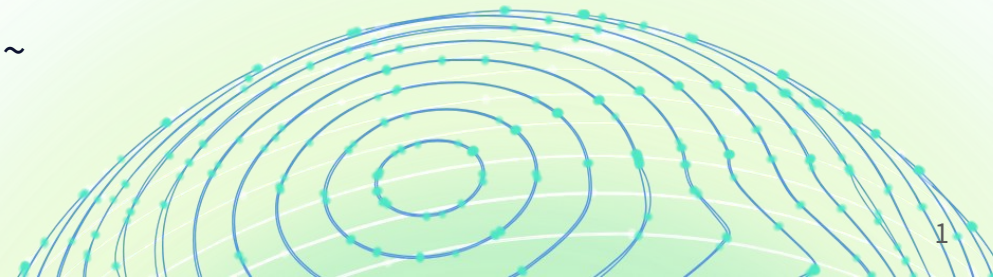
一般社団法人
AIガバナンス協会
AI Governance Association

AIガバナンスの現在地点と プライバシーガバナンスとの関係

AIガバナンス協会理事　佐久間弘明

2025.07.17 Thu
MyData Japan 2025 ~ MyData in Practice ~



自己紹介

佐久間 弘明　SAKUMA Hiroaki

一般社団法人AIガバナンス協会理事・事務局長

- 経済産業省でAI・データに関わる制度整備・運用に従事したのち、Bain & Company、Robust Intelligenceを経て現職
- AIガバナンス協会では、AIガバナンスをめぐる標準策定や政策提言などを行う。スマートガバナンスでは組織のテクノロジーガバナンス構築支援などを担当。ほか、内閣官房デジタル行財政改革会議事務局政策参与（データ利活用制度検討担当）なども務める
- 社会学の視点でAIリスクをめぐる言説の研究にも取り組む。修士（社会情報学）

AIGAのご紹介

一般社団法人
AIガバナンス協会
AI Governance Association

一般社団法人AIガバナンス協会は、AIに関わるあらゆるステークホルダーが集まるフォーラムとして、適切なリスク管理を通じてAIの価値を最大化する取組である「AIガバナンス」があたりまえのものとして定着した社会の実現をめざします。

一般社団法人AIガバナンス協会 = AIGAが重視する価値

- イノベーションの促進
- マルチステークホルダーでの信頼構築
- 社会的な価値の実現

AIGAの組織体制

業界やバリューチェーン上の立場をまたがり、多様なプレイヤーが参画

正会員社（和名五十音順）

*2025年5月現在。一部企業のロゴは未掲載。

一般社団法人 AIガバナンス協会 AI Governance Association

サマリ

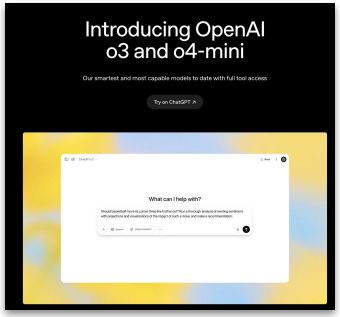
本日お話しさせていただく内容（大枠）

- 企業・組織にとっての AIガバナンスの必要性
- AI活用に付随するリスク
- AIガバナンスとプライバシーガバナンス
- 企業のAIガバナンス実装状況と課題

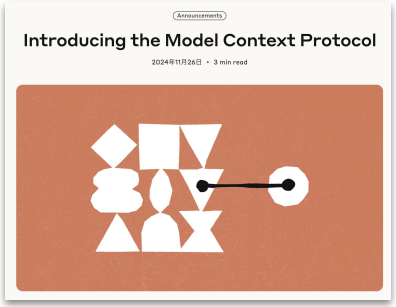
AIガバナンスの必要性

技術的な不確実性は高く、「インシデント」とされる事象も増加。
社会的な信頼の観点でも、組織はリスクと向き合う必要がある

GitHub

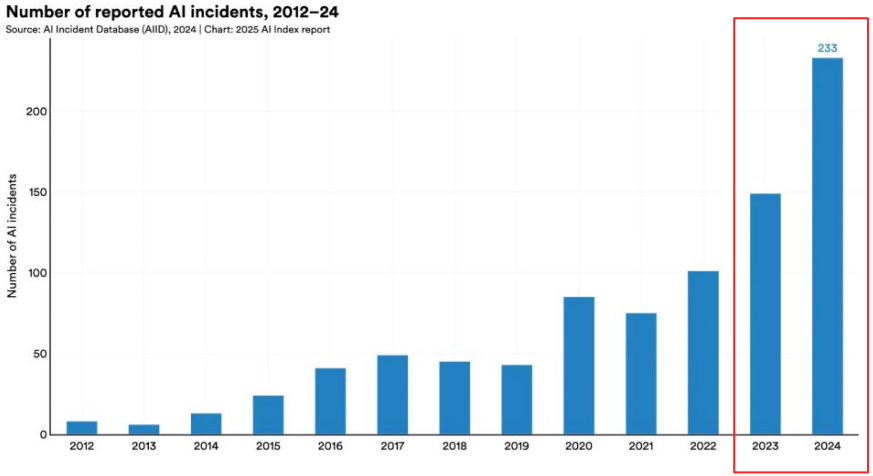


OpenAI



Anthropic

インシデントと認識される事象　*の増加



Stanford University "The 2025 AI Index Report"

* AI Incident Databaseの登録数がベース

AI法成立をはじめとして、政策動向の変化は激しい
投資家等のステークホルダーもAIリスクへの関心を強める

2025.05.28 NHK

2025.06.11 日本経済新聞

AI推進法により、**政府としての体制強化**　や**リスク対策における「PDCAサイクル」**などが規定された

- 政府へのAI戦略本部の設置や基本計画の策定
- 政府による指針の設定、リスクをめぐる調査権限

経営における「AIガバナンスの不在」は、活用の停滞を招くか、インシデント等を通じた社会的受容の失敗を帰結する

組織の経営層で「 AIに投資している」とする割合 *

95%

⟷

経営層で「 AIガバナンス構築が完了している」とする割合 *

34%

組織のうち、「データガバナンスの不足」を AI活用の障壁に挙げる割合 **

62%

▼

特にリスク許容度の低い日本企業においては、ガバナンスの雛形が無い状態が続けば ……

- リスク懸念からそもそも活用に踏み切れず、課題解決の遅れや機会損失が生じる
- ガバナンス不在のまま活用に踏み切った結果、インシデントが起き社会的受容性を損なう おそれも

AIガバナンスは適切なリスクテイクを行い、成長や社会的価値を実現するための必須要素

* Ernst & Young LPP “EY Pulse research: AI Survey” (2024.07)、500社の経営層への調査
** Precisely ”2025 Outlook: Data Integrity Trends and Insights”、米国中心の 565社を調査

AI活用に付随するリスク

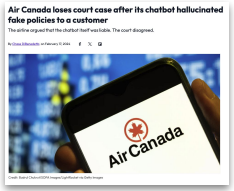
AI活用に付随するリスク

AIシステムの利活用の広がりは、セーフティ・セキュリティ両方の観点で多様なリスクにつながる

セーフティリスクの例

意図せず発生する性能や倫理面の問題

- AIの精度劣化やハルシネーションによる誤った情報の出力
- 出力へのバイアスの反映
- プライバシー侵害
- 有害コンテンツ、著作権等の権利侵害コンテンツの出力



出所: Chase DiBenedetto (2024.02.18)

出所: BBC (2019.11.12)

その他の広範囲に影響が及ぶ社会的問題

- 労働代替、格差の拡大
- AIの電力消費に伴う環境問題
- 社会の分極化
- AIのコントロール喪失

セキュリティリスクの例

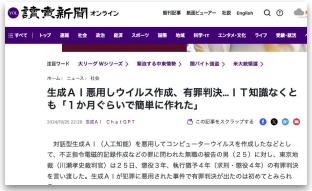
組織の組み込んだAIをインターフェースに行われる攻撃

- データポイズニング、プロンプトインジェクション等による誤作動の誘発や機密情報の引き出し
- サプライチェーンの複雑性を活かしたサイバー攻撃

出所: ZDnet (2023.07.27)

AI生成物などを活用した攻撃の増加・高度化

- 偽情報の流布
- コンピュータウイルス、マルウェア等の開発の容易化
- DDoS攻撃の容易化
- CBRN情報などの出力による、安全保障上の問題



出所: 読売新聞オンライン (2024.10.25)

一般社団法人 AIガバナンス協会 AI Governance Association

セーフティリスク例: 活用データや、評価・決定の不透明性

リスク概要と発生被害

- 2019年8月、日本IBMはAIを人事評価・賃金査定に導入すると発表。同AIは、40項目の評価要素をもとに社員の賃金査定に用いられるとされたが、これに対して同社労働組合は、**プライバシー侵害、差別、評価のブラックボックス化、そして自動化バイアスのおそれ** などを問題視し、団体交渉に発展
- この問題は東京都労働委員会において争われ、最終的には5年の係争ののち、2024年8月に以下のような労使間の和解が成立することとなった
  - **賃金査定でAIが考慮する項目全部の開示**
  - 上記項目と、**賃金規程上の評価項目との関連性の説明**
  - **不利益決定の際の、必要な AIの提案内容の開示**
  - 今後AI活用をめぐって疑義が生じた場合の、労使間での協議

AIを人事評価に活用、日本IBMが労組に情報開示へ　和解成立

有料記事

北川慧一　2024年8月2日 15時32分（2024年8月2日 17時54分更新）

日本IBMとの和解成立について公表する「日本金属製造情報通信労働組合（JMITU）」と弁護団 =2024年8月2日、東京・霞が関、北川慧一撮影

日本IBMが人工知能（AI）を利用した人事評価について、同社の従業員が加盟する労働組合が会社側にAIの詳細を説明するよう求めた労使紛争が和解した。労組が2日、記者会見で明らかにした。AIの使用するデータや評価内容を労組側に開示する内容で、AIによる雇用管理に対して国内の法規制が未整備の中、労組が透明性確保の監視役を担う。

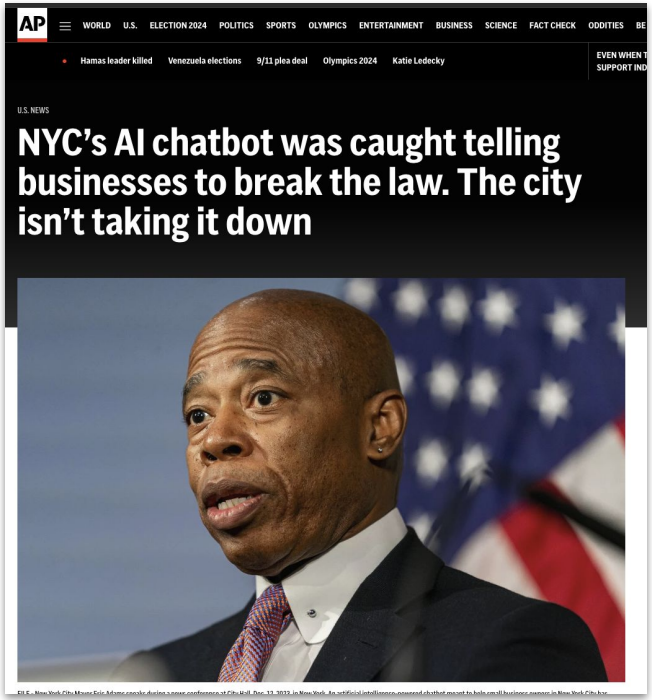
和解は1日、東京都労働委員会で成立した。労組が出した2020年4月の救済命令申立

2024年8月2日 By 朝日新聞

セーフティリスク例: 生成AIによる不適切な出力

リスク概要と発生被害

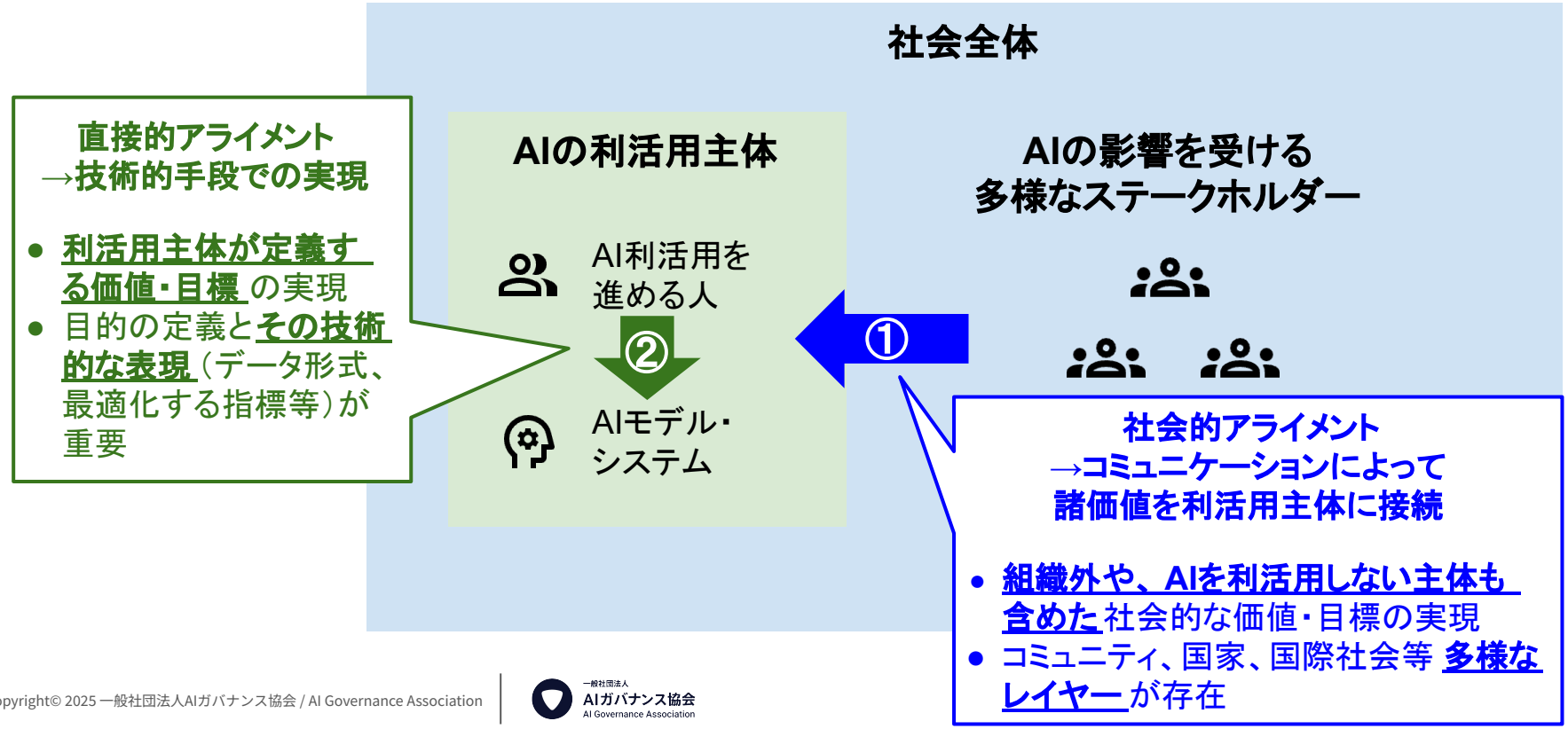
- ニューヨーク市が中小企業の経営者支援のために提供したAIチャットボット（MicrosoftのAzure基盤）が、**違法な解雇等の法令違反を勧める出力や不正確な出力** を行った
  - 出力の例:「セクハラを訴えた社員を解雇するのは合法」「レストランはネズミにかじられたチーズを顧客に提供しても構わない」
- インターフェースに下記のような工夫があったもののワークせず
  - リンクで回答を確認するように促す文章が表示されるが、**根拠となるリンクが提供されないことも多かった**
  - 出力に誤りや有害な内容等が含まれうるというディスクレイマーを用意していたものの、行政サービスであることも影響して炎上を免れなかった



2024年4月4日 By APnews

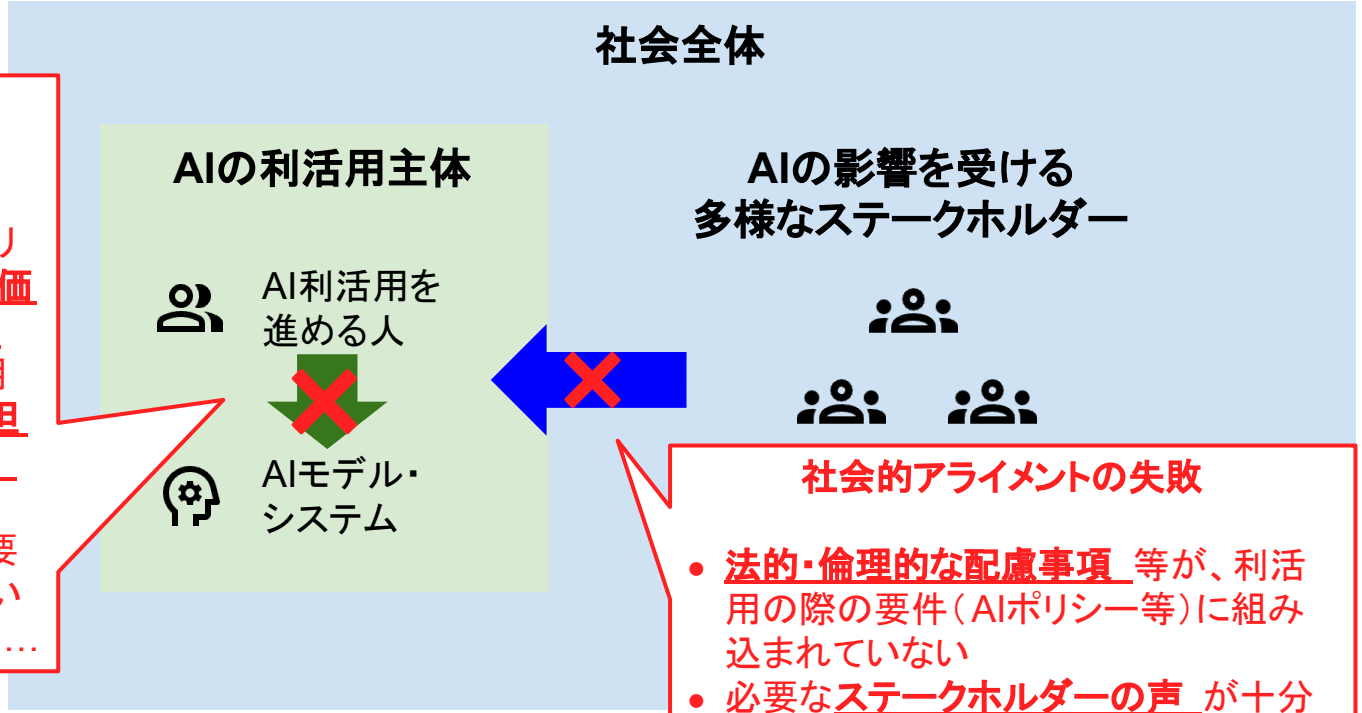
2つのAIアライメント: 社会の要請を踏まえたAI管理が必要

※Korinek, Balwit 2023より作成



AIアライメントの失敗＝技術/コミュニケーションの不全や連携の失敗

※Korinek, Balwit 2023より作成



直接的アライメントの失敗

- プライバシー、セキュリティをはじめとする**諸価値を実装するための技術的な手法**が不明
- 人・AIの間の**業務分担や責任関係の不明確さ**
- AIの不透明性により要件充足が確認できない

……

社会的アライメントの失敗

- **法的・倫理的な配慮事項**等が、利活用の際の要件（AIポリシー等）に組み込まれていない
- 必要な**ステークホルダーの声**が十分にAI利活用過程に届いていない

……

コミュニケーションと技術の連携によるガバナンスの必要性

コミュニケーションによるガバナンス (例)

- 国内外の法制度に沿ったリスクの評価プロセスや内部ポリシーの設計（含・PIA、人権DD等との連携）
- ガバナンスに関与する主体の範囲（委託先など）特定と、契約等の調整
- 情報開示等を通じた透明性確保
- 体制的なアカウンタビリティ確保

……

×

技術によるガバナンス (例)

- セキュアな技術環境の設計
- データの収集・加工における技術的なリスク対策（PETs等）
- さまざまな技術スタック（モデル選定、レッドチーミング等）によるAIモデルのバリデーション・保護

……

AIガバナンスと
プライバシーガバナンス

理念: AIガバナンスの目指す価値とプライバシーは密接に関連

OECD・AI原則	AI原則がプライバシー/データ保護の観点と関わる事項
1.1 包摂的な成長、持続可能な開発と幸福	● **AIの経済的・社会的利益とプライバシー権に対するリスクの間の比較検討** ● 管理者による**データ主体の利益を念頭に置いた行動（データ・スチュワードシップ）**　など
1.2 法の支配、人権、公平性・プライバシーを含む民主的な価値の尊重	● プライバシーやデータ保護に関するAIのリスクを特定・評価・対処すること ● プライバシー領域で重視される価値の考慮: **データ処理の適法性、目的・利用制限の原則、個人データの最小化、プライバシー・バイ・デザイン、データ保管、個人データの正確性、データ主体の権利**など ● AIモデルの正確性のために実施するデータクリーニングや重複処理（名寄せ）といった前処理と、データの品質・正確性に関するプライバシー原則の間の整合性　など
1.3 透明性と説明可能性	● データ主体に対する**透明性と説明可能性**の確保 ● EU・GDPRなどの既存の法的要件と、技術的な動向との間のすり合わせ　など
1.4 頑健性、セキュリティ、安全性	● AI領域におけるリスクベース規制とグローバルなデータ保護規制の双方への目配せが必要。**AIシステムのライフサイクルを通じたリスク対策への、プライバシー原則の統合** ● データセキュリティ（機密性、完全性など）とPETsの間の調整　など
1.5 アカウンタビリティ	● プライバシー・マネジメント・プログラムにAIシステムのライフサイクル手法を取り入れる ● 人権リスクアセスメント、PIA等のリスク評価手法の監督方法のハーモナイゼーション　など

※OECD Artificial Intelligence Papers
“AI, data governance and privacy” (2024.06.26)をベースに一部記載を省略

実務: AIライフサイクルを通じて、個人データ・プライバシー保護が必要

AIライフサイクルの主なフェーズ

データ収集	モデル構築（学習）	システム実装・運用	出力の活用
個人データの取得・第三者提供 センシティブデータの扱い	モデルのバイアス対策 データ引き出し攻撃等へのセキュリティ対策	運用中に収集するデータの取り扱い セキュリティ対策ガードレール	AIの出力を活用する際の説明や用途の限定 ユーザコミュニケーション

個人データ・プライバシーの論点（代表例）

PETs活用（匿名化、秘密計算等）

AIガバナンスとプライバシーガバナンスは不可分

企業のAIガバナンス実装状況と課題

AIGA会員各企業がAIガバナンスの取組の成熟度を自己診断するツール

AIガバナンスナビ＝AI事業者のAIガバナンス構築の取組の成熟度チェッカー

AIGAの会員企業が、政策・標準や他の会員企業の取組状況をベンチマークとして、自社の組織としてのAIガバナンス構築の取組の成熟度を自己診断し、自社の取組の強み・弱みを把握できるようにするツール

実践のスタンダード作り

- AIGA会員企業にとって、AI事業者としてAIガバナンス構築に必要な取組事項を把握する上での実践的なガイダンスを提供
- 回答集計・研究会等を通じて最新のプラクティスや課題意識を反映

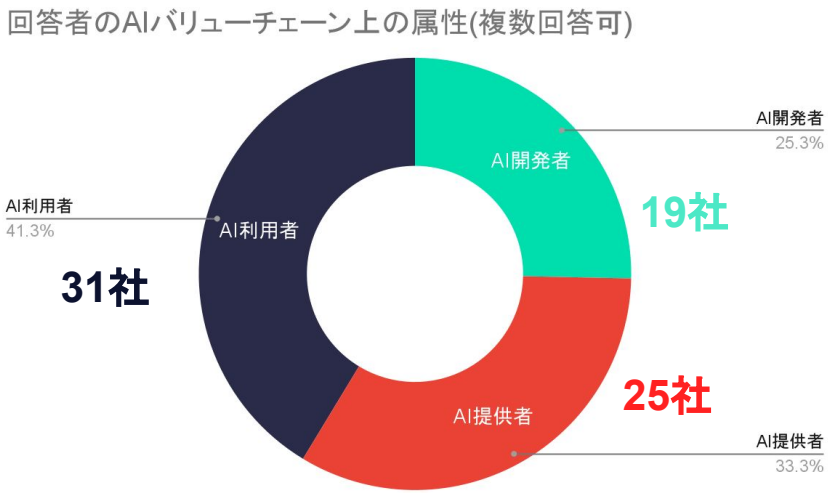
協会活動のペースメーカー

- AIGA会員で定期的に自己診断を実施し、諸産業全体としての進捗度を把握
- 項目別に自己診断の結果を分析し、全体の成熟度向上のためにフォーカスすべき議論・取組事項を特定

政策・標準との接続

- AI事業者ガイドライン等への対応関係を明確にし、企業の取組の政策への準拠を支援
- 対外的な観点で、AIGA会員全般の政策への対応状況を把握・発信

AIGA正会員から、各業界を代表する企業が自己診断に参加（4月実施）

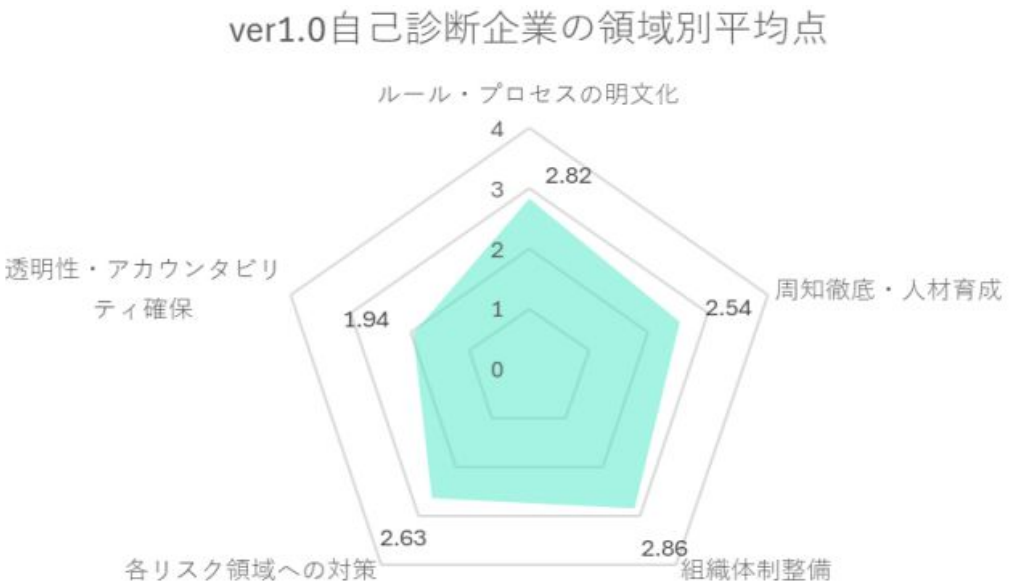


- 開発・提供・利用それぞれの立場から**計35の回答**が集まっている
- 業界としても、**IT・通信・保険・証券・銀行・インフラ・製造など**多様
- **34社/35社が生成AIを利用**したユースケースを保有している
- 出力結果が人によるチェック・修正を経ずに社外に表示・提供されるAIのユースケースがある企業は10社

一般社団法人
AIガバナンス協会
AI Governance Association

ルール整備・組織づくりが先行。一方、個別リスク対応(特に技術的対策)や透明性確保が課題

全体平均: 2.51点



- 「ルール・プロセスの明文化」「組織体制整備」については平均が2.8を超え、**社内ルールの規定化・体制の設置などの取組の進捗**が窺える
- **「透明性・アカウンタビリティ確保」についてはスコアが低く**、技術的対策や監査体制の外部連携の取組余地が大きい
- AI・生成AIの継続的な管理、ハイリスクな業務での活用は道半ばであり、**ユースケースの成熟とともに各企業に求められる対策が具体化され、低得点領域の取組も進む**と考えられる

一般社団法人 AIガバナンス協会 AI Governance Association

開発者、提供者、利用者の順に得点は低くなっている

全体平均: 2.51点

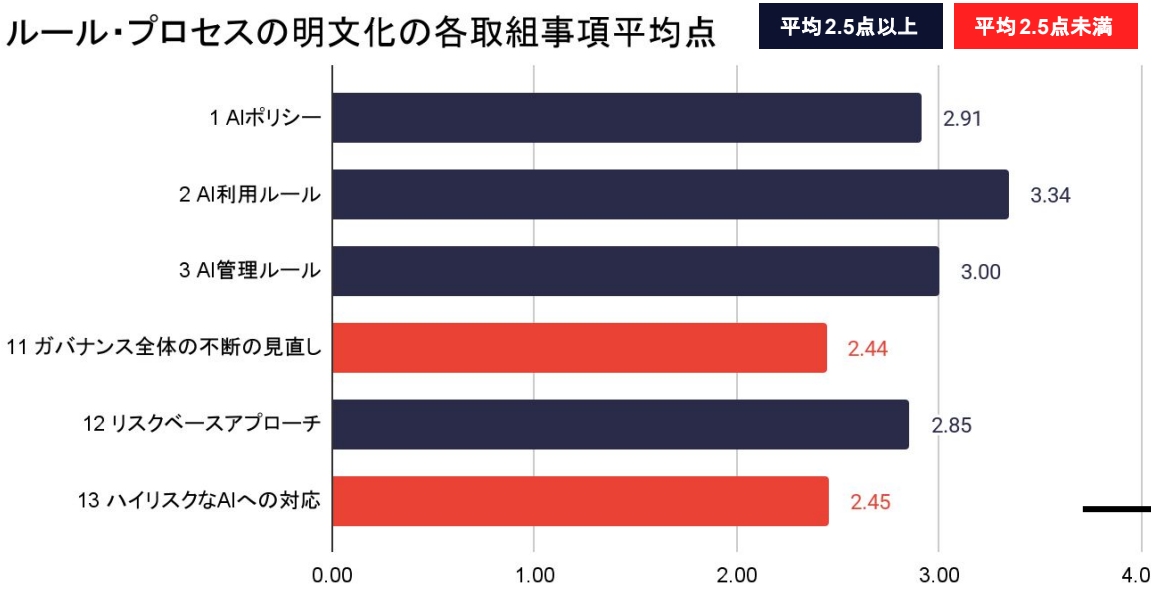
開発者平均: 2.71点(n=19)

提供者平均: 2.58点(n=25)

利用者平均: 2.44点(n=31)

- **AI利用者のみの立場**の企業の平均点は**2.2点**、**提供者・利用者の立場**の企業の平均は**2.1点**と**全体の傾向と比較して低い**評点となった
- **利用者属性（非開発者）の企業は「各リスク領域への対策」が顕著に低い**傾向あり。**他社のSaaSのAI製品を調達して利用**する場合に、**ベンダ側の安全性確保策にリスク対策を委ね、自社において対策を実装する**ことに限界があることによるものと推測される

AIの導入に向けたルール作りは進んでいるが、リスクが高いユースケースの開拓や、技術・制度動向を踏まえた見直し・更新は今後進展の余地あり

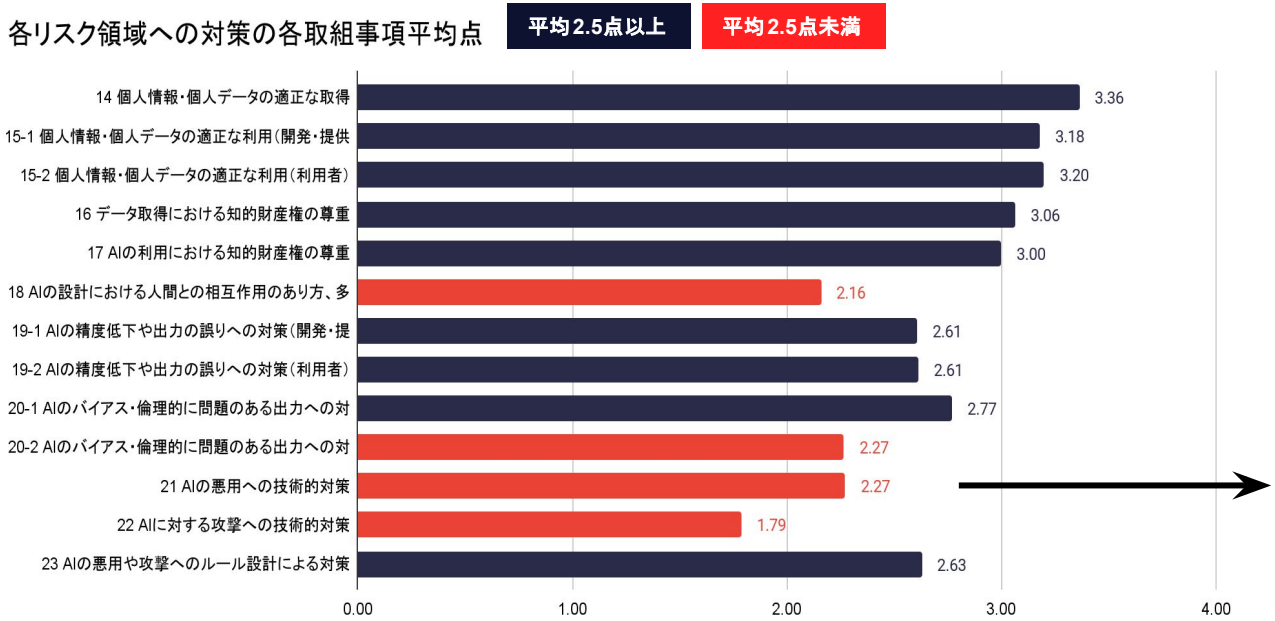


取組例（匿名）

“ハイリスクと判断された場合の対処策定は、AIの有無に関わらず必須としているが、ハイリスクと判断されたAIモデル・サービスに特化した対応は未定である。”

個人情報や著作権等法に規定されている項目については一定の整理・対策が進んでいるが、AIの悪用・攻撃への対策は利用者を中心に道半ば

各リスク領域への対策の各取組事項平均点

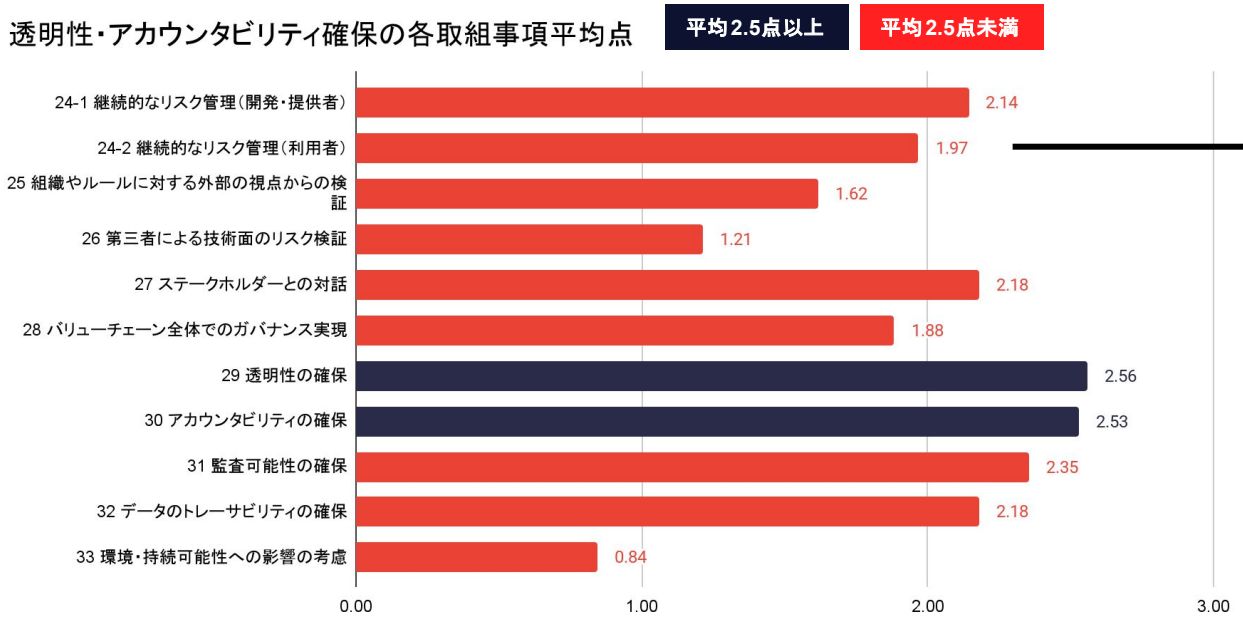


取組例（匿名）

“問題ある入出力についてフィルタリングする機能が実装されているモデル・サービスを選定し・実装しているが、現状ではサードパーティ製のファイアウォールやフィルタリングツールの導入には至っていない。”

一般社団法人 AIガバナンス協会 AI Governance Association

第三者検証などの構造的な信頼確保策や、AI利用者にとってのバリューチェーン全体を意識したガバナンスの実現に課題感



取組例（匿名）

“運用開始後も顧客および自社内でログを確認できるプロセスとなっており、顧客と定期的に運用状況を議論する場をもっている。”

“現時点では、AIシステムの企画、稼働前段階のリスク評価にとどまっている。運用開始後の定期的なリスク検証と対策検討については、運用を検討中。”

サマリ

- AIをめぐり、世界の政策潮流も激変し技術的な不確実性も高い環境。投資家もAIガバナンスに関する要求を強めており、企業・組織には**AIガバナンス構築**が求められている

- 対処すべきAIリスクの類型として**セーフティリスク・セキュリティリスク**が存在。社会実装の現場でインシデントも発生し始めており、**ユースケースごとのリスクを見極めて**対処することが必要

- AIガバナンスの文脈で、**プライバシーも重要アジェンダとして国際的に認識**されている。実際にデータ収集～エンドユーザによる活用に至るまで、**AIバリューチェーン上の様々な箇所でプライバシーをめぐる論点が存在**する。**個人への利益とリスク**をしっかり見極めて活用を進めていくことが必要

- 日本企業の現状をみると、組織作りは形式上は完了している企業が多いが**技術的なリスク対策や透明性確保にはまだまだ課題**がある。**様々な技術スタックを通じてデータ・ AIの保護を進め、利活用の地盤をつくっていく**べき

一般社団法人
AIガバナンス協会
AI Governance Association