

AI倫理と個人データの流通

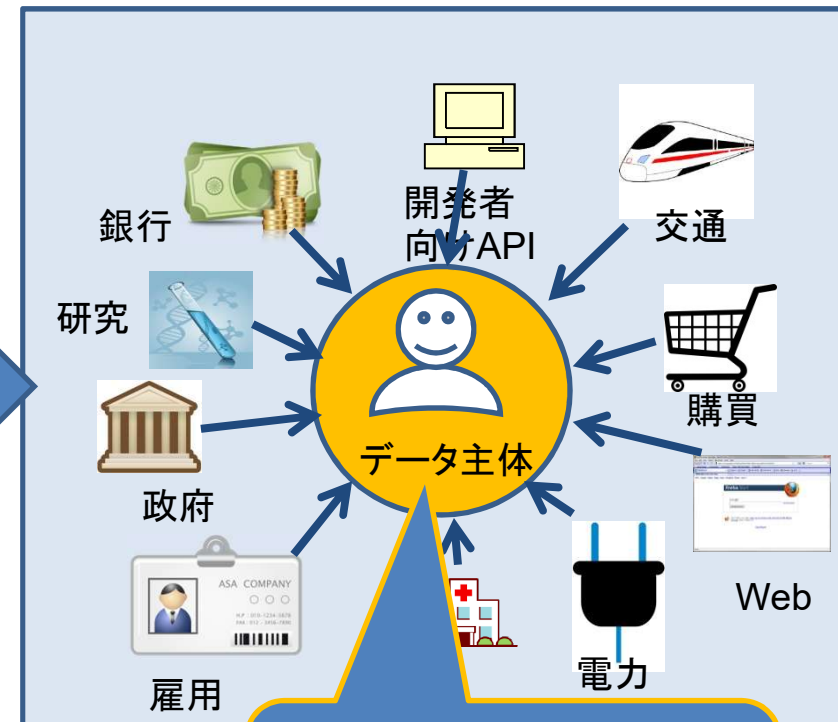
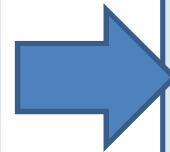
中川裕志
(理化学研究所 革新知能統合研究センター)

スライド中の図はpower point の機能でダウンロードした
creative commons のライセンスです。

個人データ管理は データ主体の個人へ: MyData

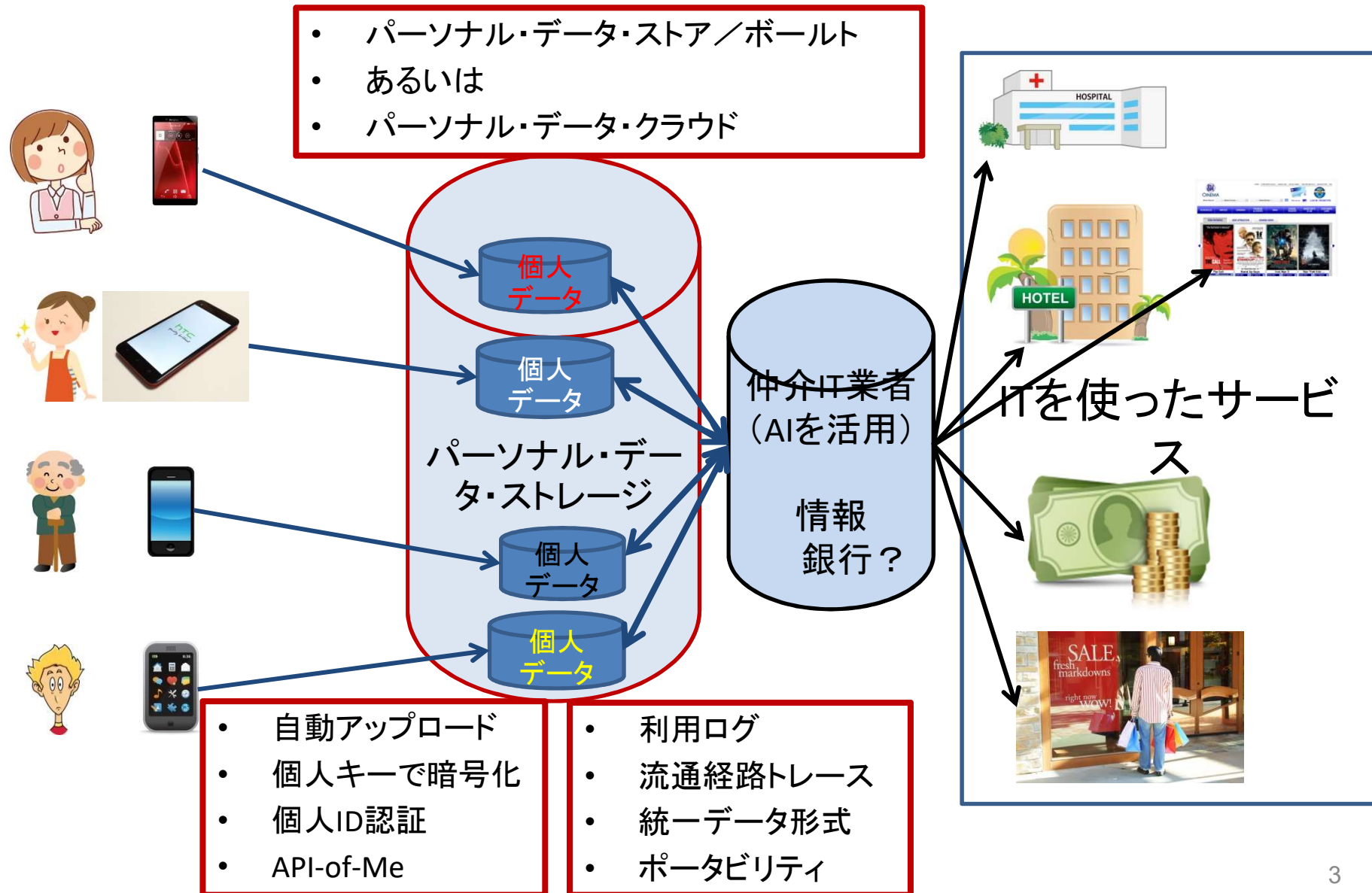


個人データを自社に囲い込んで儲ける



自分の個人データを契約によって他社に使わせる

パーソナル・データ・ストレージ (PDS)

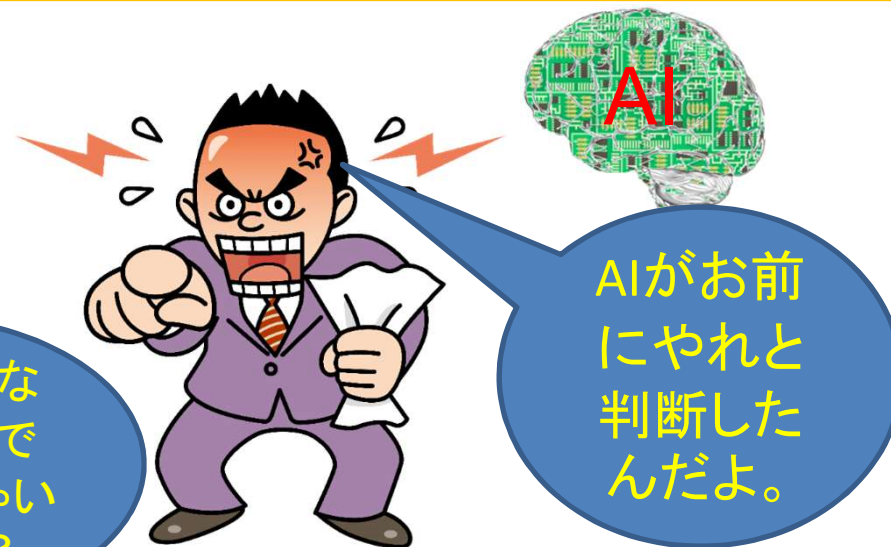


ツールAIの悪用 不法行為の責任は操作した人間

- AIの判断に従わない権利を持つ社会が必要
- 実際はどう実装？



こんな無理な仕事、なんで私がしなきゃいけないの？



AIがお前にやれと判断したんだよ。

ツールAIをあたかも自律型AIのように見立てた詐欺行為といえる。
実際起こりそうだが、法的責任者は明確

- こうして、悪用された人工知能に一般人が文句を言うことができなくなってくると、実質的に言論の自由も人権もない状況になりかねません

➤ 人工知能に対して文句を言える社会制度を考える必要があります。

➤ GDPR 22条： 計算機（人工知能）のプロファイリングから出てきた決定に服さなくてよい権利

➤ 具体的には以下のようにします。

➤ プロファイリングに使った入力データの開示

➤ 出力された決定に対する説明責任を人間が果たす。

➤ ただし、この権利の社会実装、技術による実現はかなり困難

➤ 守秘義務や企業秘密の壁もあります。

IEEE EAD version2 の法律編の示唆する 「悪用対策」

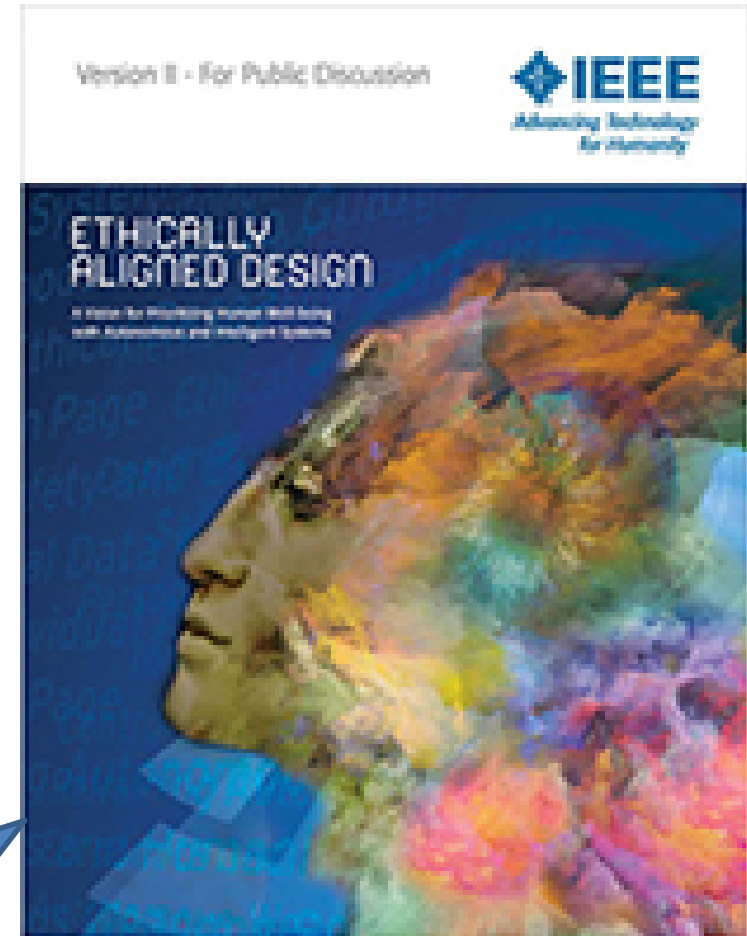
- このような悪用を看破し、見逃さぬように
- AIがその結論に至った推論、入力データを明確化できる仕掛けを予め組み込む（＝GDPR22条の実現方法）
- AIのおかしな動きに対する内部告発の制度的保証
- AIが悪さをした場合の救済策を立法化しておくこと
- 保険などの経済的救済策も重視

国内外の組織が提案している 人工知能の倫理

- 総務省 AIネットワーク社会推進委員会
 - [AI開発ガイドライン](#) OECDに提案
- OECD Principles on Artificial Intelligence 2019/5/22
- 内閣府 人間中心のAI社会原則検討会議
- EC: High Level Expert Group: Ethics Guidelines for Trustworthy AI
- IEEE Ethically Aligned Design, first edition
 - [AI](#)および開発者が持つべき倫理

IEEE Ethically Aligned Design version 2

1. Executive Summary
2. General Principles
3. Embedding Values Into Autonomous Intelligent Systems
4. Methodologies to Guide Ethical Research and Design
5. Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)
6. Personal Data and Individual Access Control
7. Reframing Autonomous Weapons Systems
8. Economics/Humanitarian Issues
9. Law
10. Affective Computing
11. Classical Ethics in Artificial Intelligence
12. Policy
13. Mixed Reality
14. Well-being



The final version was published

IEEE EAD (Final) on April 2019

- 1. Human Rights
 - A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
- 2. Well-being
 - A/IS creators shall adopt increased human well-being as a primary success criterion for development.
- 3. Data Agency
 - A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.
- 4. Effectiveness
 - A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.

IEEE EAD (Final) on April 2019

- 5. Transparency
 - The basis of a particular A/IS decision should always be discoverable.
- 6. Accountability
 - A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
- 7. Awareness of Misuse
 - A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
- 8. Competence
 - A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

EC High Level Expert Group

Ethical Guideline for Trustworthy AI

- April 2019

Trust(信頼)

- (1) 能力、誠実さ、および予測可能性を扱う一連の特定の信念
 - (2) 危険な状況下で、ある当事者が他の当事者に依存する意思
 - (証明したわけではないが...)
-
- 「信頼」はAIシステムのライフサイクルに関与するすべての人々とプロセスに帰することができます。

Trustworthy AI

- Lawful, Ethical, Robust
- 具体的要件
 1. Human agency and oversight
 2. Technical robustness and safety
 3. Privacy and data governance
 4. Transparency
 5. Diversity non-discrimination and fairness
 6. Societal and environmental well-being
 7. accountability

Recommendation of the Council on OECD Legal Instruments Artificial Intelligence

- 2019年5月22日のOECD 閣僚理事会で採択
- 強制力はないが、各国での立法の指針になる可能性大
 - 例： OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data（1980）は各国のプライバシー保護法の基礎になった



Recommendation の概要

- **Trustworthy AI**に関わる5個の価値基準の原則
 - AI 関係者に推進と実装を求める
- 1. inclusive growth, sustainable development and well-being;
- 2. human-centred values and fairness;
- 3. transparency and explainability;
- 4. robustness, security and safety;
- 5. and accountability.

Recommendation の概要

- 政策立案者に対する国内政策および国際協力に対する5個のRecommendation
- trustworthy AI のための原則
 1. investing in AI research and development;
 2. fostering a digital ecosystem for AI;
 3. shaping an enabling policy environment for AI;
 4. building human capacity and preparing for labour market transformation;
 5. international co-operation for trustworthy AI.

人間中心AIの社会原則

Social Principles of Human-Centric AI

Council for Social Principles of Human-
centric AI

内閣府, Japan

第2章

基本理念

第3章

ビジョン
(AI-Readyな社会)

第4章

4.1

人間中心のAI社会原則

4.2

AI開発利用原則
(個別原則・指針・ガイドライン・ルール等)

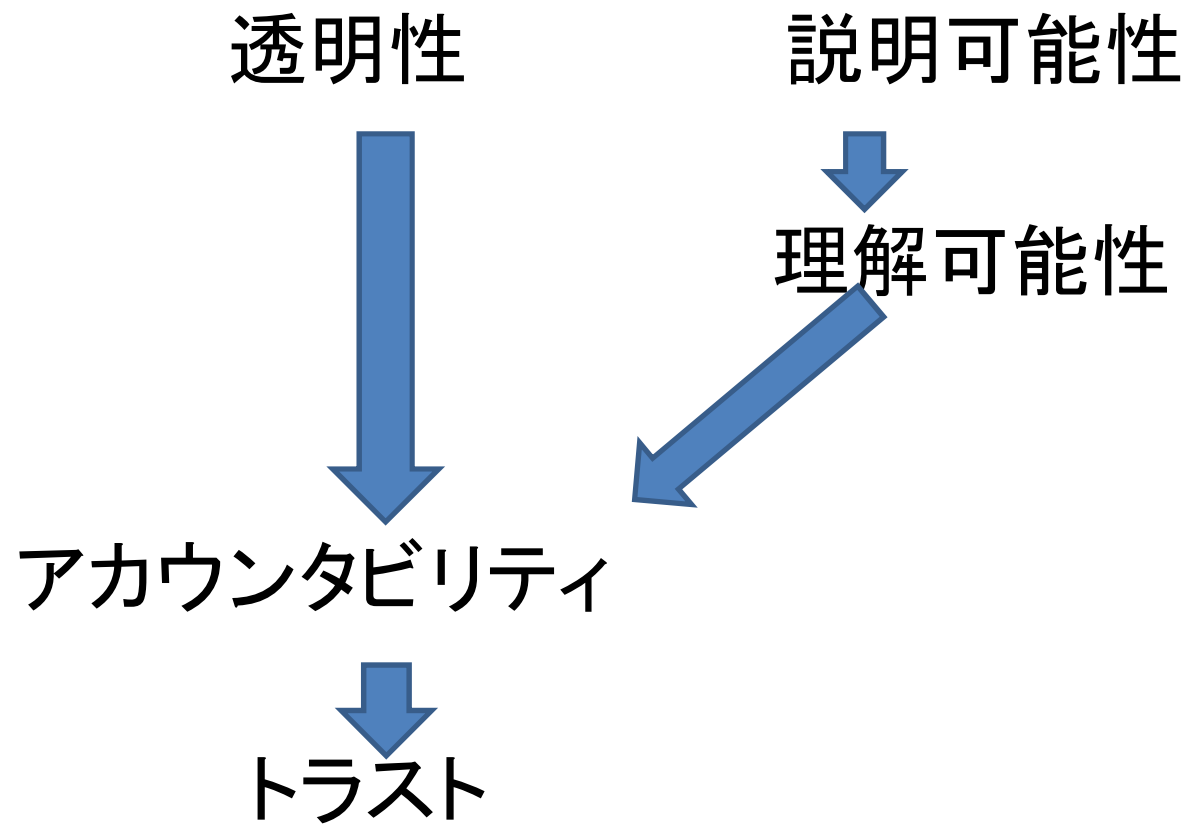
- Dignity、
- Diversity & Inclusion、
- Sustainability

- Society 5.0 → 「AI-Readyな社会」
 - 「人」、「社会システム」、「産業構造」、
 - 「イノベーションシステム(イノベーションを支援する環境)」、
 - 「ガバナンス」

AI社会原則

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

AI倫理の実装



透明性とアカウンタビリティ

- 急速にすすんでいるAIのブラックボックス化への対応策が必要
 - IEEE EADversion2 法律編によれば
- 事故時の責任の所在を法制度として明確化しておく必要がある
- 具体的には
 - 透明性 と アカウンタビリティ(説明責任)

透明性

- AIにおける学習で使われた入力教師データ と AI応用システムに投入された入力データを開示できる
- AI応用システムの出力結果
- AI応用システムにおける大雑把なデータの流れ
- AI応用システムを開発主体、出資者、

Accountability:説明責任 誤訳？

- accountabilityを日本では説明責任と訳しているが、どうもこれが誤解の元らしい。
- 「説明をする責任」ではなくて
- 「責任をだれがとるかを説明する」と言ったほうがむしろ正しいようだ。
- そして、transparency は accountabilityを支援する手段
- → AIのどの部分をtransparencyの対象にするかも自ずと見えてくる。

説明責任

- 正しい考え方
 - 入力データから結果出力の妥当性、公平性、正統性を AI応用システムの利用者が納得できる形で説明すること
 - 単なる透明性に基づく開示では不十分
 - 一般利用者には納得できない専門的な説明 ×
 - もちろんAIの動作を理解できれば理想的だが、一般利用者、多くの場合は専門家にも、なかなか困難
 - 責任者を明確に指摘できること
 - 被害者に対する補償が確保されること

説明からトラストへ：利得と損害

- 一般人に理解可能なAIの動作説明は無理
- 信じてもらうしかない！ → トラスト
- 多数に上る過去の類似例の入力データと結果出力を示して、当該結果の妥当性、公平性、正統性を納得してもらう方法
- AI応用システムの利用履歴を開示して信用を得る

注意

- 技術的にみたトラスト
= 故障せずに安定して動くこと
- AI倫理の文脈でのトラストはむしろ「社会科学
的な信用」という概念に近い

トラストはどうやって確保するか？

- 過去の学問的成果の集積を信用してもらう
 - 数学、物理学、医学、...
- 専門家をトラストしてもらうライセンス制度
 - 専門家のスキルのトラスト：医師国家試験、司法試験など
 - ライセンスする側（国など）へのトラスト
- それでも事故は起きる
 - 補償制度の確立：保険など
- これら全部を統合したシステム体系がトラストの基本

残された問題

- トラストするのは
 - 組織か （利用者からみたら）
 - 情報銀行、ITサービス企業
 - 人か （サービス業者からみたら）
 - 利用者の個人をトラストできるか
 - Self Sovereign Identity
 - データか
 - 流通する個人データか、サービスの結果か？

トラストの補足

- 利用者がサービス提供側をトラストするという局面ばかり考えてきたが
- サービス提供側が利用者をトラストするという問題もある
 - 利用者認証（多くはネットワーク越し）
 - Bad userではないことを推定
 - Self Sovereign Identity → 対応するサービスに必要なことだけ identify できればよい

個人は信頼できるID認証にアクセス できること

- 認証されたIDで金融、政府、通信などのサービスを受けることを保証するAI
- 認証はインターネットにおける個人の存在の証拠
 - － 対面認証でないので、技術的な問題が多い。
 - － 特に生体認証の危険性

old



new

	FLI Asilo ma 23	人工知能 学会倫理 委員会	ICD PPC	IEEE EAD v1,2	IEEE EAD e1	EC Trust worthy AI	人間 中心 AI	OECD AI Reco mm.
人権		○	○		○	○	○	
公平性、差別	○	○	○	○	○	○	○	○
法令遵守		○	○	△		○		
透明性			○	○	○	○		○
アカウントビ リティ			○	○	○	○	○	○
トラスト		○				○		○
自律的AI・自由 意志	○		○	○				

ICDPPC: International Conference of Data
Protection Commisioners

	FLI Asiloma 23	人工知 能学会 倫理委 員会	ICDP PC	IEEE EAD v1,2	IEEE EAD e1	EC Trust worthy AI	人間 中心AI	OECD AI Recomm.
悪用・誤用				○	○		○	
プライバシー	○	○	○	○	○	○	○	○
AIエージェント				○	○	○	○	
安全性	○	○		○	○	○		○
SDGs								○
教育				○	○		○	
独占禁止、協調							○	○
軍事利用	○			○				
幸福	○	○		○	○	○	○	○

ご清聴ありがとうございました