

教育・研究システム分科会2025年度会合

「大学DXを導くAI活用の羅針盤 — マネジメント・ガイドライン・倫理」

2025年12月23日（火）

大学DXをレスポンスブルに実装する —ELSI（倫理的・法的・社会的課題）の考え方

岸本充生（KISHIMOTO, Atsuo）

大阪大学 D3センター（2024年9月までデータビリティフロンティア機構（IDS））



大阪大学 社会技術共創研究センター（通称、ELSIセンター）センター長
Research Center on Ethical, Legal and Social Issues

社会技術共創研究センター（通称、ELSIセンター）

Research Center on Ethical, Legal and Social Issues

2020年4月設立

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

5周年

3部門と4機能

総合研究部門

実践研究部門

協働形成研究部門

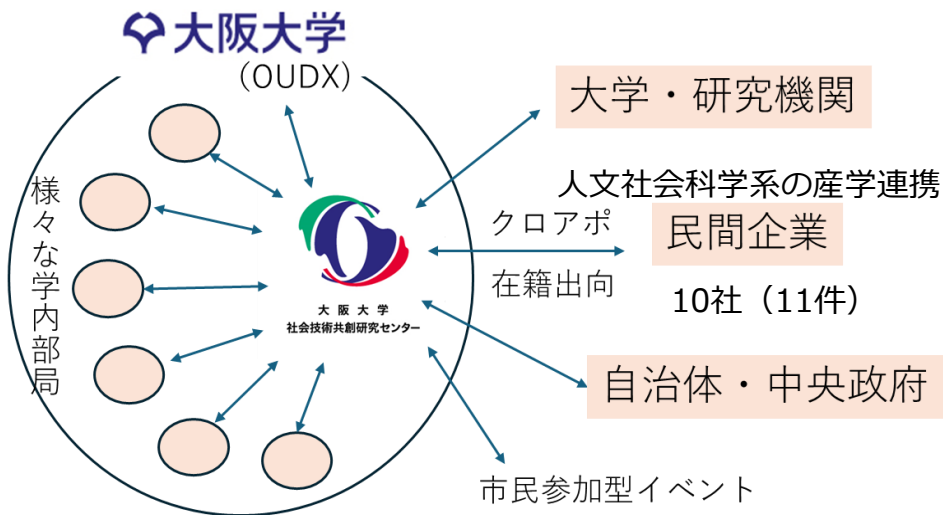
ELSI人材の育成

3部門長

+ コアメンバー（約15名）

+ 学内兼担

+ 招へい



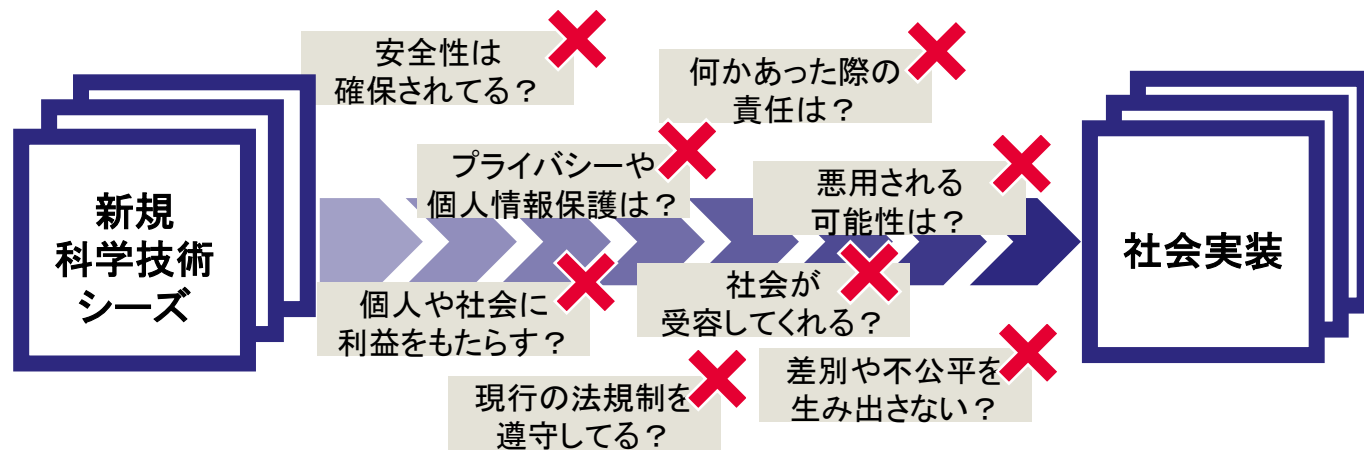
（2025年4月現在の専門分野）

情報通信法、ロボット法、AIと法、憲法学、リスク学、倫理学、情報法、デカルト哲学、情報の哲学、科学コミュニケーション、科学技術社会論、科学コミュニケーション、大学・地域連携、大学のアウトリーチ活動

人文社会科学系初の協働研究所の開所式



新しい技術の社会実装はいろんなところでつまずいてきた。



特に、パーソナルデータの利活用ではしばしば「炎上」事例が起きている。

逆の問題も…「何かあったらどうするんだ」問題

白黒はっきりするまで我が社での研究開発は中断しよう。

他国や国際機関でガイドライン作りが始まっていると聞いたので情報を集めてくれ。



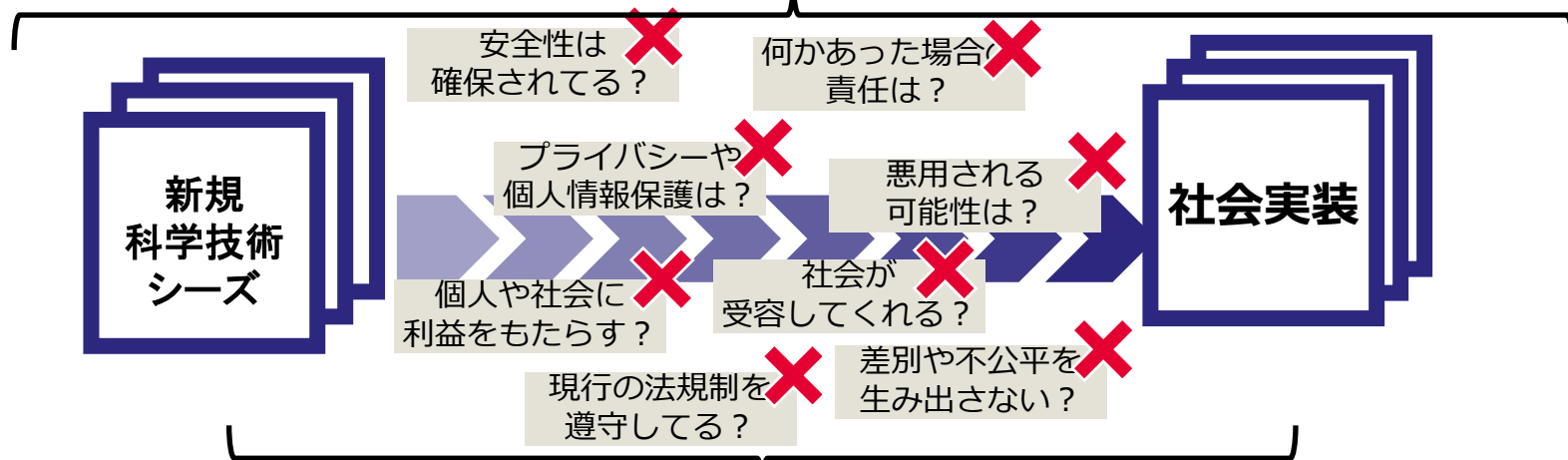
△△省や□□省は何て言っている？
同業他社はどうしてる？

何かあったらどうするんだ？
本プロジェクトは最近、〇〇が炎上した案件に似てないか？

何も新しいことができなくなる

共通の**原因**は…技術と社会の間の
ギャップを埋めるノウハウが欠如していること

責任ある研究・イノベーション
(RRI: Responsible Research and Innovation)



倫理的・法的・社会的課題
(ELSI: Ethical, Legal and Social Issues)

「ギャップを埋めるノウハウ」のことを「**社会技術**」と呼ぶ。

30年以上前に

そもそも **ELSI** という言葉は、ゲノム解析研究において生まれた。

ethical, legal, and social
issues/implications
倫理的・法的・社会的課題/含意

- 米国で1990年にスタートしたゲノム解析プロジェクトの中に「**ELSI研究プログラム**」が誕生（※当時、Issuesではなく、Implications）。**ヒトゲノムが解読された社会に生じるELSIを予想し、あらかじめ備えること**を目的に始まり、その後の「遺伝情報差別禁止法」の制定につながった。
- 外部向け研究予算の3%（のちに「**少なくとも5%**」）がELSIに関する研究に割り当てられることになり、その後、複数のELSI研究拠点が設置。
- このアプローチは、国家プロジェクトとして推進されたナノテクノロジーや脳科学などにも適用。
- 欧州では同様の研究はELSA（※AはAspectsの略）と呼ばれ、2010年代には「**責任ある研究・イノベーション（Responsible Research and Innovation : RRI）**」概念に発展。
- 日本では、主に生命医科学分野の中で、生命倫理の同義語として、あるいは、「技術以外のすべて」を包含する概念として使用されていたが、**ELSI研究への予算は長らくほとんどなかった**。
- 科学技術基本計画にも「倫理的・法的（法制度的）・社会的課題」とし第三期以来登場。

ELSIという言葉/考え方をあらゆる新規技術へ適用

AIの急速な進化とそれへ追いつきたいELSIチーム

予測・認識・分類型AI
(従来型AI)

生成AI

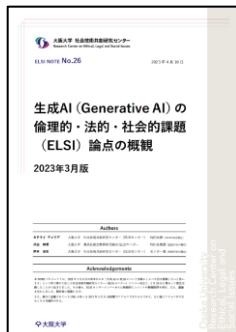
テキスト・画像・音声・動画等

AIエージェント



2024年6月21日(金) オンライン開催
第37回日本リスク学会シンポジウム
「AIのリスクを考える:
生体認証技術から生成AIまで」

日本リスク学会年次大会
2025年11月8日(土) 企画セッション
「AIエージェント」時代のリ
スクとガバナンスにおける論
点整理



?

大学の3つの機能とAIの利用とセーフガード

個別ケース

大学ルール

国ルール

教育

(教員の裁量)

例「生成AI教育ガイド」
(大阪大学SLiCSセンター)

「大学・高専における生成AIの
教学面の取り扱いについて」
(文部科学省 2023)

研究

研究倫理審査

※人を対象とする研究の場合。
ただし、被験者保護が主目的で本来
研究成果の社会実装は対象外

まだ議論が少ない。
(ELSI for “AI for Science”)

学務

(職員の裁量)

まだ議論が少ない。

※政府に対しては、「行政の進化と革新のための生成AIの
調達・利活用に係るガイドライン」(デジタル庁 2025)

大学の3つの機能とAIの利用とセーフガード

個別ケース

大学ルール

国ルール

教育

(教員の裁量)

学内外から相談

例「生成AI教育ガイド」
(大阪大学SLiCS)

ガイド作成に協力

「大学・高専における生成AIの
教学面の取り扱いについて」
(文部科学省 2023)

研究

研究倫理審査

AI時代の研究倫理審査のあり方

EUのHorizon Europeのプロジェクトに参加

まだ議論があまりない
(ELSI for “AI for Science”)

研究に着手

EdTechのELSI研究
JST RISTEXプロジェクト

学務

(職員の裁量)

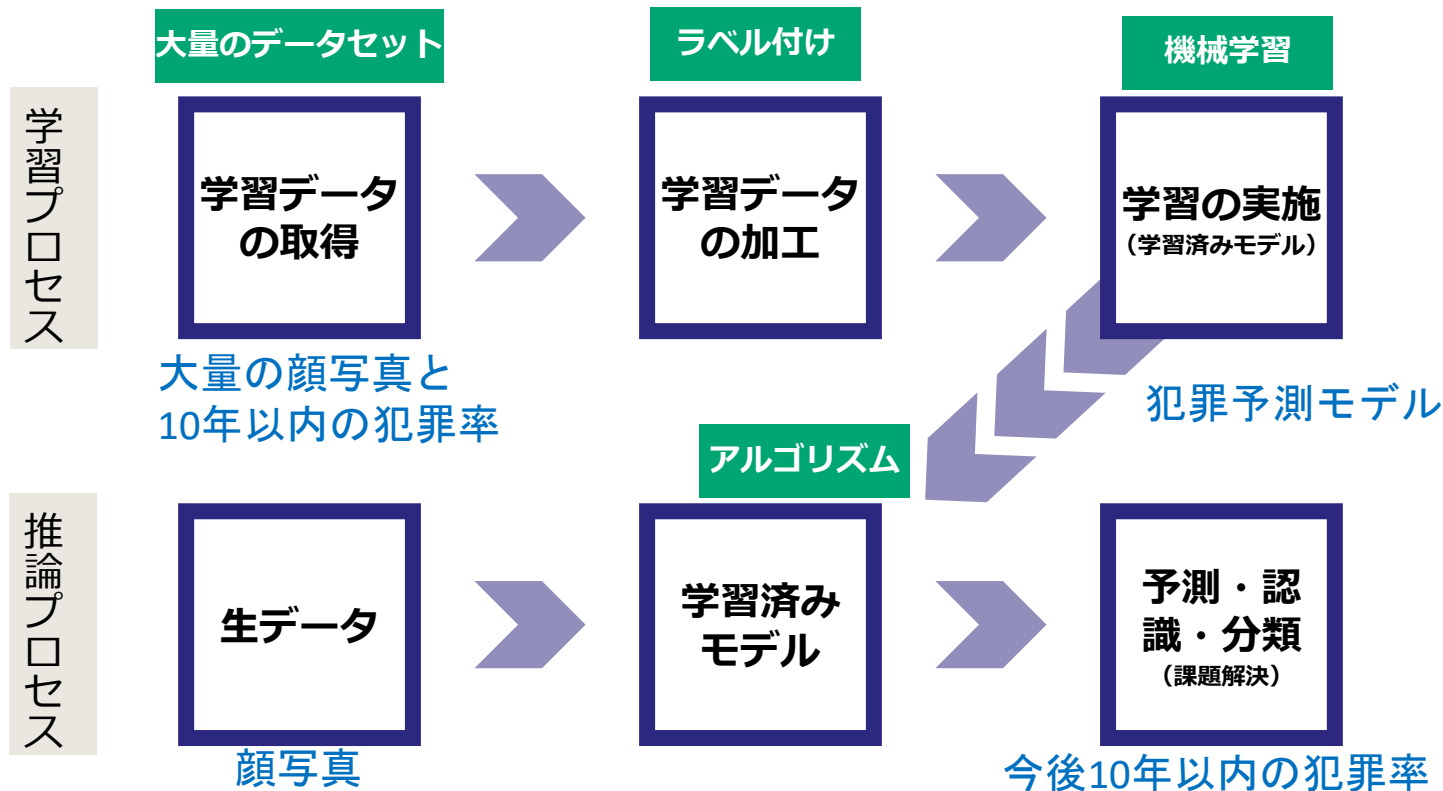
大学DXの取り組み

まだ議論があまりない。

※政府に対しては、「行政の進化と革新のための生成AIの
調達・利活用に係るガイドライン」(デジタル庁 2025)

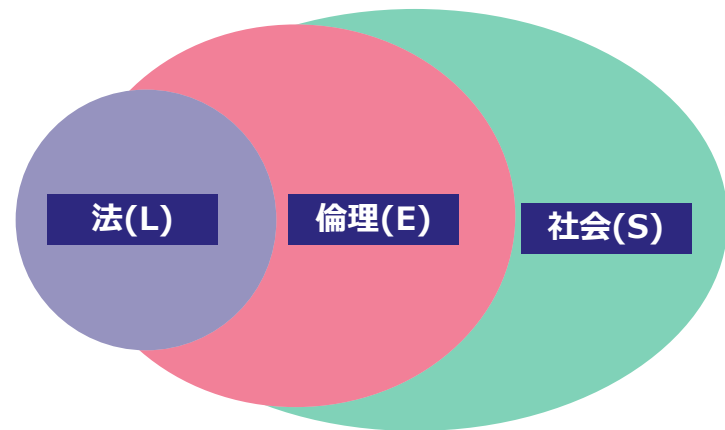
デジタル化でデータ量が激増 + 機械学習の性能が向上

⇒データセットさえできればモデル化できてしまうように見える



技術的に**できる**こと \neq 社会的にやって**よい**こと
(できるように見えること) (データ提供者や社会にとってリスクが十分に小さく、かつ)

どうやって線引きする？



法(L)

倫理(E)からの不断の見直し。

裁判所 (判例)
企業法務部

倫理(E)

社会において人々の行動を規範。安定的。法的に正当化される。

世論の動向
企業広報部

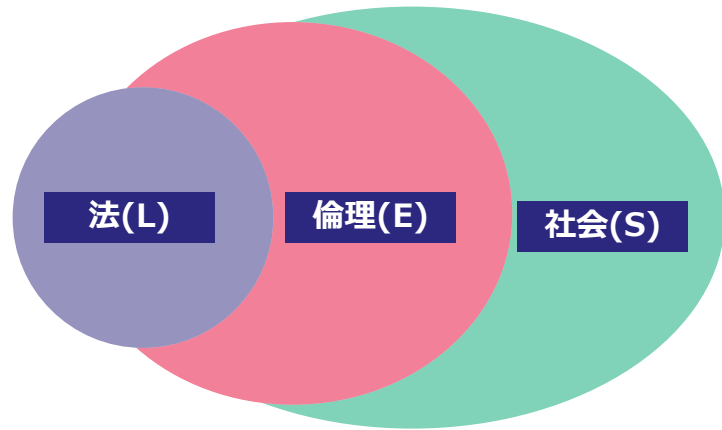
社会(S)

変化しやすい。不安定。

これまでは、法 (L) や社会 (S) を参照することでやってきた。

技術的に**できる**こと \neq 社会的にやって**よい**こと
(できるように見えること) (データ提供者や社会にとってリスクが十分に小さく、かつ、有益なこと)

どうやって線引きする？



法(L)

倫理(E)からの不断の見直し。

倫理(E)

社会において人々が依拠すべき規範。安定的。法(L)の基盤。

社会(S)

変化しやすい。不安定。

技術革新の速度が増し、法規制は後追いになる。判例も増えない。

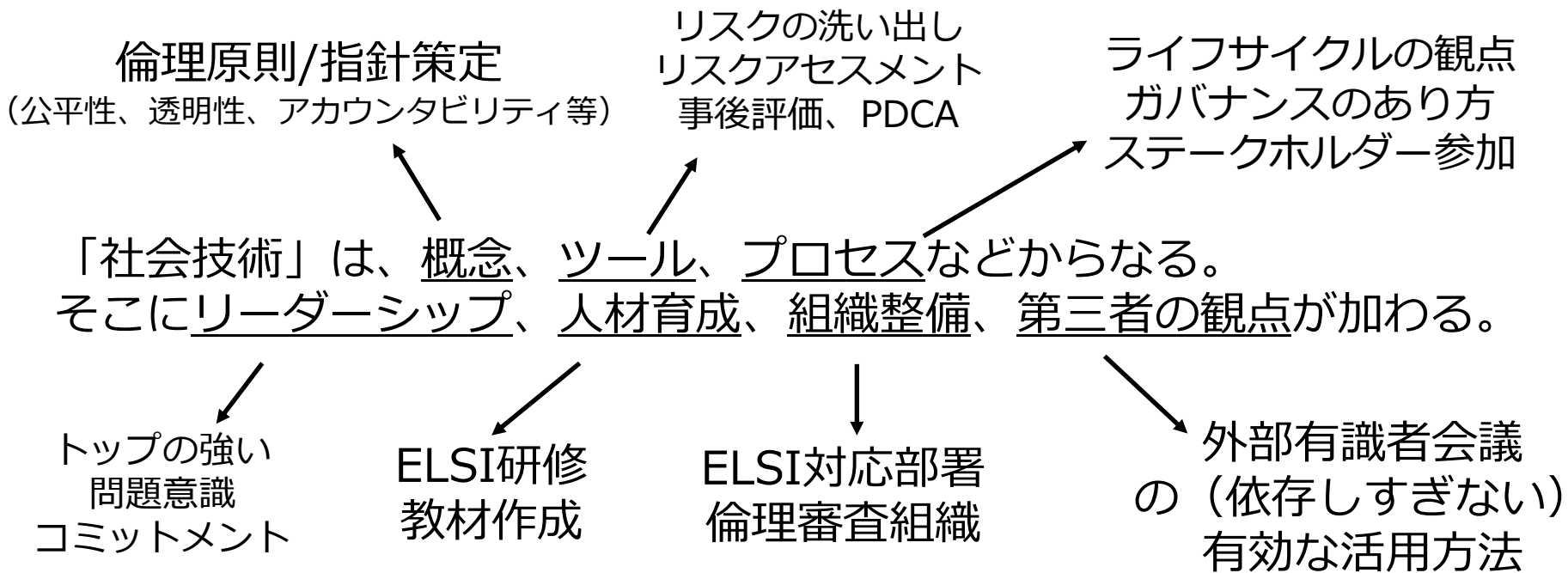
SNSに見られるように不安定で頼りにならない

不確かな時代には、羅針盤となるもの＝倫理（E）が必要。

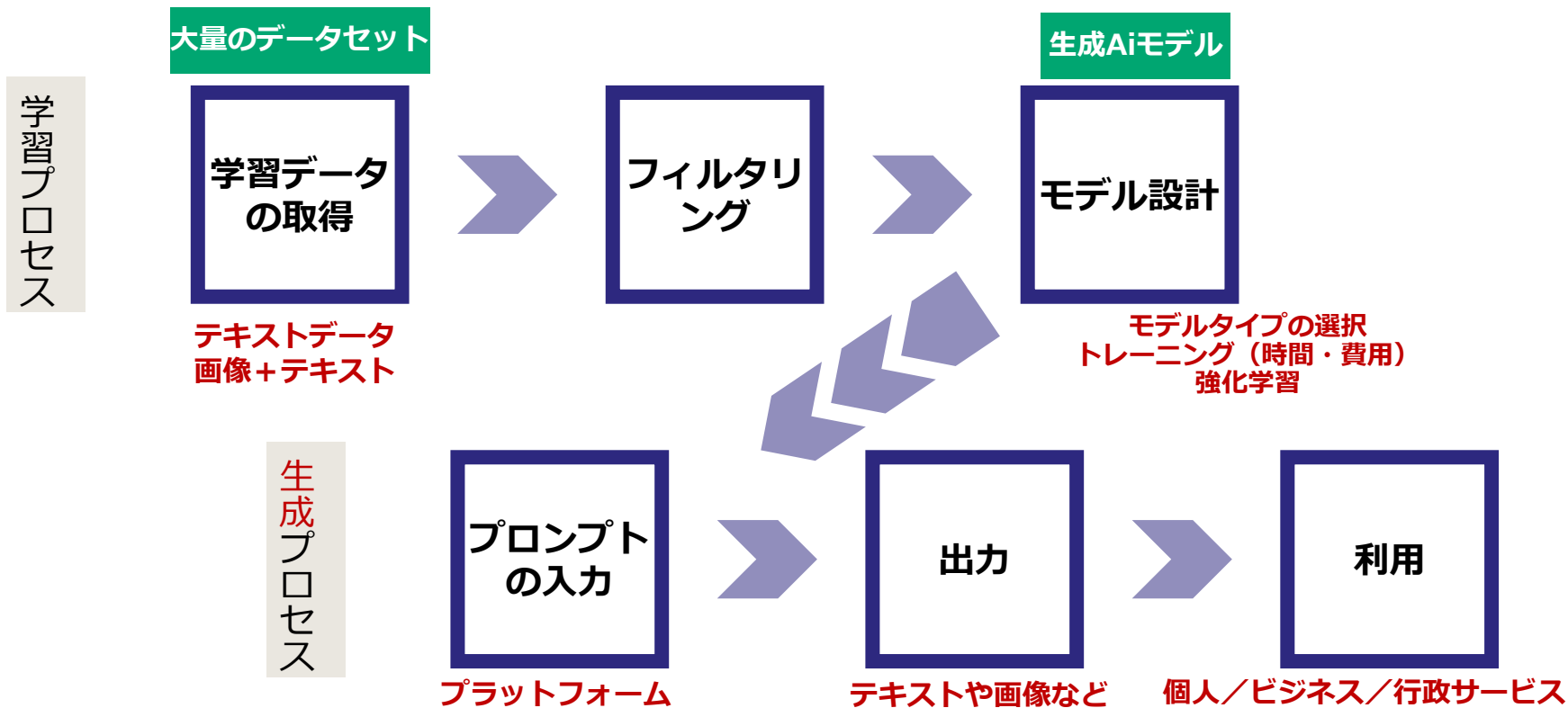
実際、国際機関やたくさんの企業がAI倫理原則/指針を策定している。

人文社会科学系の産学連携において、

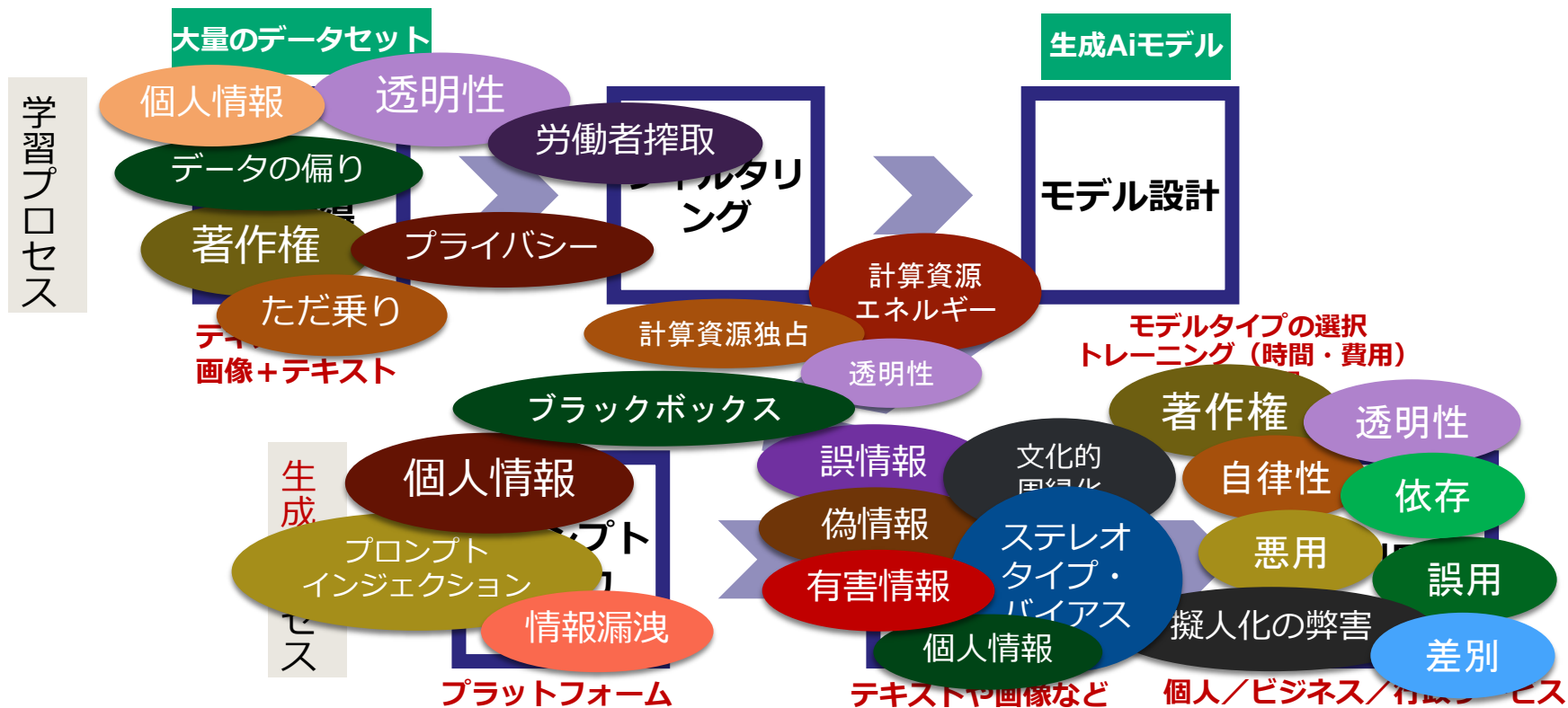
線引きのための「社会技術」の研究開発・社会実装



生成AIのケース



生成AIに指摘されているELSI



事前に何かが「安全」であることを示すには？

安全とは「許容できない**リスク**がないこと」
“freedom from risk which is not tolerable”

出典) ISO/IEC (2014) “Guide 51, Safety aspects -- Guidelines for their inclusion in standards”

「安全」（＝守りたいもの）のスコープが広がりつつある。

生命、健康、財産、
生態系、コミュニ
ティ、国土、伝統等



**人としての尊厳・自律性、
人権・自由・民主主義**

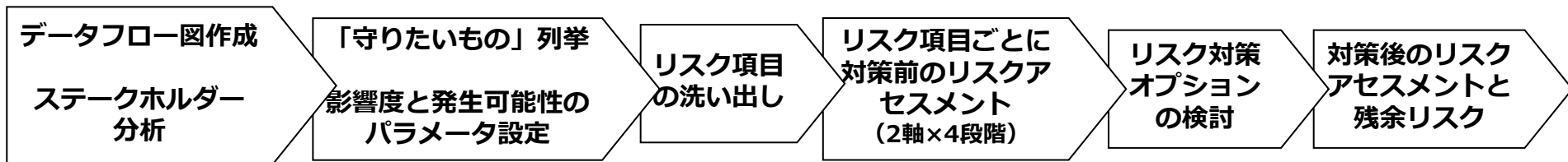
(EUのAI ActはSec.1で、環境、民主主義、法の支配を含む、安全、健康、基本的権利)

“AI Safety”も考え方は同じ

～当該AIシステムが「許容できないリスクがないこと」を示す。



AIリスクマネジメントプロセス



※残余リスクが「許容できないリスクがないこと」を示す。

誰がいつどうやってやるか？（ガバナンス）

Who + When + How

事例：大阪大学顔認証入場システムのリスクアセスメント

2024年6月から学内27か所で試行開始。1月から3月ころまでリスクアセスメントを実施。レポート作成、学内説明会で紹介。

データフロー図作成
ステークホルダー
分析

「守りたいもの」列挙
影響度と発生可能性の
パラメータ設定

リスク項目
の洗い出し

リスク項目ごとに
対策前のリスクア
セスメント
(2軸×4段階)

リスク対策
オプション
の検討

対策後のリスク
アセスメントと
残余リスク



守りたいもの	影響度 パラメータ	発生可能性 パラメータ
健康、安全	身体的影響	件数／件数比率
財産	経済的影響	件数／件数比率
人権、プライバシー	社会・精神的苦痛	人数／人数比率
尊厳	実存・精神的苦痛	人数／人数比率
自由、自律	行動萎縮の度合い	人数／人数比率
	意思決定への影響	
時間、機会	機会損失の程度	利用
利便性、期待	がっかり、期待外れ	
対人関係	交友関係への影響	

例：精神的苦痛

	指標	甚大	重大	限定的	無視可
影響度	精神的な苦痛	回復不可能なダメージ	治療が必要なダメージ	一時的なダメージ	軽微な不快感
発生可能性	指標	非常に高い	ある程度高い	一定の可能性	非常に低い
	人数比率	50%以上	10%以上	1%以上	1%未満

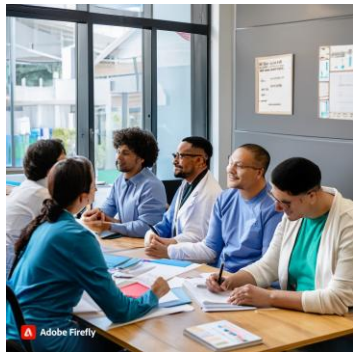
データフロー図作成

ステークホルダー
分析

「守りたいもの」列挙

影響度と発生可能性の
パラメータ設定リスク項目
の洗い出しリスク項目ごとに
対策前のリスクア
セスメント
(2軸×4段階)リスク対策
オプション
の検討対策後のリスク
アセスメントと
残余リスク

リスクの洗い出し作業がカギとなる。



リスク洗い出しワーク
ショップのイメージ

- ・メンバーの多様性が大事
- ・一段広い視点から眺めてみる（そもそも〇〇とは？）
- ・研究のライフサイクル全体を視野に
- ・データフロー図に沿って検討を進める
- ・過去の/他所での「炎上」事例を参考にする
- ・AI倫理原則などの原則を参照する
- ・多様なペルソナを想定する
- ・「使えない人」「使いたくない人」は誰か
- ・「不快に思う人」は誰か

データフロー図作成
ステークホルダー
分析

「守りたいもの」列挙
影響度と発生可能性の
パラメータ設定

リスク項目
の洗い出し

リスク項目ごとに
対策前のリスクア
セスメント
(2軸×4段階)

リスク対策
オプション
の検討

対策後のリスク
アセスメントと
残余リスク

①データの過剰収
集・保存

②プロセス面のコミュニ
ケーション（導入に至る
手続きの不備）

③内容面のコミュニケー
ション（説明不足による
誤解や不安）

④公的機関への情報連
携に関する不安

⑤ユーザビリティの
課題（変更や削除の
しにくさ）

⑥アカウントビリ
ティの課題

⑦誤認証時の不利益

⑧包摂性の課題（使え
ない人、使いたくない
人の不利益）

⑨なりすまし

⑩迷惑行為

⑪事後検証の欠落

⑫第三者の写り込み

⑬有効でない同意

影響度レベル

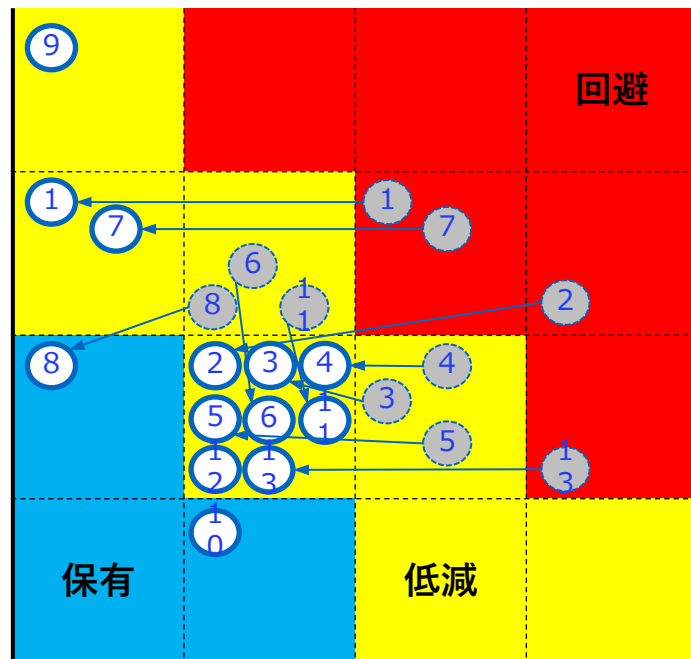
対策実施前/後のリスクアセスメント結果

4：甚大

3：重大

2：限定的

1：無視可



1：非常に低い 2：一定の可能性 3：ある程度高い 4：非常に高い

発生可能
性レベル

すでに指摘されている**AIエージェント**「リスク」の数々

➤ プライバシーに関するリスク

AIエージェントに正確で有益な仕事をしてもらうためには、人間の秘書と同様、ユーザーの様々なプライベートな情報を多く渡す必要がある。

➤ ミスアラインメントのリスク

ゴールを達成するために人間の価値観に反するような手段を採用する可能性がある。

➤ インターネットへのアクセスに伴うリスク

インターネットにアクセスして情報収集したり、予約や購入といった活動を行ったりすることが想定されるが、AIエージェントを騙すためのウェブサイトが設けられる可能性がある。

➤ 責任の所在のさらなるあいまいさ

AIエージェントが自律的な意思決定を行うために、不具合が生じた場合に、生成AI以上に責任の所在が分かりづらくなることが予想される。

➤ 自己進化するリスク

自己学習・自己進化の過程が開発者の想定範囲を超えてしまうリスクが指摘されている。

➤ AIエージェントの擬人化・依存

これはすでにAIチャットボットについて指摘されていることであるが、AIエージェントになるとさらに高度な活動が行われることでこうしたリスクが拡大することが想定される。

“AI for Science”特有のELSIはあるだろうか？

科学研究の自動化の論点

- ・ 科学的モノカルチャー化
- ・ 人間の認知能力の限界を超えた科学の出現
- ・ AI悪用のリスク
- ・ 研究成果の帰属

中村真也「科学研究とAIーサイエンスはどう変わるのかー」レファレンス, 898: 77-92, 2025.

<https://www.ndl.go.jp/jp/diet/publication/refer/index.html#no898>

課題とリスク

- ・ 環境コスト
- ・ 技術的課題：ブラックボックス性（解釈可能性の欠如）
- ・ 倫理的&セキュリティリスク：訓練データのバイアス、デュアルユース
- ・ データの課題とプライバシー懸念
- ・ システミック&社会的リスク

Fu, V. (2025). AI for Science: Opportunities, Challenges, and Future Directions. *Authorea Preprints*.

<https://www.techrxiv.org/doi/full/10.36227/techrxiv.173949768.84003950>

ご清聴ありがとうございました！

詳しい情報はウェブサイトをご覧ください。

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

ELSI NOTEは12月現在No. 64まで公開済み

https://elsi.osaka-u.ac.jp/research/research_category/elsi_note

