

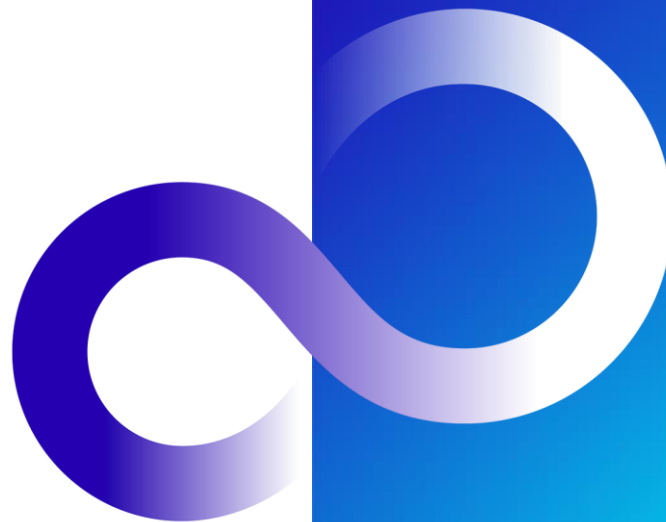
AI品質マネジメントネットワークキングダムシンポジウム

AIと社会とつながって、創る

2024-02-28

富士通株式会社

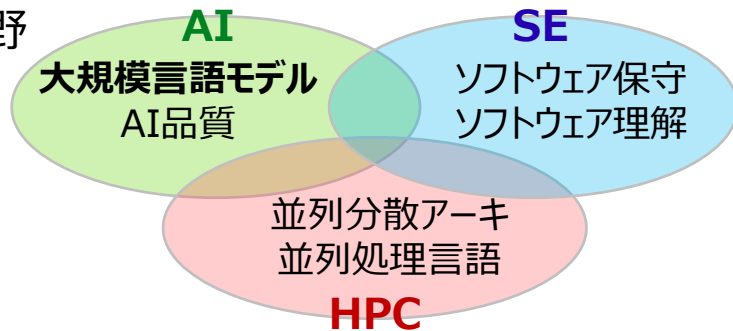
小林健一



小林 健一 博士 (情報科学)

富士通株式会社 富士通研究所 人工知能研究所
リサーチディレクター

● 研究分野



● 経歴

- 富士通研究所 および 富士通 に勤務
- 産総研 機械学習品質マネジメント検討委員 (2019～)
- 豪国立情報通信技術研究所(NICTA) 客員研究員 (2008)
- 米国 HAL Computer Systems 招聘LSI設計者 (2000)



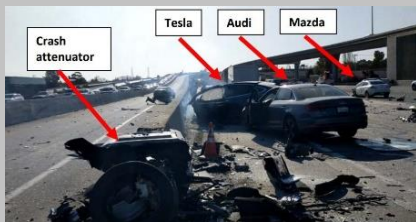
わたしたちのパーパスは、
イノベーションによって社会に
信頼をもたらし、世界をより
持続可能にしていくことです。



事例

Teslaの自動運転

自動運転用センサーが強い日差しと
トレーラーの白い車体を認識できずに衝突事故



AIの特徴

環境の変化や時間経過で、
モデルの精度は低下しうる

問題

人間の生命にかかわる問題
「AIの安全をどうやって保証する？」

出典：テスラの半自動運転死亡事故、原因「視線監視機能の不備にも」米運輸安全委が発表
https://jidouten-lab.com/u_tesla-cause-semi-autonomous-ntsb

品質保証上の考え方

安全性

異常出力値による障害発生リスクの
軽減、および発生の抑制

データ特性把握

案件にかかわる数多くのデータのうち、
何を使うかで性能が変化

機能適合性

学習データで仕様が変わるAIを
含むソフトの品質保証が未確立

要件の合意

AIはなんでもできるイメージ。実際は
トライ&エラーの中で要件合意

グーグル、対話型AIで人物の画像生成を停止 人種的な偏りが原因か (2024-02-23)

・米グーグルは22日、同社の対話型AI（人工知能）「Gemini（ジェミニ）」について、人物の画像を生成する機能を停止すると発表した。歴史上の人物として生成した画像が不適切だと批判が出ていた。最新のAI技術では人種的なバイアス（偏り）が指摘されており、対応の難しさを浮き彫りにした。

<https://www.asahi.com/articles/ASS2R4Q4TS2RUHBI00K.html>

社会に受け入れられない
歴史上の誤りや
人種的なバイアスが発生

Google、「Gemini」の人物画像生成の過剰補正 問題について説明 「AIの罫を乗り越えていく」 (2024-02-24)

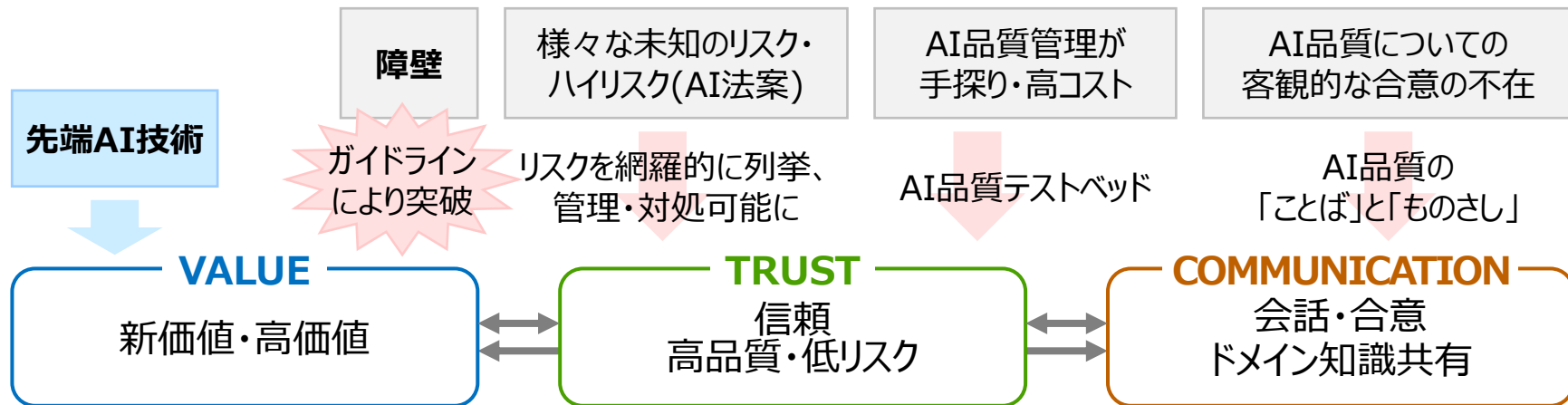
<https://www.itmedia.co.jp/news/articles/2402/24/news050.html>

実は、良かれ（多様性を確保）
と思つての行動が招いた事態

生成AI「Gemini」の失態でアルファベット株急落、時価総額13.5兆円消失 (2024-02-27)

https://forbesjapan.com/articles/detail/69386?module=toppage_1st_01

「お客様と共に、AIによって社会に価値を提供する」

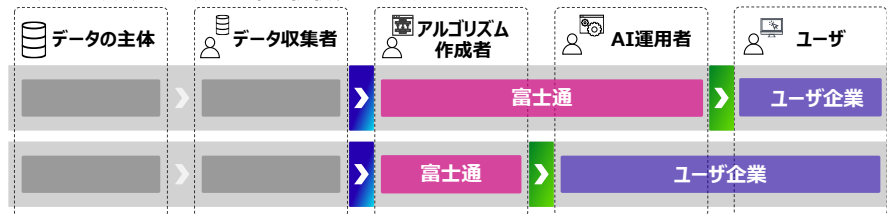


2024年の今は、AIの作り手と使い手だけの関わりでは不十分

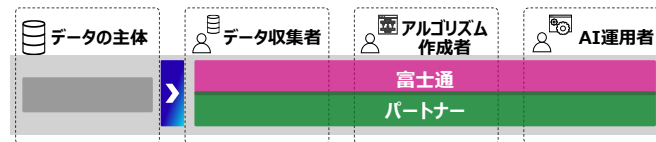
- 総務省AIネットワーク社会推進会議第5回議長ヒアリング
AIサプライチェーンに対する富士通の取組 (2022-03-24)

AIサプライチェーンの典型パターン

AIアルゴリズムを富士通が他社に提供するケース



共同研究



AIアルゴリズムを富士通が他社から購入するケース



データを購入し、ユーザに転売するケース



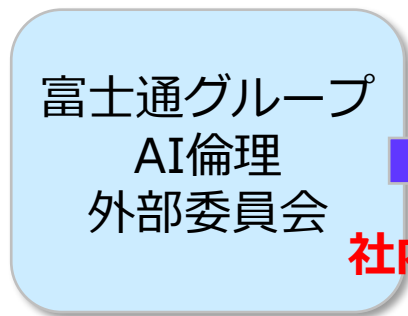
複数のステークホルダーが様々に連携しながら、AIの社会実装は実現されている

富士通グループAIコミットメントは AI4people の 5つの原則に基づき客観性を確保

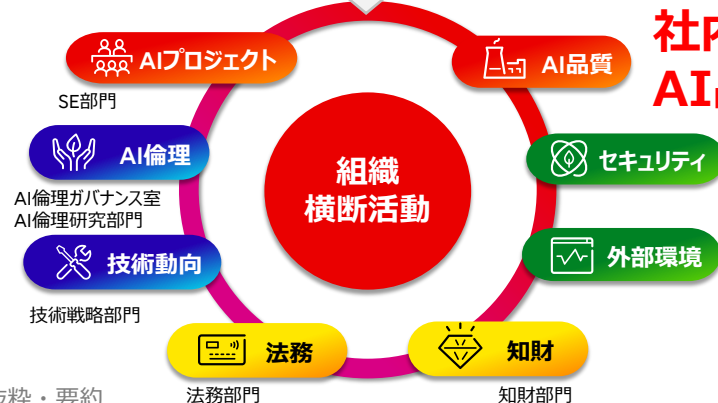
社会との約束



- 1 AIによってお客様と社会に価値を提供します
- 2 人を中心に考えたAIを目指します
- 3 AIで持続可能な社会を目指します
- 4 人の意思決定を尊重し支援するAIを目指します
- 5 企業の社会的責任として、AIの透明性と説明責任を重視します



社内のガバナンス



社内のAI品質活動

「安心安全なAIの提供に向けたガバナンスの取り組み」 AIQMシンポジウム (2023)より抜粋・要約

● 富士通の技術開発 & 公開

- AIシステムの倫理上の影響を評価する方式を開発、手順書や適用例とともに無償公開 (2022-02-21)
- 富士通の自動機械学習技術とAI公平性技術が、Linux Foundationのオープンソースプロジェクトとして始動 (2023-09-15)
- AIに潜むセキュリティリスクを分析する技術を開発 (2022-06-29)

● 富士通の知見の公開（ガイドライン活動を通じて）

- 産総研 機械学習品質マネジメントガイドライン
 - AI品質管理OSSツール Qunomon、AI品質マネジメント講座
- QA4AI AIプロダクト品質保証ガイドライン
- MLSE 機械学習システムセキュリティガイドライン

様々なステークホルダー

経産省 AI事業者ガイドライン案より

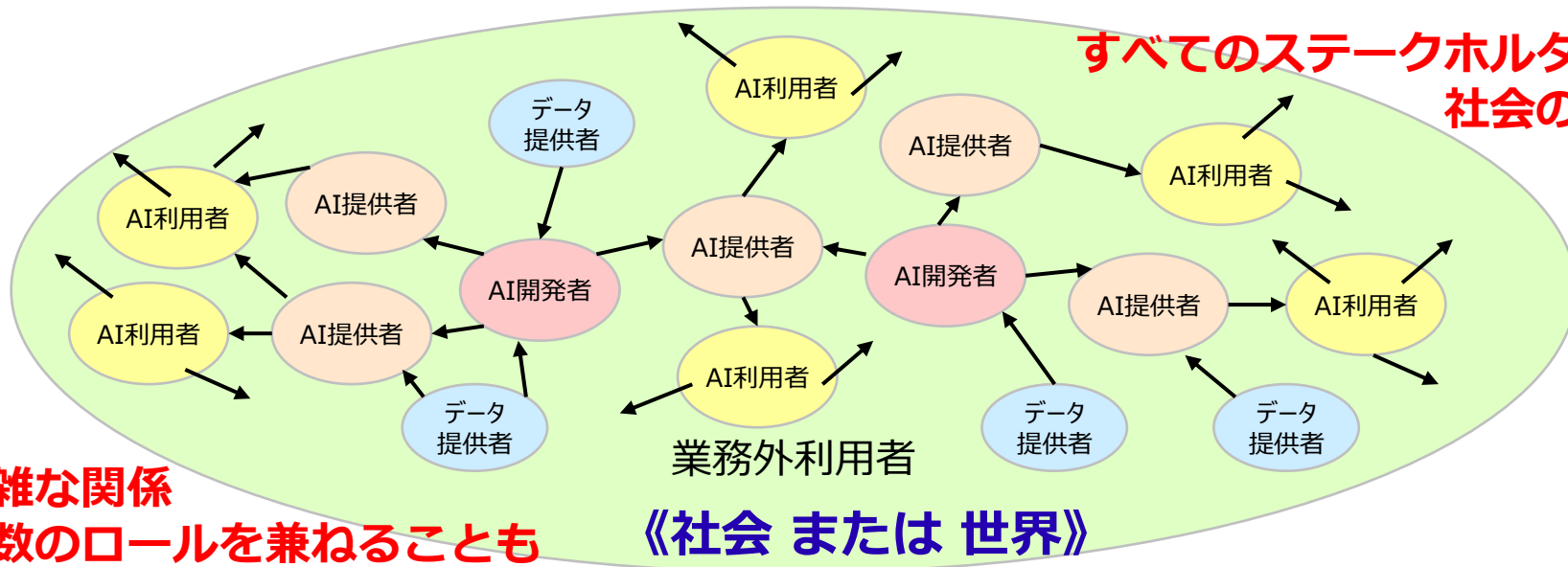


様々なステークホルダー

経産省 AI事業者ガイドライン案より



すべてのステークホルダーが
社会の一員



複雑な関係
複数のロールを兼ねることも

《社会 または 世界》

- AIの社会実装が進む中、AIの能力は常に増大
 - ➡ AIの影響と責任はどんどん拡大。リスクも拡大。
 - ➡ 生成AIの登場でこの傾向はさらに加速
- つながって、
 - 問題と解決法を共有しよう
 - ユーザー／社会の不安を払拭していこう
 - 新しい価値を、皆で創っていこう

Thank you

