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」

岸本充生（KISHIMOTO, Atsuo）

日本リスク学会　理事

大阪大学 データビリティフロンティア機構（IDS）・教授

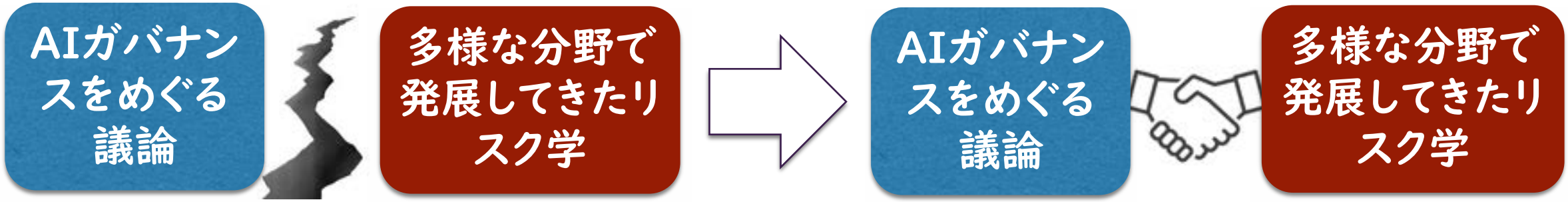
大阪大学 社会技術共創研究センター（通称、ELSIセンター）・センター長

Research Center on Ethical, Legal and Social Issues

Osaka University Research Center on Ethical, Legal and Social Issues

開催趣旨

- AIをめぐる議論は、どうしてリスク学者が出てこないのか？というくらい「リスク」概念のオンパレード。生成AIの普及後は「AI Safety」も新たなキーワードに。とはいえ単に「課題」の言い換えにすぎないことが多い。
- リスク学会関係者に向けて：AIガバナンスの議論に、他分野で培ってきたリスク学のノウハウがもっともっと貢献できるはず！
- AI関係者に向けて：AIガバナンスの議論では「リスク」「安全」概念の重要性が増している。これまでのリスク学の知見が役に立つはず！



大阪大学 社会技術共創研究センター（ELSIセンター）との共催

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

2020年4月設立

（3部門長+約15名）

3つの部門と4つの機能

- 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

新しい技術を社会実装するにあたり、倫理的・法的・社会的課題（ELSI）をあらかじめ抽出・対応することにより、責任ある研究・イノベーション（RRI）を達成する。

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

メンバーの専門分野の多様性

情報通信法、ロボット法、AIと法、国際私法、リスク学、科学社会学、臨床哲学、倫理学、社会学、情報の哲学、科学哲学、情報法、科学史・科学論、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論、音楽学・・・

阪大内の様々な部局との共創研究

中之島芸術センター、量子情報・量子生命研究センター、レーザー科学研究所、感染症総合教育研究拠点等

阪大外の様々な企業との共創研究

メルカリ、NEC、リコー、三菱電機、電通、NHK技研、クモノスコーポレーション、PwCコンサルティング等

本シンポジウムと関連するELSI NOTEシリーズ

2023年4月公開

2023年9月公開

2024年4月公開

2024年6月公開

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」

●講演

岸本充生（大阪大学）：AIリスクのガバナンスの動向とリスク学の貢献可能性

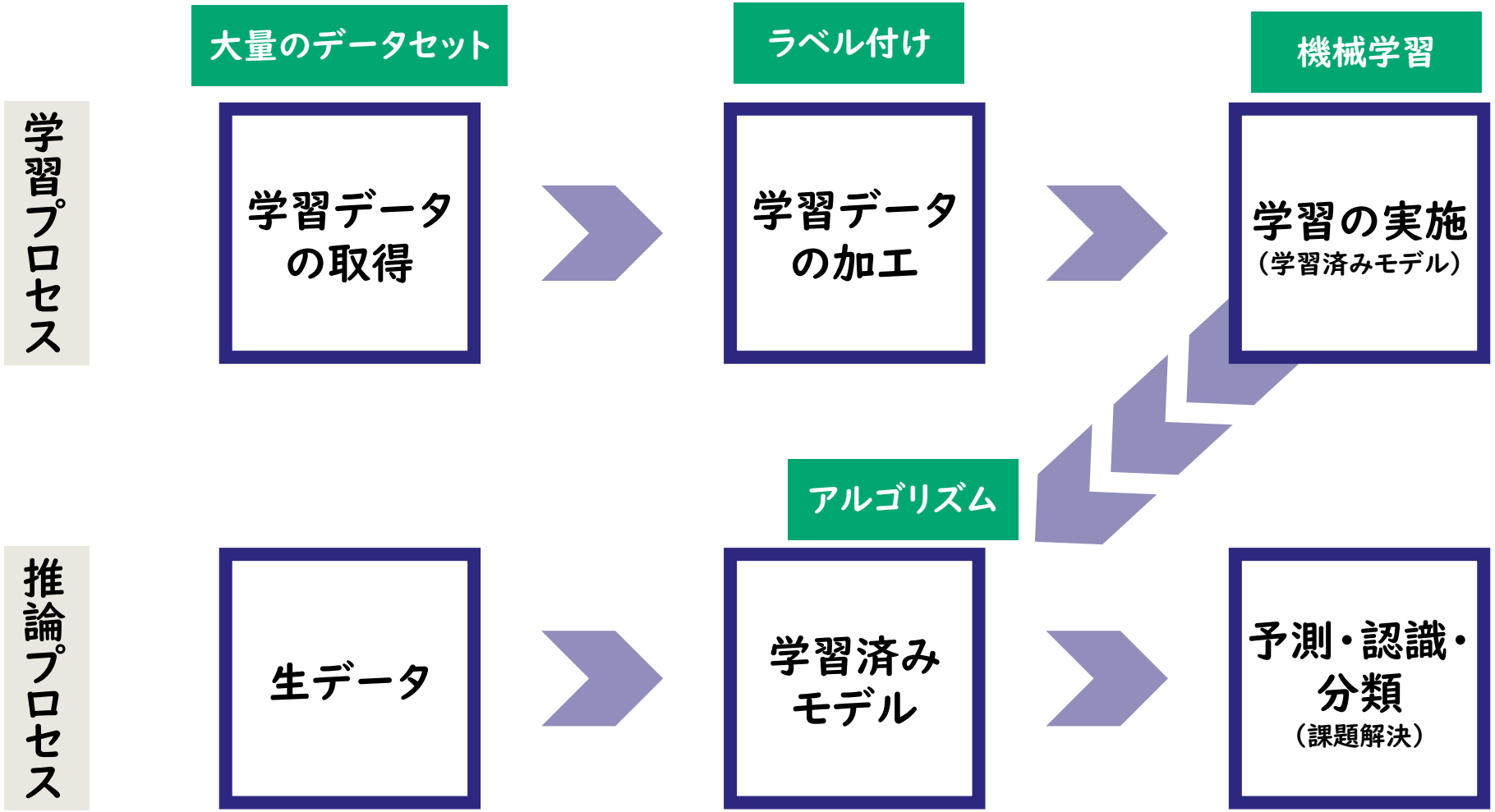
カテライ　アメリア（大阪大学）：生成AIのELSIリスクの概要

田中孝宣（大阪大学、NEC）：リスクアセスメント事例～大阪大学の顔認証入場システム～

データ利活用やAI社会実装はたびたび「炎上」してきた。

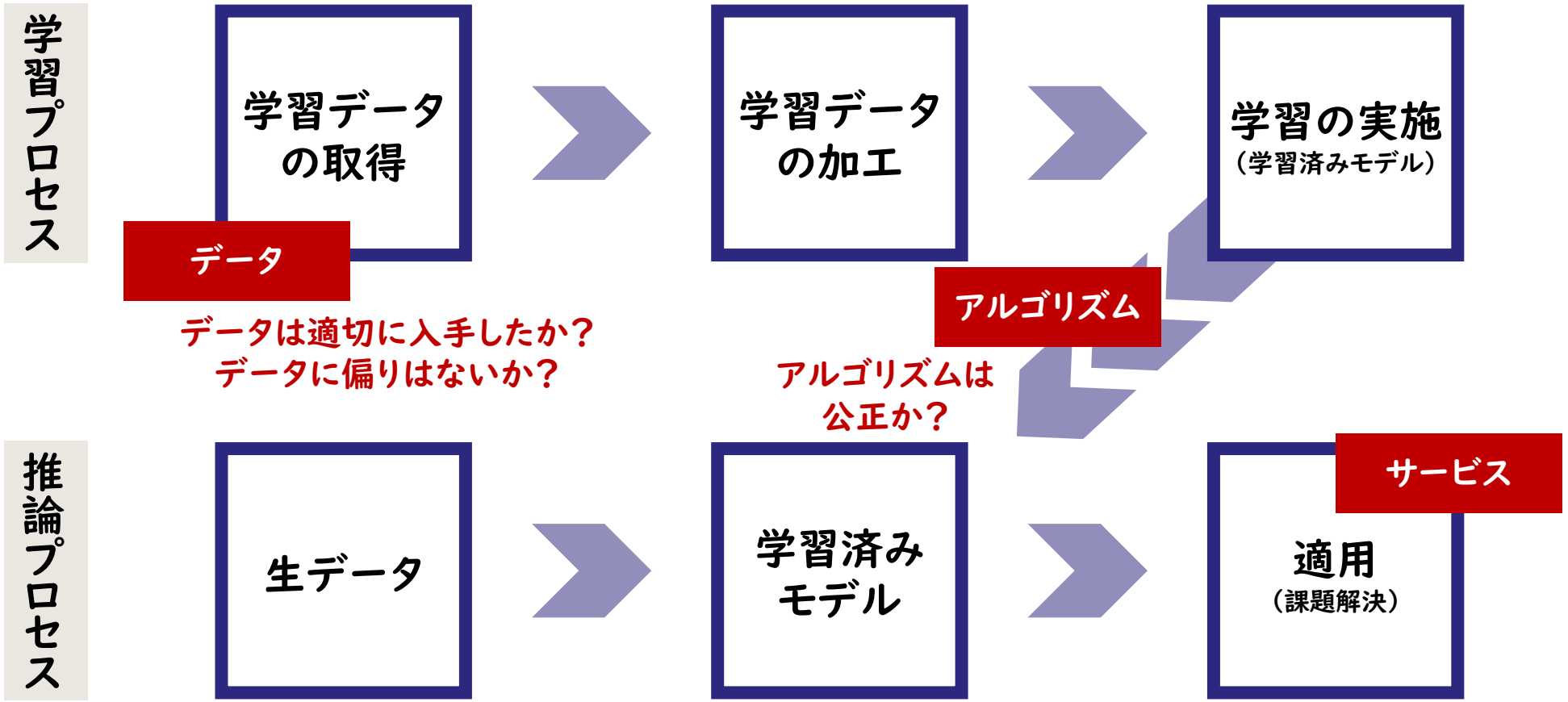
パーソナルデータとAIを使うといろんなことができる。

ように見える



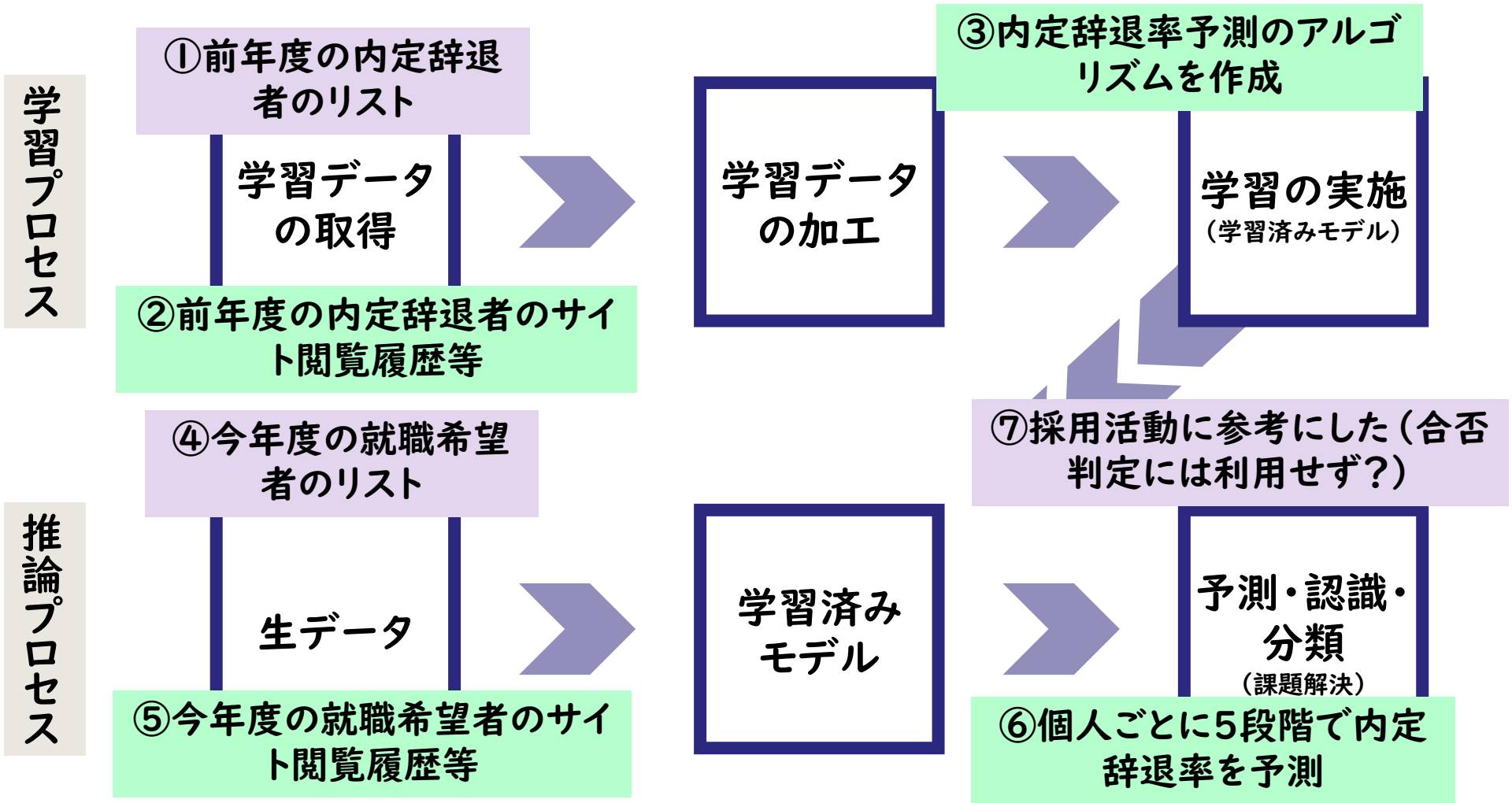
従来型のAIの開発と実装においてELSIが生じそうな箇所はどこだろうか。

予測・認識・分類タイプの



例

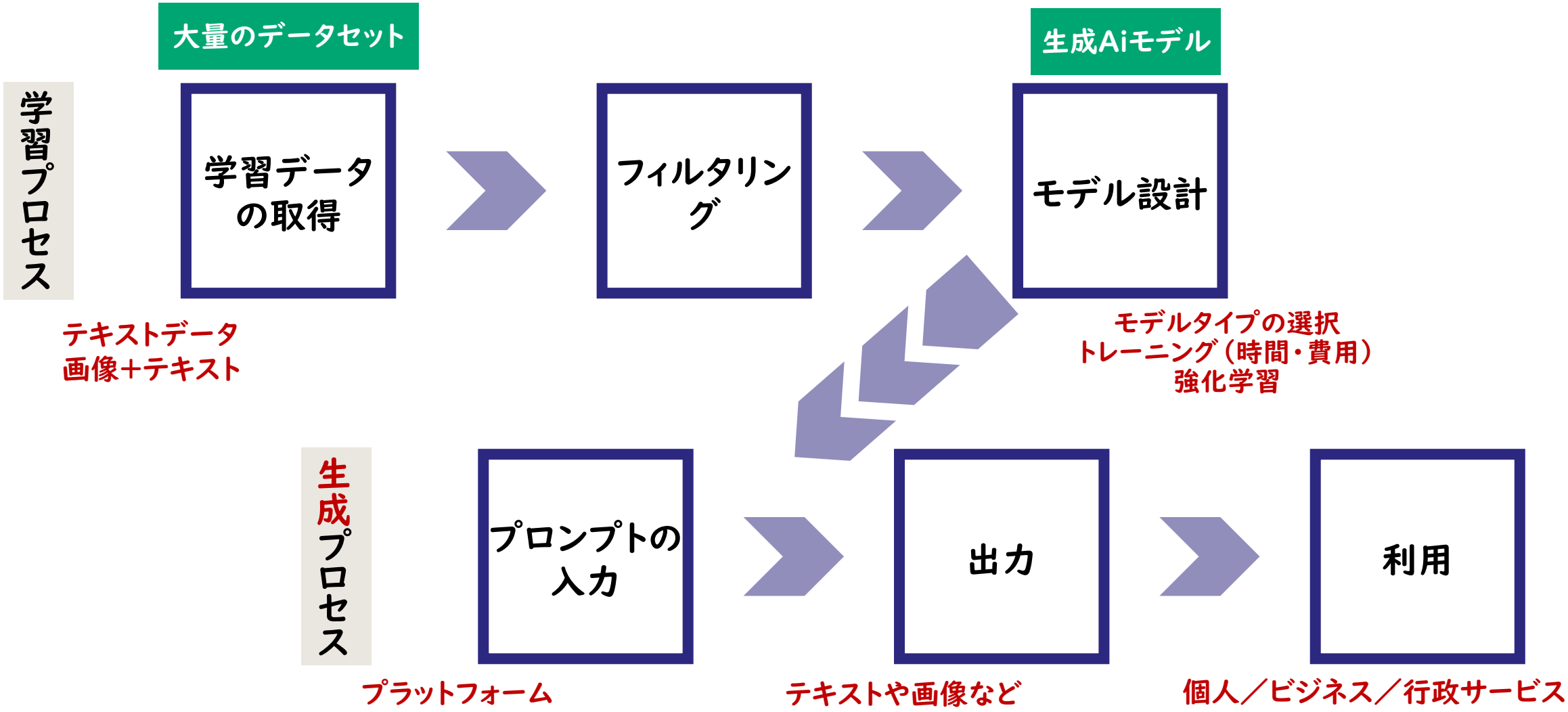
2019年のリクナビ内定辞退率予測のケース



「できる」かもしれないが、やるべきではなかった「データ利活用」ケース

生成AIの場合は？

（カテライ アメリアさんの講演参照）

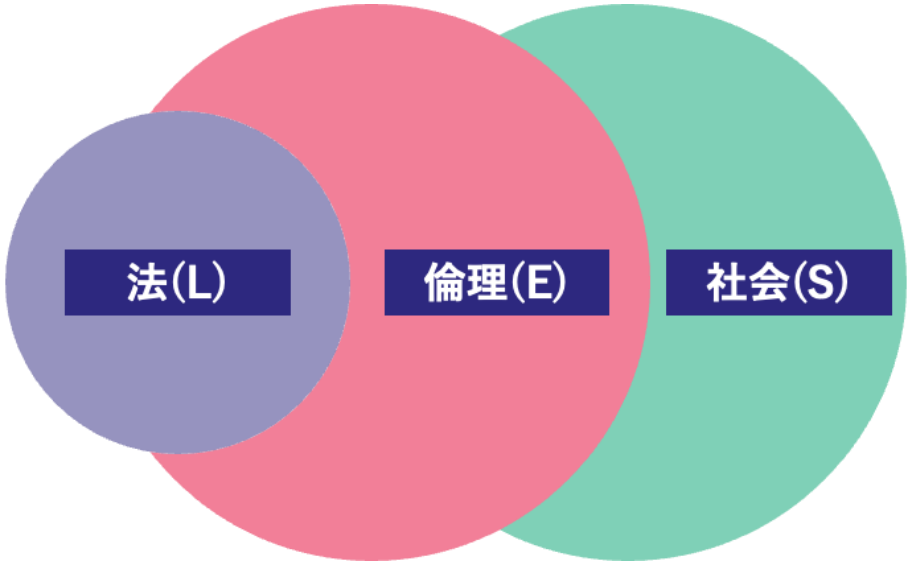


技術的にできること ≠ 社会的にやってよいこと

（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？



裁判所（判例）
企業法務部

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
範。安定的。法(L)の

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。

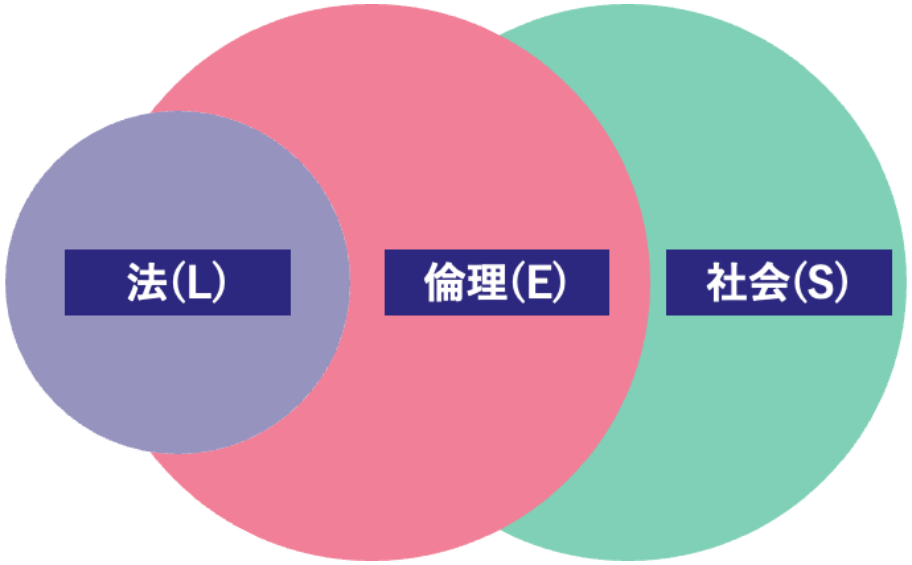
技術的にできること ≠ 社会的にやってよいこと

（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）

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。

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。

SNSに見られるように不安定で頼りにならない

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。

倫理（E）は私たちが何を守りたい（守るべき）と考えているかとも言える

EUのAI Act法案（まもなく官報に掲載）

Article 1　Subject matter

The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation.

第1条　主題

本規則の目的は、域内市場の機能を向上させ、人間中心の信頼できる人工知能（AI）の導入を促進することであり、同時に、EUにおけるAIシステムの有害な影響に対して、民主主義、法の支配、環境保護を含む、EU憲章に謳われている健康、安全、基本的権利の高水準の保護を確保し、イノベーションを支援することである。

守りたい（守るべき）対象の拡大

（＝「安全」概念の拡大）

◎リスクコミュニティにはおなじみ◎

安全（Safety）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”

- 「守りたいもの」は多くの場合、人の生命だった。
- リスクアセスメントを行い、それに基づいてリスクマネジメント。
- 許容できる／許容できないの線引き（安全目標）が必要。
- 多様なステークホルダーとのコミュニケーションによる。

◎AIコミュニティにはおなじみ◎

生成AI以降、“AI Safety”という言葉が急浮上

AISI AI SAFETY INSTITUTE

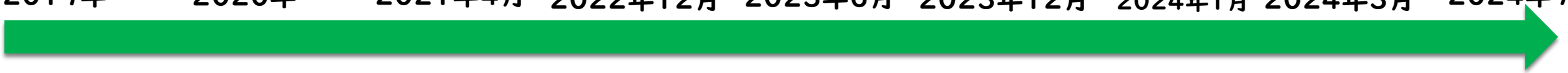
英国、米国、日本で設立

総務省と経産省から4月に公表されたAI 事業者ガイドライン（第 1.0 版）の「安全性」の項目には、*「各主体は、AI システム・サービスの開発・提供・利用を通じ、ステークホルダーの生命・身体・財産に危害を及ぼすことがないようにすべきである。加えて、精神及び環境に危害を及ぼすことがないようにすることが重要である。」*と書かれている。

→社会実装前に、多様な「守りたいもの」に対して、「許容できないリスクがないこと」を確認していく社会技術の開発が必要になる。

（田中孝宣さんの講演を参照）

EUのAI Actを中心に、AIをめぐるガバナンスの動向



AIに関するハイレベル専門家グループ (AI HLEG) 倫理ガイドライン 2019年

欧州委員会 AI白書 White Paper on AI 2020年

欧州委員会 AI Act 案を発表 2021年4月

EU理事会 AI法に関する見解を採択 2022年12月

EU議会 AI Actの修正案を採択 2023年6月

議会と理事会が政治的合意 (トリローグ合意) 2023年12月

AI Office を設立 2024年1月

EU議会がAI Actの最終案を採択 2024年3月

AI Actが官報に掲載 2024年7月

2019年3月 人間中心のAI社会原則

2020年7月 AI 原則実践のためのガバナンス・ガイドライン ver. 1.0

2022年10月 OpenAI社がChatGPTを発表

2023年3月 Future of Life Institute オープンレター公表

2023年5月 G7広島サミット「広島AIプロセス」

2023年10月 バイデン大統領 大統領令を公布

2023年11月 初のAI Safety サミット開催 (ロンドン) Bletchley宣言

2023年12月 「広島AIプロセス」包括的政策枠組み

2024年4月 「AI事業者ガイドライン (第1.0版)

2024年5月 第2回のAI Safety サミット開催 (ソウル)

リスクベースのアプローチ（Risk-based approach）

EUのAI Act（4/19案）の前文（recital）より引用

「AIシステムに対して比例的かつ効果的な拘束力のあるルールを導入するためには、明確に定義されたリスクベースのアプローチに従うべきである。そのアプローチは、AIシステムが生み出しうるリスクの強度と範囲に合わせて、ルールの種類と内容を調整するものでなければならない。したがって、特定の容認できないAI実践を禁止し、高リスクAIシステムに対する要件と関連事業者の義務を定め、特定のAIシステムに対する透明性義務を定めることが必要である。」

※実は二重の意味での「リスクベース」のアプローチになっている！
（1）リスクの大きさに応じて対応が変わる　←こっちだけ注目されがち
（2）高リスクのものはリスクマネジメントプロセスが必要←実は本命

リスクの大きさ

受け入れられないリスク
（an unacceptable risk）

→利用禁止（第5条）(a)～(h) 8項目

高リスク（a high risk）

→リスクマネジメントシステム等の要件を満たす必要（第8条～）

ある特定のAIシステム

→透明性義務

高リスクAIシステム以外のAIシステム

→行動規範とガイドライン

汎用目的AIモデル*

高インパクト能力（high impact capabilities）を有する場合（＝浮動小数点演算で測定されるその学習に使用された累積計算量が10^{25}を超える）にシステミック・リスクを伴うものと分類される。策定されるCodes of practice（実務規範）を遵守しなければならない。

*汎用人工知能（AGI）とは別の概念

第9条 リスクマネジメントシステム（Risk management system）

2. リスクマネジメントシステムは、定期的な体系的レビューと更新を必要とする、高リスクAIシステムの全ライフサイクルを通じて計画・実行される継続的な反復プロセスとして理解されなければならない。リスクマネジメントシステムは、以下のステップから構成されるものとする

(a) 高リスク AI システムがその意図された目的に従って使用された場合に，その高リスク AI システムが健康、安全または基本的権利にもたらし得る既知のリスク及び合理的に予見可能なリスクの特定（identification）及び分析（analysis）；

(b) 高リスクAIシステムがその意図された目的に従って使用され、かつ、合理的に予見可能な誤用の状況下で使用される場合に生じ得るリスクの推定（estimation）及び評価（evaluation）；

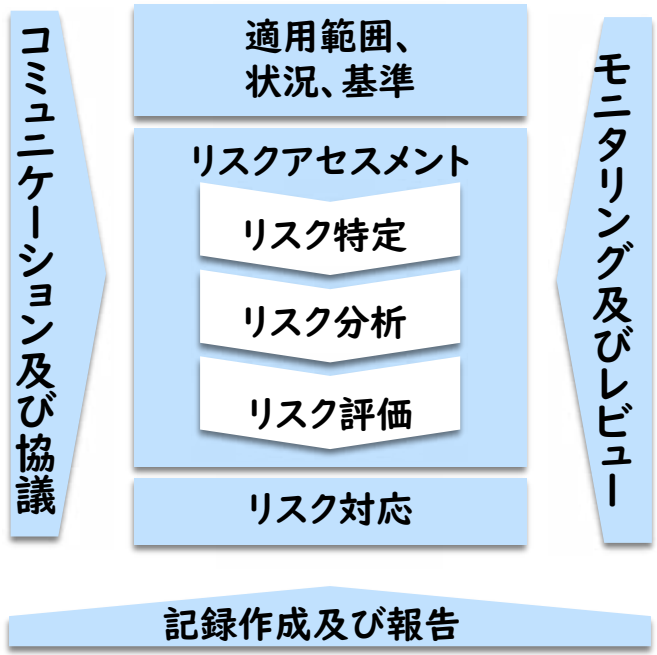
(c) Article 72にいう市販後モニタリングシステムから収集されたデータの分析に基づき、生じ得るその他のリスクを評価（evaluation）すること；

(d) (a)に従って特定されたリスクに対処するために設計された、適切かつ的を絞ったリスクマネジメント措置の採用。

（中略）

5. 2(d)に言及するリスクマネジメント措置は、各ハザードに関連する残余リスク（residual risk）、及び高リスクAIシステムの全体的な残余リスクが受容可能である（acceptable）と判断されるものでなければならない。

ISO 31000: 2018
リスクマネジメント規格
に対応している。



(65) ‘systemic risk’（システミック・リスク）←新しい概念が導入！

汎用目的AIモデルの高インパクト能力に特有のリスクで、その影響力（their reach）のために、または公衆衛生、安全、治安、基本的権利、社会全体に対する、バリューチェーン全体にわたって大規模に伝播する可能性のある実際のもしくは合理的に予見可能な悪影響のために、EU市場に重大な影響を及ぼすもの；

a risk that is specific to the high-impact capabilities of general purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain;

リスク学がこれまで開発してきた多様な概念や考え方が役に立つ。

A) リスクとベネフィット

B) リスク評価（assessment）とリスク管理（management）の役割

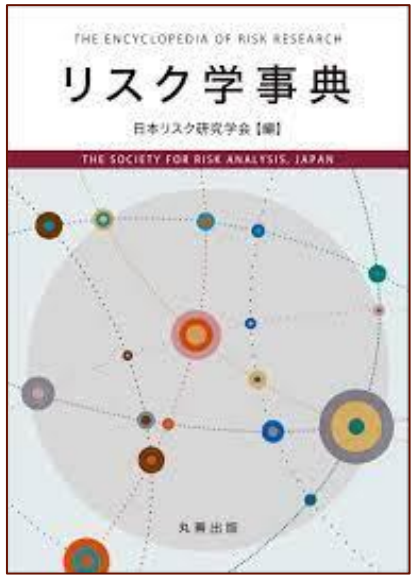
C) リスクトレードオフ（tradeoff）

D) 主観的リスク（リスク認知）と客観的リスク

E) リスクホメオスタシス/リスク相殺（homeostasis）

F) リスクコミュニケーション（communication）

G) リスクガバナンスの枠組み



丸善2019年

しかし、長期的・システミックなリスクや人類存続リスクなどの検討は不十分。

リスク学にも多様な価値を扱うためのイノベーションが求められている。

ご清聴ありがとうございました。