

AIのガバナンス 知っておきたい3つの論点 -科学論の立場から



横山広美（科学技術社会論）

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

東京大学 国際高等研究所 カブリ数物連携宇宙研究機構



横山は、物理(高エネルギー素粒子実験)出身の人文社会科学の研究者

2023.4に発足

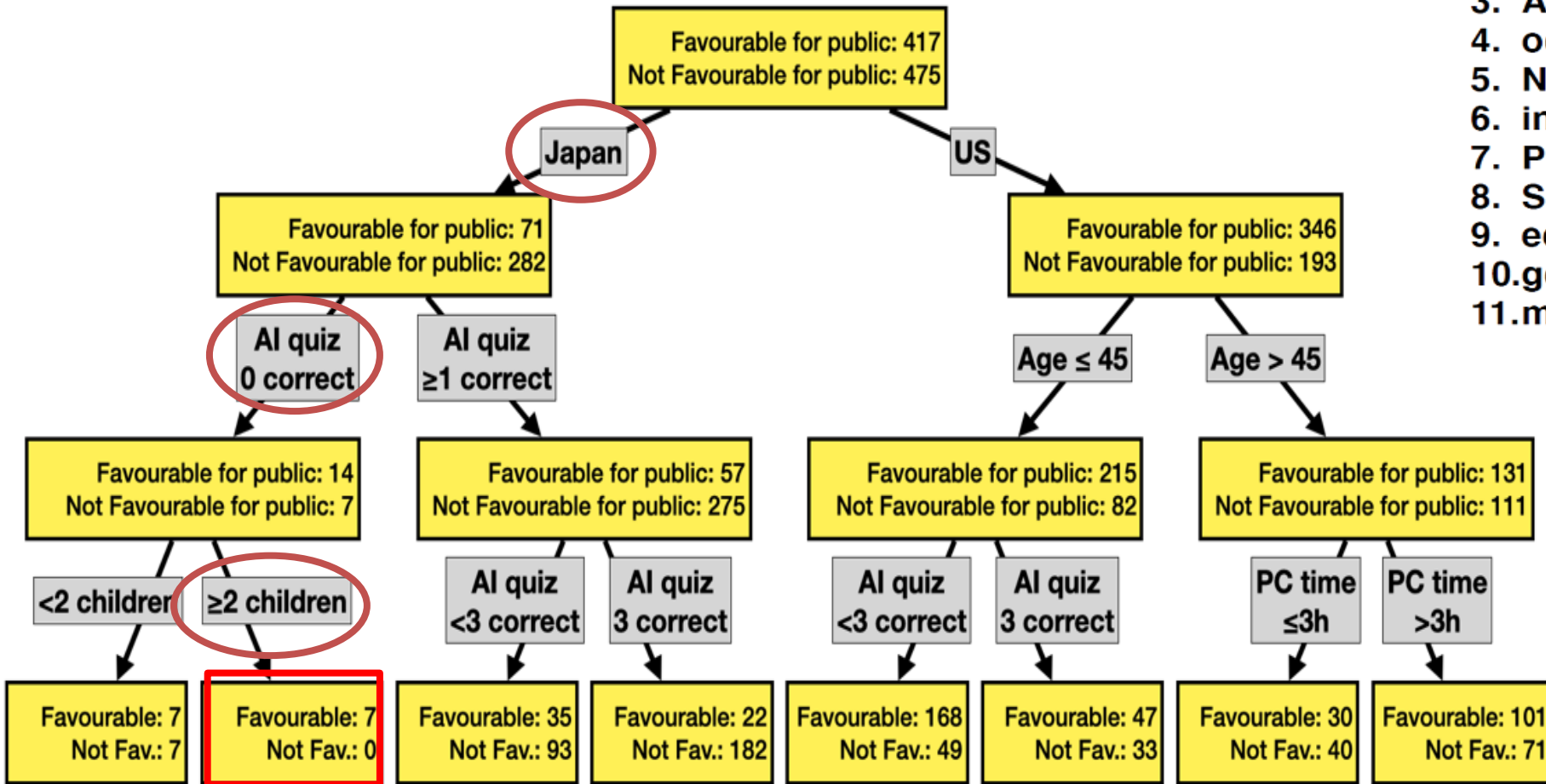
Kavli IPMU: Center for data driven discovery(CD3)



横山の参加: AI倫理の社会科学のデータもデータドリブンで解析する = AI倫理を量的研究する

AI(決定木解析)を用いて、AI倫理の量的研究を実施

Scenario (c): Autonomous Weapons



1. country: 0.307
2. age: 0.295
3. AlquizCorrect: 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

日米のAI使用に対する人々の態度を分析

←↑AI兵器: 賛否を分ける重要ファクターは 1. 国、2. 年齢(高いほど反対)、3. AIについての知識レベル(高いほど反対)

「ELSIセグメント」を提案

物理学の手法を用いた解析で意見分布をグループ化
4つの場合がもっともポテンシャルが低くなる

Plot all 13,024 points of the four scenarios

K-means clustering shows that the four groups are ideal

The exact definitions of the 4 groups are as follows:

If (Ethics + Tradition + Law < 10):

Group 1

Elseif (Ethics + Tradition + Law > 16):

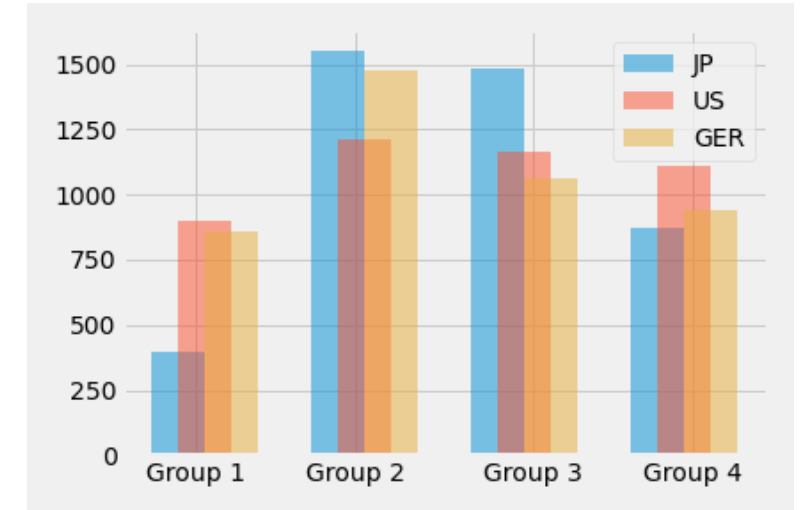
Group 4

Elseif (Law > 4):

Group 3

Else:

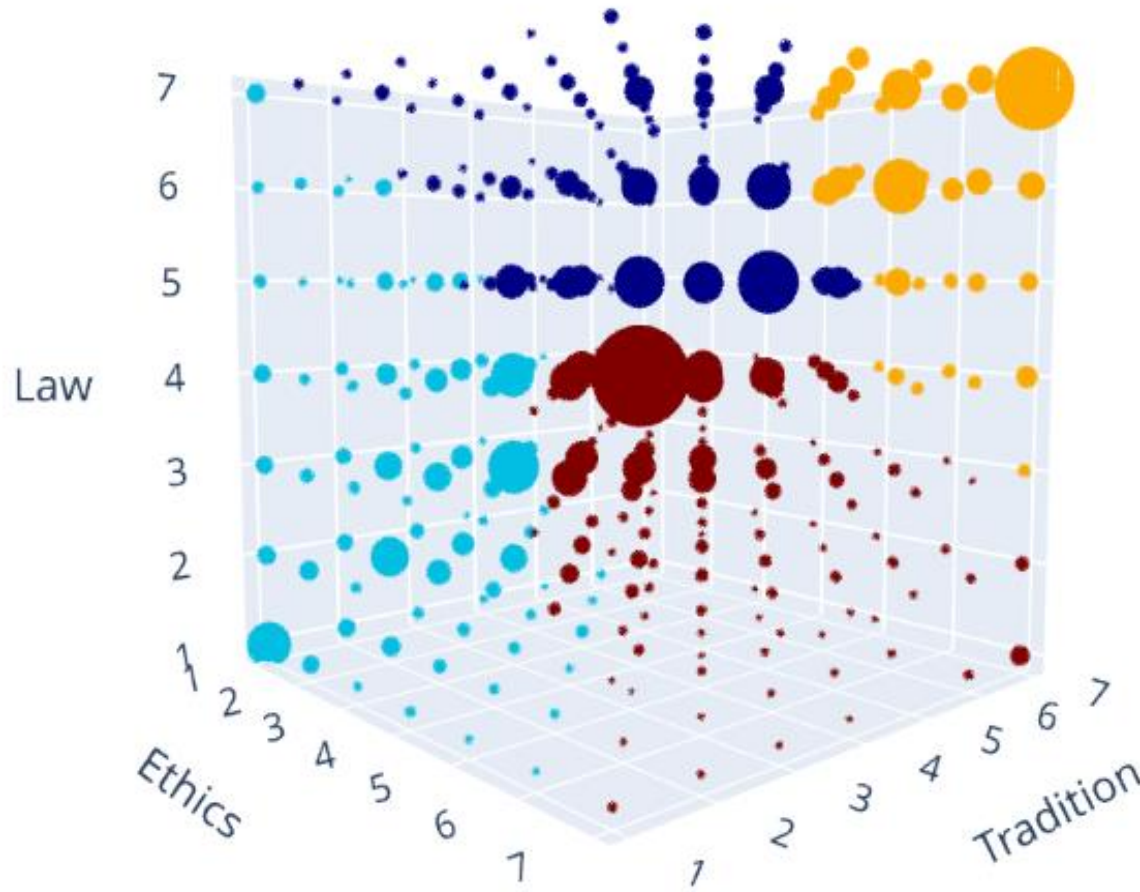
Group 2



ELSIを量的に測定し、意見の違いを可視化できる
「ELSIセグメント」を提案

Ikkatai, Y., Hartwig, T., Takanashi, & Yokoyama, HM. (2022) Segmentation of ethics, legal, and social issues (ELSI) related to AI in Japan, the United States, and Germany. *AI Ethics*

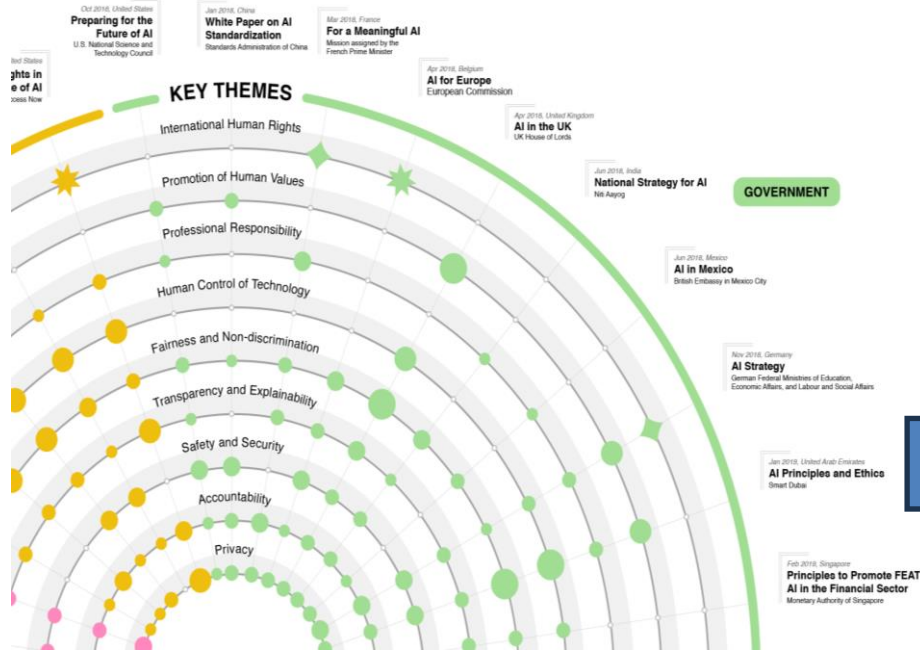
<https://doi.org/10.1007/s43681-022-00207-y>



12項目のELSIシートの中から抽出された、もっとも重要な3つの項目について、3軸として日米独の意見分布をプロット

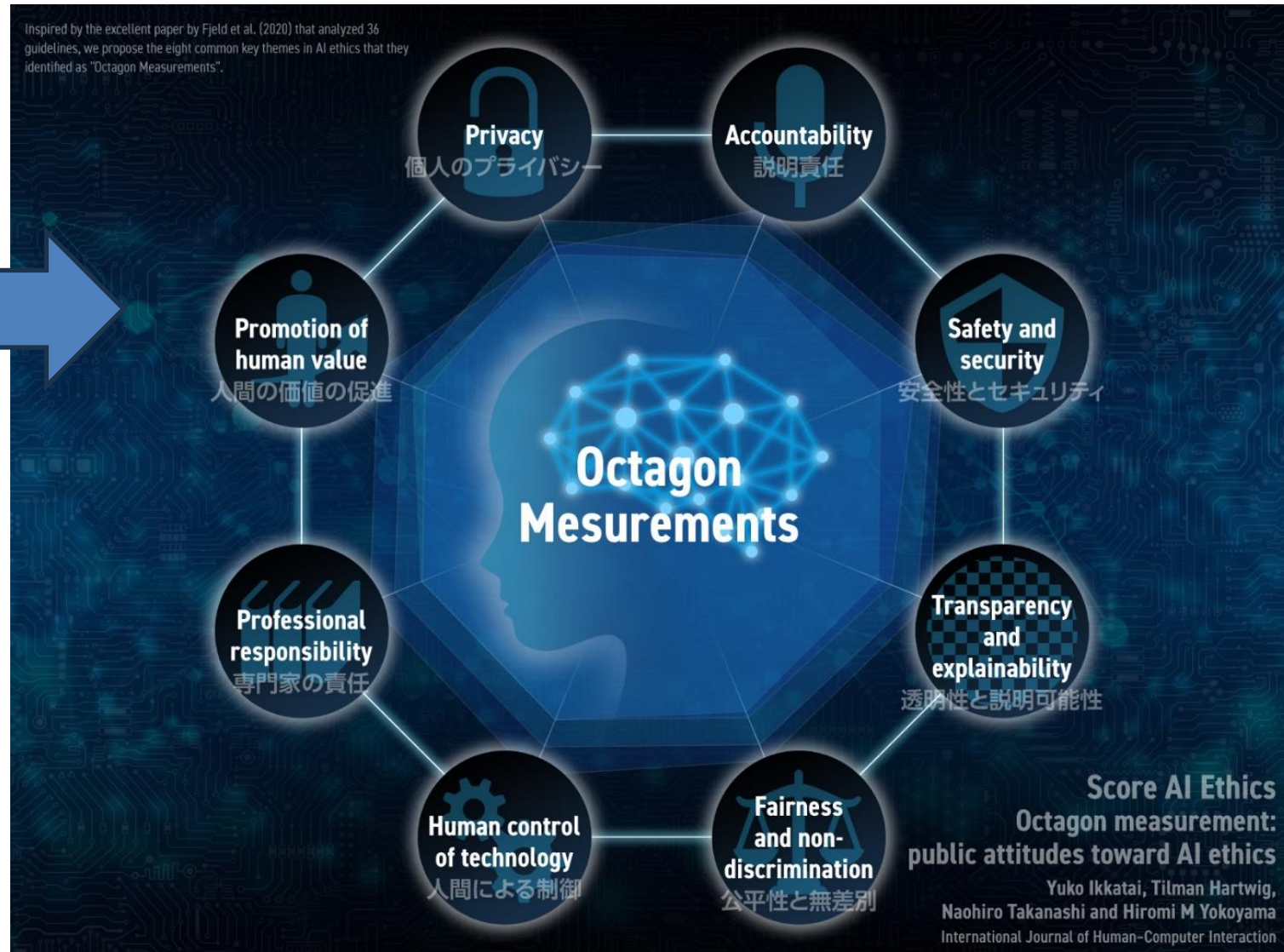
- 左下の青のグループがAI使用にポジティブ
- 右上のオレンジのグループがAI使用にネガティブ
- 中央の群青は法的対応を問題視、紫は問題視しない

「オクタゴンメジャメント」を提案



世界中のガイドラインを分析したハーバードの分析を元に、世界共通の8つの軸を測定軸に用いる「オクタゴンメジャメント」を提案

Yuko Ikkatai, Tilman Hartwig, Naohiro Takanashi & Hiromi M. Yokoyama (2022) Octagon Measurement: Public Attitudes toward AI Ethics, International Journal of Human-Computer Interaction, 38:17, 1589-1606, DOI: [10.1080/10447318.2021.2009669](https://doi.org/10.1080/10447318.2021.2009669)



ChatGPTについて

科学論の観点から、他の科学技術と違う点

- ① 多点・多目的開発
- ② 開発、社会実装のスピードが非常に早い
- ③ 世界同時経験

→ 事前にリスクを検討対応する「**予防原則**」が困難

・ 日経新聞に上記内容を寄稿

2023. 5. 9 [AI倫理の指針「AI8原則」の順守を 人工知能と社会](#)

・ Nature [Correspondence "Large language model is a flagship for Japan" Shoraro Kinoshita and Hiromi Yokoyama](#)

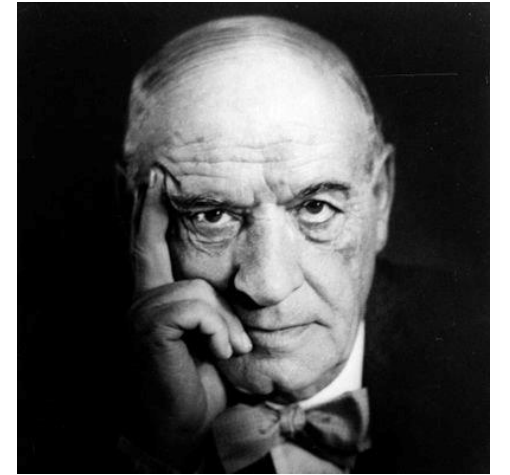
知っておきたいこと①

「作動中の科学」

- 「作動中の科学」 Science in action とは
 - 科学技術は、どんどん進化する。今日、正しいことが1年後も正しいとは限らない。
 - 科学技術は、進化することを前提に議論をする
 - つまりAIによる課題も時間によって変化する可能性

知っておきたいこと②

重要な「ローカルナレッジ」



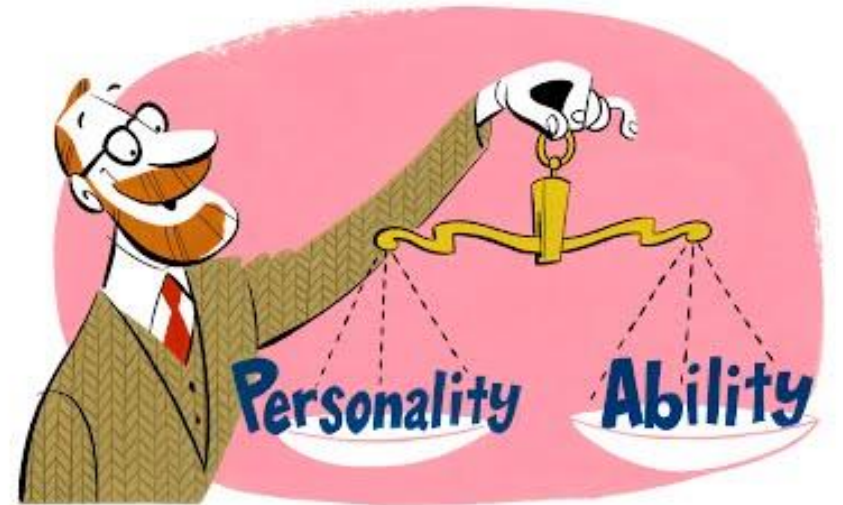
- 研究者は、安全情報に囲まれてバイアスが強い
 - 「専門家は専門しか知らないから間違える」(オルテガ)
 - 社会に実装された科学技術の影響は、しばしば想定を超える
 - ローカルナレッジという考え方(現場の方が知っている)
 - 現場からのボトムアップの提案が重要
 - ゲットプラクティスのみならず、リスクの共有が重要

知っておきたいこと③

必要な「3者の信頼」

- 科学技術の信頼を支える「3者の信頼」

1. 研究者等の信頼
2. 企業等の信頼
3. 行政の信頼



- 信頼とは

- 社会心理学/伝統的信頼モデル「**能力**」「**意図**(人柄)」
 - リスクマネージャーを念頭にSVSモデル「**価値共有**」が重要。

責任ある科学技術・イノベーション

- 責任は、各所に分散している
 - 「予防原則」のため、評価をするのは科学者
 - 実施と最終責任は行政
 - 政治的責任は、科学者には発生しない
- 皆で協力して、
「責任ある科学技術イノベーション」(RRI)の実現を