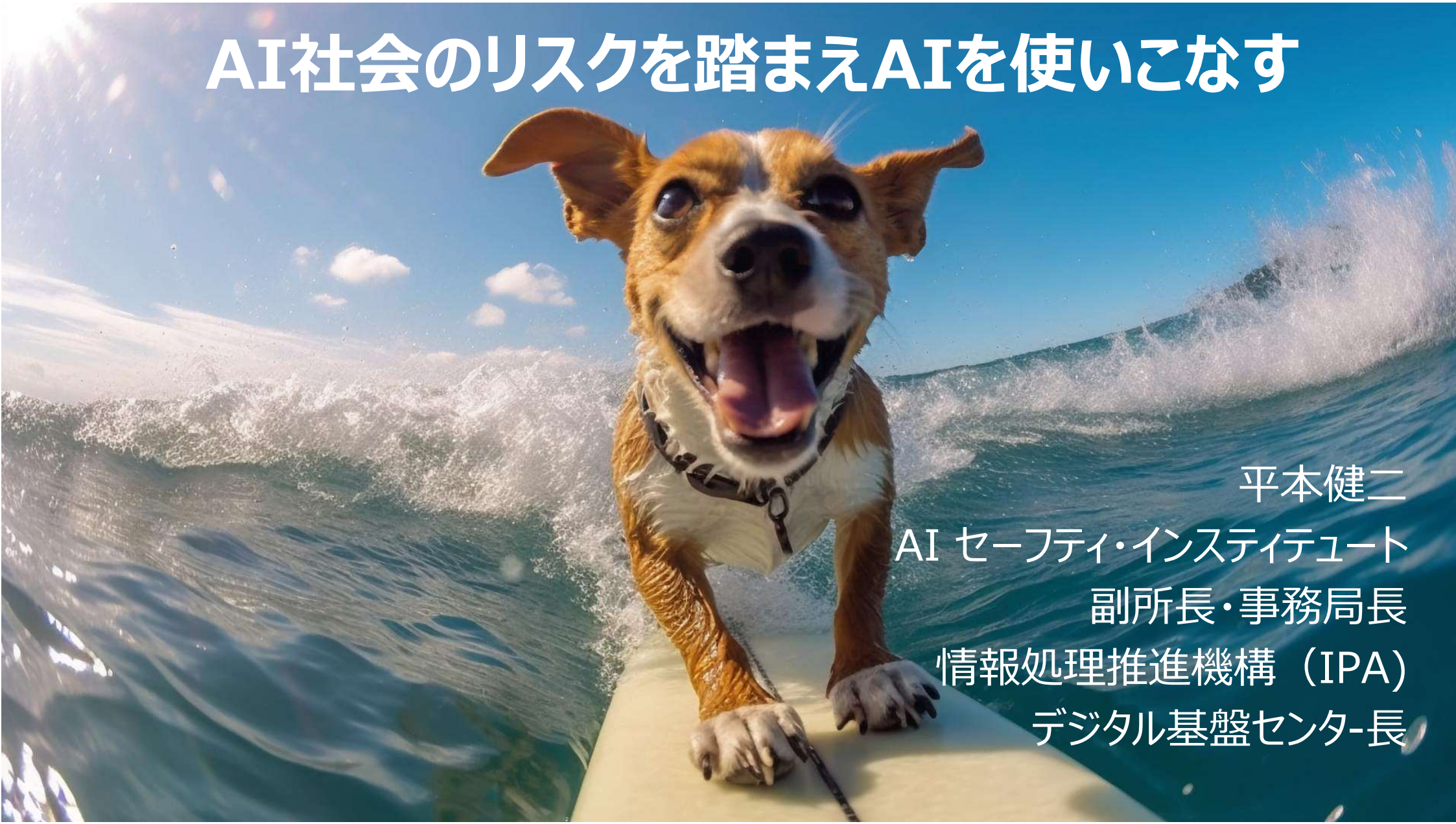
AI社会のリスクを踏まえAIを使いこなす



平本健二
AI セーフティ・インスティテュート
副所長・事務局長
情報処理推進機構（IPA)
デジタル基盤センタ-長

安心してイノベーションを起こすには、安全性が重要

- その製品やサービスは安心して使えるのか？
- 暴走してしまったらどうするのか？
- 止まってしまったらどうするのか？
- 事故が起こったらだれの責任なのか？

プラス面（ポジティブインパクト）

1.生産性と効率の飛躍的向上

- 繰り返し業務やデータ分析を自動化し、行政・企業・医療・教育などあらゆる分野で効率化を実現。
  - 例：自動翻訳、要約、カスタマーサポート、画像診断支援。

2. 新たな価値創造と産業革新

- 生成AIやマルチモーダルAIが、新しい創作・開発・デザインの可能性を開く。
- スタートアップから大企業まで、AIを基盤とした新規事業が急増。 例：AI創作支援ツール、AI探索、スマートシティなど。

3.人間の意思決定支援

- 膨大なデータをもとに、リスク分析・需要予測・政策立案支援など、より客観的で迅速な意思決定を可能にする。
- 政策形成や災害対応、気候変動予測などでも活用。

4.社会包摂・アクセシビリティの拡大

- 音声認識や自動字幕、視覚支援AIなどにより、高齢者や障がい者も含めたインクルーシブ社会の実現を後押し。
- 教育・医療へのアクセス改善。

5.知識の民主化

- 専門的知識へのアクセスが容易になり、個人が自分のペースで学び、創造できる環境が拡大。
- 教育・リスキリング・研究支援への寄与。

マイナス面（課題・ネガティブインパクト）

1.雇用構造の変化と職業喪失リスク

- 単純業務だけでなく、知的労働（文書作成、設計、分析）もAIによって自動化され、職の再定義が必要。
- 新しいスキル（AI活用・プロンプト設計・データ理解）が求められる。

2.情報の信頼性・誤情報問題

- 生成AIによるフェイクニュース、ディープフェイク、誤ったコンテンツの拡散。
- 「誰が」「何を」「どのように」作った情報かを検証するトラスト技術（プロビナンス、ウォーターマークなど）の重要性が増す。

3.プライバシー・セキュリティの懸念

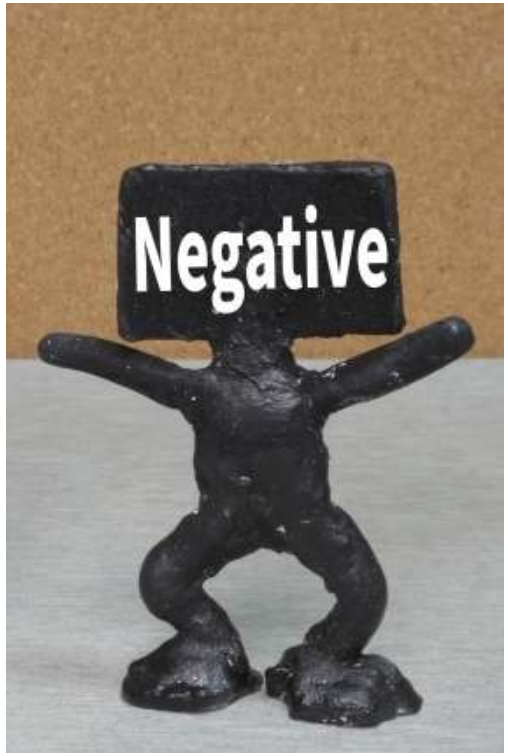
- 学習データや生成結果から個人情報が漏れるリスク。
- 大規模モデルのブラックボックス性による不透明な意思決定。

4.倫理・法制度の未整備

- 著作権、責任所在、AIによる差別・バイアスなど、法制度が追いついていない。
- 「AIガバナンス」「AI安全性（AI Safety）」の国際的な枠組みが求められる。

5.格差の拡大

- AIを使える国・企業・個人と、使えない層の間で情報・経済格差が拡大。
- データやコンピューティングリソースへのアクセス不均衡が新しいデジタル格差を生む。



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。

	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

	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事業者ガイドラインから抜粋

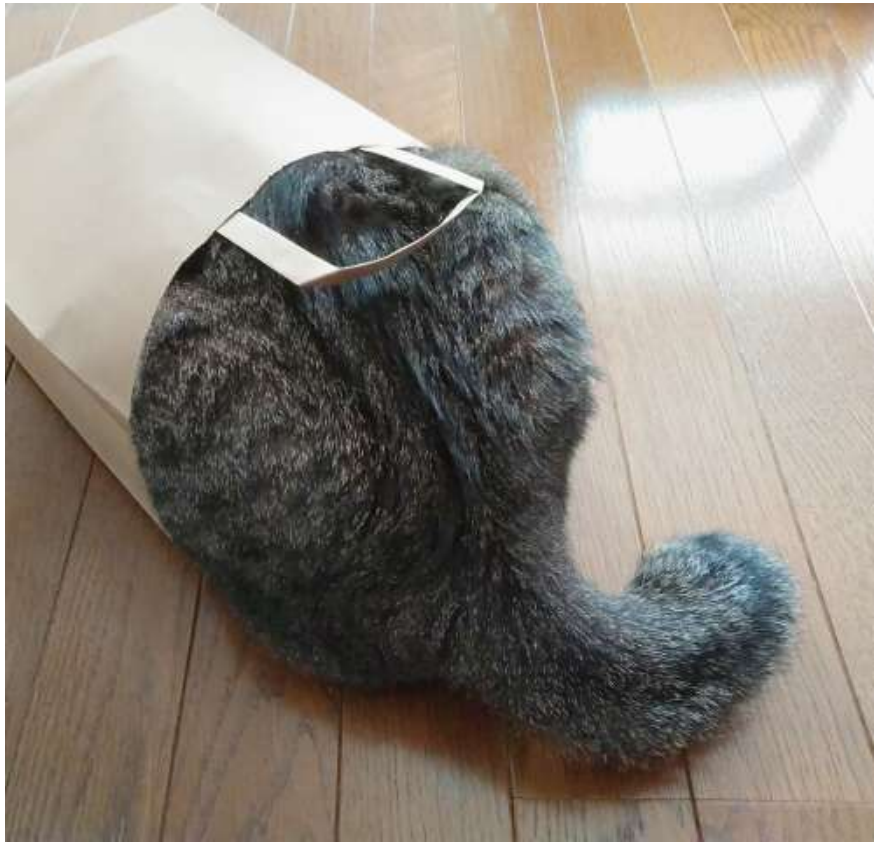
AIに関して想定されるリスク

	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

使わないリスク

「なんだかわからないから、使わない」
「今も困っていないから使わない」

周りはみんな使い始めています。



知識不足のリスク

- AIが怖くて使えないリスク
  - 意思決定者からのオーバーコンプライアンス
    - 過剰なチェック
    - 機能の制限

- AIが何でもやってくれるという幻想のリスク
  - 意思決定者からの無理な依頼
    - 無理なことへの検討と失敗
    - 必要な投資やイノベーションの阻害

AIガバナンス（リスク管理）の重要な視点

- AIリスクの状況は常にアップデートされる
- AIガバナンスの目的は、リスクをゼロにすることではない
- AIリスクは提供する側・開発する側だけでなく、あらゆる組織、個人がAIリスクと対面する
- AIガバナンスはグローバル視点で考える必要がある


IPA