富士通のAI倫理ガバナンス

2023年３月1日

富士通株式会社
AI倫理ガバナンス室 荒堀淳一
AI倫理研究センター 福田大輔

荒堀 淳一

富士通株式会社

AI倫理ガバナンス室 室長

- AI指針（AIコミットメント・後述）起草メンバー
- AIガバナンス全社方針の推進・浸透
- AIガバナンス体制の構築・推進

- 総務省AIネットワーク社会推進会議 AIガバナンス検討会 構成員
- 経団連AI活用戦略タスクフォース委員　など

Meet Arahori Here!

福田 大輔

富士通株式会社

AI倫理研究センター センター長

- AI倫理技術の開発
- AI倫理技術の社内外における実践推進

1. 富士通が目指すDXビジネス

We are Fujitsu

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

テクノロジーで**人を幸せに**します

多様な価値を信頼でつなぎ、変化に適応するしなやかさをもたらします

人や社会を第一に考えます

Fujitsu Uvance

FUJITSU

持続可能な世界のために前進します

人の生活を起点にこれからの産業の姿を描いたクロスインダストリーと、それを支えるテクノロジーの両軸で考え、富士通の持つ全てのケイパビリティを「Fujitsu Uvance」に込めました。

多様な価値を信頼でつなぎ、変化に適応するしなやかさをもたらすことで、誰もが夢に向かって前進できるサステナブルな世界をつくります。

Sustainable Manufacturing

Consumer Experience

Healthy Living

Trusted Society

Digital Shifts

Business Applications

Hybrid IT

多様な価値を信頼でつなぎ、変化に適応する しなやかさをもたらします

FUJITSU

Our Key Technologies

Computing
スーパーコンピュータ
HPC
量子

Network
クラウドネイティブ・ネットワーク
フォトニクス
光電融合

AI
説明可能なAI
信頼できるAI
ヒューマンセンシング

Data & Security
ブロックチェーン
データトラスト
デジタルアイデンティティ

Converging Technologies
最先端デジタルテクノロジー
×
人文・社会科学

産業界でのAIガバナンス浸透の課題

FUJITSU

AI倫理・AIガバナンスの
必要性が理解されにくい。

AIガバナンスの**定義が曖昧**な
ため共感されない。

ガバナンス習熟度を**数値的に**
評価する指標が未普及。

経営層のAI倫理ガバナンスにおける
必要性理解と実践に乖離がある。

画像認識、医療、人材採用分野
以外での議論が特に進んでいない。

データ、プライバシー、GDPR
にフォーカスされすぎている。

etc…

DXにおけるAIガバナンスは単純ではない

FUJITSU

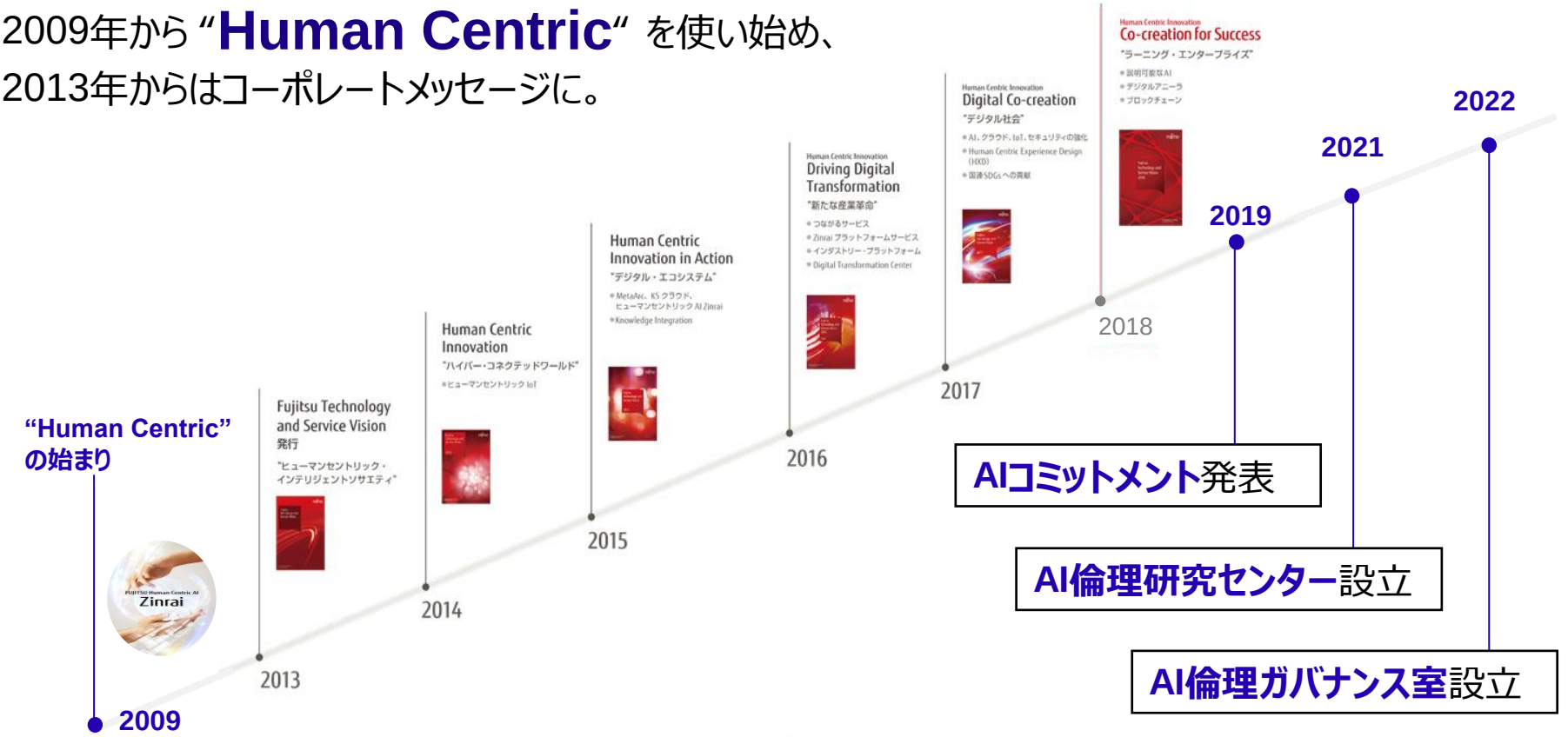
- **DX / IT活用のあらゆる場面でAIを含みうる**
  - AIの活用方法はAI単体ではなく、多様な様態がある。
    （例） SI、サービス、ソフトウェア、ハードウェア
- **AIはアルゴリズム開発者/提供者に閉じない**
  - 当社はBtoBビジネスを主に提供しているが、
    消費者と接するユーザー企業や他のAI企業と手を携えていくことが重要だと考える。

単純に倫理指針を作るだけでは足らず、
多様なAI活用シーンに適合する**柔軟な運用体制**が必要。

2. 富士通のAI倫理について

富士通のAI倫理への取り組みの歴史

2009年から“**Human Centric**“ を使い始め、
2013年からはコーポレートメッセージに。



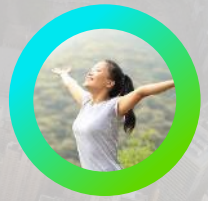
富士通のAI倫理の推進体制

Sustainable Manufacturing

Digital Shifts

Consumer Experience

Business Applications



Healthy Living

Hybrid IT

Trusted Society

サステナブルな世界を実現する 7 Key Focus Areas

自社規律・社内外浸透

AI倫理ガバナンス室

最先端テクノロジーの研究・開発・実装にまつわる倫理に関する国際的な動向、政策、法制度の動向などを踏まえ、AI倫理ガバナンスに関する全社的かつ総合的な取り組みをさらに強化するため、社長直下に2022年2月1日新設

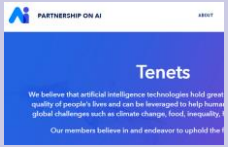
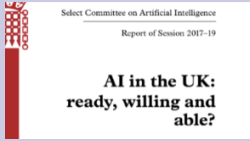
研究開発

人工知能研究所・AI倫理研究センター

サステナブルな世界の実現のために、最先端のAI技術を国内外のアカデミアと連携して開発し、AI倫理のルール作りや課題への取組みを国内外団体と議論・主導し、ユーザー企業や社会の共感を喚起

富士通グループAIコミットメントの策定

FUJITSU

PARTNERSHIP ON AI	Future of Life アシロマ	IEEE EAD	モントリオール宣言	AI、ロボット、自律型システムに関するステートメント	英国上院 AI in the UK
					

AI4People*5 原則

*ヨーロッパ初の国際的コンソーシアム

客観性を確保するために、「富士通グループAIコミットメント」はAI4Peopleの5原則を基に策定している。

FUJITSU
AIコミットメント
FUJITSU Way

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のAIの透明性と説明責任を重視します

生命倫理４原則

FUJITSU

生命倫理学　生命に関する倫理的問題を扱う研究分野

生命医学倫理４原則
（ビーチャム&チルドレス 1979）
医療現場で倫理的問題に直面した際の解決指針として
「善行」「無危害」「正義」の起源は「ヒポクラテスの誓い」

原則	説明	例
自律尊重原則	人間は自立した存在であるため、患者の意思決定を尊重し、支援する。個人的な価値観や信念、自由を尊重する	胸部大動脈瘤の手術で、破裂のリスクと対麻痺になるリスクをインフォームドコンセントしたうえで、手術をするかしないかについては、患者が選択した治療法を尊重する
善行原則	患者に最善の利益をもたらすよう、患者のために善い行いをする	術後に、徐脈のためテンポラリーペースメーカを外せない患者に対し、ペースメーカを植え込む処置をする
無危害原則	危害を与えないようにリスクを避ける。どうしても避けられない場合は、最小限にするよう努める	歩行が困難で転倒歴もある、術後せん妄の患者に対し、歩行する際にナースコールをするように説明しても１人でトイレまで歩行することがあるため、患者と家族の同意を得て、夜間のみ離床センサーを使用する
正義原則	利益・リスク・費用などの医療資源を平等に公平に提供する	病棟業務が多忙な状況であっても、緊急入院の患者を受け入れ、他の患者と同様に医療・看護を提供する

“心臓外科手術を受ける患者の看護倫理”の表を編集, Heart Vol.2 No. 7, 2012

AI倫理５原則

FUJITSU

原則	説明
自律	**AIシステムよりも人の判断や決定が上位**に位置する。AIが判断・決定・指示・指令を出したとしても、人は決定権、根拠を知る権利、従わない権利を有する。 **AIに決定を委ねる権利は当人だけが有し**、いつでもそれを撤回できる。
善行	AIシステムは、**人と共同体の幸福・福利**のために設計・開発されなければならない。 価値や富を生み出すだけでなく、公平な社会、すべての人が参加できる社会、**持続可能な社会**への貢献を含む。
無危害	AIシステムは、人に危害を加えてはならない。 危害の対象は、身体だけでなく、**尊厳、自由、プライバシー**、安心・安全も含む。
正義	AIシステムがもたらす**便益がすべての人に広く行き渡り**、負の側面が特定の人や集団に偏らないこと、および**差別がないこと**を意味する。損害に対して、補償や救済の手立てが用意されていることも含む。
説明可能	**AIシステムが下した判断の根拠がわかりやすく説明されること**、判断に問題があると人間が感じた際に、そこに至る**プロセスを再現**し、原因を特定できること、規制上・法律上の問題が生じた際に、**第三者が監査**できることを保証する。

富士通グループAI倫理外部委員会 (年2回開催)

FUJITSU

「AI倫理」は指針やチェックリストを作るだけではなく、企業全体で取り組むことが重要

1. 本委員会は**多様な社外専門家**で構成され、AI倫理の取組みを**客観的視点から評価**いただいていております。
2. 社長・副社長をはじめとする役員も出席し、さらに委員会での審議内容を「取締役会」へ報告することで、**「AI倫理」をコーポレートガバナンスの一環**に組み入れ、透明性を確保しています。

委員会メンバー

辻井 潤一　委員長
国立研究開発法人産業技術総合研究所
フェロー兼人工知能研究センター長、
東京大学名誉教授、マンチェスター大学教授

国谷 裕子　先生
ジャーナリスト、東京藝術大学理事（SDGs推進室長）

板東 久美子　先生
日本赤十字社常任理事、
公益社団法人セーブ・ザ・チルドレン・ジャパン理事 他

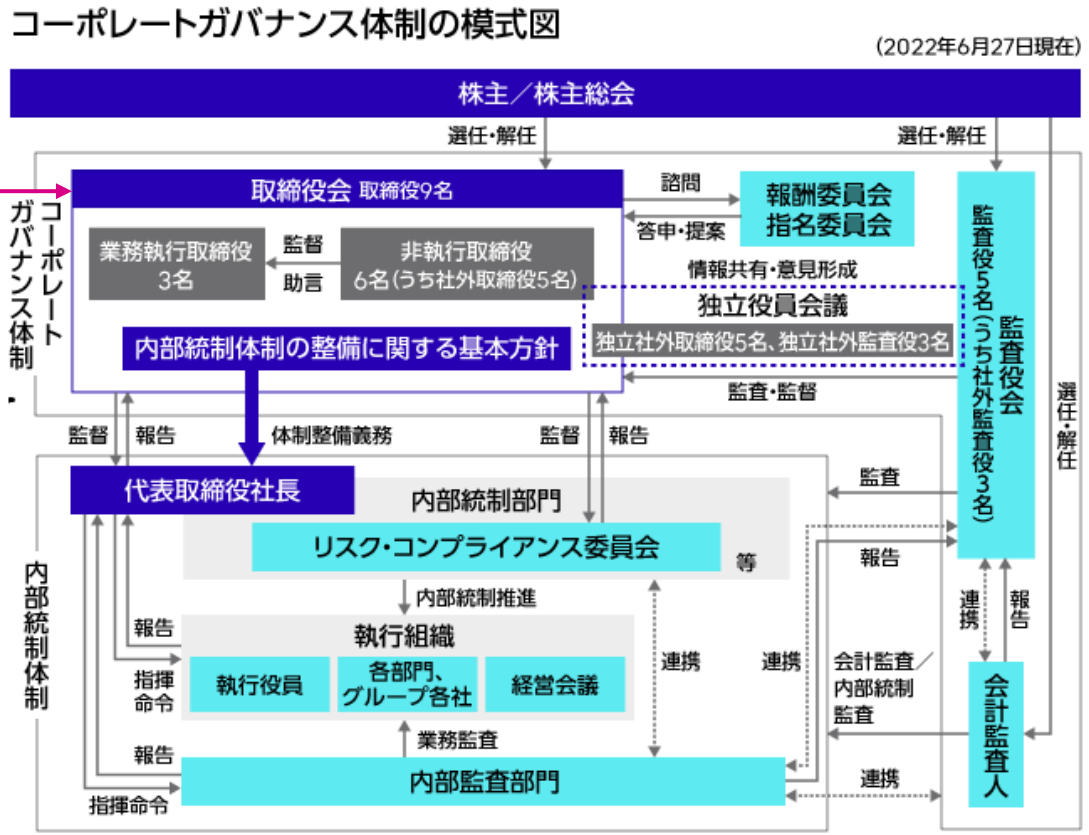
君嶋 祐子　先生
慶應義塾大学　法学部・大学院法学研究科教授、弁護士、
同大学グローバルリサーチインスティテュート（KGRI）所長、日本工業所有権法学会理事

武部 貴則　先生
東京医科歯科大学　統合研究機構教授
兼　横浜市立大学 コミュニケーション・デザイン・センターセンター長、シンシナティ小児病院　オルガノイドセンター 副センター長

湯本 貴和　先生
京都大学名誉教授　兼　中部大学客員教授

AI倫理をコーポレートガバナンスに連携

AI倫理外部委員会の
審議の結果は、
取締役会に共有される



富士通のAI倫理ガバナンス体制

FUJITSU

富士通グループAIコミットメント

開発プロセスへの組み込み

- 企画
- 開発・検証
- 導入
- 運用

各事業特性に応じた実践

サポート

自社規律・社内外浸透

- 相談体制
- 教育制度

研究開発

技術的支援

コーポレート部門

（法務・知財など）

取締役会

協調

Society

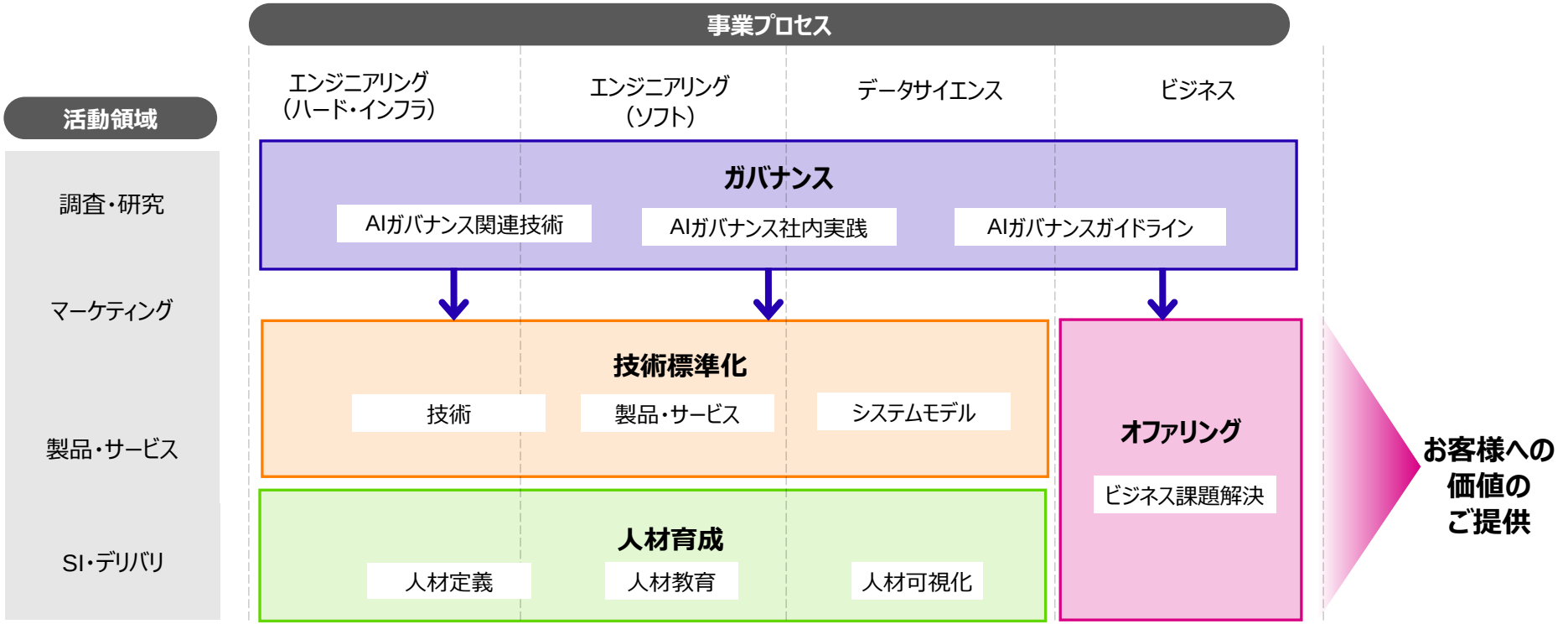
- 国際標準化
- ユーザー企業との勉強会
- 産学連携

提言

共有

富士通グループ
AI 倫理外部委員会

AIタスクフォース活動スキーム



AIタスクフォース活動・現場実践について

FUJITSU

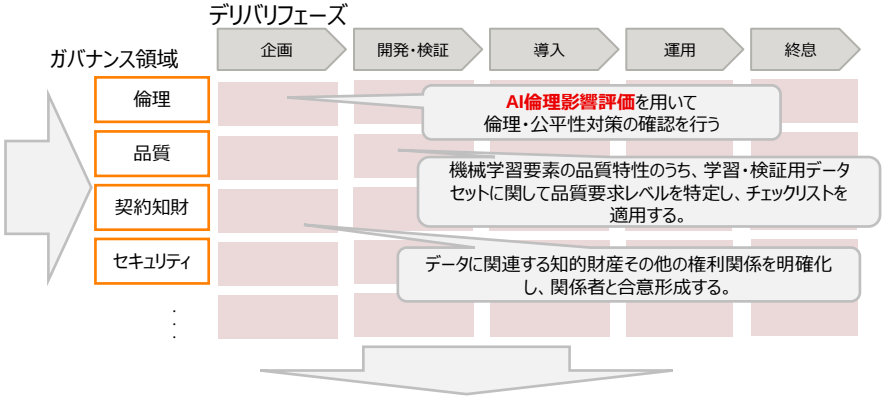
「①国内外／当社ガイドライン・専門部署知見」を**「②デリバリフェーズ／ガバナンス領域のマトリックス」**で整理し、
「③概説・実施事項およびチェックリスト」を作成。

①国内外／当社ガイドライン・専門部署知見

経済産業省	AI原則実践のためのガバナンス・ガイドライン
経済産業省	AI・データの利用に関する契約ガイドライン
産総研	機械学習品質マネジメントガイドライン
総務省、AIネットワーク社会推進会議	AI利活用ガイドライン
総務省、AIネットワーク社会推進会議	国際的な議論のためのAI開発ガイドライン案
消費者のデジタル化への対応に関する検討会 AIワーキンググループ	消費者のデジタル化への対応に関する検討会 AIワーキンググループ報告書
QA4AI	AIプロダクト品質保証ガイドライン
European Commission	Proposal for a Regulation laying down harmonised rules on artificial intelligence
European Commission	Ethics guidelines for trustworthy AI
ISO/IEC	AIガバナンス（DIS 38507）
ISO/IEC	AIマネジメントシステム標準（WD 42001）
ISO/IEC	データのトラストとデータ共有合意の国際標準化（TR 23186、DIS 23751）
Data Management Association International	The DAMA Guide to The Data Management Body of Knowledge（DMBOK）

など

②デリバリフェーズ／ガバナンス領域のマトリックス



③概説・実施事項およびチェックリスト

AIのビジネス実践加速に向けて

FUJITSU

- データサイエンティストなど特別な人だけでなく、社員全員が、データドリブンアプローチの効果的な手段として、AIを実践活用できるようになってもらいたい
- フロント部門には、データ/AI活用による価値提供の提案を加速してほしい

データ/AIの民主化を目指す！

AIに関する知識ベースの研修プログラムにとどまらず、
関係各部署と連携し、ビジネス実践をサポートする複合的な施策を提供

AI Learning Framework

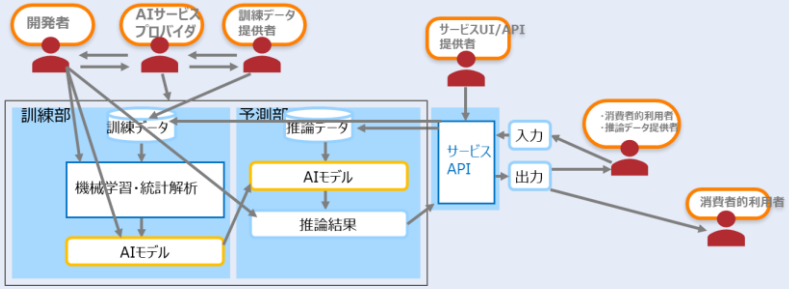
3.AI倫理実践のための技術的なご提案

AI倫理の普及に向けた課題

自分と関係がある？

- **法やガイドラインと自業務での実践に大きなギャップが存在**
  - ガイドラインの内容が抽象的
  - AI適用範囲は多岐にわたるため、網羅的な記載は不可能
- **問題の多くは無自覚なバイアスが原因**
  - データが膨大でチェックできない！
    →元データの文化圏の影響を受けた文書生成
  - 差別をしている認識はない！
    →画像の説明：男性：ビジネス、女性：外見

どこで？



なにを？

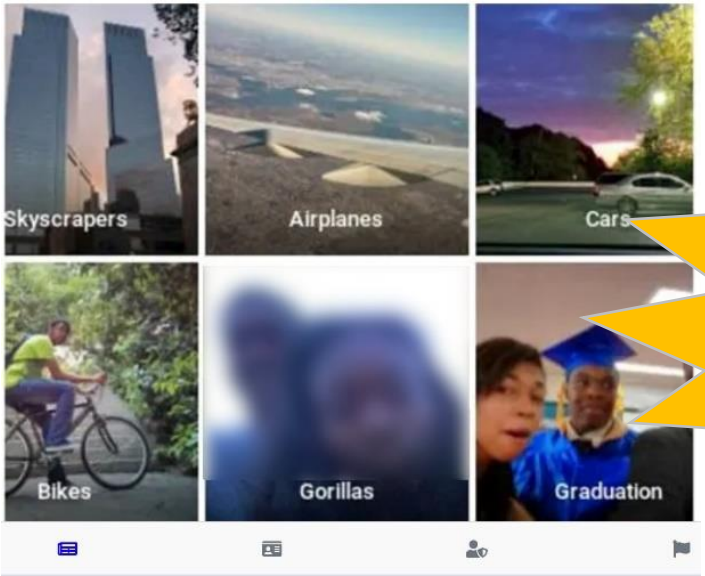
可視的多様性	
性別	年齢
国籍・人種	障害の有無
不可視的多様性	
宗教	性的指向
未既婚	経歴

- **グループ間で精度の差はないか？**
  基本的な属性間での差異
- **少数派を考慮しているか？**
  国内従業員向けAIにおける外国人従業員
- **データの質・出自は確かか？**
  主観の介入・母集団と学習データの乖離
- **特徴量の選定は適切か？**
  例）専攻科目：既に男性が多数

繰り返されるAIの倫理問題

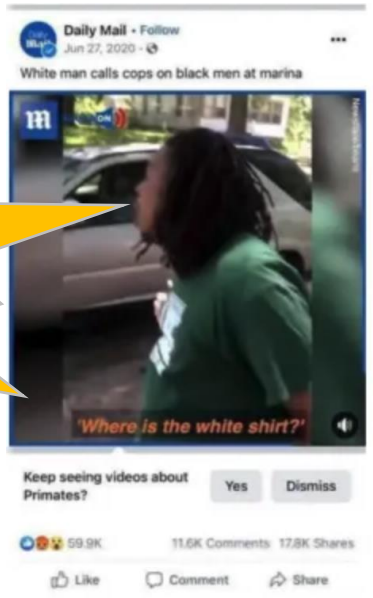
FUJITSU

○AIが写真に写っている人を
「ゴリラ」とラベルづけ
(2015年6月29日報告)



過去から
学ぶ仕組みを
作りたい！

○AIが人が写っている動画に
「霊長類」の動画を推薦
(2021年9月5日報告)



富士通のAI倫理技術

FUJITSU

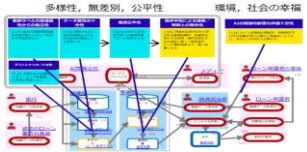
システム内のリスクを網羅的に抽出し、「人が気づかない」無自覚なバイアスを解消

	課題抽出	バイアス検知・是正
	システム内のAI倫理リスクを検出	AIのバイアスを検知・是正
従来方法と課題	専門家の経験による分析 AI倫理リスクの見逃しの可能性	単純データ属性のみで検証 バイアスの見逃し是正が不十分
富士通のAI倫理	考えられるリスクの網羅的な洗い出し	あらゆるバイアスの網羅的な緩和
	1. AI倫理影響評価	2. 交差バイアス緩和

バイアスの見落としバイアスの改善不足

無自覚なバイアスも解消

本日ご説明



	(全体)	男性	女性
(全体)			
白人			
黒人			

利用シーンによって起こる倫理リスク

リンゴの画像から品質を評価するという
単純な画像認識AIであっても…

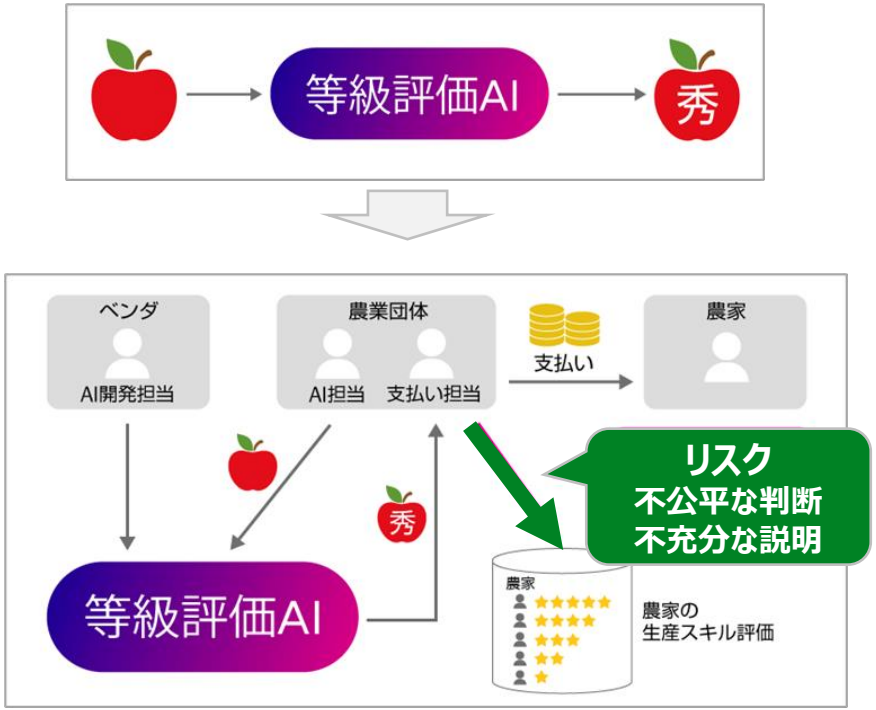
[リンゴ農家]

自分の選別作業を自動化するだけなら
問題なし

[農業団体]

リンゴを選別するだけならば良いが、農家の
ランク付けに使ってしまったら…

**→ AIシステムを誰がどう利用するかを
明らかにする**



https://www.fujitsu.com/jp/about/research/article/202112-aiethics-01.html

倫理リスクはどこにあるか?

倫理的な問題は、AIシステムの利用の仕方やAIシステムをとりまくステークホルダー、およびステークホルダー同士との関わりにおいて、初めて起きている

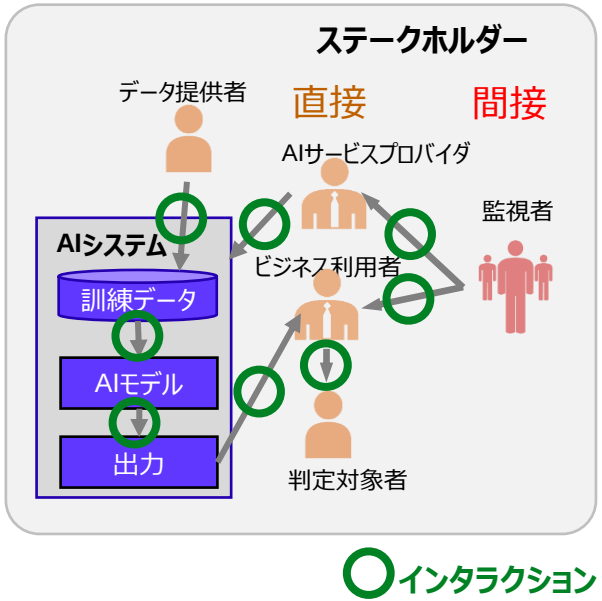
倫理リスクは、AIシステムにおける要素間の情報や制御のやりとり（**インタラクション**）に紐づけられる

※インシデント事例にて検証

→ **倫理リスクがどこで起こりうるのかを知る**

→ 何に対してどう対応すればよいかがわかる

AIシステムとステークホルダー



AI倫理影響評価手法の概要

FUJITSU

input

お客様
AIシステム仕様

なにを？に相当

欧州AI倫理ガイドライン
経産省AIガバナンスガイドライン

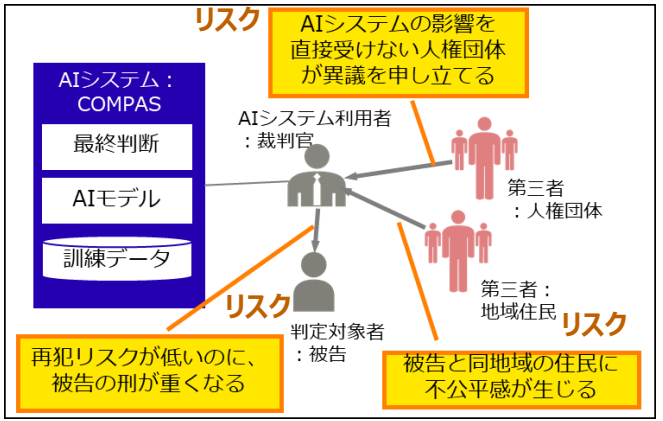
ガイドライン

どこで？に相当

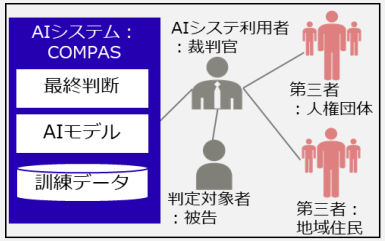
ユースケース

output

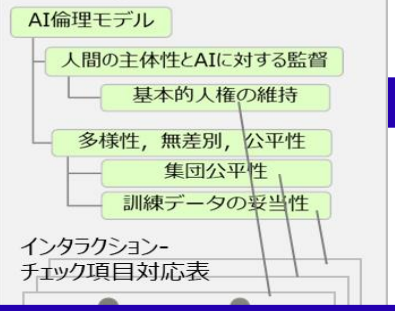
AI倫理影響評価結果



AIシステム図



AI倫理モデル



インタラクションを抽出する

インタラクションに課題を紐付ける

AI倫理の専門家でなくてもシステマチックに課題抽出が可能

参考）検証事例リスト

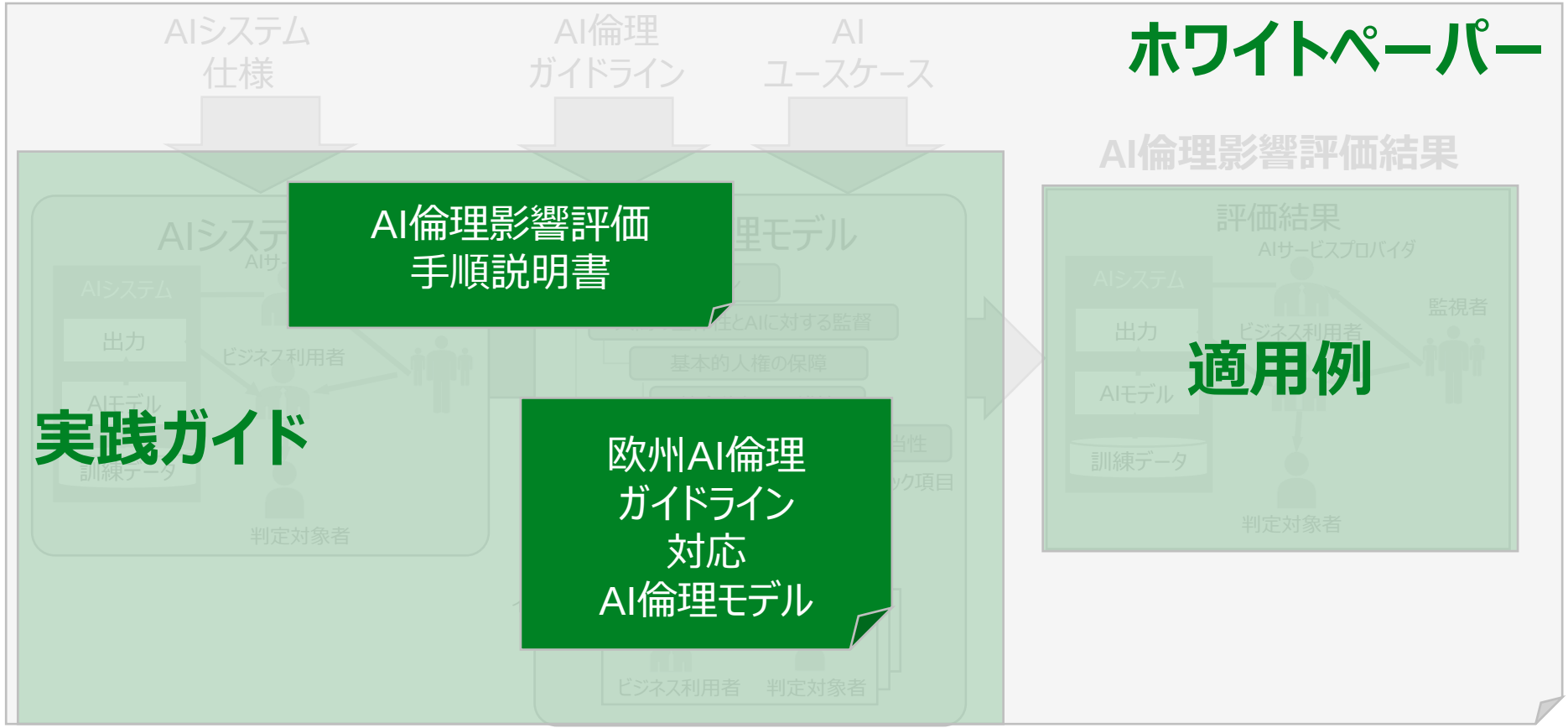
FUJITSU

過去の重大事例において、課題の抽出を確認

No.	名称	国際標準産業分類	AIタスク	データ種類	実際に発生した倫理的な問題と影響
S1-1	チャットボット	芸術・娯楽及びレクリエーション	文章理解・生成	自然言語	差別的な表現を発信した．運用を停止．
S1-2	採用AI	管理・支援サービス業	分類	テーブルデータ	採用結果に女性差別がある．対策実施も解消せず，運用を停止．
S1-3	再犯リスク予測	公務及び国防・義務的社会保障事業	分類	テーブルデータ	メディアから，再犯のリスクが高いと誤評価される比率が黒人が白人よりも高いと指摘される．
S1-4	警察による顔認識	公務及び国防・義務的社会保障事業	分類	画像	市民の顔データベースを用いて，AIが監視カメラに映った容疑者と同一人物を特定．無関係の市民が，誤認逮捕される．
S1-5	ビデオ面接を使った採用判定	管理・支援サービス業	分類	画像	プライバシー監視団体から批判を受ける．
S1-6	交通違反検知	公務及び国防・義務的社会保障事業	分類	画像	バス車体の広告画面の人物を誤って交通違反者と公表する．誤検知された人が名誉棄損で訴える．
S1-7	教師の評価	教育	分類	テーブルデータ	教職員組合がAIによる評価結果に問題があるとして訴訟を起こす．
S1-8	ロボットアームによる死亡事故	製造業	画像認識	画像・環境	作業員がロボットアームに接近，ロボットが正しく認識できず接触．
S1-9	パスポート審査	公務及び国防・義務的社会保障事業	画像認識	画像	アジア人男性の写真を認識できない．
S1-10	ショッピングセンターのセキュリティロボット	管理・支援サービス業	画像認識	画像・環境	16ヶ月の子どもと接触し怪我を負わせる．
S1-11	犯罪予測	公務及び国防・義務的社会保障事業	分類	テーブルデータ	居住地域による差別があると批判される．
S1-12	腎機能診断	保健衛生および社会事業	分類	テーブルデータ	黒人は他の人種に比べてリスクを過小評価され，適切な治療を受けられない．
S1-13	写真のタグ付け	芸術・娯楽・レクリエーション	分類	画像	SNSの投稿写真に人種差別的なタグ付けがされる．
S1-14	パーソナルアシスタント	管理・支援サービス業	分類・統計処理	テーブルデータ	関係者のスケジュールや，会議スケジュール設定時に考慮する関係者の重要度が公開され，プライバシーの問題を指摘される．
S1-15	自動翻訳	情報通信業	文章生成	自然言語	「医師」「社長」を男性に「看護師」「教師」を女性として翻訳する．
S2-1	**ローン審査AI**	**金融・保険業**	**分類**	テーブルデータ	**女性や若年事業者は年配の男性事業に比べて審査に通りにくい．**

無償公開したマテリアル

FUJITSU



なぜ無償公開したのか？

FUJITSU

- **AI倫理の重要性認知**
  - 開発した方式を「無償公開」することで大きく社会に認知をかける
  - 本方式を用いたイベント企画・コミュニティの形成を進める
  - お客様業務とAI倫理の関連性を見いだせるようにする
- **AIが信頼されるための具体的な活動を活性化**
  - 欧州AI法案に向けた対応
  - AI倫理の実践に関わる標準化
  - 公平性にとどまらない、幅広いAI倫理技術研究・開発の喚起

⇒ **多様な観点を持つ方々と議論・協業することで、内容を拡充・改善していきたい**

6/15 人工知能学会　オーガナイズドセッション
「AI倫理・ガバナンス －イノベーションと規制の狭間で－」
10/7 東京大学未来ビジョン研究センター　リスクチェーンモデルの応用事例ウェビナー
Risk Chain Model to Practice「富士通AI倫理影響評価との合同ケース検討」
11/15 GPAIセッション企画・登壇
“The ways to trustworthy AI in practice”

４.AI倫理浸透のための取組みのご提案

AI倫理を社会に浸透させるために

FUJITSU

富士通のAI倫理に関する取組み

- AI4People加盟
- 富士通グループAIコミットメント
- 富士通グループAI倫理外部委員会
- 社内ガイドライン・研修の拡充
- AI倫理影響評価手法の開発等

AI倫理ガバナンスの
さらなる浸透に向けて
ユーザー企業や消費者を巻き込んだ対応が必要

国内産業界におけるAI倫理の普及に向けて

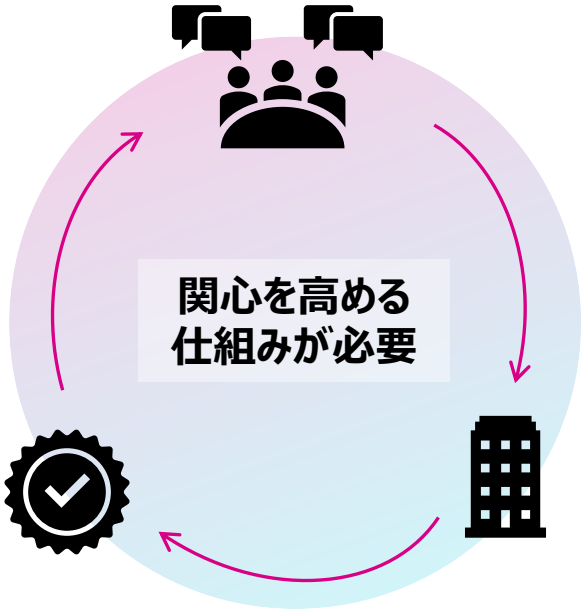
FUJITSU

AI倫理の議論のさらなる活性化を期待

産業競争力の維持・向上にはAIの利活用が必須であるところ、一部のIT企業にとどまっているAI倫理の議論を、産業界全体で活性化させる必要がある。

企業の取組みをポジティブに評価する仕組み

AI倫理に対する関心を高め、AIの利便性を最大限に享受するためには、企業の倫理に関する取組みをポジティブに評価する仕組みづくりが必要である。



Thank you