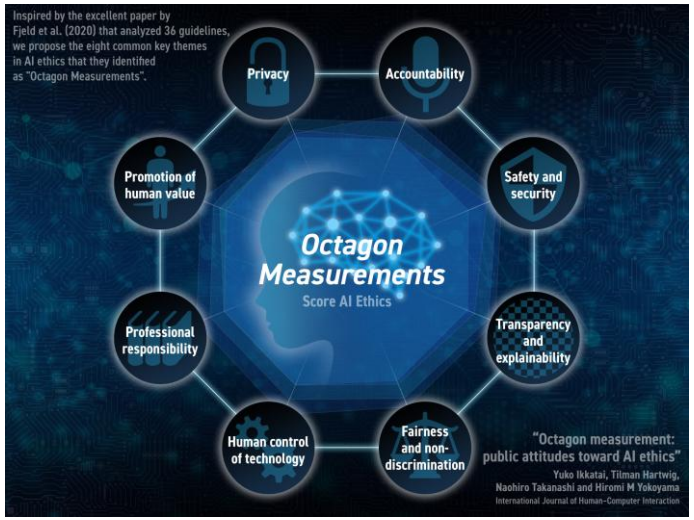
ELSI Sheet論文:ELSIの１３項目をシナリオを使った日米のデータによって決定木解析をしたところ、EとLとSの１つずつの項目が重要項目として残りました。

[AI] Ikkatai, Y., Hartwig, T., Takanashi, N., & Yokoyama, H. M. (2022). Octagon Measurement: Public Attitudes toward AI Ethics. International Journal of Human–Computer Interaction, 38(17), 1589–1606. https://doi.org/10.1080/10447318.2021.2009669

Octagon paper論文: AIガイドラインに共通する８つの項目を用いて、AI倫理のレベルを測定するオクタゴンメジャメントを開発しました。

KAVLI IPMU INSTITUTE FOR THE PHYSICS AND MATHEMATICS OF THE UNIVERSE
知の物理学 研究センター Institute for Physics of Intelligence
公益財団法人 セコム科学技術振興財団 SECOM Science and Technology Foundation

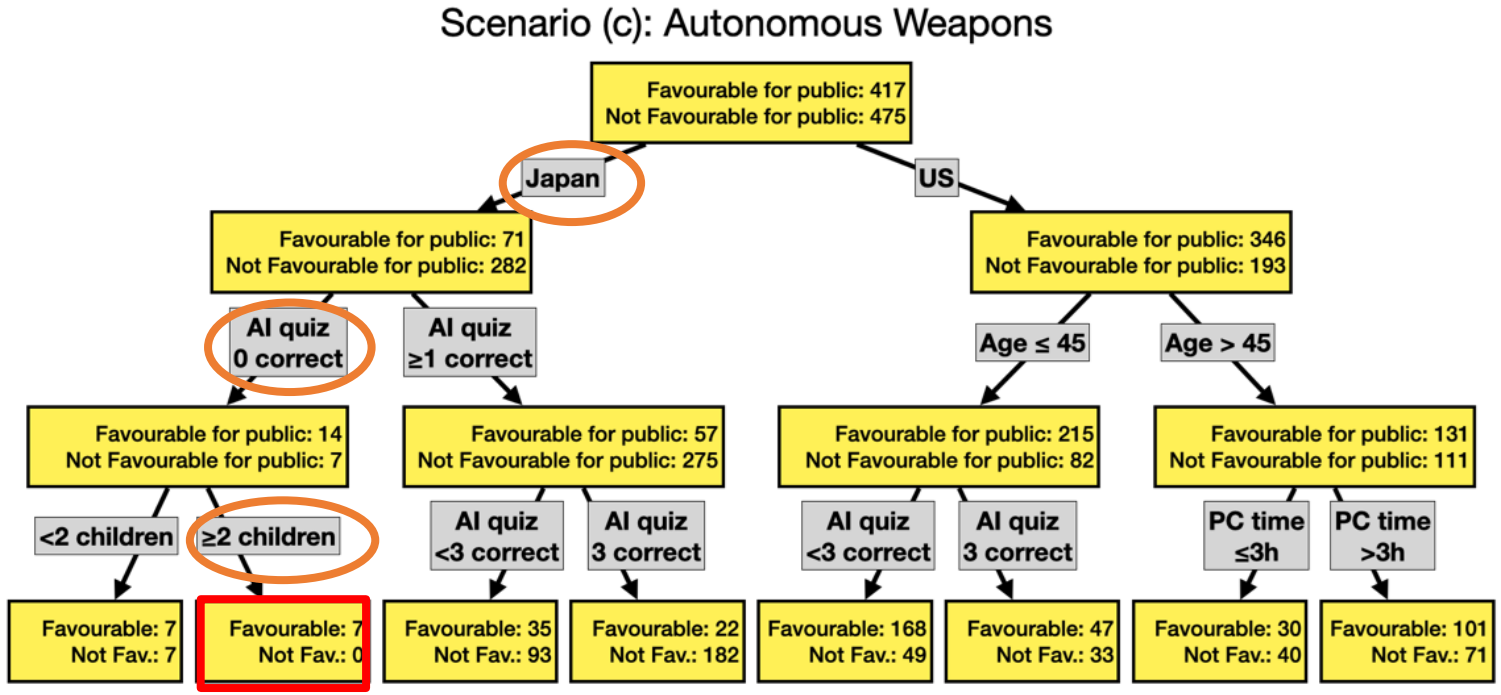




当時から戦争とAIについて懸念、国際条約はまだない（2026.8現在）

AIに対する態度について、国際比較研究を実施してきた

例：AI兵器について日本（反対）とアメリカ（賛成・反対が半々）で大きく態度が異なる



1. country: 0.307
2. age: 0.295
3. AIquizCorrect: 0.180
4. occupation: 0.065
5. Nchildren: 0.049
6. income: 0.043
7. Pctime: 0.021
8. ScienceInterestGroup: 0.018
9. education: 0.011
10. gender: 0.007
11. married?: 0.003

Hartwig, T., Ikkatai, Y., Takanashi & Yokoyama, MH. (2022). Artificial intelligence ELSI score for science and technology: a comparison between Japan and the US. *AI & Soc*. https://doi.org/10.1007/s00146-021-01323-9

Ikkatai, Y., Itatsu, Y., Hartwig, T. *et al.* The relationship between the attitudes of the use of AI and diversity awareness: comparisons between Japan, the US, Germany, and South Korea. *AI & Soc* **40**, 2369–2383 (2025). https://doi.org/10.1007/s00146-024-01982-4

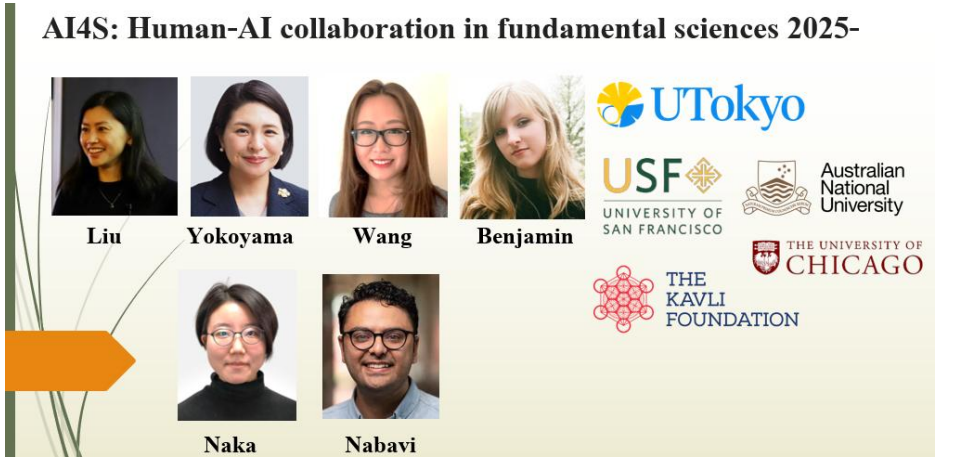
・AI活用に関する「音声」「採用」「顔」「移民」の4つのシナリオを作成し、日本、米国、ドイツ、韓国（以下、韓国）の国民の懸念を比較

・AIに関する知識のある回答者は、様々な状況において倫理への懸念がより強い

・法律に関する懸念については、日本の回答者の方がより大きな懸念を示す

生成AI以後の研究

①仕事全般への導入、②製造業への導入、
③医療AI導入、③研究への導入、
④エネルギー正義について研究
同時に多くの提言を発出



Shotaro Kinoshita & Hiromi Yokoyama, Large language model is a flagship for Japan, Nature 619, 252 (2023), doi: https://doi.org/10.1038/d41586-023-02230-3

時評
Eye on Current Affairs
2026

イラン攻撃に見るAIの軍事利用のレッドライン

科学社会学者
横山広美

アメリカによるイランへの攻撃はAIを使ったシステムによって行われた。後に米国防総省が公開した映像によると、迅速に標的を決め、簡単に操作できる様子がわかる。人の能力を超えた効率的な破壊が行われていることに、恐怖を感じる人は多いであろう。

このAIは、オープンAIから独立した新興企業のアンソロピックが提供した「クロード」である。クロードは、戦争において人の判断が入らない完全自律兵器と、米国倫理的観点を重視するアソロピックは政府のサブ掲載され、裁判になって時間がかかるため、イいられた。アンソロピックるためにオープンAIがた。世界中のユーザーはープンAIにあきれ、ア

懸念①
戦争を加速させるAI

- 国際条約がない中で、企業が政府と対峙しなければならない
- 大学の関与も低下

懸念②
データセンター問題

- How much energy will AI really consume? The good, the bad and the unknown https://www.nature.com/articles/d41586-025-00616-z

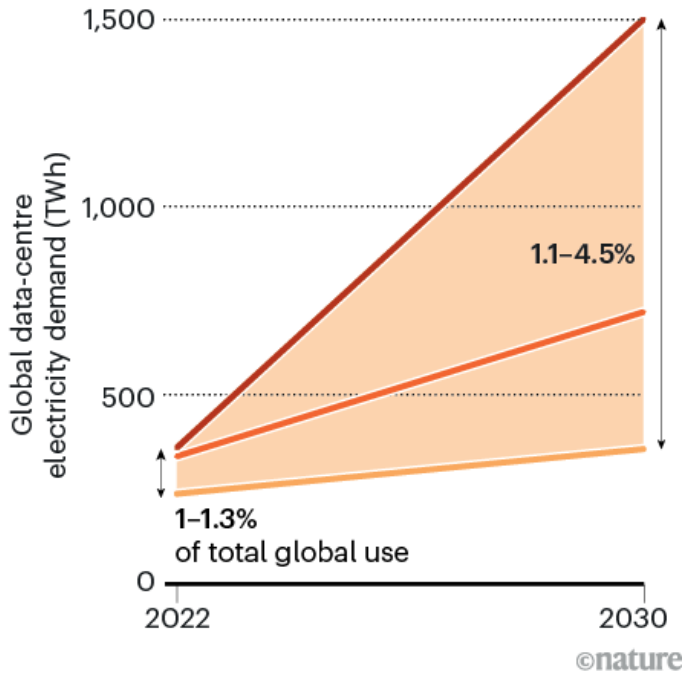
DATA CENTRES' MODEST GLOBAL IMPACT

Data centres currently use 1–1.3% of the world's electricity. But as more are built to supply artificial-intelligence models, they are expected to use up to 4.5% by 2030. The wide range shows the uncertainty in projections.

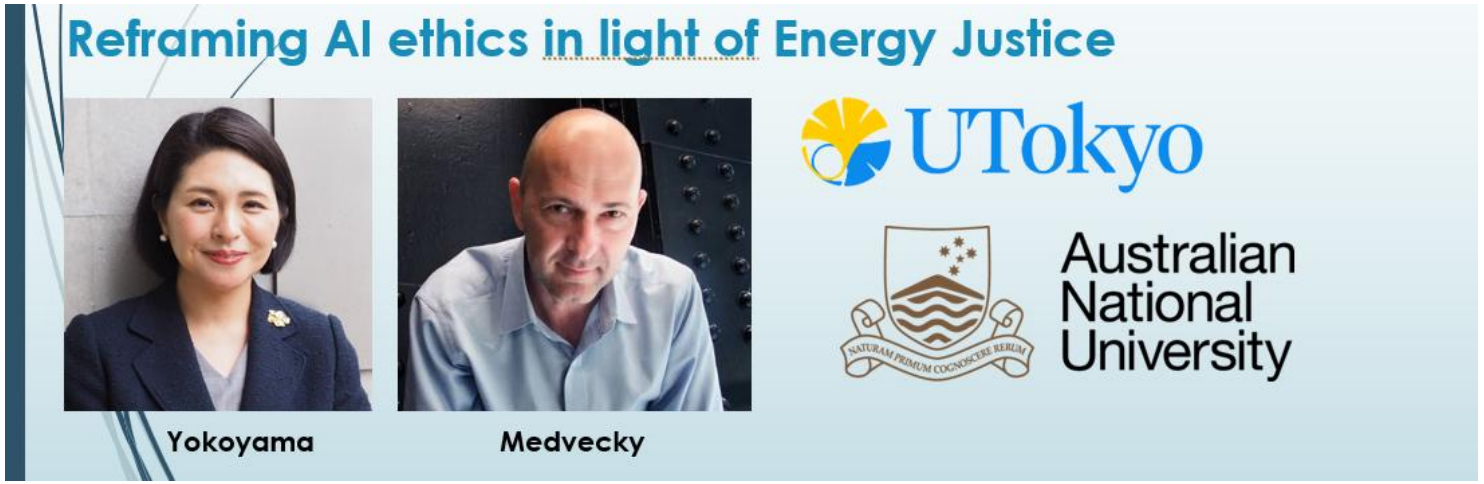
Data source

- SemiAnalysis 2024 (base case)
- IEA 2024 (high projection)
- IEA 2024 (low projection)

*IEA: International Energy Agency, *World Energy Outlook 2024*. The agency says it will release new figures in April 2025.



エネルギー政策も変化
-分配正義に問題



時評
Eye on Current Affairs 2026

AIが原発を後押しする「ワット・ビット連携」

科学社会学者 横山広美

日本に原子力を導入したのが中曽根康弘だったことはよく知られている。アイゼンハワー大統領による原子力の平和利用宣言のすぐあとに原子力予算を成立させ、1955年に制定された原子力基本法にも深くかかわった。その中曽根首相時代の国家プロジェクトが、新潟県柏崎

る砦として機能してきた。し切られた形だ。政府は再路となる道路の整備等のた電からも1000億円の基回の再稼働への道筋で民意

BBC

Thirsty data centres boom in drought-hit Mexico

29 August 2025

Share Save

Suzanne Bearne
Technology Reporter, Querétaro, Mexico

Querétaro is known for its impressive stone aqueduct

中央公論３月号（2026）時評：AIが原発を後押しする「ワットビット連携」横山広美

生成AI前
AI倫理における「3つの波」の提案

1st wave:

Singularity

2nd wave:

Bias issues

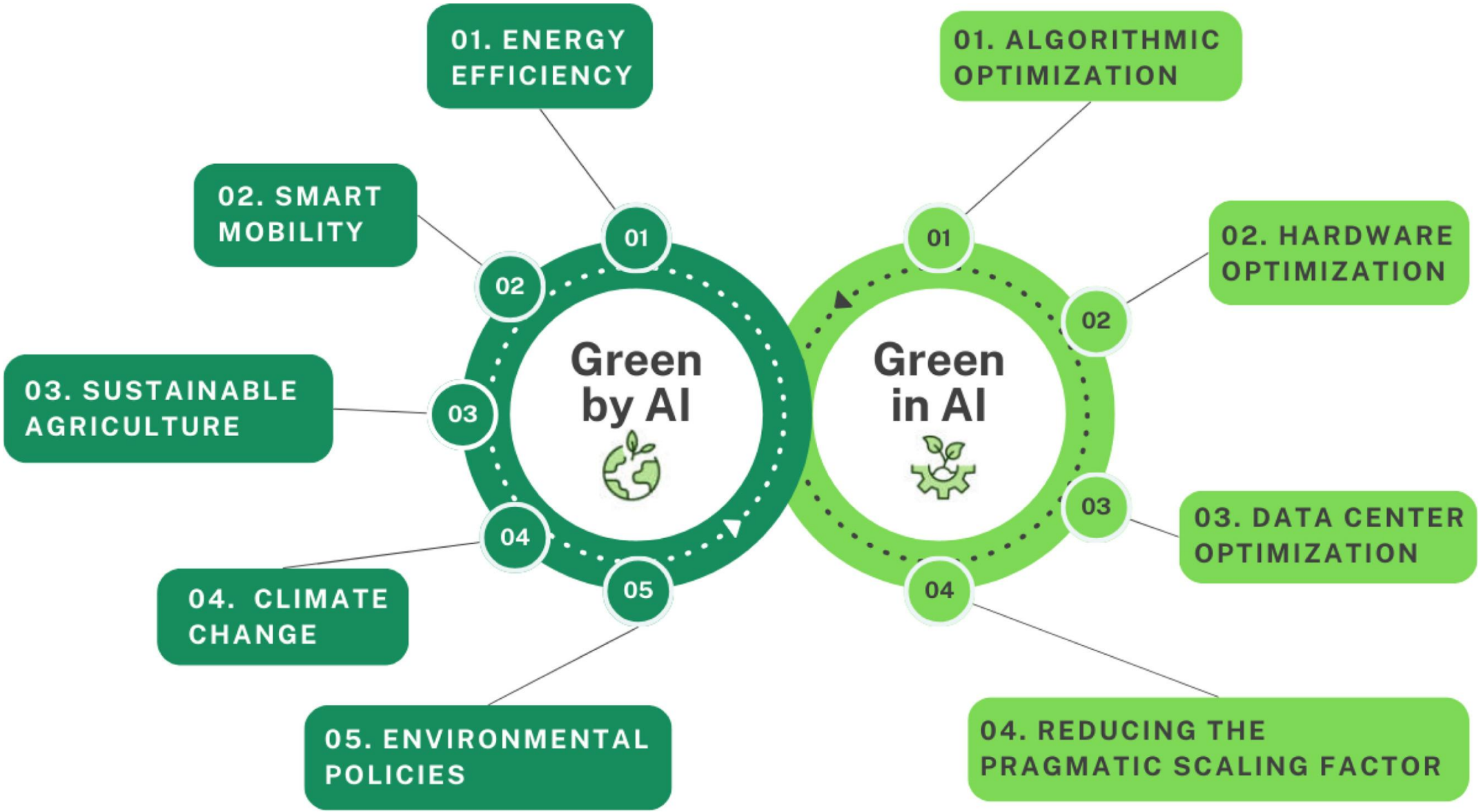
3rd wave:

Sustainability

AIライフサイクルを懸念

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*(3), 213–218. https://doi.org/10.1007/s43681-021-00043-6

Bolte, L., van Wynsberghe, A. Sustainable AI and the third wave of AI ethics: a structural turn. *AI Ethics* **5**, 1733–1742 (2025). https://doi.org/10.1007/s43681-024-00522-6



(Bolón-Canedo et al., 2024)

「正義」の視点から先端技術を捉える　横山広美氏

東京大学教授

経済教室 ✓フォロー済み

2026年4月15日 2:00 [会員限定記事]

保存

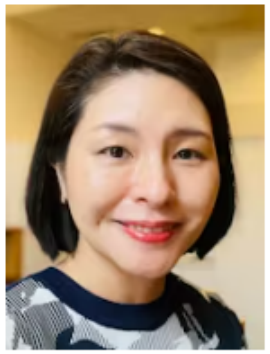
ポイント

○AIの軍事利用に規制がないという現実

○進歩に寄与する技術が状況を複雑化する

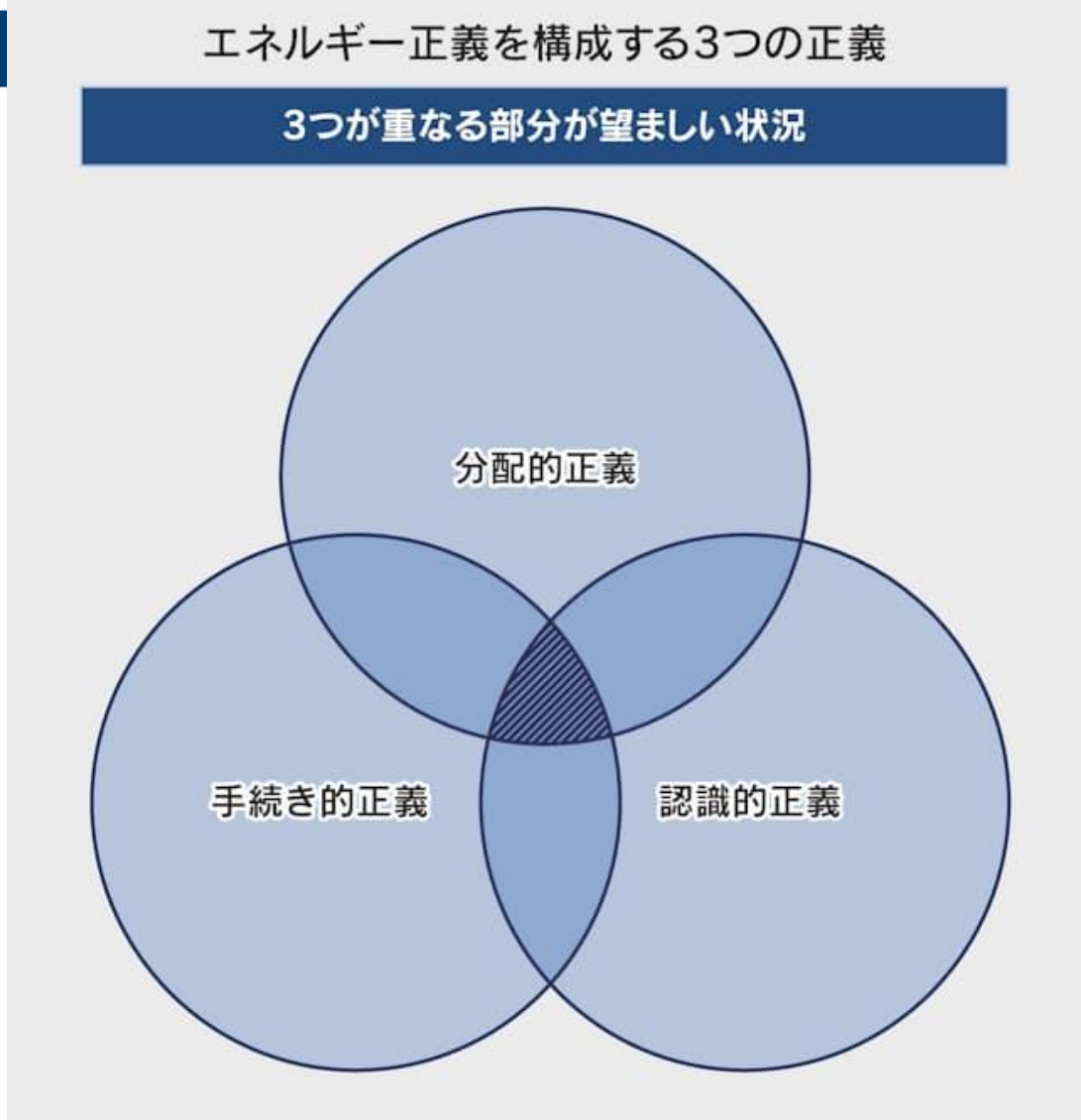
○エネルギーを倫理面から議論する時代に

核兵器、化学兵器、生物兵器に国際的な禁止条約がある。しかし人工知能（AI）の兵器としての利用にはまだない。進んだ科学技術は真っ先に軍事に使用される。科学技術で相手国を上回れば、優位に立てるからだ。我々はいま、ニュースを通じてそれを目の当たりにしている。



よこやま・ひろみ　75年生まれ。東京理科大理工卒、同大博士（理学）。専門は科学技術社会論

ウクライナ戦争、そして今般の米国・イスラエルによるイラン攻撃により、先端のAIによる情報分析、そして駆動にAIが用いられているドローン兵器が、いかに効率よく破壊を招くかを、多くの人が知ることとなった。



https://www.nikkei.com/article/DGXZQOCD095UL0Z00C26A4000000/

2025年からはAIエージェントの時代

AIエージェントは、人間に代わって判断を行う。
課題：AIのスピードが速すぎて、ヒューマンインザループがきかなくなってきている
変化：人間の仕事が、AIに指示を出し、監督することに移り変わってきている。
仕事における人間の価値は何か・・？例）AIが数学の未解決問題を解く時代、数学者の役割は・・？

対処法
①人間らしい判断をする思考を鍛える
②ルールを作る（AI法はルールがない）
③AIを人間らしくする（アライメント）
変化は止められないが、人間らしいコントロールをする

まとめ

AIの倫理は、以前のようにバイアス問題にとどまらず、戦争、エネルギーの公正と広がっている。

同時に、以前までのバイアス問題もディープフェイクなどで加速しているし、サイバーセキュリティの脅威も非常に大きくなった。

より不安定になった、加速社会で、私たちに必要な力は、基本的なことを人間が考える力。

これまでの倫理観、正義感に照らし合わせて、判断については人間が行う（政治家の意義は権力行使なのでその点は変わらないだろう）。

一般社会としては、人間として考え続け、間違っていると思うことに声を上げる。力を合わせる。