

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(AIEIA)との合同ケース検討」

日程：2022年10月07日（金）

時間：13:00-14:45

会場：オンライン(Zoom)

主催：東京大学未来ビジョン研究センター

共催：デロイトトーマツグループ Deloitte AI Institute、富士通株式会社

### はじめに

人工知能(AI)サービスの社会実装が拡大する一方で、AI の信頼性や透明性に関する問題が課題となっています。AI に係るリスクを低減するために様々なアプローチが存在していますが、その数の多さからどのアプローチをどのように使用したら良いか分かりにくい状況があります。そこで、東京大学の研究グループが開発した AI サービスの信頼性を確保するフレームワークである「Risk Chain Model(RCModel: リスクチェーンモデル)」と富士通(株)の研究チームが開発した「AI 倫理影響評価(AIEIA)」の二つのフレームワークを融合させることによって、どのようなシナジーが創出されるのかの検討を行いました。

本イベントでは、採用 AI を事例として合同ケース検討による研究発表を行い、複数の専門家によるパネルディスカッションを踏まえて、AI サービスの信頼性を実現するためのアプローチをどのように実践に落とししていくことができるのかの検討を行いました。

### 話題提供

「リスクチェーンモデルの説明」－東京大学未来ビジョン研究センター/デロイトトーマツグループ Deloitte AI Institute 松本 敬史 氏

東京大学未来ビジョン研究センター客員研究員及びデロイトトーマツグループ Deloitte AI Institute に所属する松本氏から、本研究会が掲げるテーマ及び方法論であるリスクチェーンモデルについての説明がありました。

AI サービスとリスクコーディネーション研究会では、AI サービスの社会実装が進む中で「信頼できる AI サービスマネジメント」の実現を目的としているとのことです。実現に向けた要点として「ビジネス目的と社会的信頼の両立」「様々な環境変化への対応」「ステークホルダーとの対話」を挙げていました。そして「信頼できる AI サービスマネジメント」の実現に向けて、以下3つのテーマで研究活動に取り組んでいるとのことです。

- ・方法論：リスクチェーンモデルの実践によるフレームワークの確立
- ・人材育成：AI サービスを現場でリードする人材
- ・リエゾン：マルチステークホルダーで現場リーダーを支援するコミュニケーション環境

方法論であるリスクチェーンモデルについての説明がありました。リスクチェーンモデル

は、AI サービスごとに検討すべきリスクが異なることを背景として、「AI システム」「サービス提供者」「ユーザー」を関連づけて、技術・非技術のアプローチで柔軟なリスク対策を検討するためのフレームワークであるとの説明がありました。ケースによっては利用者がリスク要因になり得るため、AI システムやサービス提供者だけではリスクを十分に防ぐことが難しいことも考えられるため、ユーザーにおけるリスク対策も検討していきます。

また、様々な環境の変化（予測対象データの分布、利用環境、利用者等）がリスクの要因になり得ることに対して、MLOps や XAI 等の様々なツールが開発されているものの、実際の現場では何のためにどう活用したらいいかが分からないということがあり、リスクチェーンを引くことによって、様々なツールの活用目的や活用方法を具体化していきます。

そして、リスクチェーンモデルを用いた一連の検討過程を様々なステークホルダーと共有することで、ステークホルダーに対する説明も果たしていくという狙いがあります。リスクチェーンモデルの検討は以下の 5 つのステップで実施されます。

Step1 AI サービスが「実現すべき価値・目的」の定義

Step2 「実現すべき価値・目的」を阻害する重要なリスクシナリオの検討

Step3 重要なリスクシナリオ毎にリスクチェーン（リスク要因の関係性）を検討

Step4 リスクチェーンで識別された構成要素ごとにリスクコントロール（リスク対策）を検討

Step5 各リスクチェーンでの検討を集約し、ステークホルダー毎の役割を整理

なおリスクチェーンモデルの利用ガイド・解説動画・ケース事例は以下の Web サイトに公開されています。

・東京大学未来ビジョン研究センター AI サービスとリスクコーディネーション研究会

<https://ifi.u-tokyo.ac.jp/projects/ai-service-and-risk-coordination/>

## 話題提供

「富士通「AI 倫理影響評価」の説明」－富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ 小野寺 佐知子 氏

富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ の小野寺氏から、富士通 AI 倫理研究センターの紹介及び AI 倫理影響評価（AIEIA）の説明がありました。

富士通株式会社では、AI が差別やプライバシー侵害のような人間の価値観・人権に悪影響を及ぼすとの問題提起や欧州 AI 規制の検討等を背景に、2019 年 3 月に「AI コミットメント」を発表し、AI 倫理の外部委員会の設置や全社的な AI 倫理教育等に取り組んできたそうです。そして、2021 年 4 月に AI 倫理研究センターを設立し、「AI 倫理影響評価（AIEIA）」という方法論を開発しています。

各国から AI 倫理原則やガイドラインが発表されているものの、AI 開発現場でどのように適用すべきかが明確になっておらず、AI 倫理に係る問題は継続して発生していることから、過去に発生した失敗事例を学習すべきと考えたことが、「AI 倫理影響評価」を開発するに至った背景であるとの説明がありました。

AI 倫理に係るリスクの捉え方は難しく、例えば「画像認識でリンゴの等級を評価する AI」を考えた場合、個人のリンゴ農家が自分の畑のリンゴを評価ならば問題にならないものの、農業団体が同じ AI を用いてリンゴの等級を評価すると、実質「リンゴ農家」のランク付けになり倫理的な問題になり得る可能性があります。このように同じ AI システムであっても「誰がどのように使うのか」を明確にすることが重要になるとのことです。

富士通が開発した「AI 倫理影響評価」では、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を重視し、AI システムにおける技術的な構成要素やステークホルダーとの関係を含めて「インタラクション」と定義しています。

AI システムにおけるインタラクションを明確にし、インタラクションの関係性に対応した「AI 倫理要件」を識別することで、どのような倫理的リスクが発生しうるかを検討できるようになります。「AI 倫理要件」は欧州等が出している AI 倫理ガイドラインを元に構造化した「AI 倫理モデル」があり、インタラクションの関係性によって導き出されるようになっています。Partnership on AI が公開している AI Incident Database から 15 件の事例を分析し、リスクを識別できることを検証しているそうです。

なお、「AI 倫理影響評価」は外部公開しており、誰でも参照できるものとなっています。

・富士通 AI 倫理影響評価 (AIEIA)

<https://www.fujitsu.com/jp/about/research/technology/aiethics/>

## 話題提供

「合同ケース評価 - 「AI 倫理影響評価」×「リスクチェーンモデル」 - 富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ 新田 泉 氏、東京大学未来ビジョン研究センター/デロイトトーマツグループ Deloitte AI Institute 松本 敬史 氏

富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ の新田氏と東京大学未来ビジョン研究センター/デロイトトーマツグループ Deloitte AI Institute の松本氏から、今回実施された「AI 倫理影響評価(AIEIA)」と「リスクチェーンモデル」による合同ケース評価で検討された内容についての説明がありました。

双方から合同ケース検討の狙いについて説明がありました。「AI 倫理影響評価 (AIEIA)」と「リスクチェーンモデル」を組合わせたケース検討を行うことで、ケース検討の深度を高めるだけでなく、双方の方法論における課題にアプローチするという狙いがありました。

松本氏から、リスクチェーンモデルにおいてはリスクシナリオの識別がステークホルダーの経験に依存するという課題があり、AI 倫理影響評価 (AIEIA) を活用することで「倫理に係るリスクシナリオ」を自動的に抽出できると考えたとの説明がありました。

新田氏から、AI 倫理影響評価 (AIEIA) はリスクを自動的に抽出できる一方で、AI 倫理モデルの参照データが倫理ガイドラインに限定されている中で、リスクチェーンモデルのケース検討に含まれる「ビジネスや社会の変化を含む幅広い影響」に係るリスクシナリオを識別し、その後の対策を含めて検討したいとの説明がありました。

検討に用いたケースは「採用 AI」になります。このケースは、既に東京大学 AI サービスとリスクコーディネーション研究会において、リスクチェーンモデルを用いたケース検討が行われており、今回実施した合同ケース検討を行った結果と比較評価が行われました。

[https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/10/RCModel\\_Case01\\_Recruitment-AI\\_JP.pdf](https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/10/RCModel_Case01_Recruitment-AI_JP.pdf)

ケース検討は、以下の流れに従い、東京大学と富士通の両研究チームが合同で実施されました。なお AI 倫理影響評価 (AIEIA) の AI 倫理モデルは欧州委員会が策定した「Ethics Guidelines for Trustworthy AI」をガイドラインデータとして参照しています。

- ① AIEIA：システム全体図におけるインタラクションを作成
- ② AIEIA：AI 倫理モデルからリスクシナリオを識別
- ③ リスクチェーンモデル：既存ケース検討で識別されたリスクシナリオと比較
- ④ リスクチェーンモデル：一部のリスクシナリオ（公平性のリスク）において AIEIA を参考にしながらかリスクチェーンを再考

・ Ethics Guidelines for Trustworthy AI

<https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

合同ケース検討の検討ステップ①～③を実施したところ、リスクチェーンモデルのみで検討されていた既存リスクシナリオ 16 件の内 7 件のリスクシナリオが AIEIA によって識別されました (図 1)。一方、以下のようなリスクは自動で識別することが難しいとの結果が見られたとのことです。

- ・ ビジネス環境／ユーザー側の変化に起因するリスク
- ・ 不正シナリオ（予測結果の目的外利用等）
- ・ 炎上リスク（風評被害）

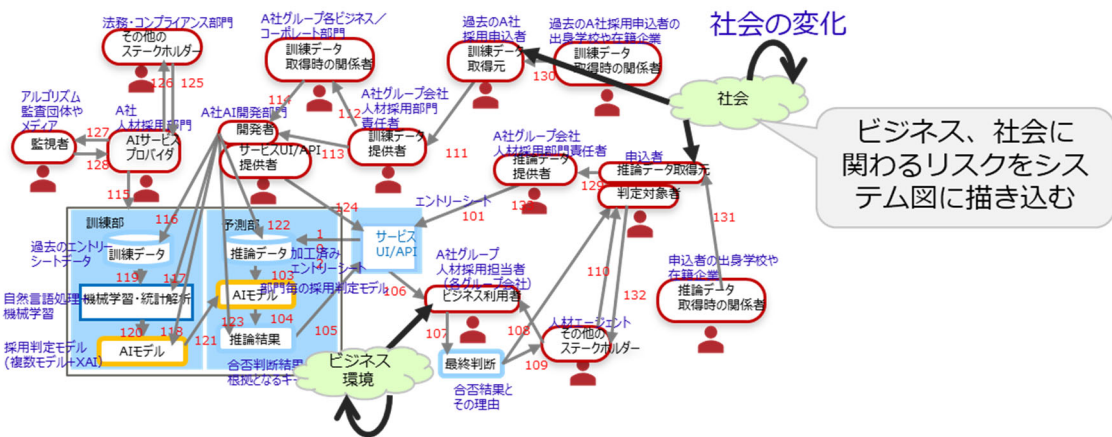
図 1. 合同ケース検討で識別されたリスク（発表資料より掲載）

実現すべき価値・目的	サービス要件と関連テクノロジー	リスクNo.	リスクシナリオ
1-1 予測性能の維持	<ul style="list-style-type: none"> <li>AIの予測精度</li> <li>AIの頑健性</li> <li>AIの説明可能性</li> </ul>	R001	適切な評価 採用する職種等により、適切な期待値を設定しなければAIの貢献を正しく評価できない
		R002	予測性能の維持 AIの予測性能が劣化したことで採用レベルが低下する
		R003	ノイズによる影響 エントリーシートで使用される文字情報が若干異なるだけで(句読点の違い等)、AIの判断結果が大きく変化する
		R004	虚偽の申込 申込内容に虚偽が含まれているが、合格と判断してしまう
		R005	過度なAI依存 人材採用担当者がAIの判断に依存しすぎること、AIの判断誤りに気が付かない
		R006	誤ったフィードバック 人材採用担当者によるAIへのフィードバック(合否のラベル設定)が不正確なことでAIの性能が劣化する
		R007	人材トレンドの変化 求める人材トレンドの変化にAIの予測傾向が対応できず、採用レベルが低下する
		R008	新たな職種 新たな職種、人材を求める際にAIモデルの予測が妥当でない可能性がある
2 採用活動に係るコストの削減	2-1 適切なサービスコスト	R009	コスト超過 サービス維持コストが超過する
3 海外グループを含めたサービス提供	3-1 サービスのローカライズ	R010	地域の会社への対応 地域の会社ごとに採用方針や申込者の傾向が異なることで、会社によって採用レベルが確保できなくなる
	3-2 広範囲な開発・サポート体制	R011	不十分な開発スピード 十分な開発体制が確保されていないことで、モデルの精度が劣化した際に再学習などの対応が行われない
4 企業の社会的責任(公平性のある採用活動)	4-1 倫理・コンプライアンス遵守	R012	判断根拠情報の不正販売 AIサービスを何度も利用することで高確率で合格判断を行うキープレース等を特定し、社外へ不正販売する
		R013	公平性 特定の国/地域/人種/性別/年齢に対して不公平な予測結果を生じさせる
	4-2 情報管理	R014	予測結果の目的外利用 AIの判断結果情報が別の目的に使用されることで、特定の人間に不利益を生じさせる
		R015	風評被害 AIの予測結果情報が外部に流出することで特定の人物に対して風評被害が発生する
		R016	プライバシー保護 AIシステムに蓄積された個人情報流出する

松本氏より、識別できなかったリスクについては、インタラクションとして「ステークホルダーの状態変化」を定義し、AI倫理モデルが参照するAIガイドラインのデータを拡充することで将来的に対応できる可能性があるのではないかとのコメントがありました。

新田氏からも、AI倫理影響評価(AIEIA)の今後の拡張性としてビジネス環境や社会に係るリスクをシステム図に書き込めるようにするという検討案について説明がありました。

図 2. AI倫理影響評価(AIEIA)の今後の拡張性



合同ケース検討の検討ステップ④において、1つのリスクシナリオ「R013 公平性」を検討したところ、AI倫理影響評価(AIEIA)で抽出された情報を参考にすることで新たなコントロール案が識別されたとのことでした。

検討ステップ②でAI倫理影響評価(AIEIA)を用いてAI倫理モデルから抽出された「リス

ク要因」を観察すると、「固有の影響・被害を想定できるリスク」と「リスク対策が機能しないリスク」の両方が抽出されており、後者のリスクを参考にリスクチェーンを見直していくと、既存のケース検討（複数の有識者と複数回検討したケース）に改善できる点が発見されたとのことです（図3）。

図3. 公平性のリスクシナリオで識別された新たなコントロール

①【リスクコントロール】 データ品質：教師ラベルの妥当性検証

AIEIA				リスクチェーンモデル				
始点	終点	I	リスク要因 (RCM結果参照前)	AIEIA→RCM	リスクシナリオ	リスクコントロール (AI System)	リスクコントロール (Service Provider)	リスクコントロール (User)
AIモデル	推論結果	104	1.採用AIが出力する可否判定結果が、保護グループ間で共有できない差がある	リスクシナリオ識別	R013 公平性：特定の国/地域/人種/性別/年齢に対して不公平な予測結果を生じさせる	①データ収集(データの偏り) ②AIモデルの妥当性検証	③システムの評価(公平性) ④情報開示・説明(AIの性能・リスク)の形態(利用者責任・役割分担)	⑤理解・リテラシー(AIの期待値・リスク)の検討(AIサービスの担当範囲) ⑥利用者の行動(最終判断)
開発者	訓練データ	116	20.過去の採用実績に不公平があるが、そのまま教師ラベルとして使っている	コントロール導出		データ品質：教師ラベルの妥当性を検証する ※AIEIA：追加コントロールを識別		
開発者	訓練データ	116	21.グループ会社や職種によってデータの品質にばらつきがあることを検証していない	コントロール導出		①データ収集(データの偏り)		

リスクチェーンモデルの検討で漏れていた項目を発見

②【リスクコントロール】 公平性の定期モニタリング

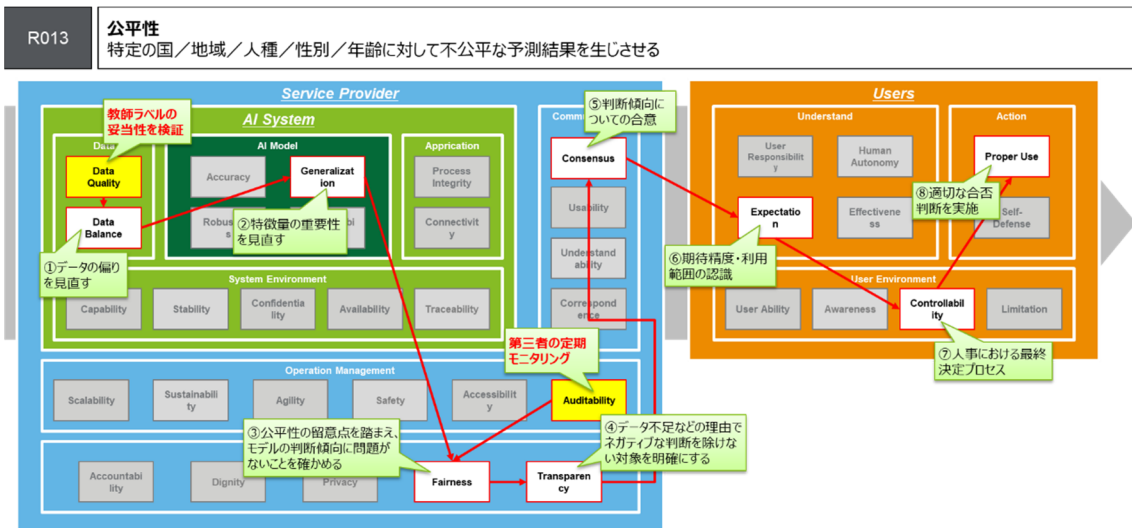
AIEIA				リスクチェーンモデル				
始点	終点	I	リスク要因 (RCM結果参照前)	AIEIA→RCM	リスクシナリオ	リスクコントロール (AI System)	リスクコントロール (Service Provider)	リスクコントロール (User)
開発者	AIモデル	118	23. 地域毎の公平性の達成基準が明確になっていない、あるいは測定できない	コントロール導出				
AIサービスプロバイダ	AIシステム	115	14. 最終判断の決定ルールを定期的に見直し、その結果として、地域毎の公平な採用の基準の変化に対応しない	コントロール導出			公平性の定期モニタリング ※AIEIA：追加コントロールを識別	
AIサービスプロバイダ	AIシステム	115	16. 地域毎の負の影響を受ける個人やグループの把握が十分でない	コントロール導出			①情報開示・説明(AIの性能・リスク)	

© 2022 Fujitsu Limited

リスクシナリオ「R013 公平性」におけるリスクチェーンの改良点として識別されたコントロールは以下の2つとのことでした（図4）。

- ・ [Data Quality] 教師ラベルの妥当性を検証
- ・ [Auditability] 第三者の定期モニタリング

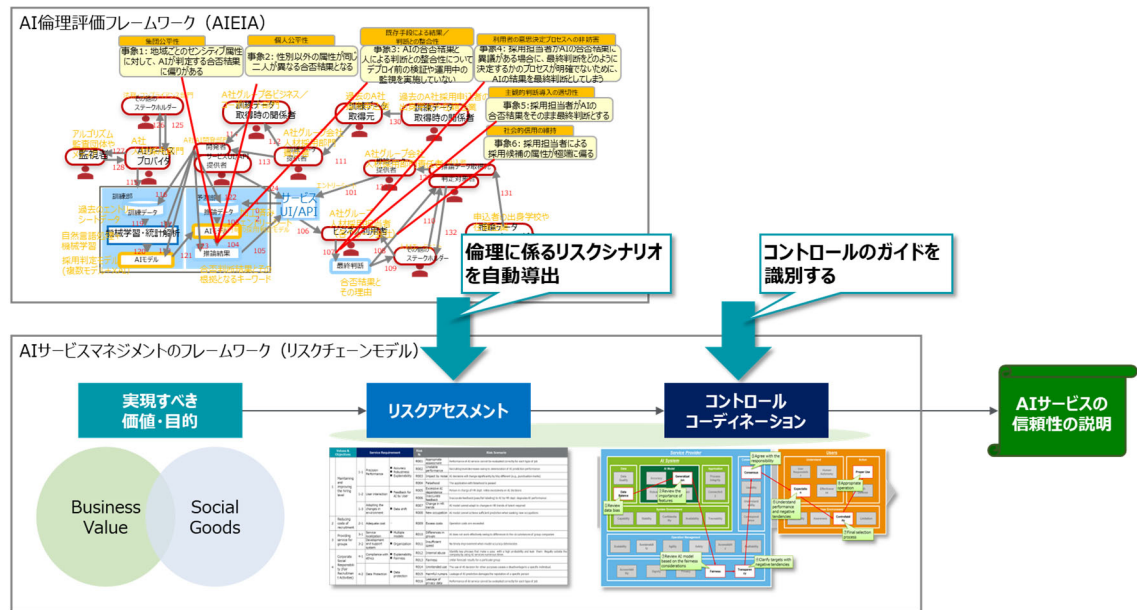
図4. リスクシナリオ「R013 公平性」におけるリスクチェーンの改良点



松本氏より、今回のケース検討を踏まえ、AI 倫理影響評価 (AIEIA) を用いながらリスクチェーンモデルの検討を進めること、以下のようにケース検討のプロセスを改善できるとの説明がありました (図 5)。

- ・リスクアセスメント：倫理に係るリスクシナリオを自動的に導出できる
- ・コントロールコーディネーション：リスク対策の検討時にガイド情報として利用できる

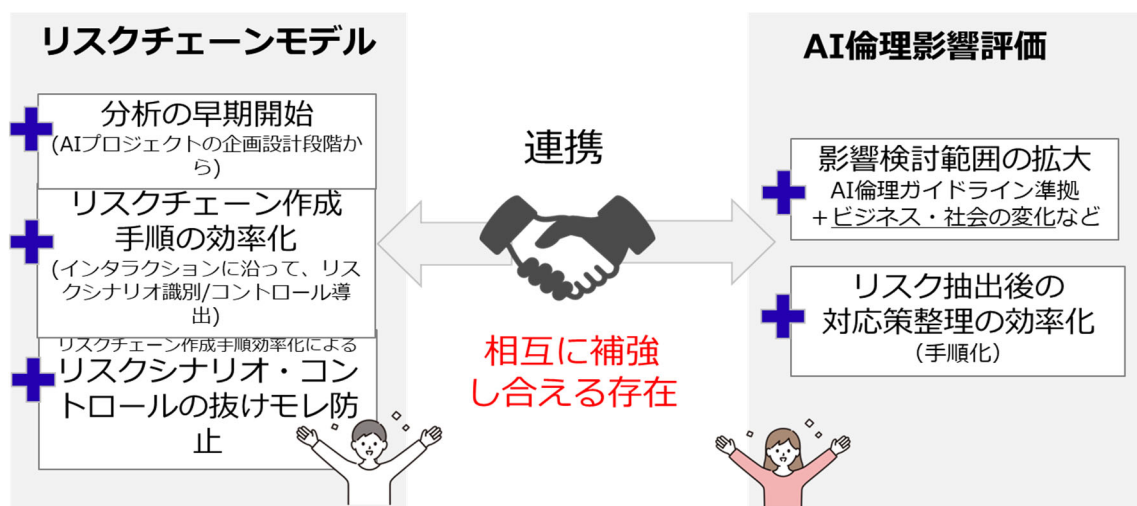
図 5. 「AI 倫理影響評価 (AIEIA)」 × 「リスクチェーンモデル」の合同ケース検討モデル



新田氏からも、リスクチェーンモデルと AI 倫理影響評価 (AIEIA) の方法論自体を相互に補強し合えるポイントについて説明がありました (図 5)。

- ・リスクチェーンモデル側が得られるメリット：分析の早期開始、リスクシナリオ識別の自動化、リスクチェーン検討時のガイド情報の提示
- ・AI 倫理影響評価 (AIEIA) が得られるメリット：影響範囲の拡大 (ビジネス・社会の変化への影響)、リスク対応策を整理する際の効率化

図 6. 相互連携によるメリットのイメージ



AI ガバナンスの実現に向けたフレームワークやツールは、世界中で多く研究開発されつつありますが、「どのフレームワークやツールを選択するか」というよりも、「様々なフレームワークやツールの良い所取りをする」というアプローチも可能であるとのことでした。

#### **パネルディスカッション**

パネルディスカッションは江間有沙准教授が進行しました。パネリストは話題提供者に以下専門家の方々を加えて実施されました。

東京大学未来ビジョン研究センター センター長 城山 英明 氏

京都大学法学研究科 特任教授/スマートガバナンス株式会社 代表取締役 CEO 羽深 宏樹 氏

富士通株式会社 研究本部 AI 倫理研究センター センター長 福田 大輔 氏

デロイトトーマツグループ Deloitte AI Institute 所長 森 正弥 氏

まず発表内容を踏まえて、専門家の方々から以下のコメントがありました。

#### **東京大学未来ビジョン研究センター センター長 城山 英明 氏からのコメント**

リスクチェーンモデルも AI 倫理影響評価 (AIEIA) も、基本的にはどちらもリスクアセスメントを起点としており共通した作業を行っているようである。しかし、リスクの対象範囲が異なり、リスクチェーンモデルはビジネス・社会を含むリスク全般、AI 倫理影響評価 (AIEIA) は倫理リスクに特化しているという点が前提となる。

今回の合同ケース検討を踏まえると、リスクチェーンモデルがリスクアセスメント時に実施していた「ステークホルダーに語ってもらうこと」では限界があることが分かった。ステークホルダーの視点で考えることは必要条件であるが、それだけでは十分ではない。

一方、AI 倫理影響評価 (AIEIA) を運用に落とし込むことについて考えると、使いこなすためにはアセッサー (評価者) の習熟度に依存することも考えられる。

双方の課題として、リスクアセスメントで識別したリスクに対して全てを十分に対応することは難しいことも考えられるため、どの範囲までリスクマネジメントしていくかという点を今後理論化していくことが必要と考えられる。

#### **京都大学法学研究科 特任教授/スマートガバナンス株式会社 代表取締役 CEO 羽深 宏樹 氏からのコメント**

今回の検討を踏まえると「アジャイル・ガバナンス」とも親和性が高いのではないかと。アジャイル・ガバナンスにおいては、複雑で不確実な社会の中で達成すべきゴール自体を含めてどのように変化に対応し、様々なステークホルダーが実施するガバナンスをどうやってコーディネートしていくかが重要となる。今後以下のような課題に対応していくことが必要と考えられる。



1. 様々なステークホルダーとのインタラクションがある中で全体最適なガバナンスを検討する際に、習熟度や価値観の異なるステークホルダー間で評価を実施し情報を共有していくために、リスクチェーンを用いてどのように方向性を併せていくか。
2. 社会環境・ビジネス環境・人々の価値観が変化していく中で、継続的にアセスメントを実施していくためにリソースやコストをどうやって確保していくか。
3. 社会全体で分散型のガバナンスを実現し継続していくために、こういったインセンティブ設計を行っていくか。

#### **富士通株式会社 研究本部 AI 倫理研究センター センター長 福田 大輔からのコメント**

AI の信頼性に係る様々な問題提起や規制が検討される一方で、中々そういった論点が社会で実践されて行かないという課題を感じており、テクノロジーによって自動化していくアプローチが必要と考えたことから、AI 倫理影響評価 (AIEIA) の開発を始めた。

リスクチェーンモデルとは親和性があり、一緒に検討を進めることが、世の中において AI 倫理やガバナンスの実践を広げることにつながると考えている。

AI 倫理影響評価 (AIEIA) を「誰がどう運用するか」という点は専門家でなければ分からない論点が多く、民主化を進めていくためにも事例を積み重ねながらシステマチックに問題を抽出していくことにアプローチしていきたい。

AI モデルの開発自体も自動化が進んできているが、信頼が伴わなければ結局市場に受け入れられないという問題を抱えているため、信頼の構築に関してもなるべく人手を削減してスピードアップし AI や DX を世の中に広げるまで進めていきたいと考えている。

#### **デロイトトーマツグループ Deloitte AI Institute 所長 森 正弥からのコメント**

デロイトグローバルでのサーベイによれば、グローバルの 76% の企業が「AI はピンポイントのソリューションではなく、社会を変化させるドライバーである」との認識を持っている。一方で、全体の 46% の企業がそのための準備ができていないと回答している。例えば、AI の運営に必要な組織マネジメントの構築・人材確保・ガバナンスポリシーの策定が出来ていないといったことが該当する。

そして、全体の 30% 以上の企業が AI に関して倫理的な問題が発生することに強い懸念を持っており、AI は社会を変えるドライバーでありながらも、そのガバナンス及びリスクコントロールが実現できていないことを課題に感じている。

AI が今後広く普及し社会を支えるインフラとなるためには、AI のガバナンス・リスクコントロールが必要であり、AI 倫理影響評価 (AIEIA) を活用して倫理に係る影響を調査することは基礎になっていくと考えられる。

リスクチェーンモデルと AI 倫理影響評価 (AIEIA) が相互補完的であるということ、ステークホルダーとは誰であるべきなのか、誰が方法論を使いこなせるのかという議論や検討を重ねることが、AI ガバナンスの確立に向けて多くの示唆を与えてくれると考えている。

専門家の方々のコメントを踏まえて、発表者から以下追加のコメントがありました。

**東京大学未来ビジョン研究センター/デロイトトーマツグループ Deloitte AI Institute 松本敬史 氏のコメント**

「採用 AI」のケースだけでも多くの有識者と 10 回以上の検討を重ねてきたが、それでも AI 倫理影響評価(AIEIA)で識別されたコントロールがあった。

ステークホルダーとの価値観を併せるというところは、参加者の多様性を確保することでバイアスの影響を薄めていくことを考えていたが、網羅性を確保する上でガイドライン情報も参考になることが理解できた。

今回検討したようなリスクシナリオは、実社会では誰かが顕在化しないように防いでくれているものも多く、インシデントの実例がなくとも具体的に検討していくことが必要なものも多いと考えている。

**富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ 小野寺 佐知子 氏のコメント**

AI 倫理影響評価 (AIEIA) としては、専門家ではない方でも倫理的なテーマに意識を持つことで AI ガイドラインの範囲で検討できるような仕組みにしていきたいと考えている。

これにより、人的リソースの負荷を下げることをねらっていききたいが、ビジネスや社会の変化に対応していくためには多様なステークホルダーの関与は必須と考えている。

リスクアセスメントからリスクマネジメントの一連の過程において、機械的にアプローチする部分と人間が対応すべき部分を見極めて、取り組みを広げていきたい。

**富士通株式会社 研究本部 AI 倫理研究センター AI 共生社会 PJ 新田 泉 氏のコメント**

ステークホルダーに語ってもらうことに限界があることと同じく、AI 倫理ガイドライン側においても必ずしも多様なステークホルダーの意見が表現されているわけではないという気づきがあった。

アジャイル・ガバナンスとの親和性という点でも今回のような取り組みを繋げていきたい。AI 倫理影響評価 (AIEIA) を一般的に使いこなしていくためには課題も多いため、ケース検討による実践を積み重ねることでブラッシュアップを図っていきたい。

その他、以下のようなトピックが議論されました。

**AI ガバナンスに係る検討を誰がどうやって始めて行ったら良いか？**

「組織全般の AI ガバナンス」と「AI サービスに特化したリスクマネジメント」によってリードすべき人が異なる。必ずしもデータサイエンティストとは限らない。

AI を詳しく理解した人に発言を求めるのか、純粋な利用者／受益者／関係者としての立場での発言を求めるのか、という点でステークホルダーからの意見を反映させる方法を検討する必要がある。

開発プロセスの企画段階でリスクアセスメントをすることができるのではないか。システム図やステークホルダーを検討する際にリスクを洗い出して開発を進めることが望ましい。一方で運用に入ってから利用者がステークホルダーの意見を反映する仕組みが必要と考えている。

### **AI ガバナンス実現に向けたインセンティブづくり**

ユーザー企業が必要性を感じにくいところがある。コストではなく、社会から称賛されるというベネフィットもアピールする必要があり、企業におけるエシックスの取組が称賛されるムーブメントと連動していきたい。

セキュリティ対策と同様に、問題が発生しないと対策が進まないようでは不十分であるため、信頼できる AI の認証制度等が必要になると考えられる。

政府から企業に向けた細則を定めていくというアプローチでは難しい。

### **AI ガバナンスを社会全体に広げていくために**

市民全てが AI について正しい理解を持つことはハードルが高いと考えられるが、「AI は過去データから学習するもの」「感情や意思を持って侵略してくるものではない」という基本的な理解は広げていくことが必要である。

グローバル平均では、3 割以上の企業が AI の適用領域も広く成果も出しているトランスフォーマーに分類されている中で、日本は未だ適用領域も成果もスターターに位置づけられている。AI の戦略的活用と AI ガバナンスは両輪であり、今回のようなアセスメントを進めていくことで、AI の適用領域を広げて成功事例を増やしていくことが必要である

アセスメントはあくまでプロセスであり、リスクを受容するかどうかの最終判断は当事者の自律性を残す形にしている。マネジメントが最終的に判断をする際に透明性を確保していくことが今回のようなケース検討の意義であり、事例を蓄積していくことが重要になる。事例を蓄積していくことで、検討すべきリスクの相場観ができてくると考えられる。

共通のプラットフォームとしては、産官学で形成していくことが必要である。

### **まとめ**

今回は「AI 倫理影響評価 (AIEIA)」と「リスクチェーンモデル」という 2 つの方法論を用いた合同ケース検討を行うことで、リスクシナリオとリスク対策の双方で検討の深度を上げることや、一部自動化による属人化の低減に寄与できることが検討出来ました。

日本における AI の社会実装を広げていくためにも、ビジネスや社会の変化に対応しながら信頼できる AI ガバナンスを実現するための方法論は必要です。

今回のようなケース事例を蓄積していくことが、AI ガバナンスの社会実装を広げていくことにつながると考えられます。



(上段左より) 城山氏、羽深氏、松本氏、新田氏  
(下段左より) 小野寺氏、森氏、福田氏、江間氏