AI社会の歩き方

2024-10-17

独立行政法人情報処理推進機構（IPA）
デジタル基盤センター長
AIセーフティ・インスティテュート　（AISI）
副所長・事務局長

平本　健二


実は崖を
登ってないか？


しかも
すごい速さで？

日々進歩するAIをうまく使いこなしながら、

しかも安全に安心して使える環境整備が必要です。

AI（Artificial Intelligence）とは

人工知能とは

- 人工知能学会記事の「教養知識としてのAI」の中で以下のように説明されている。

実は，「人工知能とは何か」という問いに対する答えは，単純ではない．人工知能の専門家の間でも，大きな議論があり，それだけで*1* 冊の本となってしまうほど，見解の異なるものである．そのような中で，共通する部分を引き出して，一言でまとめると，「人間と同じ知的作業をする機械を工学的に実現する技術」といえるだろう．

教養知識としてのAI〔第1回〕AIってなに？
https://www.ai-gakkai.or.jp/resource/ai_comics/comic_no1/

AIの3つのステージ

	Artificial Narrow Intelligence (ANI)	Artificial General Intelligence (AGI)	Artificial Super Intelligence (ASI)
Description	Programmed to perform specific tasks, such as facial recognition and voice assistance, to criteria set by humans	Programmed to perform more than one task—and learns from each task with the general intelligence of a human	Surpasses human intelligence. Outperforms humans in every task with its ability to think on its own
Characteristics	Lacks consciousness and self-awareness	Shows human-like awareness and expression	Displays understanding and beliefs of its own
	Outperforms humans in some tasks	**Competes with humans in all tasks**	**Outperforms humans in everything**

“WHITE PAPER Artificial General Intelligence”, BCG （September 2024）
https://media-publications.bcg.com/BCG-Artificial-General-Intelligence-Whitepaper.pdf

Open AIの定義するAGIの5段階スケール

Level 1the kinds of conversational chatbots available today

Level 2 tackles human-level problem solving

Level 3 is all about agents, where systems can take actions

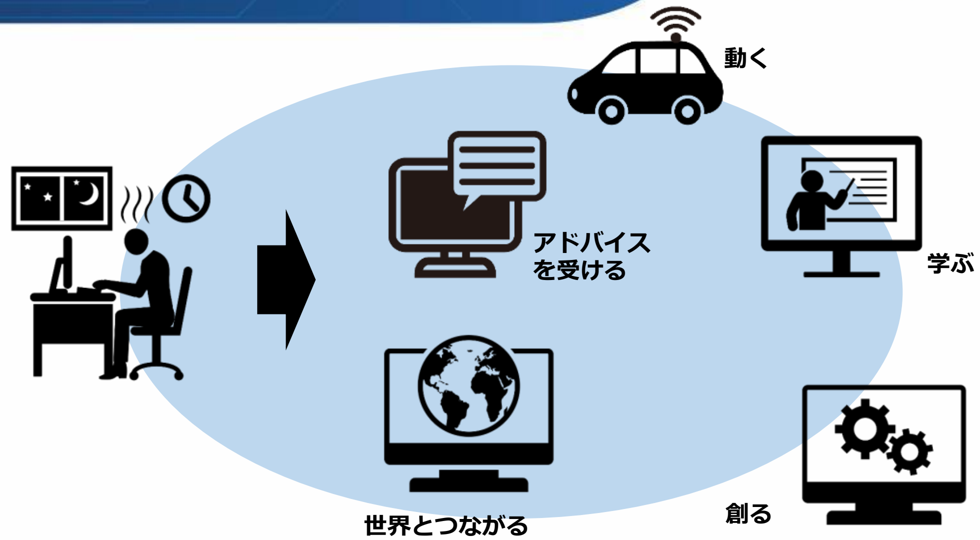
Level 4 moves to AI that can aid in invention

Level 5 to AI that can do the work of an entire organization

”OpenAI Scale Ranks Progress Toward ‘Human-Level’ Problem Solving” ,Bloomberg （2024-07-11）
https://www.bloomberg.com/news/articles/2024-07-11/openai-sets-levels-to-track-progress-toward-superintelligent-ai?sref=b0SdE1lu

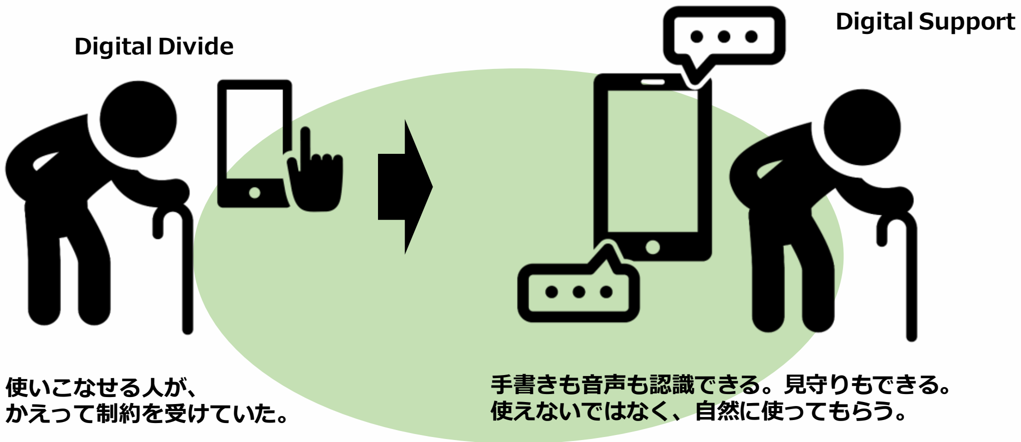
デジタル技術は、

1.伸ばせる力、人の可能性を最大化できる。



2.人の仕事を奪うものではない。

3.誰にも優しい。

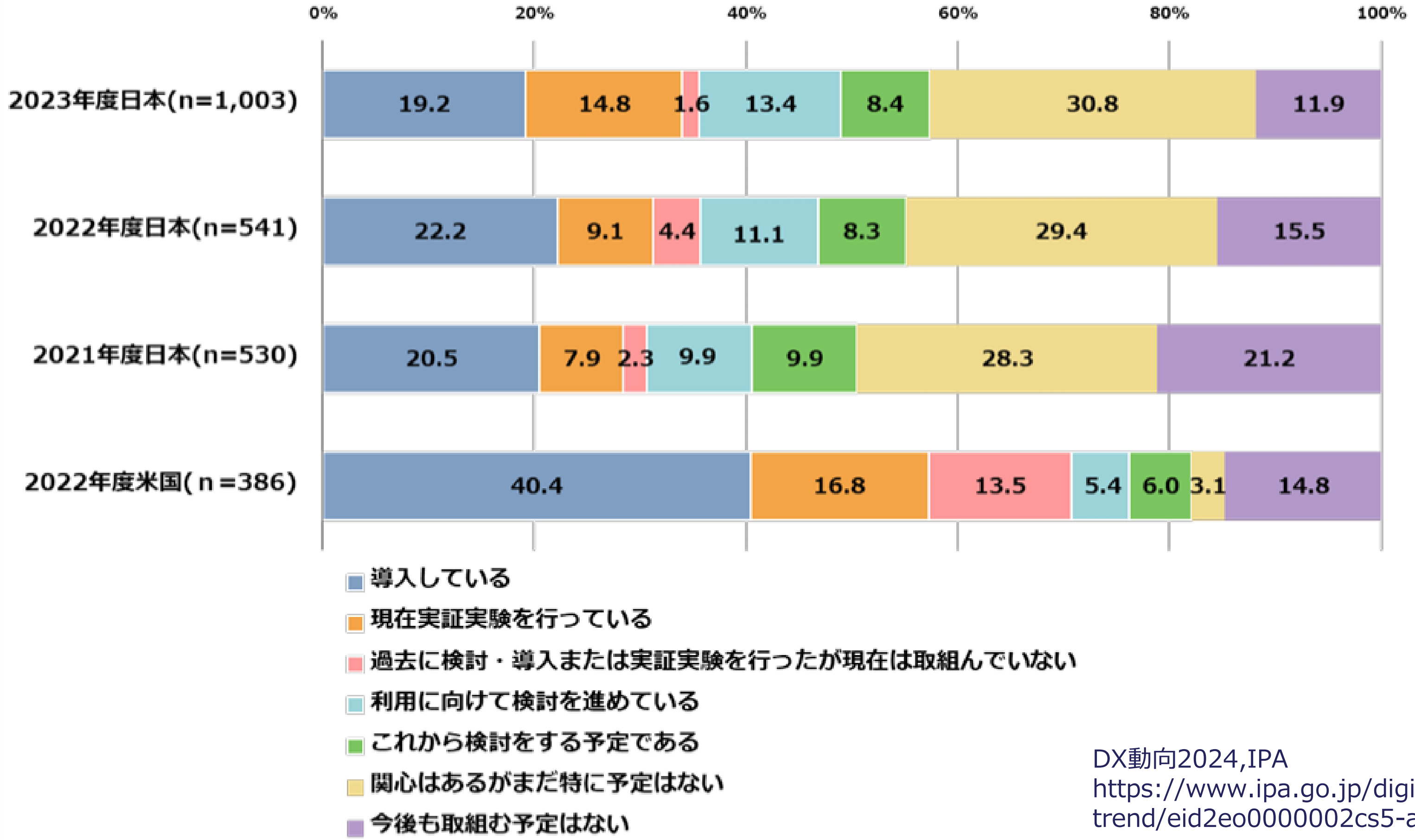


AIの導入状況

AIの導入状況（経年変化および米国との比較）

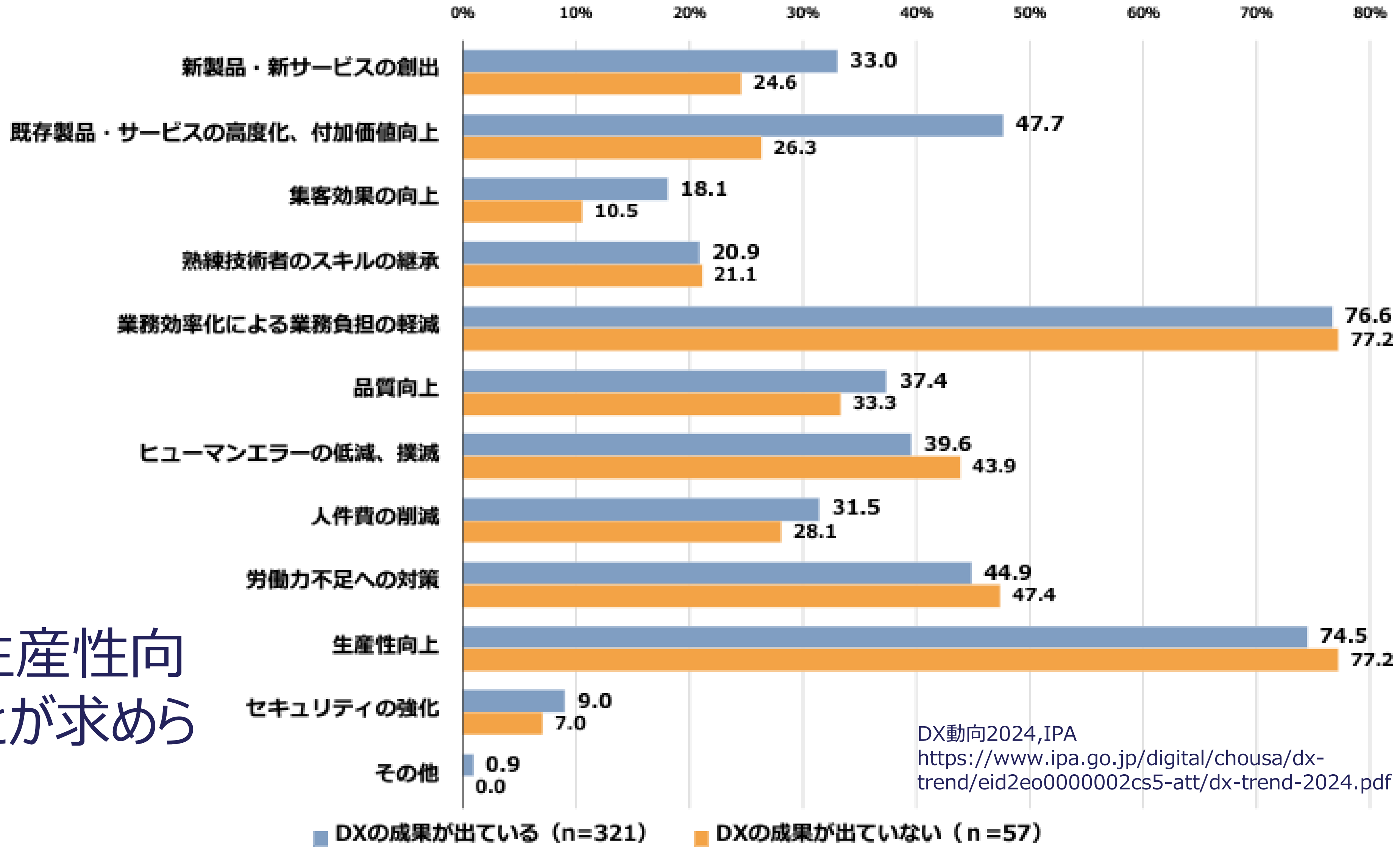
IPA

- 日本の「導入している」の回答割合は19.2%であり、2022年度水準とあまり変わらず、同40.4%である米国とは乖離が大きい。



DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

AIの導入目的（DX成果別）

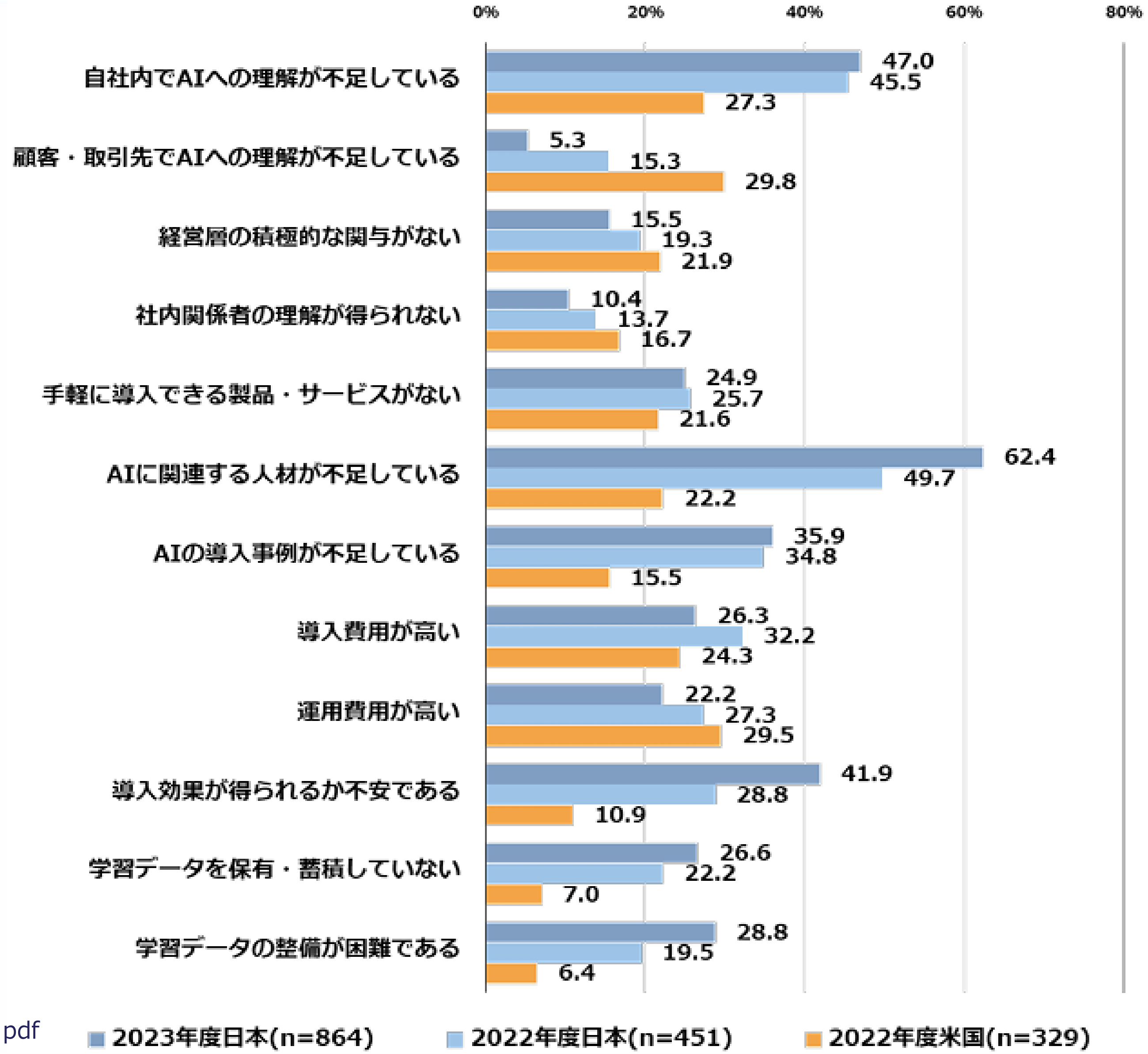


DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

- 業務効率化や生産性向上に寄与することが求められている

AIの導入課題（経年変化および米国との比較）

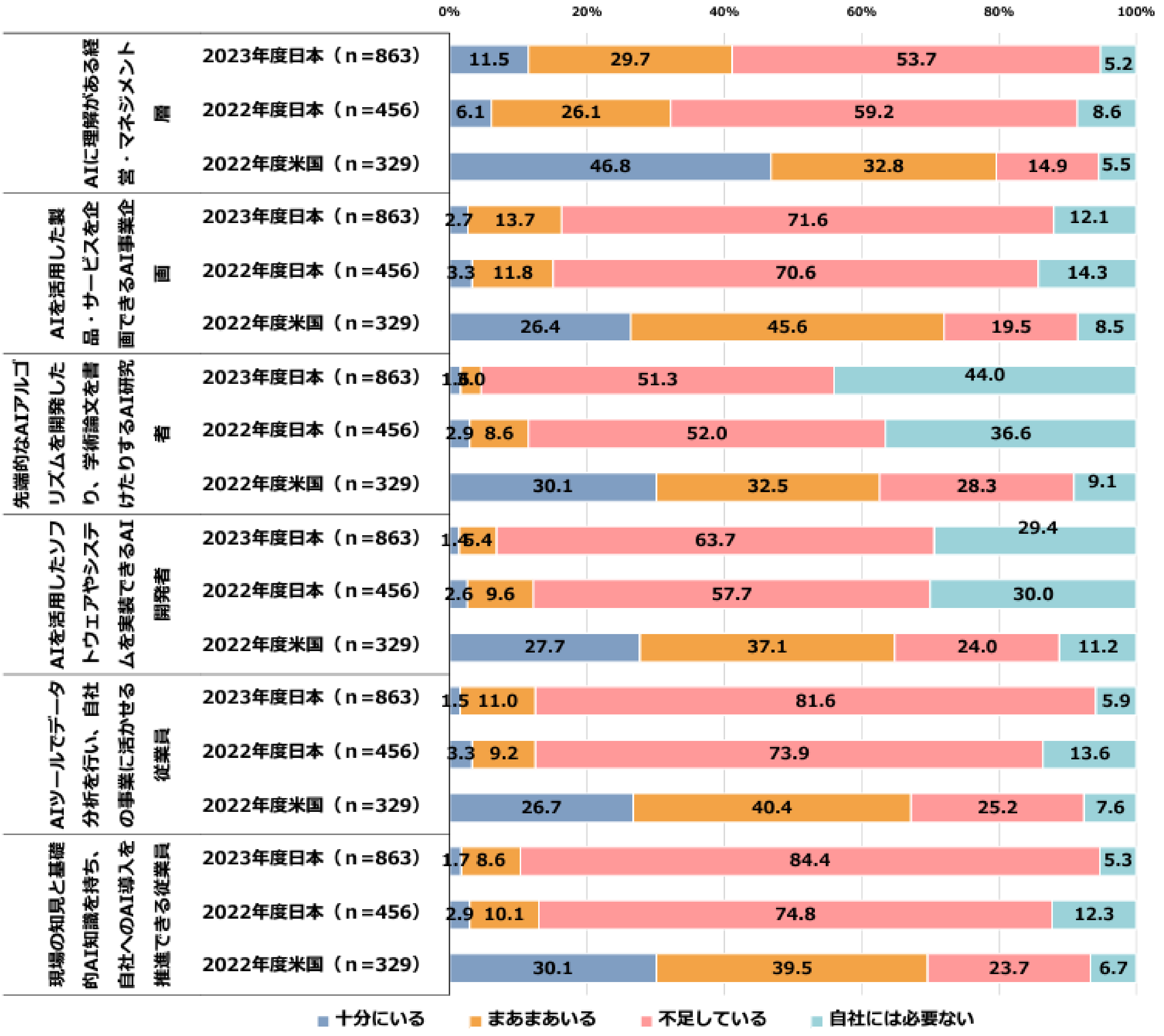
- AIに対する理解や人材が不足していることが課題



DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

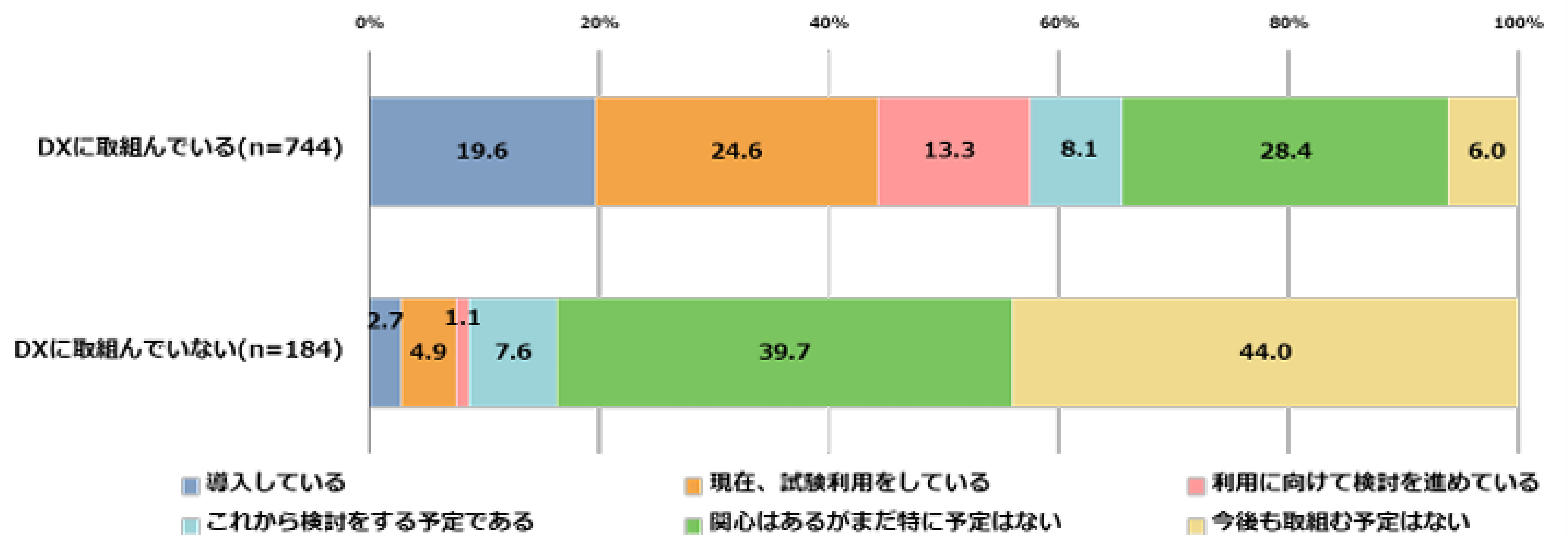
AI人材の充足度
（経年および米国との比較）

「AIを活用した製品・サービスを企画できるAI事業企画」「AIツールでデータ分析を行い、自社の事業に活かせる従業員」「現場の知見と基礎的AI知識を持ち、自社へのAI導入を推進できる従業員」において、「不足している」の回答割合が高い。



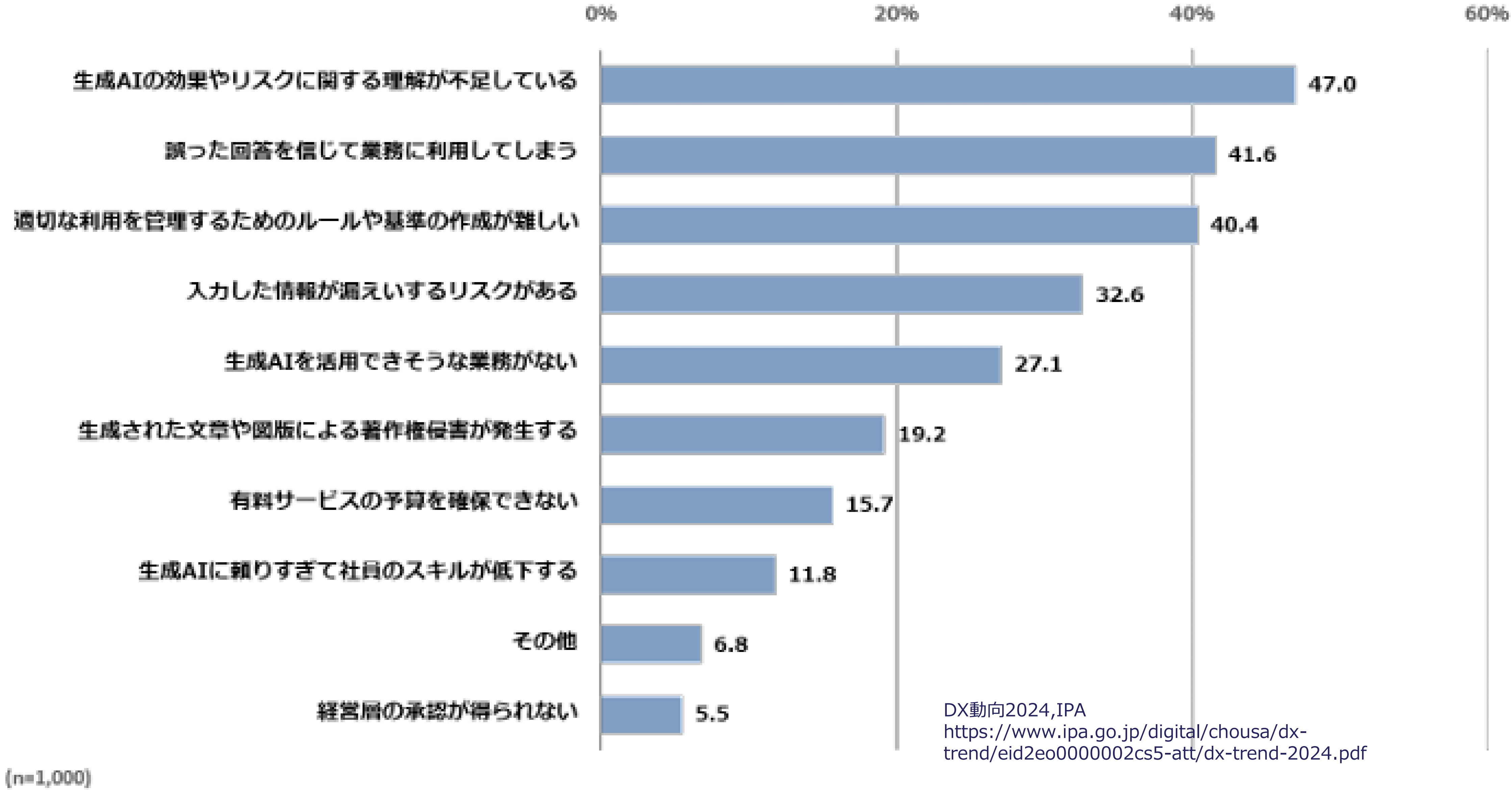
DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

生成AIの導入状況（DX取組状況別）



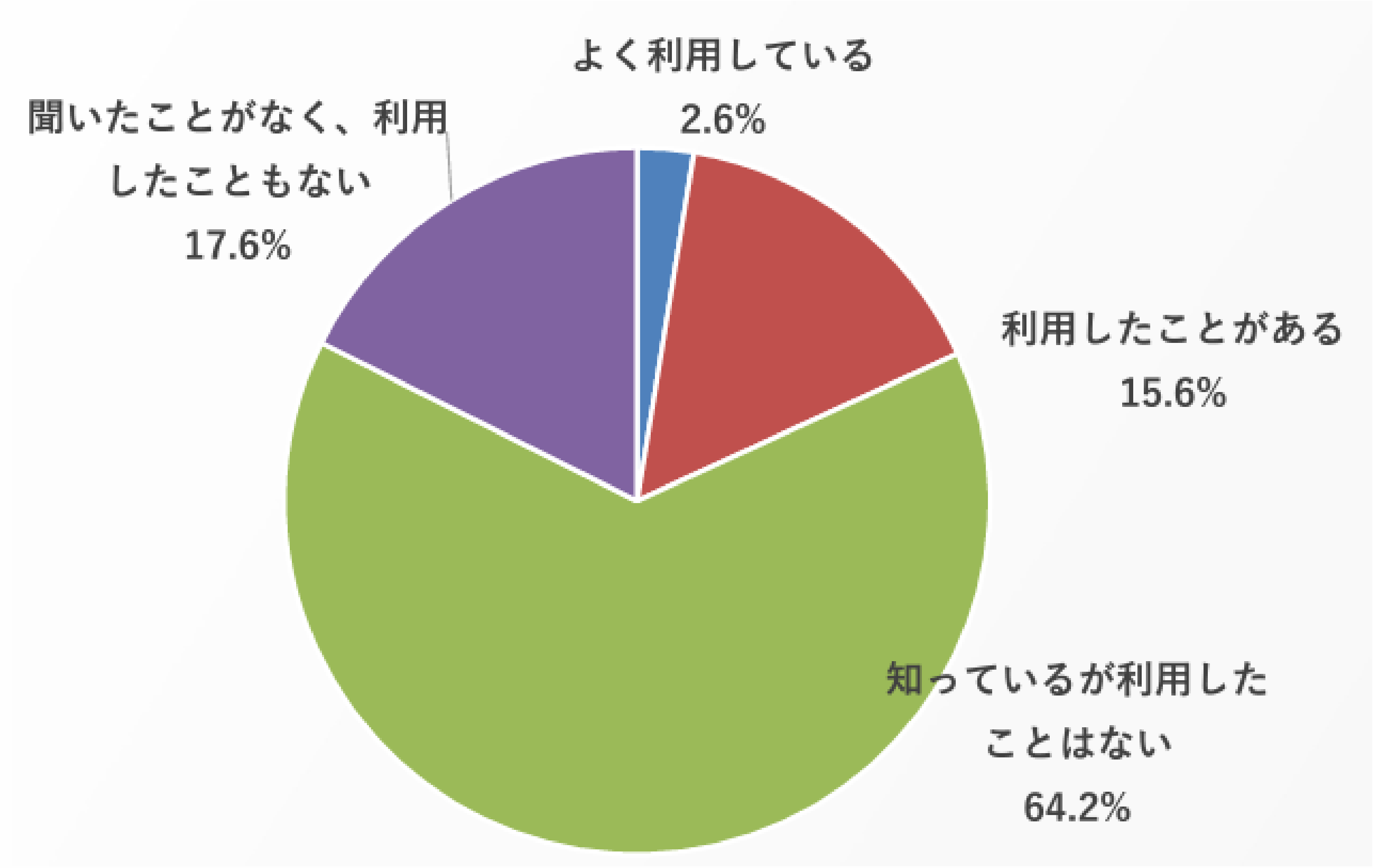
DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

生成AI導入の課題



DX動向2024,IPA
https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf

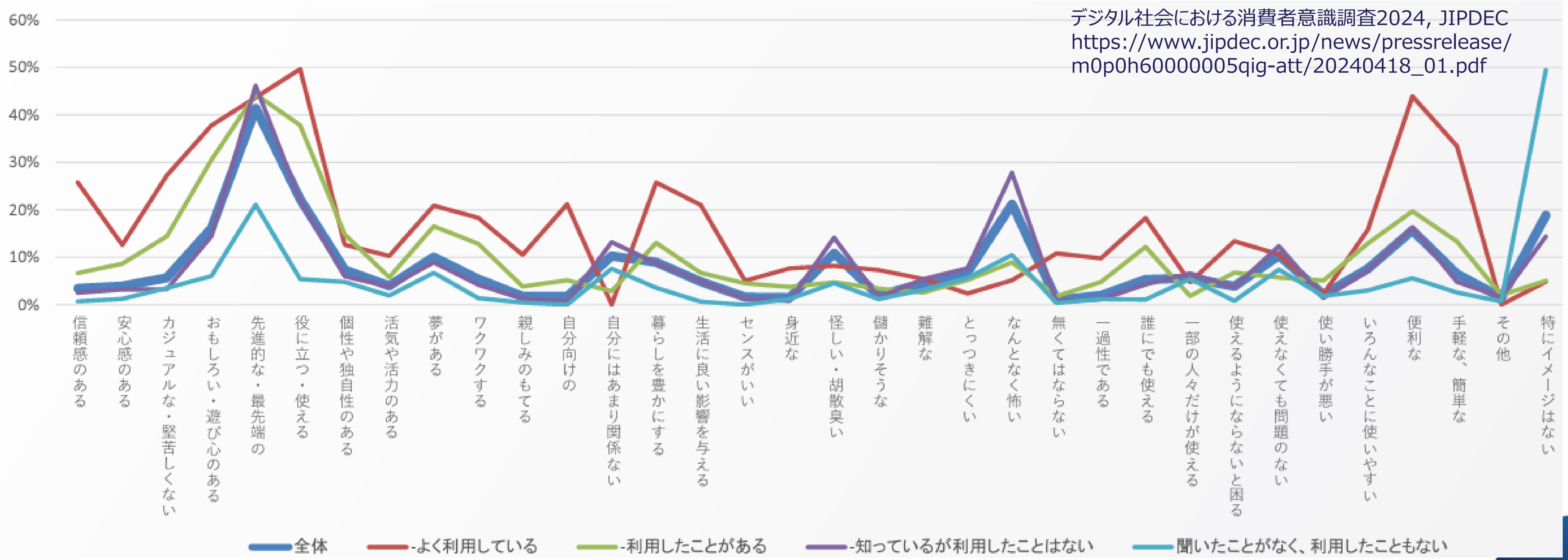
プライベートでの「生成AI」の利用



デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

生成AIのイメージ

- あなたは「生成AI」について、どのようなイメージを持っていますか。（複数回答）

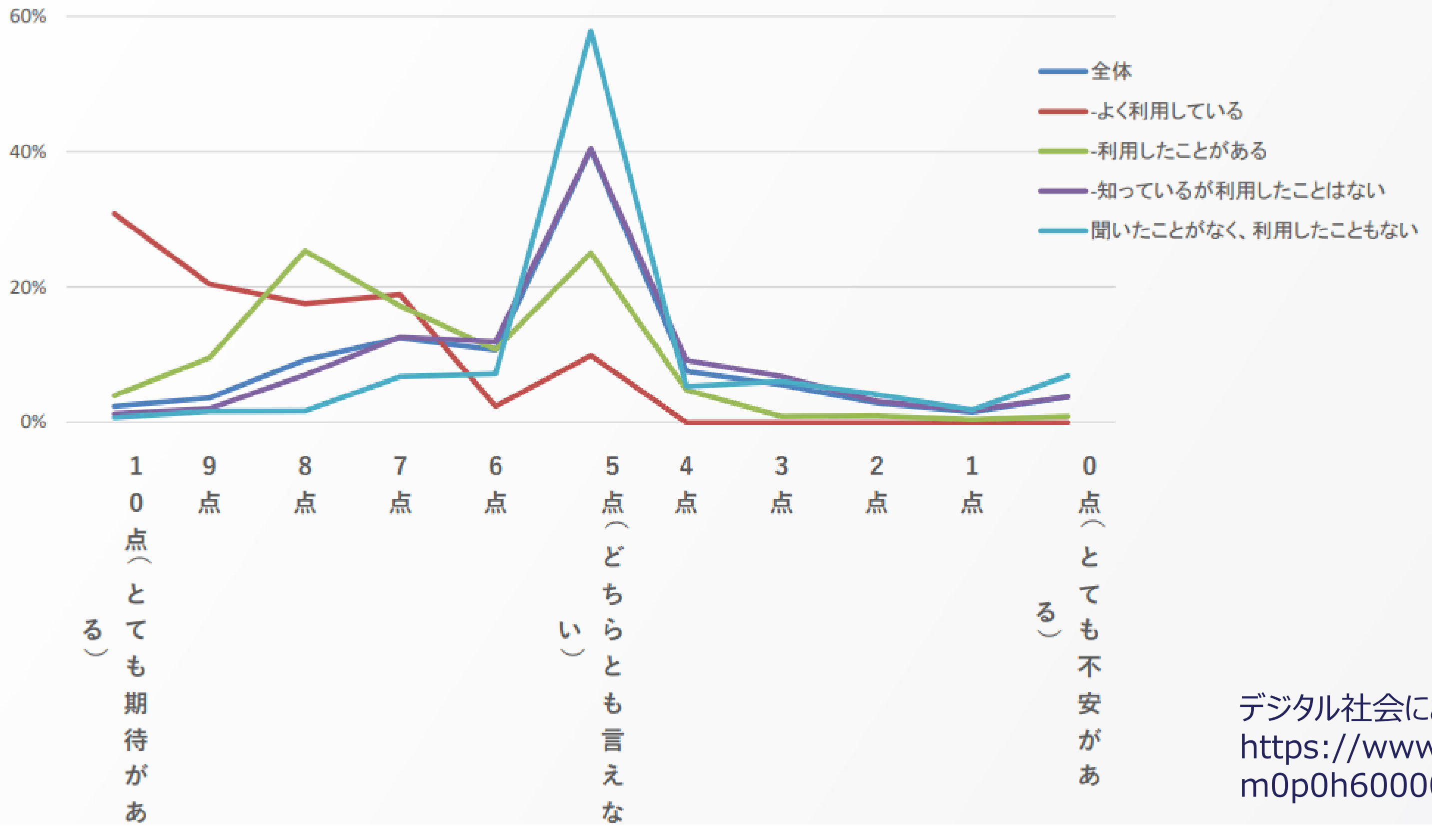


デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

個人の受け止め

- 期待に比べて不安を感じている人は少ない。

あなたは、「生成AI（人工知能）」を使ったサービスの普及に対して、「期待」と「不安」のどちらのイメージがありますか。10点（とても期待がある）～0点（とても不安がある）の間で、お気持ちに近いものを1つお選びください。(n=1449)



【傾向】

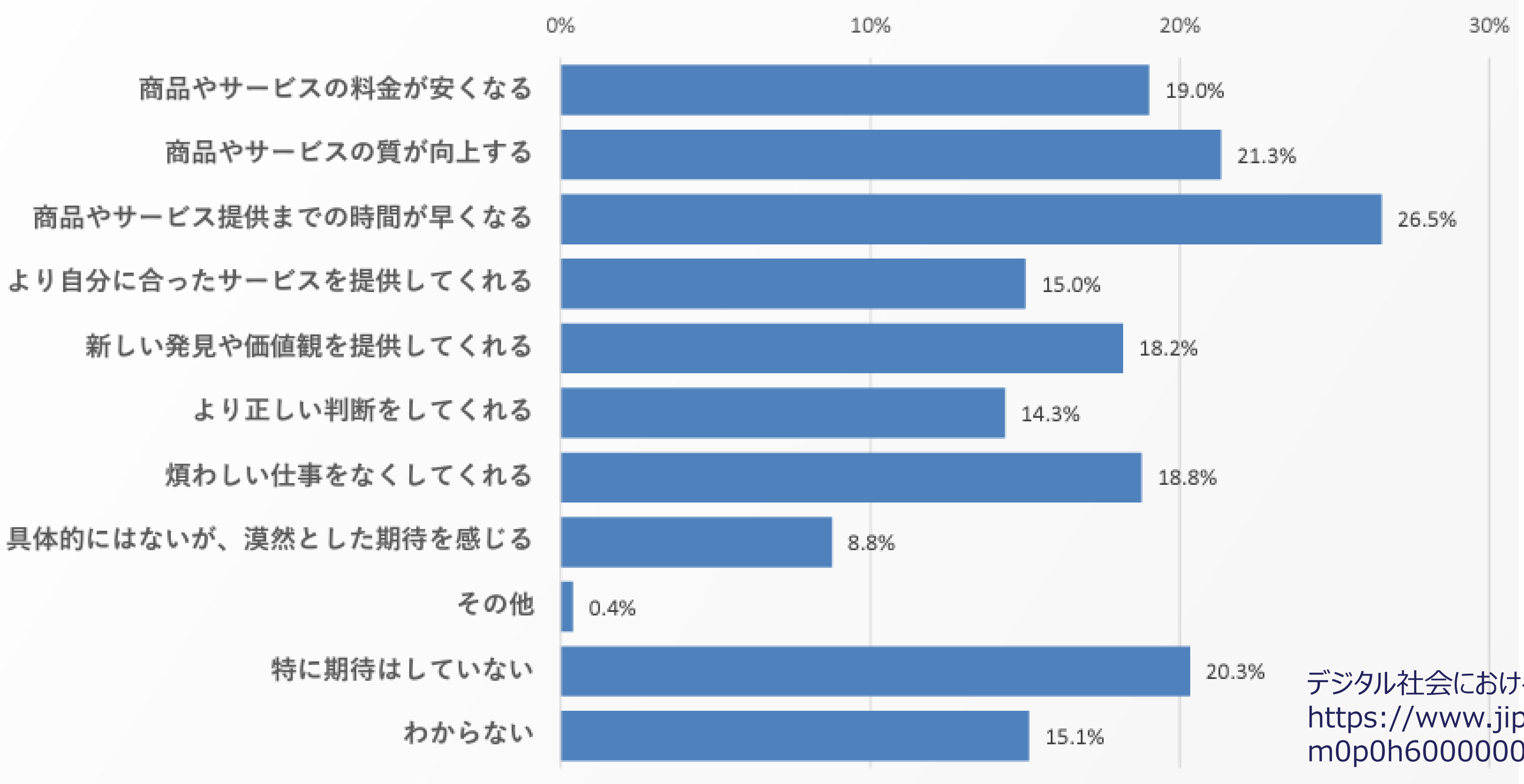
性年代別：
若年層男性の期待度が高い

生成AI利用経験：
期待に関しては差があるものの、利用の有無に関わらず不安イメージは少ない。

デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

個人の受け止め（期待要素）

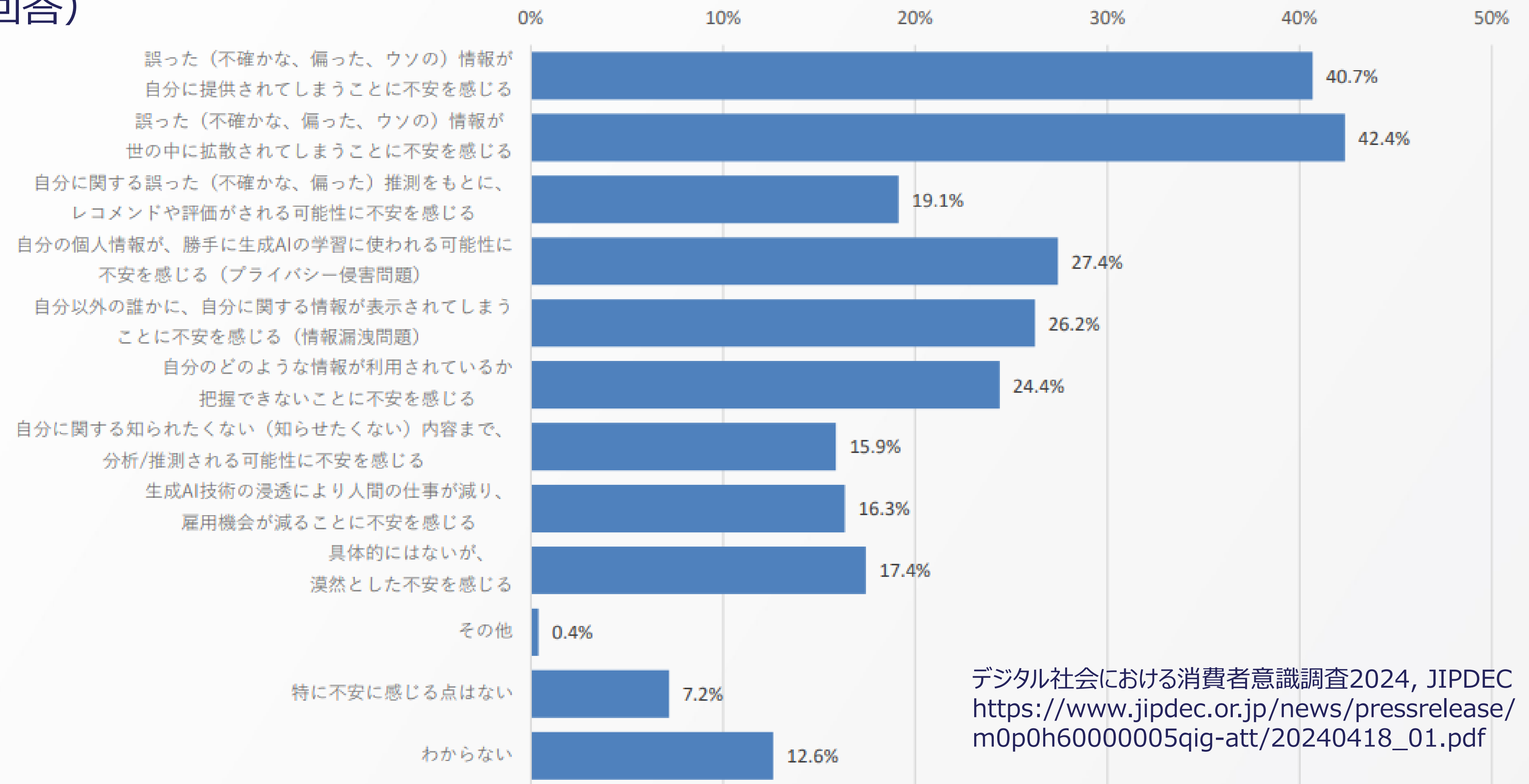
- あなたは、企業が「生成AI」を用いたサービスを提供することに、期待を感じますか？（複数回答）



デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

個人の受け止め（不安要素）

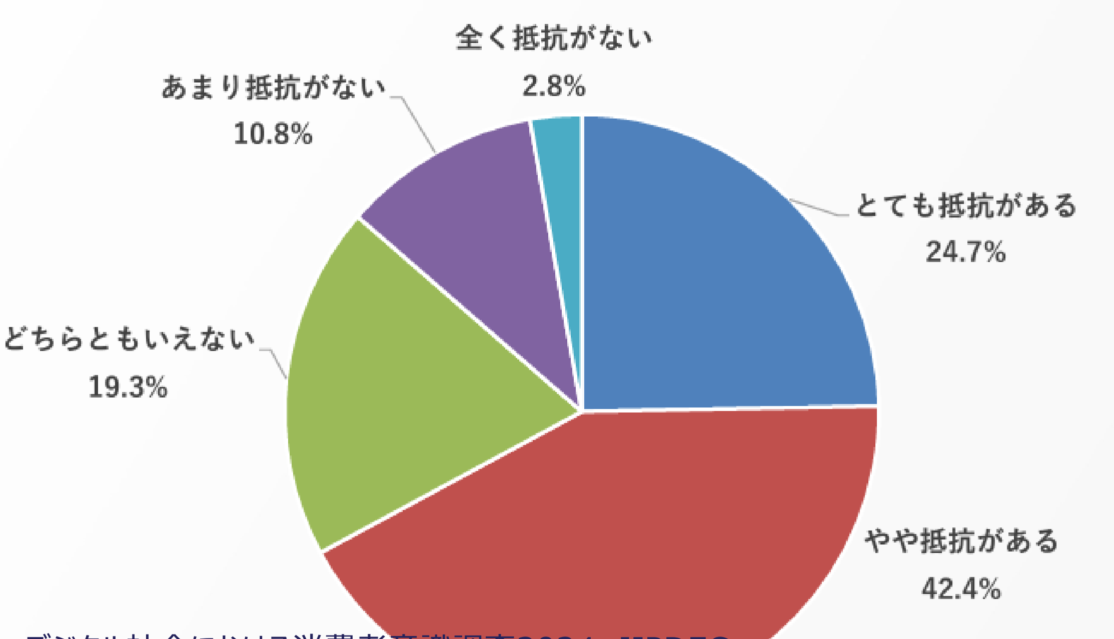
- あなたは、企業が「生成AI」を用いたサービスを提供することに、不安を感じますか？（複数回答）



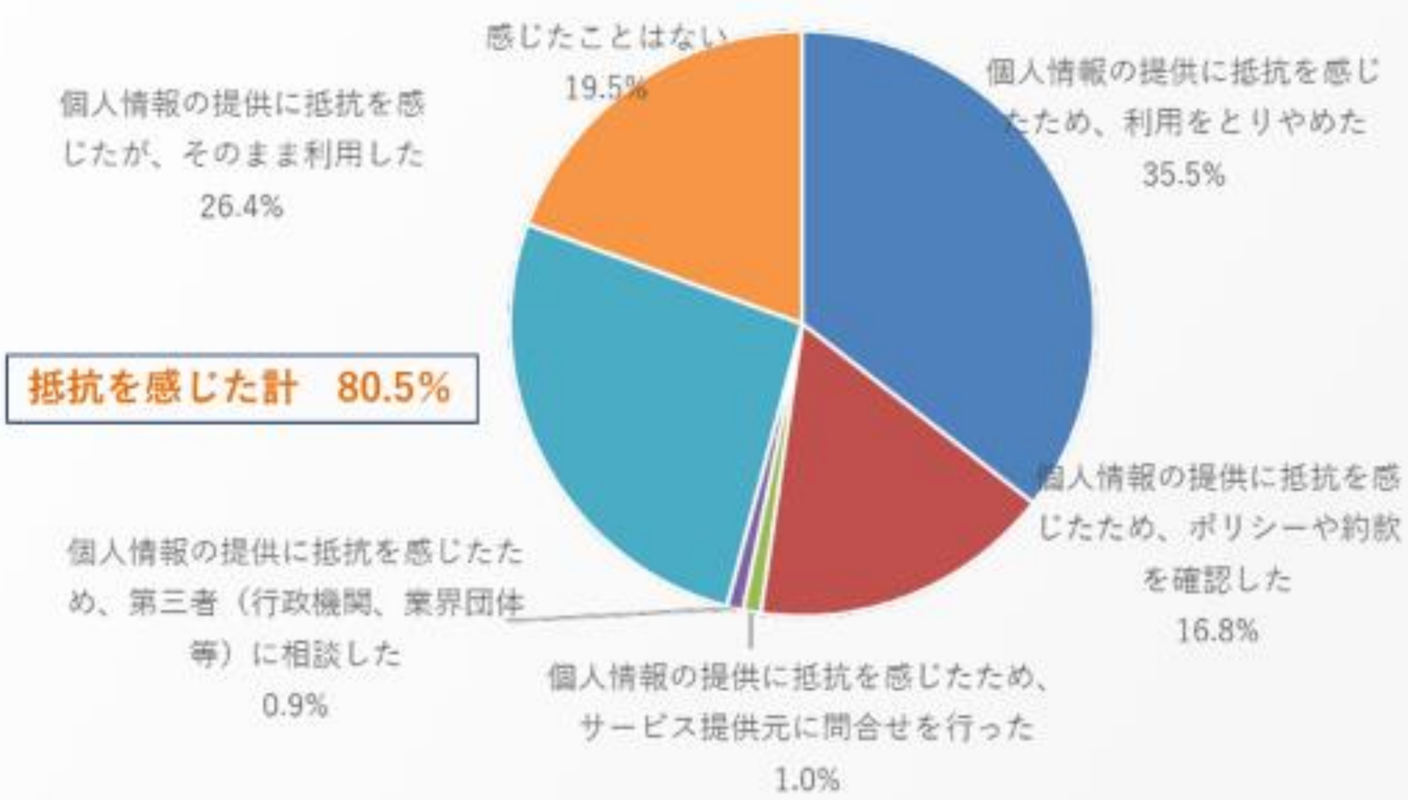
デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

個人情報やプライバシー情報に関する意識

- あなたは、Webサービスやアプリケーションを利用する際に、自分や家族の個人情報やプライバシーに関する情報を提供することに、どの程度抵抗がありますか。



- 抵抗を感じた際にあなたがとった行動として最も多いものを１つお選びください。



デジタル社会における消費者意識調査2024, JIPDEC
https://www.jipdec.or.jp/news/pressrelease/m0p0h60000005qig-att/20240418_01.pdf

AIをうまく使おう

AIを上手に使うためのポイント

- 目的を明確にする
  - 何を達成したいのか、具体的な目標を設定することで、AIの適用範囲や方向性が明確になります。
- データの質を確保する
  - AIはデータに基づいて学習します。正確で質の高いデータを使用することが重要です。
- 適切なツールを選ぶ
  - 目的に応じて、適切なAIツールやプラットフォームを選びましょう。これにより、効率的に問題を解決できます。
- AIに丸投げせず、人間が関与する
  - AIが提供する結果や提案はチェックし、人間の判断を加えることで、より正確な意思決定が可能になります。
- 継続的な学習と適応
  - AI技術は急速に進化しています。最新の技術やトレンドを学び、システムを適応させることが重要です。

AIに関する情報の収集

- AIに関しては政府や企業から多くのガイドやレポートが公表されています。国内のガイドに閉じず、海外の情報も積極的に活用することが重要です。
  - 動向
    - AISI https://aisi.go.jp/
    - OECD.ai https://oecd.ai/en/
    - 内閣府AI戦略会議 https://www8.cao.go.jp/cstp/ai/ai_senryaku/ai_senryaku.html
    - 国連AI Advisory Body https://www.un.org/ai-advisory-body
  - 技術
    - YouTube等のネット上のリソース
- 特に事例は最先端技術の使い方を知るのに重要です。世界の事例から多くのことを学ぶことができます。

こんなことには気を付けよう

AIに関するリスクを知っておくことが重要

- AIのリスクを知っていることで、リスクを回避したり軽減しやすくなります。また、リスクをわかっているからこそ、リスクを避けてイノベーションを推進することができます。

- リスクには、第三者からの攻撃等の悪意のリスクだけでなく、AIが誤回答をする場合や、誤った使い方で問題を起こす誤使用等の悪意のないリスクがあることに注意が必要です。

AIに関して想定されるリスク

- AIには、Physical, Social, Economical, Psychologicalなリスクがある。

	AI事業者ガイドライン　共通の指針	主なリスク
1) 人間中心	① 人間の尊厳及び個人の自律 ② AIによる意思決定・感情の操作等への留意 ③ 偽情報等への対策 ④ 多様性・包摂性の確保 ⑤ 利用者支援 ⑥ 持続可能性の確保	・人間の尊厳及び個人の自律を損なうリスク （プロファイリング時の配慮の必要性等） ・AIにより意思決定・感情の操作をされてしまうリスク ・偽情報などのリスク ・多様性や包摂性が確保されないリスク ・地球環境への影響のリスク
2) 安全性	① 人間の生命・身体・財産、精神及び環境への配慮 ② 適正利用 ③ 適正学習	・動作が止まる、低下するリスク ・意図しない動作のリスク ・ステークフォルダがリスクを知らないリスク ・目的外に利用してしまうリスク ・学習データに十分な品質がないリスク ・学習データのコンプライアンスリスク
3) 公平性	① AIモデルの各構成技術に含まれるバイアスへの配慮 ② 人間の判断の介在	・バイアスによる公平性を損なうリスク ・潜在的なバイアスが発生するリスク ・人間の介在が不足するリスク ・バイアスの評価プロセスが不十分なリスク

AIに関して想定されるリスク

	AI事業者ガイドライン　共通の指針	主なリスク
4) プライバシー保護	① AIシステム・サービス全般におけるプライバシーの保護	・プライバシーを侵害するリスク
5) セキュリティ確保	① AIシステム・サービスに影響するセキュリティ対策 ② 最新動向への留意	・不正操作のリスク ・AIシステム自体へのセキュリティ侵害へのリスク ・不正データが使われるリスク
6) 透明性	① 検証可能性の確保 ② 関連するステークホルダーへの情報提供 ③ 合理的かつ誠実な対応 ④ 関連するステークホルダーへの説明可能性・解釈可能性の向上	・検証ができないリスク ・ステークホルダーに十分な情報提供がされないリスク ・合理的でない情報提供を求められるリスク
7) アカウンタビリティ	① トレーサビリティの向上 ② 「共通の指針」の対応状況の説明 ③ 責任者の明示 ④ 関係者間の責任の分配 ⑤ ステークホルダーへの具体的な対応 ⑥ 文書化	・トレーサビリティ情報が入手できないリスク ・共通の指針への対応状況が報告されないリスク ・責任が明確にならないリスク ・ステークホルダーと適切なコミュニケーションが取れないリスク ・各種情報をドキュメンテーションできていないリスク

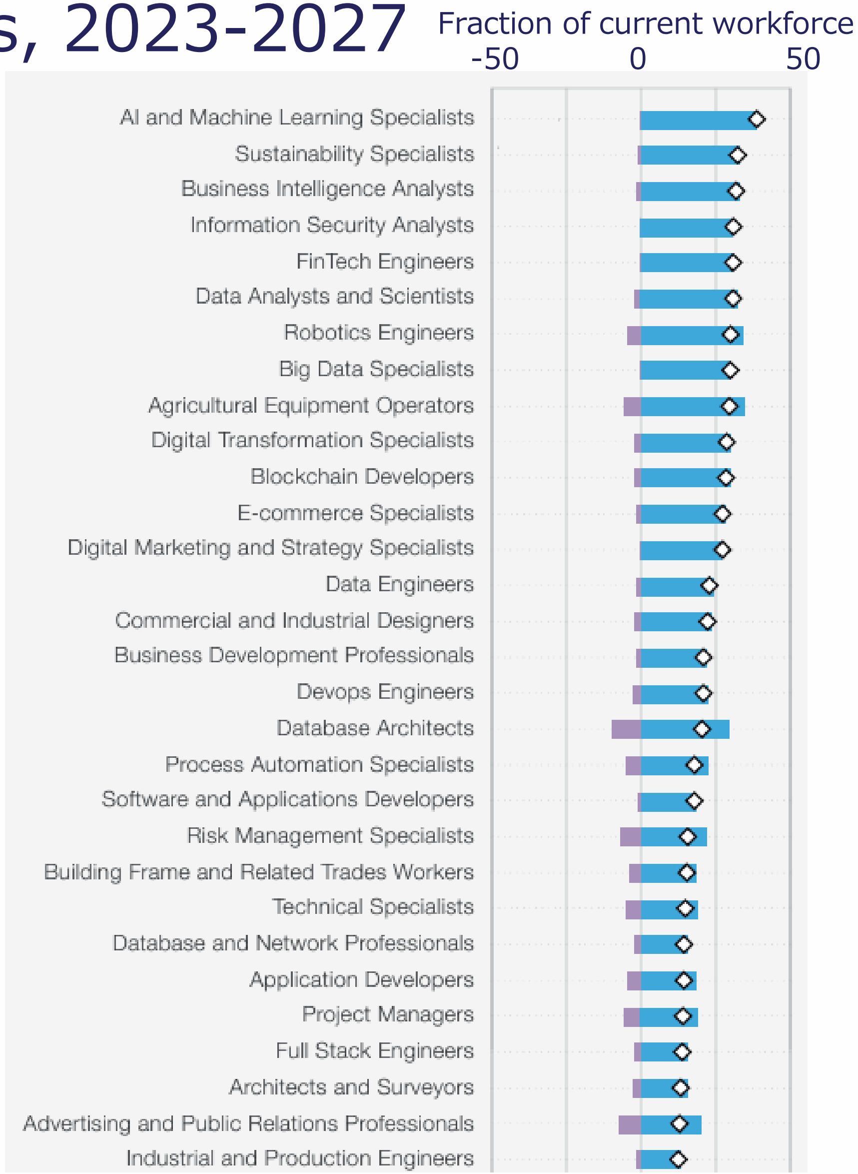
AIに関して想定されるリスク

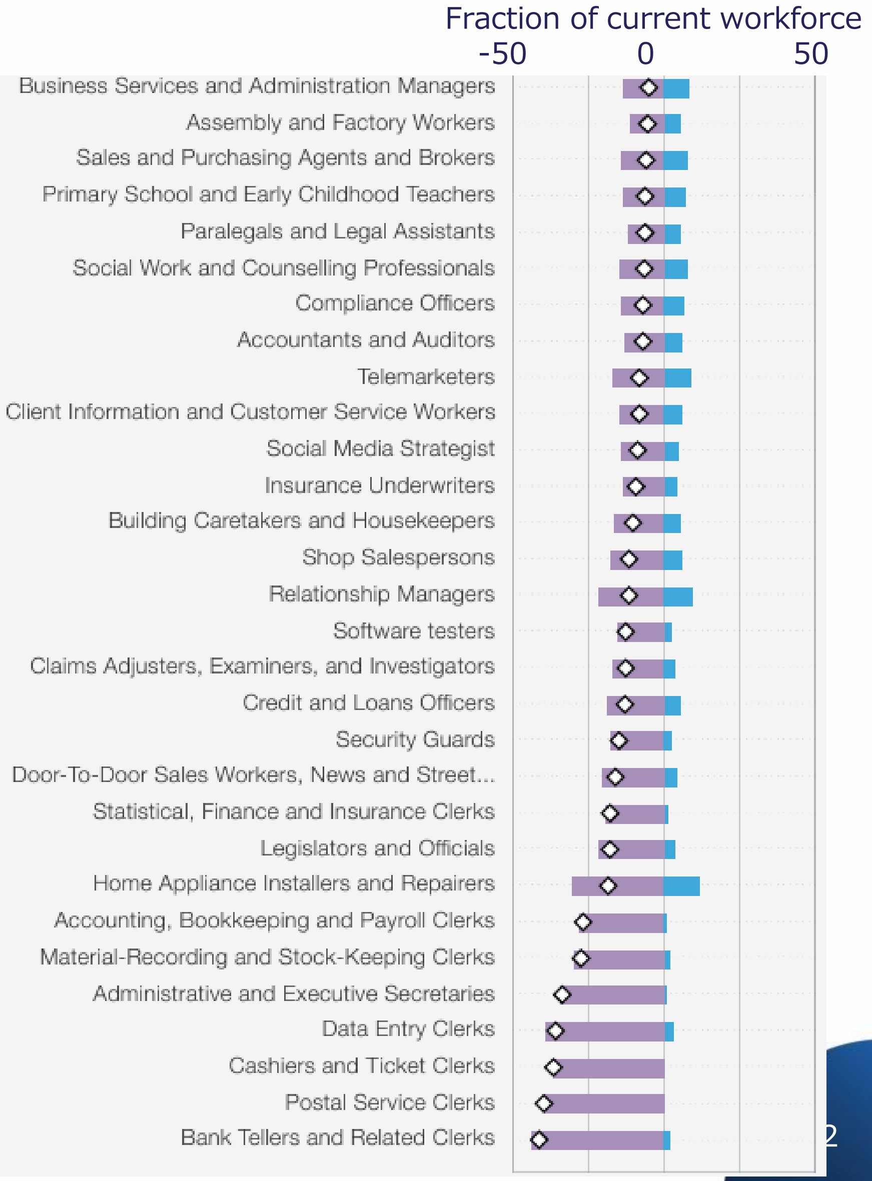
	AI事業者ガイドライン　共通の指針	主なリスク
8) 教育・リテラシー	① AIリテラシーの確保 ② 教育・リスキリング ③ ステークホルダーへのフォローアップ	・AI利用者が判断能力を持たないリスク ・AIにより雇用が奪われるリスク ・ステークホルダーが技術などの進化に追随できないリスク
9) 公正競争確保		・AIに関して公正な競争が阻害されるリスク
10) イノベーション	① オープンイノベーション等の推進 ② 相互接続性・相互運用性への留意 ③ 適切な情報提供	・AIのイノベーションが阻害されるリスク ・相互運用性が確保されないリスク ・AIに関する情報が十分に伝達されないリスク

社会的リスクとして注目されるAIの雇用への影響

AIによる雇用への影響

New jobs and lost jobs, 2023-2027





"The Future of Jobs Report 2023" World Economic Forum（30 April 2023）
https://www.weforum.org/publications/the-future-of-jobs-report-2023/

- スキルやタスクベースで考える必要がある。
- 人も組織も必要なスキルを身に着け、ピボットし続ける必要がある。

AIに対する政策的対応

各国のAI安全性確保への取組み

AISI Japan AI Safety Institute

- **米国**
  - NIST（国立標準技術研究所）にAISIを設立
  - 基本は民間主導、民間企業とのコンソーシアム（AISIC）との協働を強力に推進
  - 人員規模は30人程度。80名位を目指し推進中
- **英国**
  - DSIT（科学イノベーション技術省）にAISIを設立
  - 政府主導で、AIの安全性に関する評価やTestingを強力に推進
  - 規模は100名体制。技術者を多数雇用予定。また、今夏にサンフランシスコオフィスを開業予定
- **EU**
  - EC（欧州委員会）にあるAIオフィスで、利活用に加え、安全性も推進。AI法の整備と推進も担う
  - 60人程度の規模
- **カナダ**
  - 国内機関の協力のもとAISI設立を準備中
- **シンガポール**
  - 南洋理工大学（NTU）内のデジタルトラストセンターがシンガポールのAISIを指定
  - 大規模言語モデル（LLM）の国際標準化を目的とした安全性評価テストツールの提供等を実施
- **オーストラリア**
  - 国立の研究所がAISI機能を担う
- **韓国**
  - 2024年度末を目指しAISIを準備中
  - アジアのハブを目指す

AIに関する国際機関の取り組み

AISI Japan AI Safety Institute

- UN
  - AI Advisory Bodyを設置しAIについて議論
  - GOVERNING AI FOR HUMANITY最終レポート公開（2024年9月）
  - 2024年9月の国連未来サミットの成果文書である「未来への協定」の付属文書である「グローバル・デジタル・コンパクト」で、デジタル協力と人工知能（AI）ガバナンスに関する包括的な世界的枠組みを整備。
    - コンパクトの中核は、すべての人々の利益のために、テクノロジーを設計、利用、規制するというコミットメントであり、AIに関する以下のコミットメントが含まれる。
      - 国際科学パネル並びにAIに関するグローバル政策対話を含めたロードマップにより、AIを統治する。
- OECD
  - The OECD Artificial Intelligence Policy Observatoryで、AIに関するレポートやツール、インシデントの情報を収集
  - Global Partnership on AI (GPAI)とも連携

日本におけるAISIの設立

AISI Japan AI Safety Institute

- 2023年5月
  - 岸田総理大臣が「広島AIプロセス（※）」を提唱
    ※G7広島サミットで提唱された生成AIに関する国際的なルールの検討を行うためのプロセス
- 2023年10月
  - 広島AIプロセス「国際指針」及び「国際行動規範」（※）に合意
    ※生成AIを含む高度なAIシステムに関する国際的な指針と行動規範
- 2023年11月
  - 英国主催AIセーフティサミットを開催
- 2023年12月
  - 「広島AIプロセス包括的政策枠組み」等に合意
  - 岸田総理大臣がAIセーフティ・インスティテュート設立を表明
- 2024年2月14日
  - IPA（情報処理推進機構）にAIセーフティ・インスティテュート（AISI）を設立

所長　村上　明子

出典：
広島AIプロセス <https://www.soumu.go.jp/hiroshimaaiprocess/documents.html>
AI Safety Summit 2023 <https://www.gov.uk/government/topical-events/ai-safety-summit-2023>
AI戦略会議 <https://www.kantei.go.jp/jp/101_kishida/actions/202312/21ai.html>
AIセーフティ・インスティテュート <https://aisi.go.jp/>

AISIの概要

AISI Japan AI Safety Institute

- AISIの位置づけ
  - 今後、官民が協力して、AIの安全安心な活用が促進されるよう、AIの開発や利用をする全ての関係者がAIのリスクを正しく認識し、ガバナンス確保などの必要となる対策をライフサイクル全体で実行できるようにしていく必要がある。
  - また、これらの取組を通じ、イノベーションの促進とライフサイクルにわたるリスクの緩和を両立する枠組みを実現していく必要がある。
  - AISIは、上記を実現するための**官民の取組を支援する機関**である。
- 取組方針
  - 技術がグローバルかつ目まぐるしく進歩していることから、国内、国際的な関係機関と協調して取組を推進していく。

AISIの役割とスコープ

- 役割
  - 政府への支援として、AIセーフティに関する調査、評価手法の検討や基準の作成等の支援を行う
  - 日本におけるAIセーフティのハブとして、産学における関連取組の最新情報を集約し、関係企業・団体間の連携を促進する
  - 他国のAIセーフティ関係機関と連携する

  ※自ら研究開発する組織ではない
- スコープ
  - AIによる以下の事象や検討事項の中で、諸外国や国内の動向も見ながら柔軟にスコープを設定し取組を進めていく。
    - 社会への影響
    - ガバナンス
    - AIシステム
    - コンテンツ
    - データ

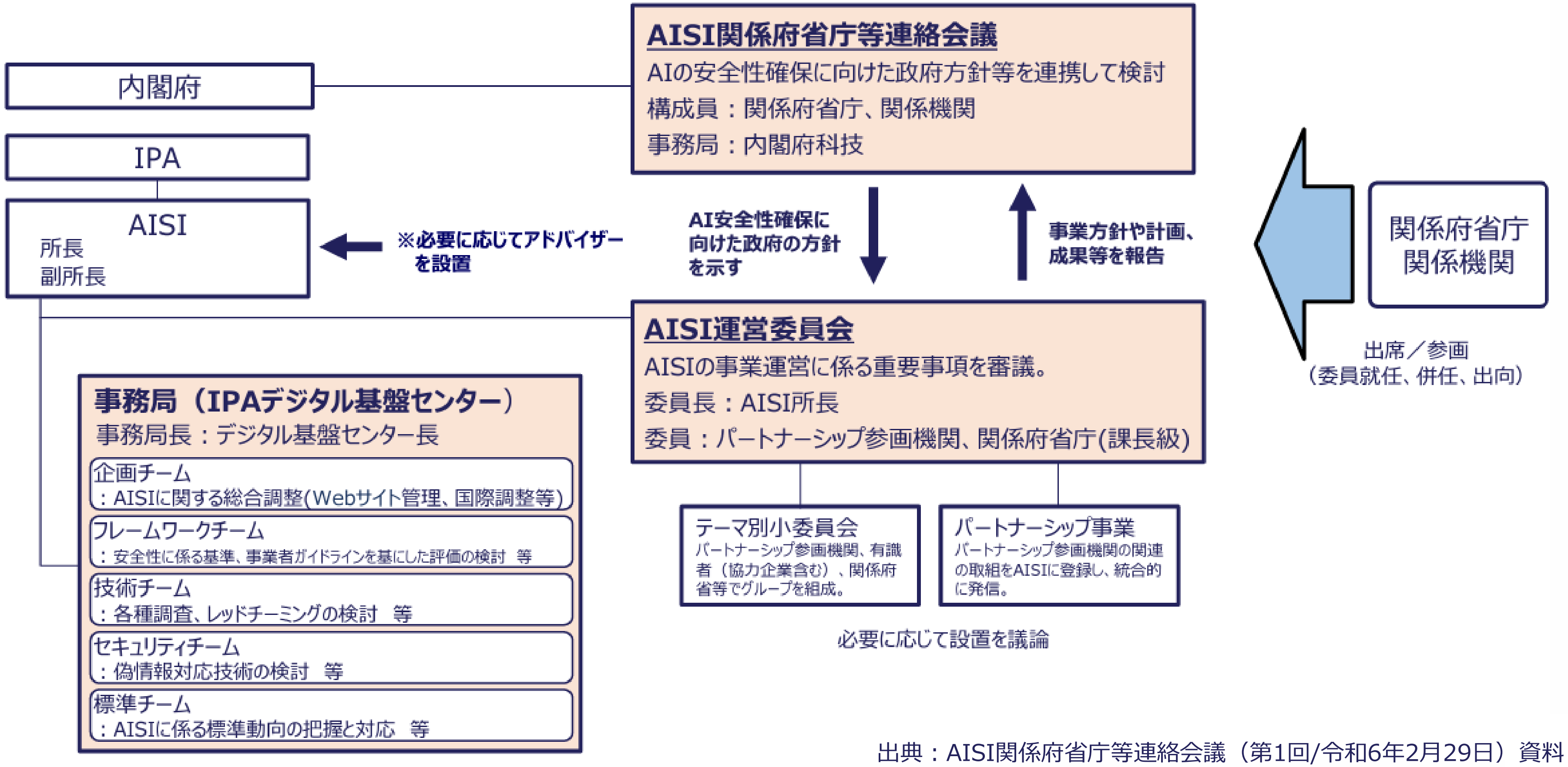
実現に向けた業務

AISI Japan AI Safety Institute

1. 安全性評価に係る調査、基準等の検討
   ① 安全性に係る標準、チェックツール、偽情報対策技術、AIとサイバーセキュリティに関する調査
   ② 安全性に係る基準、ガイダンス等の検討
   ③ 上記に関するAIのテスト環境の検討
2. 安全性評価の実施手法に関する検討
3. 他国の関係機関（英米のAI Safety Institute等）との国際連携に関する業務

AISIの推進体制

- 内閣府を事務局とする「AISI関係府省庁等連絡会議」を設置し、重要事項を審議（年間２～３回の開催を予定）。AISIの中に、AISI所長を委員長とする「AISI運営委員会」を設置（月１回の開催を予定）。
  - 運営委員会の下に、必要に応じて、「テーマ別小委員会」や「パートナーシップ事業」（研究機関等の関連の取組みをAISI事業として発信）を設置。



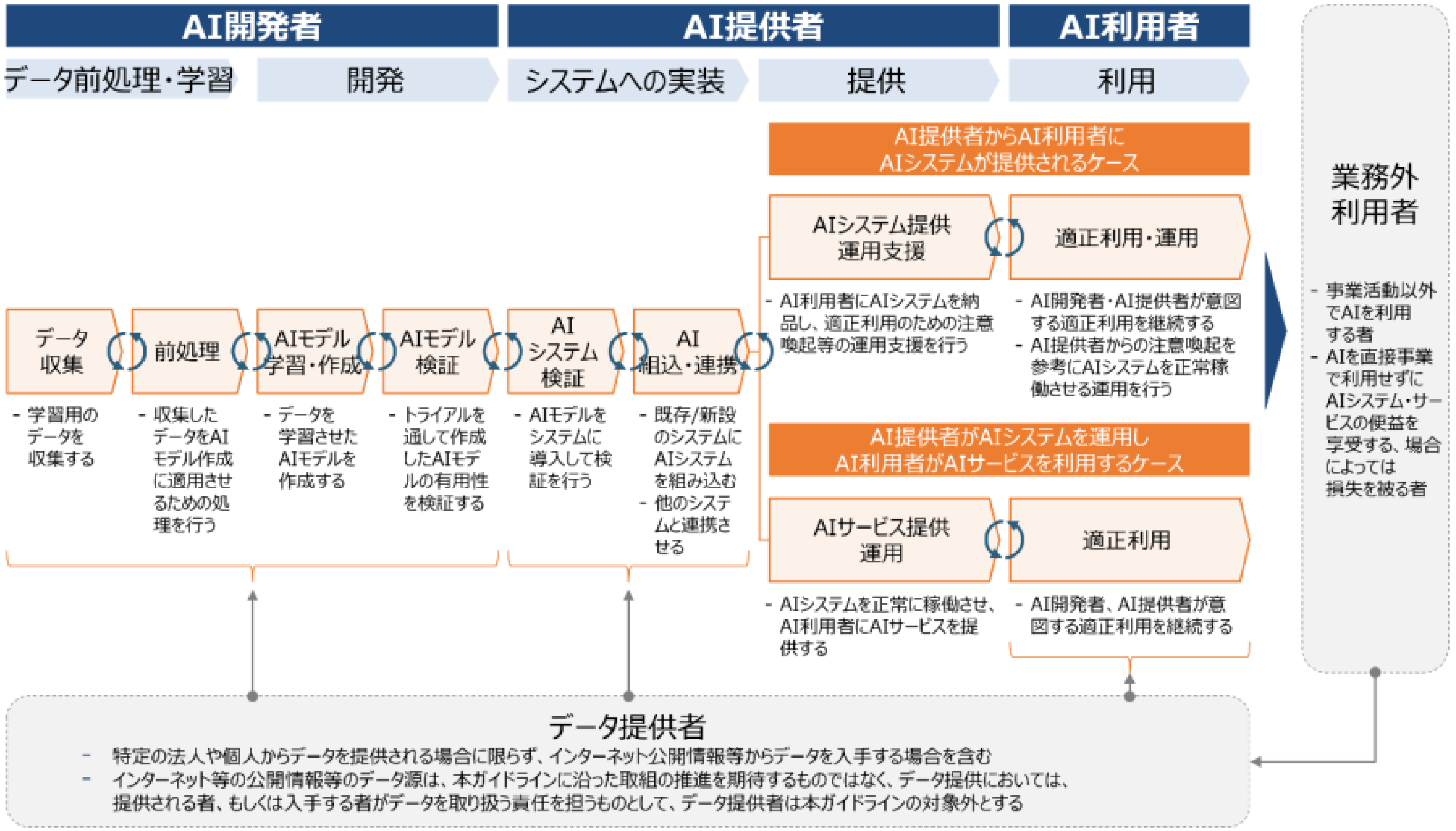
【関係府省庁】
・内閣府（科学技術・イノベーション推進事務局）
・国家安全保障局
・内閣サイバーセキュリティセンター
・警察庁
・デジタル庁
・総務省
・外務省
・文科省
・経産省
・防衛省

【関係機関】
・情報処理推進機構（IPA）
・情報通信研究機構
・理化学研究所
・国立情報学研究所
・産業技術総合研究所

出典：AISI関係府省庁等連絡会議（第1回/令和6年2月29日）資料２

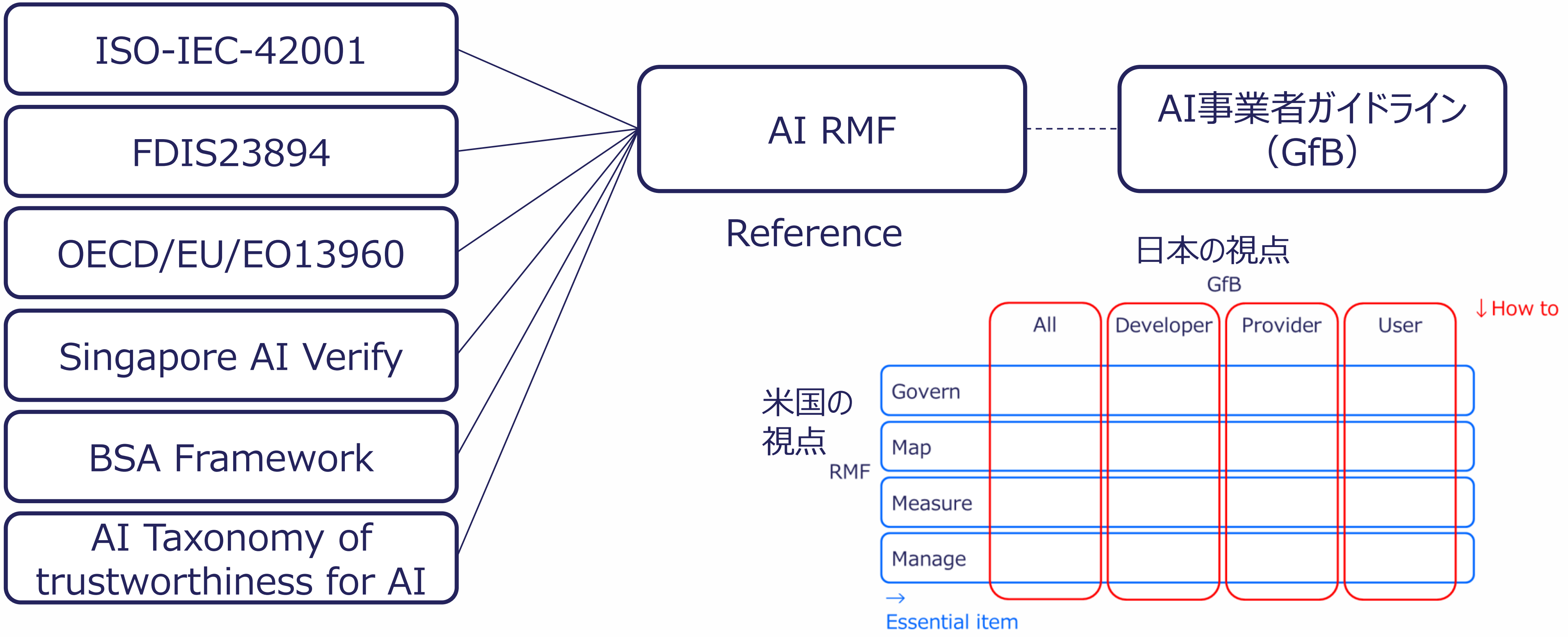
AI事業者ガイドライン

- AI活用の流れの中で、各ステークホルダが対応すべきことを明確化



日米クロスウォーク

- 米国NISTのAI Risk Management Framework(RMF)と日本のAI事業者ガイドライン(Guidelines for Business; GfB)の相互関係を確認
  - 米国のAI RMFをリファレンスに各国ガイドライン等との確認も可能



AIセーフティに関する評価観点ガイドの公開

- AI事業者ガイドライン「C. 共通の指針」において各主体が取り組む事項とされているもののうち、下記6つの事項を、AIセーフティを向上するうえで重視するべき重要要素とし、AIセーフティ評価の観点を導出

重要要素	概要説明
①人間中心	AIシステム・サービスの開発・提供・利用において、全ての取り組むべき要素が導出される土台として、少なくとも憲法が保障する又は国際的に認められた人権を侵すことがないようにすること。また、AI が人々の能力を拡張し、多様な人々の多様な幸せ（well-being）の追求が可能となるように行動すること。
②安全性	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ること。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。
③公平性	AIシステム・サービスの開発・提供・利用において、特定の個人ないし集団への人種、性別、国籍、年齢、政治的信念、宗教等の多様な背景を理由とした不当で有害な偏見及び差別をなくすよう努めること。また、各主体は、それでも回避できないバイアスがあることを認識しつつ、この回避できないバイアスが人権及び多様な文化を尊重する観点から許容可能か評価した上で、AIシステム・サービスの開発・提供・利用を行うこと。
④プライバシー保護	AIシステム・サービスの開発・提供・利用において、その重要性に応じ、プライバシーを尊重し保護すること、及び関係法令を遵守すること。
⑤セキュリティ確保	AIシステム・サービスの開発・提供・利用において、不正操作によって AIの振る舞いに意図せぬ変更又は停止が生じることのないように、セキュリティを確保すること。
⑥透明性	AIシステム・サービスの開発・提供・利用において、AIシステム・サービスを活用する際の社会的文脈を踏まえ、AIシステム・サービスの検証可能性を確保しながら、必要かつ技術的に可能な範囲で、ステークホルダーに対し合理的な範囲で情報を提供すること。

AIセーフティに関するレッドチーミング手法ガイド

- 本ガイドは、AIシステムの開発や提供に携わる者が、対象のAIシステムに施したリスクへの対策を、攻撃者の視点から評価するためのレッドチーミング手法に関する基本的な考慮事項を示す。
  - 国内外における検討や先行事例を勘案し、国際整合性を考慮した上で、現段階でレッドチーミングを実行する際に重要と思われる事項を示す。

種別	記載項目の例
What （レッドチーミングとは何か）	➢「レッドチーミング」の定義やスコープ ➢ 本書が対象とするAIシステム
Why （なぜレッドチーミングを実施するか）	➢ レッドチーミングの目的 ➢ レッドチーミングの重要性・期待される効果
Who （誰がレッドチーミングを実施するか）	➢ どのような役割の者がレッドチーミングを実施するか
When （いつレッドチーミングを実施するか）	➢ レッドチーミングの実施時期
Where （どこでレッドチーミングを実施するか）	➢ 自組織が実施するか、第三者（サードパーティ）が実施するか
How （どのようにレッドチーミングを実施するか）	➢ レッドチーミングの実施計画の立て方や、実施する際の準備事項 ➢ レッドチーミング実施に際して想定する脅威
想定読者	AI開発者・AI提供者　開発・提供管理者　事業執行責任者　※左記のうち、レッドチーミングの企画・実施に関与する者が想定読者。

AIセーフティに関するレッドチーミング手法ガイド【目次】

AISIの今後の取り組み予定

AISI Japan AI Safety Institute

- 日本の強み分野の特定
  （例）ホリゾンタル：データ品質、ロボティックス、ＩｏＴ等
  バーティカル：工場・プラント、ロボット、ヘルスケア等、特定分野での具体的な検討
- 情報発信の強化
  - Web等での発信強化
  - 国際会議のタイミング（11月、２月）のアウトプット検討
  - 関係機関の取り組み状況の収集
- 調査研究の拡大
  - マルチモーダルAIの検討等
- 独法、国研等とのパートナーシップ【順次】
  - パートナーシップ協定の確定、 参画依頼。
- ベストプラクティスや関連事例集の収集・公表【随時】
  - 各省庁、パートナーシップ参画機関における事例の収集
  - 関係機関のご協力を得つつ、AISIとしても国内外へ情報発信
- 民間企業との協力関係整理

データからAIまでのストーリー作り

AIの本格的な導入

- 個社からデータスペースまで、データの価値をフルに活用

データスペースの実現

- データをアセットや知識として使いこなせる社会・経済圏を構築

データ連携

- クロスインダストリ、クロスボーダーでのコネクタ型の連携へ

データマネジメント

- データ構造化から、品質の高いデータの安定的な供給へ

IPA

AISI Japan AI Safety Institute