

■AI倫理指針の作成、普及への貢献

- IEEE EAD ver.2, 1eの作成(中川、江間)
- 内閣府 人間中心AI社会原則の作成(中川、江間)
- 総務省 AIネットワーク利用原則(中川、江間、成原)

■主要AI倫理指針の傾向分析(左から古い順に並べた)扱われている項目を色付けしている

	FLI Asilomar 23	人工知能学会倫理委員会	ICDPPC(privacy commsioner conf.)	IEEE EAD v1,2	IEEE EAD e1	人間中心AI社会原則	EC Trustworthy AI	OECD Recommendation.	Whitehouse Guidance
公平性、非差別	■	■	■	■	■	■	■	■	■
透明性									
アカウントビリティ									
トラスト									○
自律的AIの制限	■		■	■					×
悪用・誤用									
プライバシー	■	■	■	■	■	■	■	■	■
AIエージェント									
教育									
独占禁止、協調									
軍事利用	■			■			■		

- 1.公平性、非差別、プライバシーは常に重視
- 2.透明性、アカウントビリティ、トラストは重視され始めた
- 3.軍事利用はナイーブに禁止と言い切れない政治的問題
- 4.独占禁止は国家、プラットフォームを意識
- 5.AIエージェントは将来課題としてIEEE EADが重視
- 6.米国政府は企業に対する開発倫理指針案を出している。ここでは、便益がリスクを大きく上回ることによって、public trustを確保するように示唆されており、他の倫理指針に対しては異色である。

■AI倫理指針関連の普及活動

- 人工知能学会 倫理委員会主催で大会にて企画セッションを行い、発表(中川)
- 15以上の学会などで招待講演(中川)
- 企業のAI倫理指針作成に協力(中川)

■ プライバシー保護技術

- 第5回PWSCUPを開催(菊池、中川)



- 本年度は地理的移動履歴の匿名化をタスクにした
- 実データは使わず、実データの統計的性質を保存したランダム生成したデータの利用(法的問題を避けるため)
- 20チームが参加(台湾、カナダを含む)
- 優勝は台湾チーム

■ プライバシー保護法制

- GDPRなどヨーロッパの法制度の紹介活動例えば学術会議での講演(板倉、池貝)
- 多数の政府委員会においてプライバシー保護、データ流通の政策立案に貢献(板倉、池貝、寺田)

■ AIの公平性 Fairwash(荒井)

- 機械学習モデルの公正性と説明性に着目し、差別的な判断をする不公平なモデルに説明可能AIの技術を適用し、あたかも同モデルが公平であるかのような説明を生成できることを示した(不公平なモデルを公平であるかの「洗浄(washing)」)。具体的には、精度と公正性の2項をパラメータでバランスした目的関数に対し、適合度の高い順にモデルを列挙し、説明者にとって都合の良い説明を選択できるようにした。本成果はICML2019に採択された(阪大、UQAMらとの共同研究)。

■ 機械学習と公平性のシンポジウム(中川、江間)

- 2020年1月9日 一橋講堂:人工知能学会、電子情報通信学会、情報処理学会の3研究会も共催

■ 論文、著作活動

- 著作
 - 裏側から見るAI(中川・単)、AI倫理(河島・共)、AI時代の自律性(河島・共)、AIxクリエイティビティ(河島・共)、データ戦略と法律(板倉・共)、アメリカプライバシー法(板倉・共)、HRテクノロジーで人事が代わる(板倉・共)
- 論文
 - 情報ネットワーク・ローレビュー、情報法制学会誌などで多数