

Title	企業間信頼構築のためのリスク情報伝達支援ツールの開発 : AIサービス・製品を題材にして
Sub Title	Development of risk information communication support tools for building trust between companies : on the issue of AI services and products
Author	高橋, 久実子(Takahashi, Kumiko) 白坂, 成功(Shirasaka, Seikō)
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2023
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2023年度システムエンジニアリング学 第361号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002023-0035

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

修士論文

2023 年度

企業間信頼構築のための
リスク情報伝達支援ツールの開発
： AI サービス・製品を題材にして

高橋 久実子
(学籍番号：82233282)

指導教員 白坂 成功

2024 年 3 月

慶應義塾大学大学院システムデザイン・マネジメント研究科
システムデザイン・マネジメント専攻

論 文 要 旨

学籍番号	82233282	氏 名	高橋 久実子
論 文 題 目： 企業間信頼構築のためのリスク情報伝達支援ツールの開発：AI サービス・製品を題材にして (内容の要旨)			
現代社会は新技術の進展に伴い予測不可能なリスクが常に存在し、それらに対応するためのリスクマネジメントの重要性が増している。特に AI について、新たな倫理的、法的、社会的なリスクを生み出す可能性についてと警鐘を鳴らす専門機関は、従来の製品・サービスに対するリスクマネジメントでは通用せず、ステークホルダー連携に基づくリスクマネジメントへの変革が必要であり、ステークホルダー間での対話促進と共通理解を動的にアップデートする体制を構築することが重要だと指摘している。しかしながら、その実践においては AI リスクそのものへの理解やステークホルダーとの信頼構築に基づく連携に向けたノウハウが不足しており、困難に直面していると実際の企業から意見が得られた。 そこで、本研究では AI 技術に対する社会的な信頼を高めることを社会的要請として、AI を使用した製品・サービスに対して適切なリスクマネジメントを図ることの主体となる AI システム開発企業、及び AI サービスプロバイダー企業間の信頼関係の維持・向上を本研究の上位目的とし、企業による情報提供のあり方及び社外環境の変化、ユーザーのリスク性向を踏まえ、リスク情報伝達に係る戦略の検討支援ツールを開発することを本研究の目的とした。提案ツールは、信頼行動とは信頼する者が知覚したリスクに対してリスクテイクを行うことであるという信頼とリスクの関係のモデルを前提とした。AI に対するリスク性向を把握したうえで、リスク情報がどのように伝達されているか、及び、社会的にどのような周知がなされているかに基づきリスク知覚状況を整理し、その整理結果に基づき、リスク情報提供のあり方の改善に向けたアクションを検討するためのステップ及びガイドで構成されたものである。 評価にあたっては、被験者 10 名に対し提案ツールを使用した実験を行い、被験者の記述内容とツール使用前後のアンケートを分析した。その結果、リスク知覚状況の整理やリスク情報提供の改善に向けた検討について有効性が示唆された。一方で、企業のリスク情報提供に対する前向きな意識を醸成することには有効性が確認できなかったことからさらなる検証が求められる。リスクコミュニケーション専門家による評価に基づき、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方を適切に改善できることについて有効性が示され、提案の妥当性が確認された。			
キーワード (5 語) AI (人工知能)、ガバナンス、リスクマネジメント、リスクコミュニケーション、信頼構築			

SUMMARY OF MASTER’S DISSERTATION

Student Identification Number	82233282	Name	Kumiko Takahashi
Title: Development of Risk Information Communication Support Tools for Building Trust between Companies: on the Issue of AI Services and Products			
Abstract Today, unpredictable risks are always present with the advancement of new technologies, and risk management to deal with these risks is becoming increasingly important. Professional organizations have warned about the potential for AI to create new ethical, legal, and social risks, pointing out that conventional risk management for products and services will not work, and that a change to risk management based on stakeholder collaboration is needed. They point out that it is important to establish a system that dynamically updates the promotion and common understanding among stakeholders. However, the study also found that companies are facing difficulties in implementing this change due to a lack of understanding of AI risk itself and know-how for cooperation based on building trust with stakeholders. Therefore, this study aims to maintain and improve trust relationships between AI system developers and AI service providers, who are the main actors in appropriate risk management for AI-based products and services, as a social requirement to enhance public trust in AI technology. The objective of this study is to develop a tool to assist companies in considering strategies for communicating risk information, based on the way information is provided by companies, changes in the external environment, and the risk propensities of users. The proposed tool is based on a model of the relationship between trust and risk, in which trust behavior is risk-taking in response to perceived risk by those who trust. The model consists of steps and guides to organize the risk perception situation based on how risk information is communicated and how it is communicated to the public, and to consider actions to improve the way risk information is provided based on the results of the organizing process. For the evaluation, we conducted an experiment using the proposed tool with 10 subjects and analyzed their descriptions and questionnaires before and after using the tool. The results suggest that the tool is effective in organizing the risk perception situation and considering actions to improve risk information provision. On the other hand, further validation is required because the effectiveness of the tool in fostering a positive attitude toward risk information provision in companies could not be confirmed. Based on the evaluation by practical experts in risk communication, the effectiveness of the proposal was demonstrated regarding the ability to appropriately improve the way risk information is provided to maintain and improve trust among companies, and the validity of the proposal was confirmed.			
Key Word (5 words) AI (Artificial Intelligence), governance, risk management, risk communication, trust building			

目次

第1章	序論	1
1.1	研究の背景と社会的課題	1
1.2	先行研究が示す課題及び先行研究における不足	2
1.3	研究の目的と新規性	3
1.4	用語の定義	4
1.4.1	AI	4
1.4.2	リスク	4
1.4.3	リスクマネジメント	4
1.4.4	リスクコミュニケーション	4
1.5	本論文の構成	5
第2章	先行研究	6
2.1	本章の概要	6
2.2	AI リスクマネジメントに関する先行研究	6
2.3	信頼構築及び信頼毀損に関する研究	7
2.4	リスク情報提供に関する研究	9
2.5	本章のまとめ	10
第3章	ツールの提案	11
3.1	提案の全体像と対象者	11
3.1.1	提案の全体像	11
3.1.2	提案に対する要求	11
3.1.3	対象となる企業、使用フェーズ	15
3.2	提案設計の大枠：信頼のメカニズムの活用	16
3.3	AI に対するリスク性向把握プロセスの設計	17
3.3.1	先行研究及びインタビューに基づく機能要求定義	17
3.3.2	ステップの提案	21
3.4	リスク知覚状況の把握プロセスの設計	24
3.4.1	先行研究に基づく機能要求定義	25
3.4.2	ステップの提案	27
3.5	リスク性向、リスク知覚状況を踏まえ知覚リスク適正化に向けた情報提供の改善 プロセスの設計	36
3.5.1	先行研究及びプロトタイプ実施に基づく機能要求定義	36
3.5.2	ステップの提案	41
3.6	設計したステップの統合	44
3.7	手順書、提案の使い方	46

3.7.1	手順書	46
3.7.2	プラットフォーム	48
3.7.3	ツールの使い方	49
第 4 章	検証と妥当性確認	51
4.1	検証・妥当性確認の方法	51
4.2	検証・妥当性確認における評価計画	53
4.2.1	筆者作成の評価基準による評価	53
4.2.2	リスクコミュニケーション専門家による評価	55
4.2.3	ケーススタディの類似事例との比較による評価	55
4.3	実験概要	55
4.3.1	被験者	56
4.3.2	実験内容	57
4.4	検証の評価結果	70
4.4.1	ツールの理解性・利用性に関する検証結果	70
4.4.2	「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」 に関する検証結果	71
4.4.3	「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」 に関する検証結果	73
4.4.4	「企業のリスク情報提供に対する具体的なアクションを支援できること」に 関する検証結果	80
4.4.5	「企業のリスク情報提供に対する前向きな意識を促進すること」に関する検 証結果 86	
4.5	妥当性確認の評価結果	90
4.5.1	筆者作成の評価基準による評価結果	90
4.5.2	リスクコミュニケーション専門家による評価結果	98
4.6	考察	101
4.6.1	検証結果の考察	101
4.6.2	妥当性確認結果の考察	102
4.6.3	本提案ツールの限界	106
第 5 章	結論と今後の展望	108
5.1	結論	108
5.2	今後の展望	108
	謝辞	110
	参考文献	111
	Appendix	114

図目次

図 1-1 本研究の新規性.....	3
図 2-1 T.K.Das ら（2004）信頼とリスクのフレームワークを基に筆者作成.....	8
図 2-2 Robert B Cialdini（2014）による隠蔽の影響図を基に筆者作成	9
図 3-1 提案全体像.....	11
図 3-2 AI 製品・サービスのサプライチェーン	16
図 3-3	24
図 3-4 AI システム構成要素の構造化.....	30
図 3-5 リスク情報提供状況の可視化フレームワーク	31
図 3-6 社会的リスク喚起状況と顕在化する事例を構成する要素モデル	32
図 3-7 社会的リスク喚起状況と顕在化する事例を構成する要素モデルを通したステップ実施イメージ	33
図 3-8 リスク知覚状況マトリクス	34
図 3-9 4 象限への振り分けの考え方	35
図 3-10 4 象限への振り分け作業イメージ	35
図 3-11 各 AI リスク項目に関するリスク知覚状況（プロトタイプ）	37
図 3-12 各象限について改善に向けたアドバイスが記載されたフレームワーク（プロトタイプ）	38
図 3-13 象限ごとのリスク知覚状況に対する注意事項	42
図 3-14 リスク情報提供内容・方法の改善の検討ステップのフレームワーク（第一象限を例に）	43
図 3-15 リスク情報提供内容・方法の改善の検討ステップのフレームワークの記入手順.....	44
図 3-16 リスク情報提供内容・方法の改善の検討ステップのフレームワークの記載例（第一象限を例に）	44
図 3-17 提案ツールのステップ構成.....	45
図 3-18 概要及び趣旨.....	47
図 3-19 作業の流れ.....	47
図 3-20 アウトプットイメージ.....	48
図 3-21 Miro ボード上のツール全体像.....	50
図 4-1 ケーススタディにおける着眼点	64
図 4-2 STEP3 における検討フレーム	77
図 4-3 インタビューに使用した STEP5 成果物一覧表	85
図 4-4 ISO31000 リスクマネジメントに基づく研究の境界の整理	107

表目次

表 3-1 提案に対する要求の整理.....	13
表 3-2 General Attitudes toward Artificial Intelligence Scale (GAISS)の項目	18
表 3-3 Attitude Toward Artificial Intelligence scale (ATAI) の項目	19
表 3-4 インタビュー対象者一覧.....	19
表 3-5 大企業の担当者の回答.....	20
表 3-6 ベンチャー企業の担当者の回答	20
表 3-7 AIAS-5 (AI Attitude Scale)	22
表 3-8 標準的な AI リスク項目の整理.....	28
表 3-9 AI リスク項目一覧表（採用 AI を対象とした記入イメージ）	29
表 3-10 AI システム構成要素の説明	30
表 3-11 プロトタイプ参加者一覧.....	36
表 3-12 プロトタイプを通し得られたフィードバック	39
表 3-13 ツールの全体所要時間の目安	49
表 4-1 提案ツールの検証・妥当性確認の設計	52
表 4-2 筆者作成の STEP5 成果物の評価基準.....	54
表 4-3 被験者一覧.....	57
表 4-4 事前アンケートの質問項目	59
表 4-5 ケーススタディの AI システムの内容	64
表 4-6 ケーススタディにおける前提条件としての AI リスク一覧表（事前提供資料）	66
表 4-7 提案ツールの理解性・利用性に関するアンケート結果	71
表 4-8 質問項目「実務上でステークホルダーのリスク性向（リスクに対する考え方や 前向きさ等）の把握の方法を理解ができている」のアンケート結果	72
表 4-9 Wilcoxon の符号付き順位検定の結果.....	73
表 4-10 質問項目「リスク情報を整理するために何を調査すべきかが理解ができてい る」のアンケート結果.....	75
表 4-11 Wilcoxon の符号付き順位検定の結果	75
表 4-12 質問項目「適切にステークホルダーに対してリスク情報を提供するための客 観的な状況の把握・整理ができる」のアンケート結果	76
表 4-13 Wilcoxon の符号付き順位検定の結果	76
表 4-14 STEP3 の記載内容の一覧	79
表 4-15 インタビューに基づく STEP3 にて新たな気づきや洞察を得られた件数	80
表 4-16 質問項目「（ステークホルダーに対する）リスク情報提供の検討に必要なプロ セスを理解ができてい」のアンケート結果	82

表 4-17	Wilcoxon の符号付き順位検定の結果	82
表 4-18	質問項目「(ステークホルダーに対する) リスク情報提供に向けて必要なアクションを実行できる」のアンケート結果	83
表 4-19	Wilcoxon の符号付き順位検定の結果	84
表 4-20	インタビューに基づく STEP5 のインサイト件数	86
表 4-21	質問項目「リスク情報をステークホルダーに提供することの重要性を理解している」のアンケート結果	88
表 4-22	Wilcoxon の符号付き順位検定の結果	88
表 4-23	質問項目「リスク情報をステークホルダーに提供することのモチベーションがある」のアンケート結果	89
表 4-24	Wilcoxon の符号付き順位検定の結果	89
表 4-25	筆者作成の STEP5 成果物の評価基準	90
表 4-26	各被験者の STEP5 の成果物に対する評価表 (被験者 ID:01 の例)	92
表 4-27	被験者の STEP5 成果物に対する評価結果	94
表 4-28	筆者作成の評価基準に基づく評価結果の分布	96
表 4-29	STEP5 成果物の記入文字数と評価結果の一覧	97
表 4-30	STEP5 成果物に関する評価結果のリスク対応、コミュニケーション、その 2 つの観点の合計値、STEP5 成果物の文字数の相関関係	97
表 4-31	専門家によるリスク対応の観点からの評価結果	99
表 4-32	専門家によるコミュニケーションの観点からの評価結果	100
表 4-33	評価結果一覧	101
表 4-34	検証結果一覧	102
表 4-35	妥当性確認の一覧	103
表 4-36	被験者 ID:07 に対するインタビュー結果	104
表 4-37	被験者 ID:09 に対するインタビュー結果	105

第1章 序論

1.1 研究の背景と社会的課題

現代社会は、予測不可能なリスクが常に存在し、それらに対応するためのリスクマネジメントの重要性が増している。我が国においては、Society 5.0¹の概念が提唱されているが、サイバー空間とフィジカル空間が融合する Society5.0 においては、「コト」を中心としたサービス・ビジネスの構築がより進み、産業の SoS（システムオブシステムズ）化によって異なる目的を持つシステム同士が接続・連携してサービスを提供することとなる。さらに、人工知能（以下、「AI」という）により、必要な情報が必要な時に提供されるようになり、今までにない新たな価値を生み出すことが目指されている。しかし、こうして形成されるシステムの全体は複雑化する。具体的には、落合ら（2019）の指摘する接続されたシステム同士の相互作用によって新たな機能・振る舞いとなる創発特性や、Centre for Data Ethics and Innovation（2020）の整理する AI 特有のリスクである使用方法・学習内容によって予期しない機能性の変化が生じることで、予見・予防が困難な事態も発生し、市民の安全安心を脅かすリスクは高まるといったことが挙げられる。そのような複雑で変化の速い先進テクノロジーを生かすために、私たちは絶えず変化する環境、技術、そして目標に適応し、先進テクノロジーがもたらすリスクに対応するための従来とは異なる新たなリスクマネジメントの仕組みを必要としている。

Society5.0 の実現に向けて様々なテクノロジーの活用が見込まれるが、本研究では、近年驚くほどのスピードで活用が進む AI に焦点を当てた。Centre for Data Ethics and Innovation（2020）の整理する代表的な AI 特有のリスクには、前述の環境変化の考慮が必要であることのみならず、使用フェーズのインシデント検知が困難であること、人権などを含むより広い観点が必要であることが謳われている。そのような AI 特有のリスクの性質を踏まえ、松本敬史・江間有沙（2020）は、AI に対するリスクマネジメントは、多岐にわたるステークホルダーの積極的な関与を必要とし、例えば、データ提供や AI モデル開発を外部委託している場合は AI サービス提供者だけでリスクの十分な低減が難しいことを強調している。NIST（2021）は、従来のような一企業単独で完結するリスクマネジメントから、複数のステークホルダー間の連携に基づいたリスクマネジメントのフレームワークを提示している。

しかし、Kai Jia & Nan Zhang（2022）の主張するように、現存する AI のリスクマネジメントやガバナンスに係るガイドラインはあくまでガバナンス対応に実質的に焦点を当ててのみで、ガイドライン全体として重要な AI リスクをすべてカバーするこ

¹ 内閣府 Society 5.0 とは (https://www8.cao.go.jp/cstp/society5_0/index.html)

とができるのかを理解することなく、多様な原則を提案している。AIに係るステークホルダー企業がAI原則の必要性和有効性を遵守するためには、AIがどのようなリスクを生み出すかを明確に理解し、それに基づく取組を実施することが必要である。AIシステム開発企業にヒアリングしたところ、実際に、リスクマネジメントにおいてAIサービスプロバイダー企業等との連携の重要性を認識しているものの、その実践においてはAIリスクそのものへ理解やステークホルダーとの信頼構築に基づく連携に向けたノウハウが不足しており、困難に直面しているとの意見があった。

1.2 先行研究が示す課題及び先行研究における不足

これまでのAIを対象としたリスクマネジメントやガバナンスに係る研究分野においては、AIの活用促進には社会からの信頼が必要であり、リスクに関する透明性を担保する重要性も強調されている。OECD²の取りまとめた信頼できるAIの責任ある普及についての5つの価値に基づく原則（以下、「OECDのAI原則」という）を発端として、その原則を採択した各国で独自のAIガバナンスやリスクマネジメントのあり方の検討が進められている。我が国においても、2019年3月、OECDのAI原則の検討に参画した統合イノベーション戦略推進会議により、「人間中心のAI社会原則」を公表がなされた。「人間中心のAI社会原則」には、社会全体が主体となり実現すべきAI社会原則が定められるとともに、この原則を踏まえて、AIの開発・運用等の当事者となる事業者が、各自のAIの開発・運用等の目的や方法等に応じ、実施すべき目標（AI開発利用原則）を自ら定め、遵守すべきであることが示されている。さらに、「人間中心のAI社会原則」に基づくガイドラインの設計が様々な主体によって進められている。総務省の報告書³によれば、このようなOECDによるAI原則に基づく取組は、各国政府や研究者らが公表しているAI関連のガイドラインは約40本に及ぶという。その数からも分かるように、国際的に様々な主体がAI関連のガイドラインについて、様々な観点や主張を持ち、取りまとめを行っている。しかしながら、当事者となるAI開発企業によると、これらのガイドラインは具体的な手法やユースケースを提示しておらず、すぐに活用可能なノウハウではないことを課題として認識している。

また、OECDのAI原則は、AI技術に対する社会的な信頼を高めることが、その技術の責任ある普及と社会的、経済的利益の最大化に不可欠であると指摘している。そして、一度失われた信頼は完全に回復することが難しい（Lewicki, 2016）という信頼毀損

² OECD Recommendation of the Council on Artificial Intelligence
(<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>)

³ 総務省 AI ネットワーク社会推進会議「報告書 2022～「安心・安全で信頼性のある AI の社会実装」の更なる推進～」(2022年7月25日)

の研究分野における一般的な認識を考慮すると、AI 技術の社会実装を迅速に進めるためには、AI サービスや製品に関わるステークホルダー間の信頼関係の維持と、信頼の毀損を未然に防ぐことが重要と言えよう。しかし、信頼分野の研究は、主に信頼回復のアプローチに関する構成要素・評価に焦点を当てたものが多く、信頼毀損の未然防止のための具体的な方策については提供されていない。この点は、AI 技術の普及にあたっても重要な考慮事項であり、今後の課題として残るものである。

1.3 研究の目的と新規性

本研究では、AI システムに係るステークホルダーのうち、リスクマネジメントの根幹を為す AI システム開発企業と AI サービスプロバイダー企業に焦点を当て、これらの企業間の信頼関係の維持・向上を促進させるための、企業間のリスク情報提供に着目した「リスク情報提供のあり方検討支援ツール」を提案する。

先行研究と照らし合わせた本研究の新規性を図 1-1 に示す。これまでの AI ガバナンス・リスクマネジメント分野で整理された知見を基に、信頼をキーワードに AI システム開発企業と AI サービスプロバイダー企業間のコミュニケーションに焦点を当て、実際に企業が活用可能なツールとして具体化するところに新規性がある。

AI 技術に対する社会的な信頼を高めることを社会的要請として、AI を使用した製品・サービスに対して適切なリスクマネジメントを図ることの主体となるリスクマネジメントの主体となる AI システム開発企業とサービスプロバイダー企業との間の信頼関係の維持・向上に貢献することを本研究の上位目的とし、企業による情報提供のあり方及び社外環境の変化、ユーザーのリスク性向を踏まえ、リスク情報伝達に係る戦略の検討支援ツールを提案することを本研究の目的とする。

	先行研究の内容	残課題/先行研究による課題
AIリスク マネジメント	ステークホルダー間でリスクに関して対話促進と共通理解を動的にアップデートする体制構築が重要(松本,2014)	マネジメントに必要な機能は先行研究で示されているが、 対話に向けた具体的なツールや手法は提示されていない
信頼構築及び信頼 毀損	リスクの知覚程度が信頼構築過程で重要な要素 (T.K.Das,2024) であり、隠蔽は想定より有害な損益を生む (Robert B, 2014)	リスク情報の隠蔽による悪影響を上層幹部を含む 企業全体がより深く理解する必要

図 1-1 本研究の新規性

1.4 用語の定義

1.4.1 AI

人工知能（AI: artificial intelligence）については明確な定義は存在しないものの、一般社団法人人工知能学会によると「大量の知識データに対して、高度な推論を的確に行うことを目指したもの」（一般社団法人 人工知能学会設立趣意書からの抜粋）とされている。従来の機械学習では、精度を左右する特徴量の設計を人間が行う必要があったのに対し、新しい機械学習方法である「ディープラーニング（深層学習）」では、入力されたデータをもとに、コンピュータが自ら特徴量を導き出すことができる。ディープラーニングのように学習に使う特徴量を自力で獲得することができる画期的な人工知能アルゴリズムの登場、データ量の増大、コンピュータの計算性能向上などにより、技術開発のスピードが加速している。また、近年においては、あらかじめ膨大な量の情報から深層学習によって構築した大規模言語モデル（LLM（Large Language Models））に基づき、ある単語や文章の次に来る単語や文章を推測し応答する生成 AI の開発・活用が急速に進められているが、本研究において生成 AI は検討の対象外とする。

1.4.2 リスク

ISO31000 の定義に基づき、リスクとは目的に対する不確かさの影響とする。影響とは、期待されていることから乖離することをいう。影響には、好ましいもの、好ましくないもの、又はその両方の場合があり得る。

1.4.3 リスクマネジメント

ISO31000 の定義に基づき、リスクマネジメントとはリスクについて組織を指揮統制するための調整された活動とする。リスクマネジメントプロセスには、方針、手順及び方策を、コミュニケーション及び協議、状況の確定、並びにリスクのアセスメント、対応、モニタリング、レビュー、記録作成及び報告の活動に体系的に適用することが含まれる。

1.4.4 リスクコミュニケーション

厚生労働省によると、リスク分析の全過程において、リスク評価者、リスク管理者、消費者、事業者、研究者、その他の関係者の間で、情報および意見を相互に交換するこ

とである。リスク評価の結果およびリスク管理の決定事項の説明を含む。

1.5 本論文の構成

本論文の構成について説明する。

1 章の緒論では、本研究の背景及び、AI に係るガバナンス及びリスクマネジメントにおける問題点を論じた上で、本研究の目的を提示した。その後、先行研究では解決されていない課題に触れると共に、本研究の新規性について言及している。加えて、本研究における用語を定義した。

2 章では、AI に対するリスクマネジメント、信頼構築・信頼毀損、リスク情報提供に関する説明と先行研究を整理した。上記の 3 つの観点から既存研究を整理し、信頼とリスクの関係メカニズムについて概要を説明した上で、本研究における論点を示した。

3 章では、本研究で提案したツールの提案内容を示した。提案ツールは、全部で 5 つのステップに分かれており、AI プロバイダー企業の AI に対するリスク性向を把握したうえで、リスク情報が AI 開発企業からどのように伝達されているか、及び、社会的にどのような周知がなされているかを整理し、AI サービスプロバイダー企業のリスク知覚状況を検討し、その整理結果に基づき、リスク情報提供のあり方の改善に向けたアクションを検討する構造である。構造に併せてツールの使い方や想定使用者についても説明した。

4 章では、提案ツールの効果を確認するための評価について説明した。ケーススタディに基づき提案ツールを使用した 10 名の被験者に対するアンケート及び提案ツールの使用結果である成果物を分析し、評価を行った。

5 章では、本研究の結論を明らかにし、今後の展望について論じた。

第2章 先行研究

2.1 本章の概要

本章では、AI に係るリスクマネジメントとそのマネジメント対象として言及されるリスク項目に関する先行研究を整理する。また、上位目的としたリスクマネジメントの主体となる AI システム開発企業とサービスプロバイダー企業との間の信頼関係の維持・向上に貢献することに関して、企業間の信頼についての考え方を整理する。それらを踏まえ、信頼という観点から AI リスクマネジメントのあり方について概観し、本研究の位置づけを述べる。

2.2 AI リスクマネジメントに関する先行研究

AI リスクに関する先行研究では、AI が新たな倫理的、法的、社会的なリスクを生み出すと提唱している (Floridi & Cowls, 2019)。他の先行研究では、より具体的な課題について詳細に論じているが、その中でも主に 3 つの課題が挙げられている。第一に、技術的な説明不可能性、すなわち「ブラックボックス」は、法的な説明責任を確認しなければならないときに、異なる利害関係者の境界を曖昧にする恐れがあること (Liu et al.)。第二に、AI の自己強化メカニズムは、偏見、差別、エコーチェンバーなどの既存の社会問題を悪化させる恐れがあること (Nelson, 2019)。第三に、AI の自律性は、特にかつて人間にのみ権利が与えられていた分野において、倫理的あるいは法的な問題を引き起こす恐れがあること (Balkin, 2018)。これらの AI の特性に基づく、潜在的なリスクを認識することで、関連するステークホルダーが自ら主体的に規制するよう動機づけるために様々なガイドラインが施行されている。政府、民間セクター、市民社会、アカデミア等の異なる主体による合計 123 件の AI ガイドラインを分析した Kai Jia (2022) によると、リスクに関して次の 8 つの論点がガイドラインの中で頻繁に繰り返されているという。

- ① 透明性と信頼
- ② バイアスと公平性
- ③ プライバシー
- ④ インピュテーション
- ⑤ Autonomy
- ⑥ ロバスト性
- ⑦ ソーシャルセキュリティ

これらの AI システムに関連するリスクに対処するためには、従来のリスク管理者のみならず、コンピュータサイエンス、バイオインフォマティクス、データプライバシーの専門家などの複数の主体が協力する必要性を強調している (Banja 2020)。

NIST (2022) の提唱する AI リスクマネジメントに関するフレームワークである AI RMF でも、リスクマネジメントシステムに必要な機能として、AI のライフサイクルステージ全体に対して、①Govern 機能、②MAP 機能、③Measure 機能、④Manage 機能の 4 つの機能が必要であるとの整理が行われており、それらの機能の実行のためには、ライフサイクルステージの各ステージを跨いだステークホルダーの連携の重要性を謳っている。

特に、我が国においては、データの取得・AI モデルの開発・サービス提供の全てを一社で実現している米国の大企業とは異なり、サービスの提供者・AI モデルの開発者・実行環境の提供者・データの提供者が全て異なるケースが多く、AI サービス提供者が複数のリスク要因を包括的に検討するためには、ステークホルダー間で対話するための枠組みが必要となる。さらに、その協力関係における重要な要素として、AI サービス提供者、AI 開発者、利用者の間での対話促進と共通理解を動的にアップデートする体制を構築することが重要だと指摘されている (松本敬史,2014)。

2.3 信頼構築及び信頼毀損に関する研究

Mayer (1995) は提案する信頼の定義において、リスクは信頼のモデルに不可欠な要素であると論じている。脆弱であろうとする意思 (すなわち、信頼すること) にはリスクはないが、脆弱であろうとする意思の行動表出にはリスクが内在している。信頼と信頼行動の根本的な違いは、リスクを引き受ける「意思」と実際にリスクを「引き受ける」ことの違いである。つまり、信頼とはリスクを引き受ける意思であり、行動的信頼とはリスクを引き受けることであると整理している。この整理において、「Perceived Risk (知覚されるリスク)」という概念が重要な役割を果たしている。「Perceived Risk (知覚されるリスク)」は、主体が信頼関係において感じるリスクの度合いを指し、信頼関係が成立するためには、信頼する者が信頼される者に対して何らかのリスクを負う意思が必要であり、信頼があるからこそ主体はリスクを取る意思決定をするということである。

T.K.Das ら (2004) はこの行動的信頼の定義に則り、信頼とリスクの関係を理解するためのフレームワークを示している (図 2-1)。このフレームワークでは、信頼とリスクの関係において、リスク性向や信頼性向などの個人特性が果たす役割も考慮されている。リスク性向とは、リスクを取る意欲のことであり、信頼性向とは、他人を信頼する一般的な傾向のことである。このフレームワークは、これらの特性が主観的信頼と知

覚的リスクの両方に影響を与え、リスクテイク行動に関与する可能性があることを示唆している。

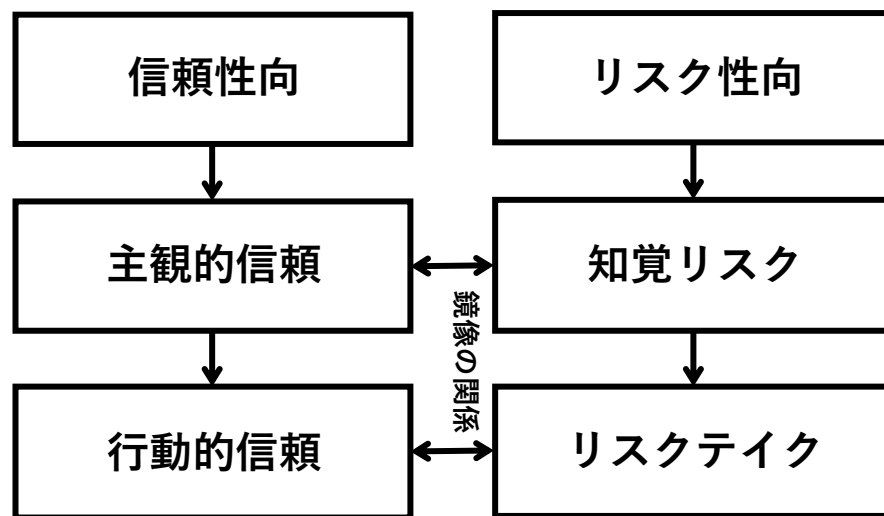


図 2-1 T.K.Das ら（2004）信頼とリスクのフレームワークを基に筆者作成

製品がますます複雑になり、技術がますます進歩するにつれて、企業間のパートナーシップにおいて、信頼は「ほとんどの協力関係をつなぎとめる接着剤」（Lewicki RJ,Bunker BB.1996）と言われる程の不可欠な要素となりつつある。しかし、ひとたび信頼が失われると、企業間のパートナーシップは、ますます苛酷な相互作用や、警戒心、用心深さ、用心深さの雰囲気蔓延する滑り台のような坂道を下っていく可能性が高く（Ariño & de la Torre, 1998; Doz, 1996）、信頼が元のレベルに完全に回復することはあり得ない事態であるとも言われている（Lewicki RJ,Bunker BB.1996）。

さらに、図 2-2 Robert B Cialdini（2014）による隠蔽の影響図を基に筆者作成に示すメカニズムのように、隠蔽をすることによる悪影響は単なる長期的な利益の低下のみならず、様々な企業活動に影響を与え、多方面でのコスト増加に繋がることが示されている。つまり、企業が戦略的に信頼関係を構築するためにリスクを隠蔽する行為は、その企業組織事態と利害関係者に従来認識されているよりもはるかに有害な損害がその企業自体にももたらされることを示唆している（Robert B Cialdini, 2014）。

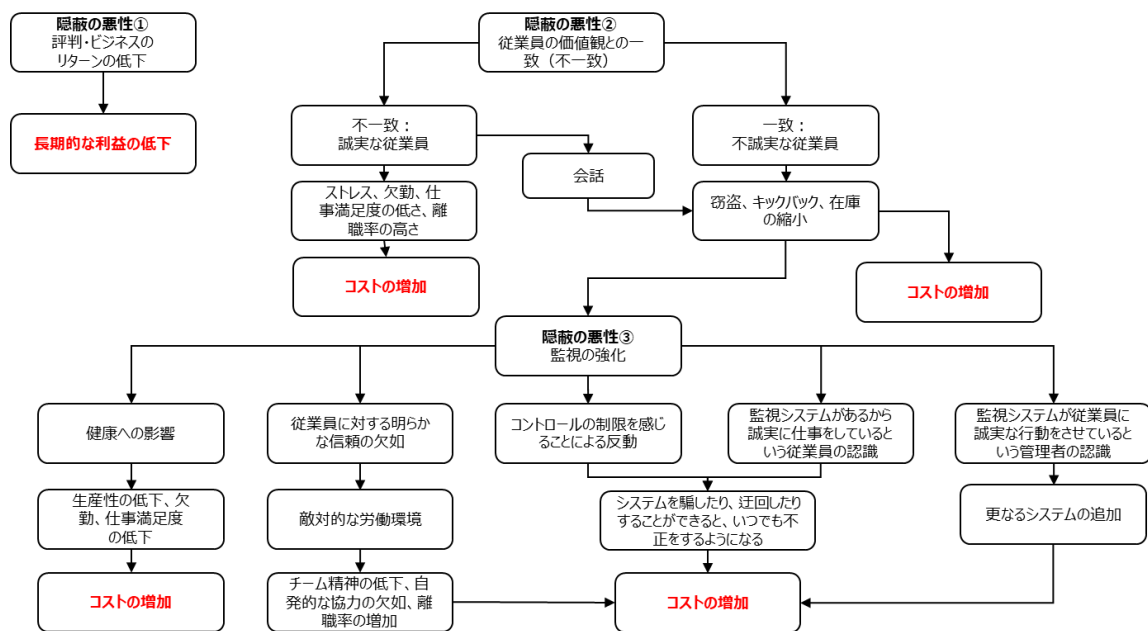


図 2-2 Robert B Cialdini (2014) による隠蔽の影響図を基に筆者作成

2.4 リスク情報提供に関する研究

リスクコミュニケーションは複雑で発展途上の分野であり、社会政治的文脈の中で科学技術的リスクを理解し、効果的に伝えることを包含している(Powell, 2005)。Renn (1992)によると、リスク管理を担う専門家と一般市民との間のギャップを埋めるためには、メッセージの認識、態度や行動の変化、対立の解決に焦点を当てた対話が不可欠であると言われている。

しかし、リスク情報を一般市民が理解できないとしてもリスク情報を伝えることは必要不可欠である。木下富雄 (2008)によると、リスク情報は、たとえそれが十分理解されなくても、リスクコミュニケーションには必要不可欠な要素である。市民は、内容が十分理解できなくても、技術情報が全くなければ落ち着かず、自分たちは騙されているのではないかと等と、軽く見られているのではないかと感じる、と言われている。

そのため、専門家や政策立案者は、市民に向けたコミュニケーションにおいて、技術的な情報の提供のバランスをとることが重要であるとされている。Lewenstein (2003)や Bucchi (2008)によると、情報の提示方法が複雑すぎると、市民はそれを理解することが困難になるが、一方で、情報が不足していると、彼らは不信感を抱き、結果として政策や技術に対する支持が低下する恐れがある。従って、リスク情報の伝達においては、専門家が専門用語を適切に簡略化し、理解しやすい形で提供することが求められる。これにより、市民は自分たちが重要な決定プロセスに参加していると感じ、より公平で透明性のある方法で技術的な意思決定が行われると認識することができるとされてい

る。また、木下富雄（2008）によれば、日本におけるリスクコミュニケーションの実践においては、特に以下の三つの重要な原則が強調されている。

- ① 情報の透明性：リスクコミュニケーションにおいては、対象のポジティブな側面だけでなく、ネガティブな側面についても可能な限り開示することが重要である。これにより、情報の全面的な透明性を担保し、信頼を築く基盤を作ることができる。透明性は、市民や関係者が自身の状況を正確に理解し、情報に基づいた意思決定を行うために不可欠となる。
- ② 双方向のコミュニケーション：効果的なリスクコミュニケーションは、一方的な情報の伝達ではなく、関係者間の双方向の対話として機能するべきである。これにより、すべての関係者が自分の意見や懸念を表明し、互いの理解を深めることができる。
- ③ 問題解決に向けた共同の模索：リスクコミュニケーションの目的は、単に相手を説得することではなく、関係者全員が問題解決に向けてより良い解決法を共同で模索することにある。これは、持続可能で受け入れられる解決策を見つけるために、多様な視点や専門知識を統合することを意味する。

2.5 本章のまとめ

先行研究の整理から、信頼行動とは、信頼する者が知覚したリスクに対してリスクテイクを行うことを意味し、これは信頼とリスクの関係を考慮する上で重要な概念であり、対象リスクをどの程度ユーザーが知覚（理解・認識）しているかが、この過程において重要な要素となると言えることが分かった。また、適切に知覚されていないリスクの顕在化は、組織に多大な悪影響を及ぼす可能性があることも示唆された。この点について、Robert B（2014）などの先行研究が指摘している通り、隠蔽や不正確な情報提供がリスクの誤認識を引き起こし、結果として組織に対する信頼を損なうことがある。

以上の通り、リスクマネジメント分野におけるステークホルダー間での連携の重要性や、これまでの信頼分野における信頼とリスクの関係や、リスクコミュニケーション分野におけるリスク情報伝達における原則についての先行研究を整理した。

第3章 ツールの提案

3.1 提案の全体像と対象者

3.1.1 提案の全体像

本研究において、AI を搭載したサービスや製品の開発を実施する企業に所属し、AI サービス・製品の開発になんらかの立場で関与する従業員が、AI サービスプロバイダー企業のリスク知覚状況を把握し、リスク情報提供戦略を検討するためのツールを提案する。

先行研究で示された T.K.Das ら (2004) による信頼とリスクの関係を理解するためのフレームワークを基に、提案を3つのプロセスで構成した。まず、① (AI プロバイダー企業の) AI に対するリスク性向を把握するプロセス、そして、リスク情報が AI 開発企業からどのように伝達されているか、及び、社会的にどのような周知がなされているかを整理することで、② (AI サービスプロバイダー企業の) リスク知覚状況を把握するプロセス、そして、適切な信頼関係を構築する (=適切なリスクテイクをしてもらう) には、知覚リスクの適正化が必要であるため、①②の整理結果に基づき、③リスク情報提供のあり方の改善を検討するプロセスである。(図 3-1)

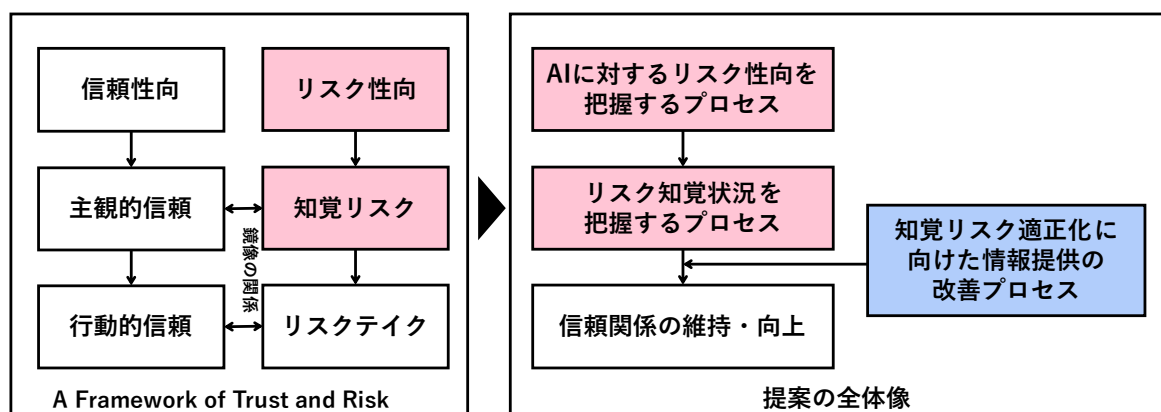


図 3-1 提案全体像

3.1.2 提案に対する要求

企業間の信頼関係の維持・向上を促進することを目的に、信頼する行為とは「信頼する者が知覚したリスクに対してリスクテイクを行うこと」とであるという信頼とリスク

の関係性を前提として、対象となる企業のリスク知覚状況を把握し、リスク情報伝達のあり方を検討するツールを提案するにあたり、先行研究で示された T.K.Das ら (2004) による信頼とリスクの関係を理解するためのフレームワークを基に、提案に対する要求を以下の3つのそれぞれにまとめた。AI に対するリスク性向の把握については、先行研究では GAISS が時間を必要としすぎること、ATAI が極端な結果を強調し過ぎることが課題とされているため、要求は企業が実務上容易に利用でき、リスクに対する考え方を把握できる手法であると整理した。続いて、リスク知覚状況の把握については、消費者が十分かつ正確な情報に基づき意思決定を行うためには、提供される情報が客観的に十分である必要があると指摘されているため (川端 2006)、要求はリスク知覚状況を明示的な情報を通じて客観的に整理できるようにすることと整理した。最後に、知覚リスク適正化に向けた情報提供の改善に関しては、企業が競合他社に対する自身の弱点や失敗を隠蔽する傾向にあるという課題があるため (Dichu Bao ら 2021)、これに対する要求として、リスク知覚状況に応じた方針を示すことによって、企業のリスク情報提供に対する前向きな意識を促進し、その改善を支援することと整理した。

表 3-1 提案に対する要求の整理

プロセス	先行研究で提示された課題	要求
AI に対するリスク性向の把握	GAISS は、実施に時間がかかるため実用的ではなく、ATAI は極端な結果を強調しており、一般的な態度を評価する文脈での使用可能性に関する課題	・ リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと
リスク知覚状況の把握	川端 (2006) によると消費者が十分かつ正確な情報に基づいて自らの意思によって契約を締結することができるような状況が作出されるのに、客観的に十分な情報が与えられているかという点の考慮が必要	・ 明示的な情報で整理することでリスク知覚状況を客観的に整理できること
知覚リスク適正化に向けた情報提供の改善	Dichu Bao ら (2021) によると、企業は状況によって自身の弱点や失敗を競合他社に知られることを避けるためにリスク情報を隠蔽する傾向にあることが課題	・ リスク知覚状況に応じて、企業のリスク情報提供に対する前向きな意識を促進すること ・ 企業のリスク情報提供のあり方に対する改善を支援できること

(1) AI に対するリスク性向の把握に対する要求

T.K.Das ら (2004) によると、主観的信頼と反比例の関係にある知覚リスクは主体のリスク性向に影響を受けるため、リスクの知覚状況を適切に把握するためにはリスク性向を考慮する必要がある。

しかし、Simone Grassini (2023) によると、人々のリスク性向を評価する既存の尺度のほとんどは AI を特に取り上げておらず、特に AI に対する態度を評価する目的で開発された尺度は少ないことが指摘されている。また、Schepman and Rodway (2020) は、2 因子構造 (AI に対する肯定的態度、否定的態度) を持つ 20 項目の尺度である General Attitudes toward Artificial Intelligence Scale (GAISS) を導入したものの、実施に時間がかかるため実用的ではないことが課題として指摘されている。また、Sindermann ら (2021) による、2 因子構造 (受容と恐怖) を特徴とする簡潔な 5 項目

尺度である Attitude Toward Artificial Intelligence scale (ATAI) は、「人工知能は人類を滅ぼす」、または、「人工知能は人類に利益をもたらす」といった項目に基づく、極端な結果を強調しており、一般的な態度を評価する文脈での使用可能性に関する課題が残ることに言及している。そこで、第一のプロセスへの一つ目の要求を「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」とする。

(2) リスク知覚状況の把握に対する要求

リスク知覚状況、及び、リスク認知状況の把握の困難性については、木下 (2002) が指摘するように、リスクは客観的な確率事象であり、リスク認知はその主観的な認識であるから、両者の間にはしばしばギャップが生じやすいことがある。また、リスク認知を捉える際の複雑さと難しさとして、個人差がリスク認知に大きく影響するという点が指摘されている。そのため、同じ情報に対しても個々人のリスク認知が大きく異なる可能性があることを前提にリスク情報の伝達を行わねばならないことが重視されている。つまり、どのようにリスク知覚もしくは認知をしているか、という点を把握するためには個人個人の知覚状況を深く聞いていかねばならず、企業にとっては現実的な観点から実行は困難である。

そこで、リスク知覚状況、及び、リスク認知状況の把握に関する議論について、より客観的な指標を取り入れている法学分野を参考とした。消費者契約法における説明義務、情報義務として、川端 (2006) によると消費者が十分かつ正確な情報に基づいて自らの意思によって契約を締結することができるような状況が作出されるのに、客観的に十分な情報が与えられているかという点が重視されている。そこで、第二の提案への要求を「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」とする。

(3) 知覚リスク適正化に向けた情報提供の改善

リスク情報伝達のあり方に係る代表的な学術分野であるリスクコミュニケーション分野において、木下 (2002) はリスクコミュニケーションの本質とは一過性の情報提供ではなく、「ステークホルダー間の信頼の構築に基づく共考の哲学とツール」と述べている。また、藤垣 (2019) によると、福島原発事故の際に、時々刻々と状況が変化する原子力発電所事故の安全性に関する事実を 1 つに定めること、統一することに重きをおいていたことが問題であったと指摘され、幅のある情報発信を行うことの重要性が指摘されている。つまり、社会的合理性を担保するためには、安全を強調する内容のみに

偏る統一した一つの情報を提供することのみに固執するのではなく、意思決定に必要な情報の開示と選択肢の多様性の保証が必要である、ということだ。しかし、意思決定に必要な情報の開示と選択肢の多様性の保証は企業にとってはハードルの高い情報提供である。Dichu Bao ら（2021）によると、企業は状況によって自身の弱点や失敗を競合他社に知られることを避けるためにリスク情報を隠蔽する傾向にあることが課題として指摘されている。そこで、第三の提案への要求を「リスク知覚状況に応じて、企業のリスク情報提供に対する前向きな意識を促進すること」とする。

また、リスクコミュニケーションの指針は文部科学省「リスクコミュニケーション案内」等があるものの、リスクコミュニケーションの実践に必要な抽象度の高い機能が示されているのみで、管理対象となるリスクやシステムに応じて、その機能に対する実現手段は指針の参照者が自ら検討せねばならず、様々な企業や組織が機能の実装には頭を抱えているのが実情である。そこで、第三の提案への二つ目の要求を「企業のリスク情報提供のあり方に対する改善を支援できること」とする。

3.1.3 対象となる企業、使用フェーズ

本研究では AI システム開発企業と AI サービスプロバイダー企業間の信頼関係に着目するため、AI システム開発企業が開発した AI システムに係るリスクを AI サービスプロバイダー企業にどの様に情報提供すべきかを検討する際に利用してもらうことを想定している。

また、本ツールを AI 製品・サービスのサプライチェーンのフェーズのいずれを対象とするかという点については、AI システム開発のプロジェクト発足時には、リスク特定はなされてはおらず、リスク情報提供の検討を実施できるほどの情報が整理されている状態ではなくツール使用の必要性がそもそもない段階であるため、図 3-2 に示す通り、本ツールを使用する AI 製品・サービスのサプライチェーンのフェーズは、運用フェーズにおける本番運用段階として設定した。

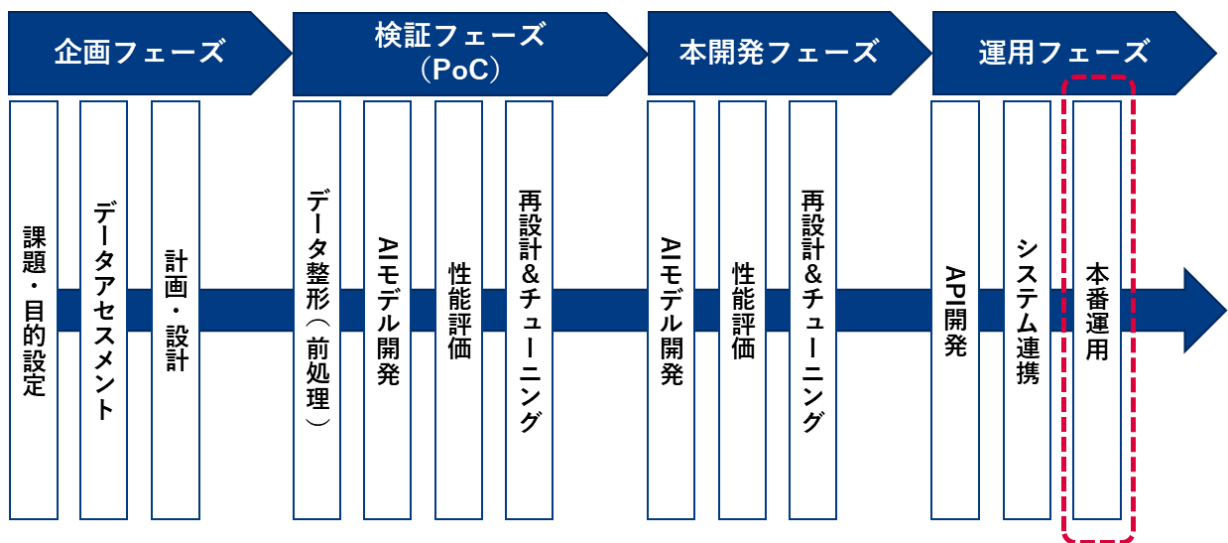


図 3-2 AI 製品・サービスのサプライチェーン

3.2 提案設計の大枠：信頼のメカニズムの活用

本提案の設計には、先行研究で示された T.K.Das ら（2004）による信頼とリスクの関係を理解するためのフレームワークを大枠として用いることにした。これは、Mayer（1995）の提案する、行動的信頼とはリスクを引き受けることであるとの整理に則ったものである。このフレームワークでは、信頼とリスクの関係において、リスク性向や信頼性向などの個人特性が果たす役割も考慮されている。リスク性向とは、リスクを取る意欲のことであり、信頼性向とは、他人を信頼する一般的な傾向のことである。このフレームワークは、これらの特性が主観的信頼と知覚的リスクの両方に影響を与え、リスクテイク行動に関与する可能性があることを示唆している。

このフレームワークの特徴は「Perceived Risk（知覚されるリスク）」という概念が重要な役割を果たしている点である。「Perceived Risk（知覚されるリスク）」は、主体が信頼関係において感じるリスクの度合いを指し、信頼関係が成立するためには、信頼する者が信頼される者に対して何らかのリスクを負う意思が必要であり、信頼があるからこそ主体はリスクを取る意思決定をするということである。

そのため、本提案の設計において、信頼関係を構築する相手（本研究においては AI サービスプロバイダー企業）が知覚するリスク状況を適切に把握することは、信頼構築に向けたリスク情報提供戦略の検討の支援材料となると考え、「Perceived Risk（知覚されるリスク）」の状況把握を中核に置いた設計とした。本提案は、T.K.Das ら（2004）による信頼とリスクの関係を理解するためのフレームワークを参考として、3つのプロセスによる構成とした。1つ目のプロセスは、「Perceived Risk（知覚されるリスク）」に影響を与えるリスク性向を把握するプロセスとした。2つ目のプロセスは、信頼関係

を構築する相手（本研究においては AI サービスプロバイダー企業）が知覚するリスク状況を把握するプロセスとした。3 つ目のプロセスは、把握した信頼関係を構築する相手（本研究においては AI サービスプロバイダー企業）が知覚するリスク状況を基に、信頼関係の維持・向上に向けてリスク情報提供のあり方の改善に向けたアクションを検討するプロセスとした。

- ① リスク性向を把握するプロセス
- ② 知覚するリスク状況を把握するプロセス
- ③ リスク性向、リスク知覚状況を踏まえ知覚リスク適正化に向けた情報提供の改善プロセス

3.3 AI に対するリスク性向把握プロセスの設計

T.K.Das ら（2004）による信頼とリスクの関係を理解するためのフレームワークでは、信頼とリスクの関係において、リスク性向や信頼性向などの個人特性が果たす役割も考慮されている。リスク性向とは、リスクを取る意欲のことであり、信頼性向とは、他人を信頼する一般的な傾向のことである。このフレームワークは、これらの特性が知覚されたリスクに影響を与え、リスクテイク行動に関与する可能性があることを示している。そのため、把握した信頼関係を構築する相手（本研究においては AI サービスプロバイダー企業）が知覚するリスク状況を適切に把握するために、リスク性向をまず把握するプロセスの提案を検討する。このプロセスの設計においては、AI に対するリスク性向把握に対する要求「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」に応える提案を検討する。

まず初めに、先行研究に提示された AI のリスク性向を把握することに係る課題の整理を基に、要求を整理した。その整理された要求からツールにおけるステップを検討し、企業インタビュー及び専門家インタビューを行いながら提案の設計を実施した。

3.3.1 先行研究及びインタビューに基づく機能要求定義

2.3 で述べた通り、Simone Grassini（2023）によると、人々のリスク性向を評価する既存の尺度のほとんどは AI を特に取り上げておらず、特に AI に対する態度を評価する目的で開発された尺度はかなり限られた数であるとの課題が指摘されている。まず、既存の AI に対する態度を評価する目的を持つ 2 つの尺度から、要求を検討した。一つ目の既存の AI に対する態度を評価する目的を持つ尺度は、Schepman and Rodway（2020）は、2 因子構造（AI に対する肯定的態度、否定的態度）を持つ 20 項目の尺度の General Attitudes toward Artificial Intelligence Scale (GAISS)である（表 3-2）。しか

し、これは 20 項目で構成される尺度であり、実際の調査の実施に時間がかかるため実用的ではないことが課題として指摘される。また、二つ目の既存の AI に対する態度を評価する目的を持つ尺度は、Sindermann ら (2021) による、2 因子構造 (受容と恐怖) を特徴とする簡潔な 5 項目尺度である Attitude Toward Artificial Intelligence scale (ATAI) である (表 3-3)。本尺度は、「人工知能は人類を滅ぼす」、または、「人工知能は人類に利益をもたらす」といった項目に基づく、極端な結果を強調しており、一般的な態度を評価する文脈での使用可能性に関する課題が残ることに言及している。そこで、第一の提案への一つ目の要求を「企業の実務上で利用しやすいこと」とした。

表 3-2 General Attitudes toward Artificial Intelligence Scale (GAISS)の項目

- | |
|--|
| <ol style="list-style-type: none">1. 日常的な取引であれば、私は人間よりも人工知能システムとやりとりしたい。2. 人工知能はこの国に新たな経済的機会をもたらすことができる。3. 人工知能システムは、人々がより幸福を感じるのを助けることができる。4. 私は人工知能ができることに感銘を受けている。5. 私は人工知能システムを日常生活で使うことに興味がある。6. 人工知能は人々のウェルビーイングに良い影響を与えることができる。7. 人工知能はエキサイティングだ。8. 人工知能エージェントは、多くの定型的な仕事では従業員よりも優れているだろう。9. 人工知能には多くの有益な応用がある。10. 人工知能システムは人間よりも優れたパフォーマンスを発揮できる。11. 社会の多くは、人工知能に満ちた未来から恩恵を受けるだろう。12. 私は自分の仕事に人工知能を使いたいと思っている。13. 組織は非倫理的に人工知能を使用する。14. 人工知能システムは多くの誤りを犯すと思う。15. 私は人工知能が不吉だと思う。16. 人工知能が人間を支配するかもしれない。17. 人工知能は危険だと思う。18. 人工知能の将来の利用法を考えると、私は不快感に震える。19. 人工知能がどんどん使われるようになれば、私のような人間は苦しむことになる。20. 人工知能は人々をスパイするために使われる。 |
|--|

表 3-3 Attitude Toward Artificial Intelligence scale (ATAI) の項目

1. 人工知能が怖い。
2. 私は人工知能を信頼している。
3. 人工知能は人類を滅ぼす。
4. 人工知能は人類に恩恵をもたらす。
5. 人工知能によって多くの雇用が失われる。

「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」を実現する要素を具体化するために、実務上で利用しやすいと感じる尺度（調査項目）、利用しにくいと感じる尺度（調査項目）について企業インタビューを実施した。企業インタビューは、2社から1名ずつのAIサービス・製品の開発に関与する方として、大企業のプロジェクトマネージャーとスタートアップ企業のCTO/人材育成を担う人を募った。

表 3-4 インタビュー対象者一覧

企業形態・規模	AI サービス・製品の開発 における役割	AI サービス・製品の分野
大企業（1000名以上）	プロジェクトマネージャー	人材採用 AI
スタートアップ企業（10名以下）	CTO、人材教育	マテリアルインフォマティクス

インタビューでは、以下のような質問を行った。

- ① 開発するAIシステムを使用してサービスを実施するAIサービスプロバイダー企業のAIに対するリスク性向、AIに対する態度を把握したいと考えるか。
- ② 現在、AIに対するリスク性向、AIに対する態度を把握している場合、どのように実施しているか。それに対する課題意識はあるか。
- ③ アンケート調査を実施する場合、どのようなアンケート調査であれば実務上、実施可能であるか。

得られた回答は表 3-5、表 3-6 に示す通りである。大企業の担当者はリスク性向と態度の把握を非常に重要と考え、リスク管理と期待値の調整のために定期的なミーティングと直感を用いているが、AIの急速な進化による状況の変化を課題としている。一方で、ベンチャー企業の担当者はプロジェクトの円滑な進行のために重要と捉え、顧客との接点を通じてリスク性向を把握しており、このプロセスが不均一であることを課題としている。両者とも、実務上実施可能なアンケート調査の重要性を認識していることが分かった。

表 3-5 大企業の担当者の回答

- ① AI サービスプロバイダー企業の AI に対するリスク性向と態度の把握の重要性に対する回答：非常に重要。サービスプロバイダーの AI に対する理解度とリスク許容度を把握することは、リスク管理と期待値の調整に不可欠である。
- ② 現在の把握方法と課題意識：定期的なミーティングでのディスカッションを通じて直感に基づき把握している。AI の急速な進化に伴い、リスクと態度が絶えず変化することが課題だと捉えている。
- ③ 実務上実施可能なアンケート調査：簡潔で分かりやすい質問で構成されていることが重要ではないか。アンケート結果をどのように考慮したらいいのかまでの指南がないと使いにくいと感じる。

表 3-6 ベンチャー企業の担当者の回答

- ① AI サービスプロバイダー企業の AI に対するリスク性向と態度の把握の重要性に対する回答：重要である。サービスプロバイダーの AI に対するリスクの合意形成はプロジェクトを円滑に進めるために重要なプロセスであるため、そのベースとなるリスク性向やリスクをそもそもどの程度理解しているかを知ることがはやるべきだと感じている。リスクに対する認識があまりに低く、AI 技術に対する知識があまりない場合には、プロジェクト遂行上のリスクに繋がる恐れがあるため、仕事を受けないことすらある。
- ② 現在の把握方法と課題意識：セールスからカスタマージャーニーなどそれぞれのタッチポイントでリスクに対する誤解が無いように伝達する際に、顧客のリスク性向はある程度把握できているはずである。ただし、顧客と接点を持つ社員次第ではあり、社員の実施する顧客に対する説明等のコミュニケーションの仕方に斑があるため、社員によってはリスク性向や態度の深堀が難しいとも感じている。
- ③ 実務上実施可能なアンケート調査：労力をかけずに実施と活用ができること。

以上の回答から、機能要求を以下の3つに整理した。

- (a-1) 明確かつ簡潔な質問構成であること
- (a-2) AI リスク性向を多角的に把握できること
- (a-3) 調査結果の適用に関するガイダンスがあること

「明確かつ簡潔な質問構成であること」は、大企業とベンチャー企業の担当者の双方が、質問が明瞭で簡潔であることを重視していることがインタビューから把握できた。質問が明瞭で簡潔であることによって、回答者が質問の意図を正確に理解し、適切に回

答することが可能になることを意図するだけでなく、回答を受け取った企業の担当者がどのような経験、知識レベルであったとしても質問の意図を理解することを容易とするため、必須な要素であると考える。

「AI リスク性向を多角的に把握できること」について、大企業の回答からは、サービスプロバイダーの AI に対する理解度とリスク許容度の把握はリスク管理と期待値の調整に重要だと述べており、ベンチャー企業の回答からは、プロジェクト遂行上のリスク管理上の必要性が述べられている。それらの管理や調整においては AI システムの様々な観点からの検討が必要であり、多様な角度でのリスク性向を把握する質問を含む必要がある。

「調査結果の適用に関するガイダンスがあること」に関しては、大企業とベンチャー企業の担当者の双方がアンケート調査の結果を活用しやすいことに対する要望を述べていた。そのため、調査結果を実際のリスク管理やリスク情報提供の検討におけるプロセスにどのように組み込むかに関する明確な指針を提示することが必要であると考ええる。

3.3.2 ステップの提案

先行研究において提示された課題とヒアリングに基づく 3 つの機能要求を基に AI に対するリスク性向の把握プロセスのステップを以下のように提案する。

(1) AI に対するリスク性向を把握するステップ

機能要求である「明確かつ簡潔な質問構成であること」及び「AI リスク性向を多角的に把握できること」を満たす AI に対するリスク性向を把握することが可能な尺度の調査を実施した結果、Simone Grassini (2023) の整理する AIAS 尺度 (表 3-7) が適切であると判断し、選定した。AIAS 尺度とは、認識される利益、潜在的リスク、AI 技術の利用意向など、認識される有用性と技術が社会や人類に与える潜在的影響に焦点を当てることで、AI に対する態度を捉えるように設計されている。Simone Grassini (2023) によると、AI 技術の急速な発展や実装の時期における、AI に対する社会的態度の変化を縦断的に評価するために用いることが今後の期待として示されている。

表 3-7 AIAS-5 (AI Attitude Scale)

- | |
|--|
| ① 私は AI が私の生活を向上させてくれると信じている。
② AI は私の仕事を向上させてくれると信じている。
③ 私は将来、AI 技術を使うと思う。
④ AI 技術は人間にとって脅威だと思う（逆項目）。
⑤ AI 技術は人類にとってプラスだと思う。 |
|--|

「①私はAIが私の生活を向上させてくれると信じている。」という項目については、この項目は、技術受容モデル (Technology Acceptance Model: TAM) (Davis, 1989) から導き出されたもので、知覚された有用性（すなわち、個人が技術を使用することでパフォーマンスが向上すると信じる度合い）が技術受容の主要な決定要因であると仮定している。AI の文脈では、この項目は、生活の質を向上させるという AI の潜在的な利益に関する個人の信念を評価する。このような理論は、近年 AI の分野で応用されている（例えば、Sohn and Kwon, 2020 を参照）。

「②AI は私の仕事を向上させてくれると信じている。」という項目は、①と同様に技術受容モデル (Technology Acceptance Model: TAM) (Davis, 1989) に基づくものであり、仕事領域における AI の潜在的利益に関する個人の信念を反映している。この項目は、Wijayati ら（2022）によって議論されている AI 技術への期待に沿って、一般的に人間の労働条件を改善するといった AI の役割を取り上げている。

「③私は将来、AI 技術を使うと思う。」というこの項目は、前述の TAM (Davis, 1989) や Venkatesh ら（2003）による UTAUT (Unified Theory of Acceptance and Use of Technology: 技術の受容と利用に関する統一理論) に基づくものである。この行動意図は、実際の技術利用の近接的な決定因子と考えられており、この項目は、将来 AI 技術を採用し利用する個人の意図を測定するために設計されている。

「④AI 技術は人間にとって脅威だと思う（逆項目）」というこの項目は、AI 技術の潜在的なリスクと否定的な結果に関する個人の懸念を評価する。この項目は、超知的 AI のリスク (Barrett and Baum, 2017) やテクノロジーに関連するリスク認知 (Frey and Osborne, 2017) に関する文献に立脚している。これらの理論的立場は、転職（職業の喪失）、プライバシーの喪失、あるいは実存的脅威など、AI がもたらす潜在的な危険に対する個人の懸念を理解することの重要性を強調している。

「⑤AI 技術は人類にとってプラスだと思う。」というこの項目は、AI 技術が社会や人類全体に与える影響に関する個人の総合的な評価を捉えるものである。この項目は、AI の社会的意味合いに関する先行研究 (Brynjolfsson and McAfee, 2014) に基づいている。

本尺度が「明確かつ簡潔な質問構成であること」という機能要求を満たしていると

判断した理由として、まず分かりやすい表現を用いているという点が挙げられる。各項目は、簡潔かつ具体的な表現を使用しており、例えば、「私は AI が私の生活を向上させてくれると信じている。」や「AI 技術は人間にとって脅威だと思う。」という表現は、AI に対する個人の信念や懸念を直接的に問うもので、理解しやすいと考えられる。また、これらの質問は、Technology Acceptance Model (TAM) やその他の関連理論に基づいており、AI 技術のリスク性向を把握するにあたり、信頼性を高く、明確であると判断した。

続いて、本尺度が「AI リスク性向を多角的に把握できること」という機能要求を満たしていると判断した理由として、これらの質問は AI のポジティブな影響（例えば、生活や仕事の向上）、将来の使用意向、そして潜在的なリスクや社会的・人類的影響といった多様な側面を包含している点が挙げられる。例えば、「AI 技術は人間にとって脅威だと思う」という逆スコア項目は、AI の潜在的な危険やネガティブな側面に対する認識を評価することで、技術に関する懸念や不安を捉える一方で、「AI 技術は人類にとってプラスだと思う」という項目は、AI が社会や人類全体にもたらす肯定的な影響に関する個人の総合的な評価を問うものである。これらのことから、本尺度はリスク性向（AI に対する態度）をネガティブ・ポジティブの双方の観点、個人的視点・社会的視点の複数の観点をかけ合わせた多角的な把握を志向していると考えた。

本尺度を用いて AI サービスプロバイダー企業の担当者に対するアンケート（5 件法）を実施することで、AI サービスプロバイダー企業の AI に対するリスク性向を把握するステップとした。各アンケート項目はライカート尺度を用いて評価され、1 から 5 までのスコアが割り当てられているため、項目に対する回答の平均値を算出し、集計されたデータを用いて回答者の AI に対する態度の指標を得ることができるよう設計した。

(2) AI に対するリスク性向からリスク情報提供の方針を把握するステップ

機能要求である「調査結果の適用に関するガイダンスがあること」を満たすために、AI に対するリスク性向を把握した後のステップをリスク情報提供のあり方の検討に対する指針を示すように設計を行った。図 3-3 に示す通り、調査結果の適用がより容易にリスク情報提供のあり方と結びつけた検討が行えるように、実施した AI に対するリスク性向の調査結果を基にリスク情報提供に対する方針を結び付けた設計とした。

項目に対する回答の平均値を算出し、3 以上の場合は、AI リスク性向の傾向として「AI リスクに対して積極的である」、3 を下回る場合には「AI リスクに対して消極的で

ある」という2つの分岐とした。その分岐それぞれに対して、リスクコミュニケーション分野における先行研究の Finucane et al.(2000)で示された感情ヒューリスティックのモデルを参考に指針を設計した。Finucane et al.(2000)によると、高便益あるいは低リスク情報が提示された場合には、低リスク、高便益の評定がなされ、低便益あるいは高リスクの情報が提示された場合には、高リスク、低便益と評定することが示されている。そのため、「AI リスクに対して積極的である」場合には、リスク情報提供として、「AI サービスプロバイダー企業が過剰にリスクを少なく見積もって認識している可能性があるため、リスク情報提供の際にはよりリスクを強調することが重要となる」という指針を、「AI リスクに対して消極的である」場合には、リスク情報提供として、「AI サービスプロバイダー企業が過剰にリスクを大きく見積もって認識している可能性があるため、リスク情報提供の際にはより便益・対策を強調することが重要となる」という指針を示した。

AI 態度尺度 (AI Attitude Scale; AIAS -5)

- ① 私はAIが私の生活を向上させてくれると信じている。
- ② AIは私の仕事を向上させてくれると信じている。
- ③ 私は将来、AI技術を使うと思う。
- ④ AI技術は人間にとって脅威だと思う（逆項目）。
- ⑤ AI技術は人類にとってプラスだと思う。

▼ 回答の平均値が3以上の場合

AIリスクを取ることにに対して積極的

- AIサービスプロバイダー企業が過剰にリスクを少なく見積もって認識している可能性があるため、リスク情報提供の際にはよりリスクを強調することが重要となる

OR

▼ 回答の平均値が3より小さい場合

AIリスクを取ることにに対して消極的

- AIサービスプロバイダー企業が過剰にリスクを大きく見積もって認識している可能性があるため、リスク情報提供の際にはより便益・対策を強調することが重要となる

図 3-3 AI 態度尺度に基づくリスク情報提供の方針

3.4 リスク知覚状況の把握プロセスの設計

T.K.Das ら (2004) による信頼とリスクの関係を理解するためのフレームワークでは、「Perceived Risk (知覚されるリスク)」という概念が重要な役割を果たしている。「Perceived Risk (知覚されるリスク)」は、主体が信頼関係において感じるリスクの度合いを指し、信頼関係が成立するためには、信頼する者が信頼される者に対して何らかのリスクを負う意思が必要であり、信頼があるからこそ主体はリスクを取る意思決定をするということである。

そのため、本提案の設計において、信頼関係を構築する相手（本研究においては AI サービスプロバイダー企業）が知覚するリスク状況を適切に把握することは、信頼構築に向けたリスク情報提供戦略の検討の支援材料となると考えた。そこで、「Perceived

Risk（知覚されるリスク）」の状況把握、つまり、リスク知覚状況の把握プロセスが本ツールの中核を成す設計とした。

3.4.1 先行研究に基づく機能要求定義

Kim, S. D. (2012)が述べるように、ドナルド・ヘンリー・ラムズフェルド元米国防長官が「Unknown-Unknowns（未知の未知）」(Rumsfeld, 2002)に言及した後、人々はリスクの性質を理解し説明するために、知識の象限、すなわち、既知の既知、未知の未知、未知の既知、既知の未知を使い始めた。Cleden, (2009)によると、リスクの典型的な分類は、リスク事象の発生に関する知識のレベル（既知または未知）と、その影響に関する知識のレベル（既知または未知）に基づいているおり、以下の4つの可能性が導かれる。

- I. Known-knowns（既知の既知）：リスク事象とその影響がともに既知である状況を指す
- II. Known-unknowns（既知の未知）：リスク事象は既知だが、その影響は未知である状況を指す
- III. Unknown-unknowns（未知の未知）：リスク事象もその影響もともに未知である状況を指す
- IV. Unknown-knowns（未知の既知）：リスク事象は未知だが、その影響は既知である状況を指す

本研究では、リスク事象の発生に関する知識のレベル（既知または未知）と、その影響に関する知識のレベル（既知または未知）を参考にリスク知覚状況の整理を図る設計とした。リスクの典型的な分類は、リスク事象の発生とその影響に焦点を当てているが、本研究の趣旨として対象となる AI サービスプロバイダー企業のリスク知覚状況を把握するためには、対象となる知識の汎用度と共有の有無の観点からの解釈が必要だと考えた。知識の汎用度とは、リスクに関わる情報が一般的な知識として活用可能な状態で広まっているかという観点であり、知識の共有の有無とは、リスクに関わる情報が AI 開発企業から知識として共有されているかという観点である。その解釈に基づくと、4つのリスクの類型について以下のような整理が図られる。

- I. Known-Knowns
 - (ア) 情報の汎用度 (Known)：対象リスクに関する情報（リスク顕在化による影響や対策等）が一般的な知識として広まっている状態。
 - (イ) 情報の共有の有無 (Known)：対象リスクに関する情報がステークホルダー

一間で共有されている状態。※本論文の場合、AI システム開発企業がリスク情報を提供した AI サービスプロバイダー企業が知っているという状況を指す。

II. Known-Unknowns

- (ア) 情報の汎用度 (Known): 対象リスクに関する情報 (リスク顕在化による影響や対策等) が一般的な知識として広まっている状態。
- (イ) 情報の共有の有無 (Unknown): 対象リスクに関する情報がステークホルダー間で共有されていない状態。※本論文の場合、AI システム開発企業が AI サービスプロバイダー企業に対象リスクに関する情報を提供していない状況を指す。

III. Unknown-Unknowns

- (ア) 情報の汎用度 (Unknown): 対象リスクに関する情報 (リスク顕在化による影響や対策等) が一般的な知識としては広まっていない状態。
- (イ) 情報の共有の有無 (Unknown): 対象リスクに関する情報がステークホルダー間で共有されていない状態。※本論文の場合、AI システム開発企業が AI サービスプロバイダー企業に対象リスクに関する情報を提供していない状況を指す。

IV. Unknown-Knowns

- (ア) 情報の汎用度 (Unknown): 対象リスクに関する情報 (リスク顕在化による影響や対策等) が一般的な知識としては広まっていない状態。
- (イ) 情報の共有の有無 (Known): 対象リスクに関する情報がステークホルダー間で共有されている状態。※本論文の場合、AI システム開発企業がリスク情報を提供した AI サービスプロバイダー企業が知っているという状況を指す。

上記を踏まえ、機能要求を以下の2つに整理した。

(b-1) リスク情報の共有の有無 (対象リスクに関する情報がステークホルダー間で共有されているかどうか) を把握できること

(b-2) リスク情報の汎用度 (対象リスクに関する情報 (リスク顕在化による影響や対策等) が一般的な知識としては広まっているかどうか) を把握できること

3.4.2 ステップの提案

(1) 企業由来のリスク情報提供状況を一覧するステップ

機能要求の一つである「リスク情報の共有の有無を把握できること」を満たすために、AI 開発企業から AI プロバイダー企業に対してどのようなリスク情報提供を実施しているのかを整理、一覧化するステップを設計した。まず、標準的な AI リスク項目を、AI に対するリスクマネジメントに関するガイドラインや先行研究を基に作成した。(表 3-8)

AI リスク項目一覧表として標準的な AI リスク項目を整備したリストを提供し、そのリスクについて AI 開発企業として認識しているリスク（項目「どのようなリスクが存在するか」）、説明の実施状況（項目「説明の実施」）、関係する AI システム構成要素について記入を行うものである。(表 3-9) AI システム開発者は、自らの AI システムの構成要素を適切に把握することで、適時適切なリスク情報提供が可能になると考えられるため、「関係する AI システム構成要素」を記入する列を作成した。「関係する AI システム構成要素」は、落合浩一郎ら（2019）や松本敬史、江間有沙（2020）を参考に、図 3-4 に示す整理と表 3-10 に示す解説を整備した。

AI リスク項目一覧表に記入した情報を基に、図 3-5 に示すフレームワークに基づき、AI システム開発企業から AI サービスプロバイダー企業へのリスク情報提供がなされたか（リスクについて説明したか否か）について、AI リスク項目をそれぞれ「説明済」もしくは「未説明」の枠に振り分けることで、リスク情報提供状況の可視化を行う。

表 3-8 標準的な AI リスク項目の整理

AI バロメーター	N 社ガイドライン	グローバル AI ガイドライン	AI リスクチェックンモデル	NIST AIRMF
差別につながる偏見	1)意思決定システムにおける偏り	透明性	公平性	データがコンテキストや用途を適切に表現していない
説明力不足	2)差別の助長	バイアス	頑健性	データ量の増加と複雑化
規制の再構築	3)ハラスメントの助長（妊娠・介護等）	公平性	説明責任	学習中の変更による性能の根本的な変化
サイバー攻撃	4)サービスへの不平等なアクセス	プライバシー	適正利用	データセットが本来のコンテキストから切り離される
制度に対する信頼の喪失	5)プライバシーの侵害	救済制度の不足		AI システムの規模と複雑さ
同意の仕組みの不具合	6)透明性の欠如	低いヒューマンインザループ		事前学習済みモデルの使用によるリスク
サービスへの不平等なアクセス	7)同意の仕組みの不足	ロバスト性		創発特性に対する故障モード予測の難易度
透明性の欠如	8)低いヒューマンインザループ（人間の介在）	社会保障		データ集計機能強化によるプライバシーリスク
プライバシーの侵害	9)救済機会の不足			データ、モデル、コンセプトのドリフト
技術・データ成熟度の低さによる影響	10)自己決定権の侵害			不透明感の増加と再現性への懸念
過剰なデータ保持	11)個人の尊厳の侵害			ソフトウェアのテスト基準の未整備
プラットフォームとデータの独占	12)安全性の低下			AI ベースのソフトウェアテストの困難さ
誤報	13)情報セキュリティの脅威拡大			
低いヒューマンインザループ				
AI への信頼喪失				
低精度				
AI ツールへの過度な信頼				

表 3-9 AI リスク項目一覧表（採用 AI を対象とした記入イメージ）

ID	AIリスク項目	どのようなリスクが存在するか	説明の実施（済or未）	関係するAIシステム構成要素
1	意思決定における偏り・偏見	AIが過去のデータに基づいて意思決定を行うため、過去の採用プロセスにおける偏見が組み込まれるリスク	AIが過去のデータに基づいて判断するため、歴史的な偏見が組み込まれる可能性があり、適合率重視の評価指標が偏見を強化するリスクがあるため、人間の介在を運用プロセス上における工夫や、多様性を考慮したデータセットの作成を行う必要がある。	・提供サービス上の利用方法 ・学習データのバランス ・パラメータ・特徴量・重みづけ
2	説明責任への対応の不足・透明性の欠如	選定理由・基準が不明確である場合、採用の申込者に判断根拠を納得してもらえないリスクがある。	エントリーシートの中で判断に強い影響を与えたキーワードがハイライトされるが、その選定理由・基準が不明確である場合、採用の申込者に判断根拠を納得してもらえない可能性がある。	・説明可能性 ・パラメータ・特徴量・重みづけ ・フィードバック ・問合せ対応
3	情報セキュリティの脅威拡大	個人情報が含まれるデータの漏洩や不正アクセスが発生するリスクがある。	プライベートクラウド環境によるデータ管理において、個人情報が含まれるデータの漏洩や不正アクセスが発生した場合は、社会的な批判を強く受ける可能性がある。	・開発責任者・管理体制 ・実運用データ ・リスクマネジメント・監査体制
4	同意の仕組みの不足	採用の申込者が自分のデータがどのように使われるかを理解せずにサービスを利用するリスクがある。	採用の申込者のデータの利用において、明確な同意プロセスがないと、同意なく個人情報を含めた情報を使用したことで批判を受ける可能性がある。	・実運用データ ・問合せ対応
5	利用における不平等性	一律の評価指標による選考によって、特定のグループに対するサービスの不公平な提供となるリスクがある。	－ ※説明未実施	・ベネフィット・インパクト ・提供サービス上の利用方法 ・学習データのバランス ・パラメータ・特徴量・重みづけ
6	プライバシーの侵害	個人情報が含まれるエントリーシートの使用を行うため、個人情報の不正使用や漏洩のリスクがある。	政府の発行する個人情報保護規則等に準拠し、適切なデータ保護対策を講じない場合、個人情報の不適切な取り扱いだと見なされる可能性がある。	・開発責任者・管理体制 ・実運用データ ・事業者によるモニタリング・メンテナンス
7	技術・データの成熟度の低さによる影響	正解率70%という期待精度であり、不完全な技術やデータによる誤判断のリスクがある。	－ ※説明未実施	・学習データの概要 ・学習手法 ・パフォーマンス精度・限界
8	品質保証とテスト基準の不備	不適切な品質保証やテスト基準により、予期せぬエラーが発生するリスクがある。	正解率70%を下回った場合にはAIモデルの自動更新を行わないなど、厳格な品質基準を設けることで、AIシステムの品質を担保する必要がある。	・変更の適正手続 ・事業者によるモニタリング・メンテナンス ・リスクマネジメント・監査体制
9	人間の介在の不足・過度なAI依存	AIの判断に対する過度な依存により、人間の判断が排除されるリスクがある。	最終的な合否判断は人間が行い、AIは参考情報として位置付けないと、人間の判断が排除される可能性がある。	・利用目的 ・ベネフィット・インパクト ・提供サービス上の利用方法
10	運用上の救済制度の不足	不適切なAI判断に対する救済措置が不足している場合、採用の申込者に不満が生じるリスクがある。	不適切な判断があった場合のレビュープロセスを設け、救済措置を提供する等の運用に不備が存在すると、採用の申込者に不満が生じる可能性がある。	・提供サービス上の利用方法 ・リスクマネジメント・監査体制 ・問合せ対応
11	個人の尊厳・権利の侵害	AIの判断により採用の申込者の尊厳や権利が侵害されるリスクがある。	－ ※説明未実施	・利用目的 ・ベネフィット・インパクト ・パラメータ・特徴量・重みづけ ・リスクマネジメント・監査体制
12	過剰なデータ保有・独占	過去1年分のバージョンの保存が過剰なデータ保有によるプライバシー侵害につながるリスクがある。	データの最小化原則に従い、必要最低限のデータのみを保有することで、過剰なデータ保有ではない運用とすることがある。	・開発責任者・管理体制 ・実運用データ
13	不適切な利用（悪意を含む）	個人情報の不正販売や、虚偽の内容に基づく採用の申込が行われるリスクがある。	システムへのアクセス管理と使用履歴の監視を行うことで、個人情報の不正販売への対策を行う。虚偽の内容に基づく採用の申込は、ESチェック後の面接の中で社員が直接確認することで、対策を行う必要がある。	・事業者によるモニタリング・メンテナンス ・リスクマネジメント・監査体制 ・問合せ対応
14	予期せぬ事態への対応の不足	AIシステムの不測の事態に対応できないリスクがある。	緊急時のプロトコルを設定し、障害発生時の迅速な対応を確保することで、AIシステムの不測の事態に対応できないリスクに対応できる運用体制・プロセスとする必要がある。	・開発責任者・管理体制 ・変更の適正手続き ・事業者によるモニタリング・メンテナンス ・リスクマネジメント・監査体制
15	社会・文化・環境への悪影響	（未検討）	－ ※説明未実施	・利用目的 ・ベネフィット・インパクト

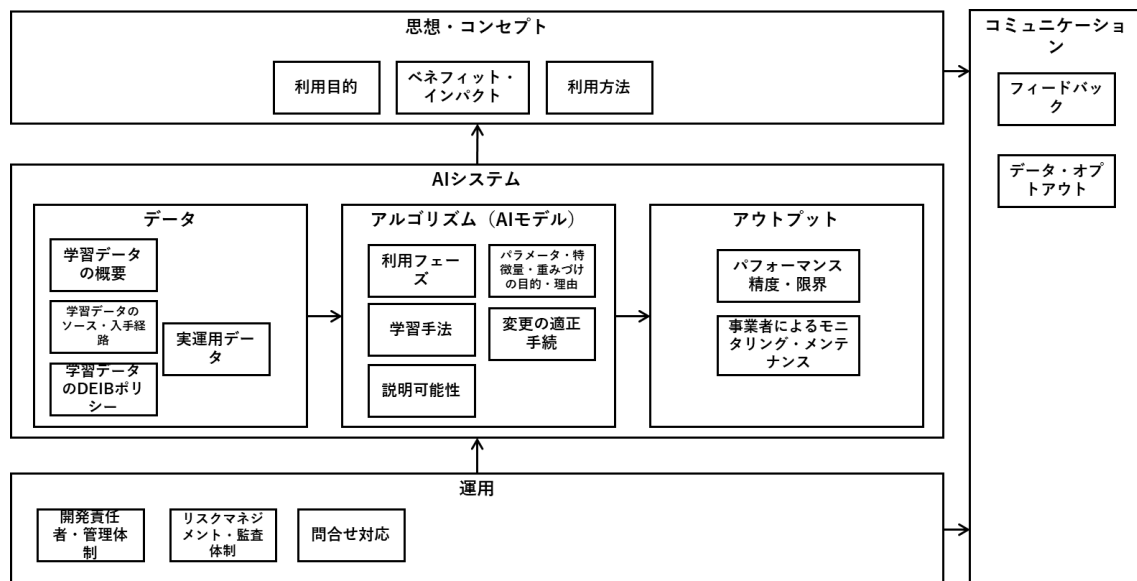


図 3-4 AI システム構成要素の構造化

表 3-10 AI システム構成要素の説明

ID	Lv1	Lv2	説明
A	プロダクト思想	利用目的	AIを利用する目的
B		ベネフィット・インパクト	AIを利用することによって得られる便益（ベネフィット）、影響（インパクト）のアセスメント
C		利用方法	AI が活用されているプロセス、AI が担う役割（意思決定そのものを行うのか、人間の意思決定をサポートするのか等）
D	開発チーム	開発責任者・管理体制	管理責任部門・責任者の所在、委託先の管理体制・方法
E	データ	学習データの概要	学習データ・テストデータの概要
F		学習データのソース、入手経路	学習データ・テストデータの所在、入手方法・入手経路
G		学習データのバランス	学習データ・テストデータの多様性、想定利用者との整合性
H		実運用データ	実際に用いられるデータの概要（AI は実際に何をData として Output を抽出するのか）、利用規約の運用等
I	AIアルゴリズム	利用フェーズ	利用されるアルゴリズムのフェーズ（研究、テスト、実運用等）
J		学習手法	学習手法 AI（エキスパートシステム、ルールベース、機械学習、深層学習等）
K		説明可能性	アルゴリズム自体の説明可能性
L		パラメータ・特徴量・重みづけとその目的・理由	合理的に理解可能な程度に抽象化・言語化されたパラメータ
M	アウトプット	変更の適正手続	アルゴリズムの変更に係る適正手続、変更による影響、変更の目的等
N		パフォーマンス精度・限界	アウトプットの有意確率、誤謬率、偽陰性・偽陽性等
O		事業者によるモニタリング・メンテナンス	アウトプットの品質に対するモニタリング・メンテナンス体制（頻度・方法・体制）、検出件数
P	利用者によるコントロール	フィードバック	利用者によるアルゴリズムやデータセットへのフィードバックの有無、可否及び方法
Q		データ・オプトアウト	学習データからのオプトアウト方法の有無
R	マネジメント体制	リスクマネジメント・監査体制	AI を用いることに対するリスクに対するアセスメント、最小化又はマネジメント対応、AI の開発・提供に対する監査体制及び実施状況
S		問合せ対応	問い合わせ先・不服申立窓口、カスタマーエクスペリエンス体制（頻度・方法・体制）、一定期間経過後の苦情件数

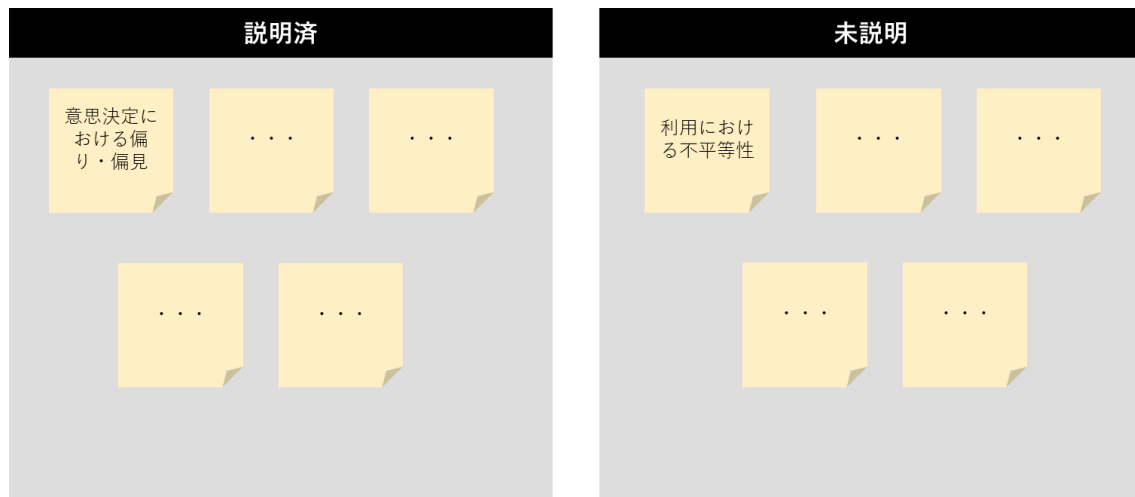


図 3-5 リスク情報提供状況の可視化フレームワーク

(2) 社会的リスク喚起状況と顕在化事例をマッピングするステップ

機能要求の二つ目である「リスク情報の汎用度を把握できること」を満たすために、リスクに関わる情報が一般的な知識として活用可能な状態で広まっているかという観点でツールの使用者が調査を実施できるようステップを設計した。リスク情報の汎用度としては、社会的にリスクの顕在化による影響や対策が知見として蓄積されリスクに対する注意喚起等がなされていることや、リスクが顕在化した事例によってどのようなリスクがありどのような影響があるのかが一般的な知見になっていることとして整理を行った。そこで、以下のような観点での調査を実施することで、自社の開発する AI システムに対するリスク知覚に影響を与える要因を整理し、リスク情報の汎用度を把握する設計とした。

- ・ 企業に関連する規制・ガイドライン動向：新たなガイドライン等が導入され、それに準拠していないことで、その企業への信頼を損なう要因
- ・ 企業の競合他社や所属する業界におけるインシデント・事故の発生状況：企業が、競合他社や同業界の企業で発生したリスクや脅威に対して適切な対策を取らないことで、その企業への信頼を損なう要因
- ・ AI 技術（そのもの）に関連する世論・見解動向：社会全体の倫理観やプライバシーに対する意識が高まった結果の水準に対して十分な対策を講じていないことで、その企業に対する信頼を損なう要因

より調査を実施しやすくする工夫として、AI システムに関わる社会的リスク喚起

状況と顕在化する事例を構成する要素モデルを作成した。(図 3-6) この要素モデルに基づき、調査を実施することで、自社の開発する AI システムに対するリスク知覚に影響を与える要因をより網羅的に把握することが可能となる。要素モデルは、左から右に「自社の提供する AI システムが利用される分野」、「自社の提供する AI システムによるサービス」、「自社の利用する AI 技術」という構成になっており、それぞれについて調査を実施しながらリスク知覚に影響を与える要因に具体化していく流れである。例えば、「自社の提供する AI システムが利用される分野」について、「関連する業界団体・機関」を検討し、その業界団体や機関が発行している「ルール・ガイド」を特定する、という流れである。列ごとに一番下の行にキーワード等の調査結果を記載できたら、記載したキーワード・内容を基に、関連する AI リスク項目一覧リストのリスク項目の枠に内容を転記することで、各 AI リスク項目についてリスク情報の汎用度を把握できるような設計とした。(図 3-7)

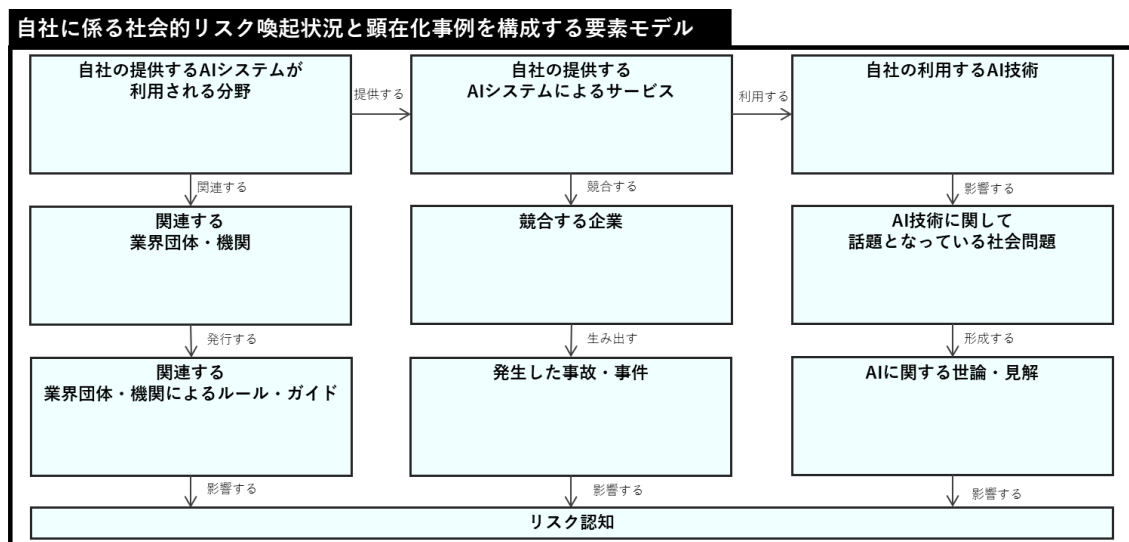


図 3-6 社会的リスク喚起状況と顕在化する事例を構成する要素モデル

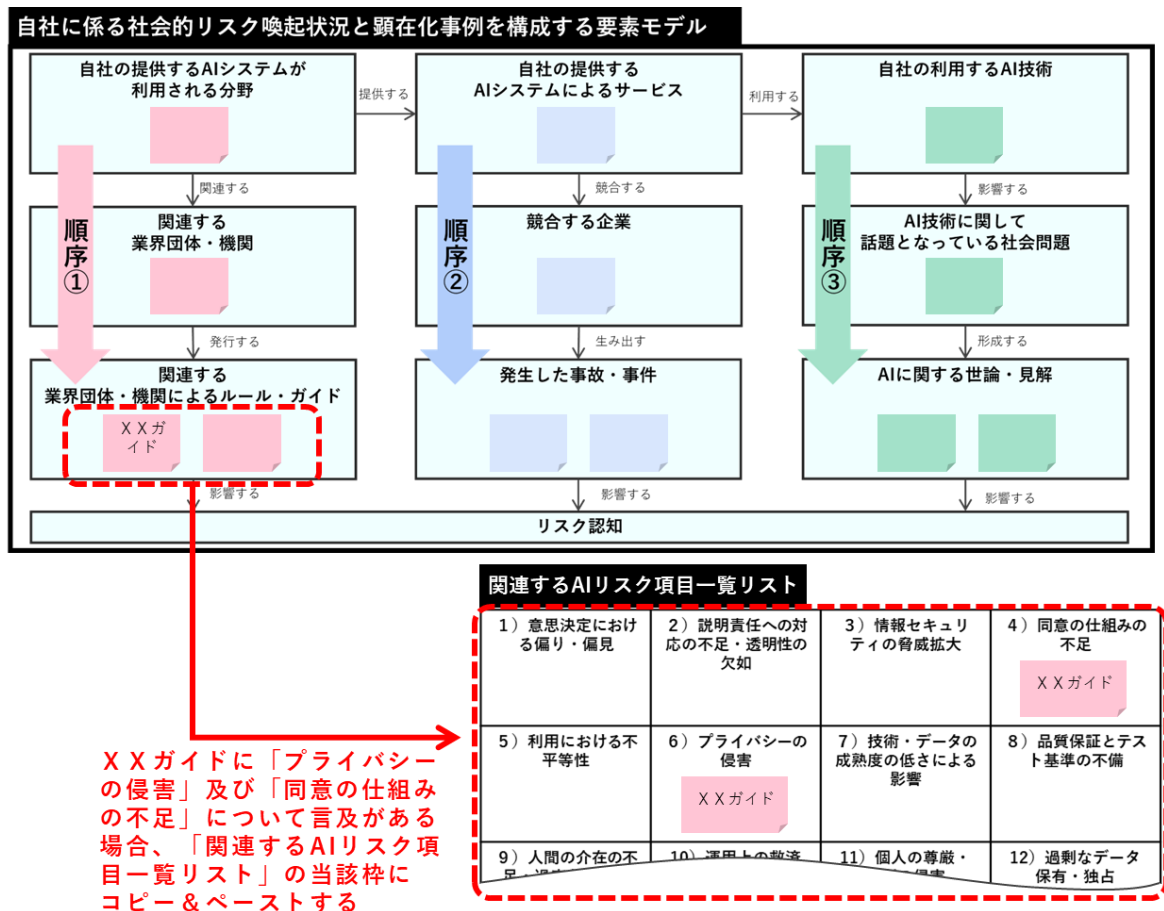


図 3-7 社会的リスク喚起状況と顕在化する事例を構成する要素モデルを通したステップ実施イメージ

(3) リスク知覚状況を整理・把握するステップ

企業由来のリスク情報提供状況を一覧化するステップと社会的リスク喚起状況と顕在化事例をマッピングするステップで実施した結果を概観できるよう、AI リスク項目を以下の 4 象限に名前を付けて解説を付記したうえで、整理するステップを設計した。(図 3-8)

- I. **【複数の情報源によるリスク知覚形成】** 企業から情報伝達もされており、組織外部で発生事例があり、ユーザーは顕在化した際の影響等について理解できている
- II. **【社会的喚起依存のリスク知覚形成】** 企業から情報提供はされていないものの、企業外部では注意喚起や発生事例等がありリスク認知はされている

- III. 【リスク知覚の非形成】企業内部から情報提供もされていないし、企業外部での発生事例もない
- IV. 【企業主導によるリスク知覚形成】企業から情報伝達したが、企業外部で注意喚起や発生事例等がなく、ユーザーは顕在化した際の影響等について理解できていない

図 3-9 及び図 3-10 に示す通り、STEP 2 の成果、及び STEP 3 の成果を用いることで 4 象限のいずれに各 AI リスク項目を振り分けるべきなのかを判断することができる。具体的な判断基準としては、検討対象となる AI リスク項目が、STEP 2 において説明済みに振り分けられたのか、もしくは、未説明に振り分けられたのか、という点、そして、STEP 3 において社会的にリスク喚起がなされているか顕在化事例があるかを調査した結果、対象となる AI リスク項目に関係のある事象・事例が存在するか、存在しなかったかという点である。

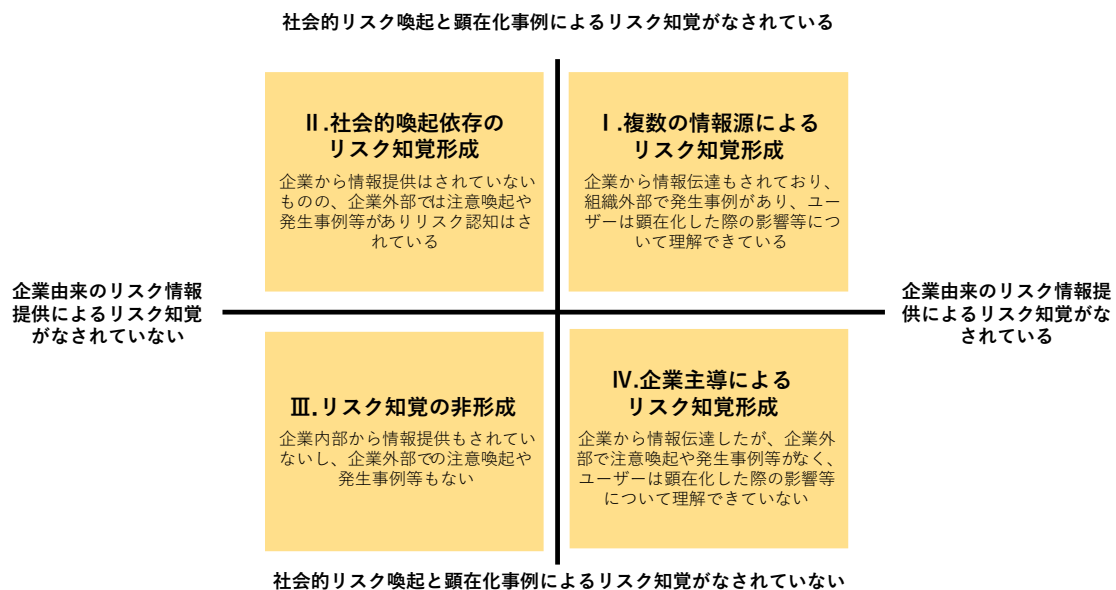


図 3-8 リスク知覚状況マトリクス

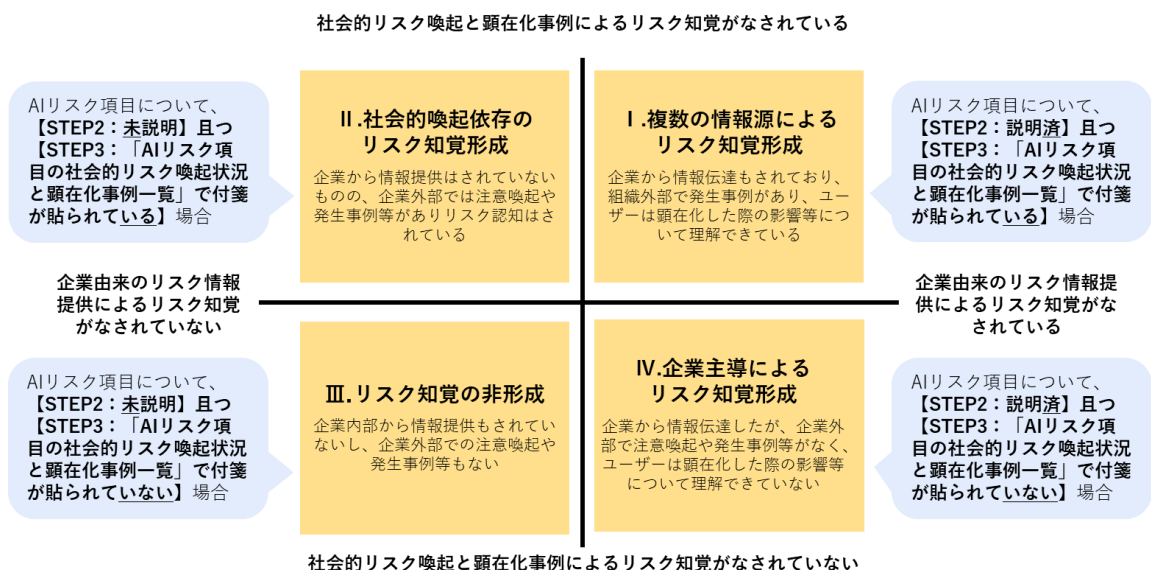


図 3-9 4 象限への振り分けの考え方

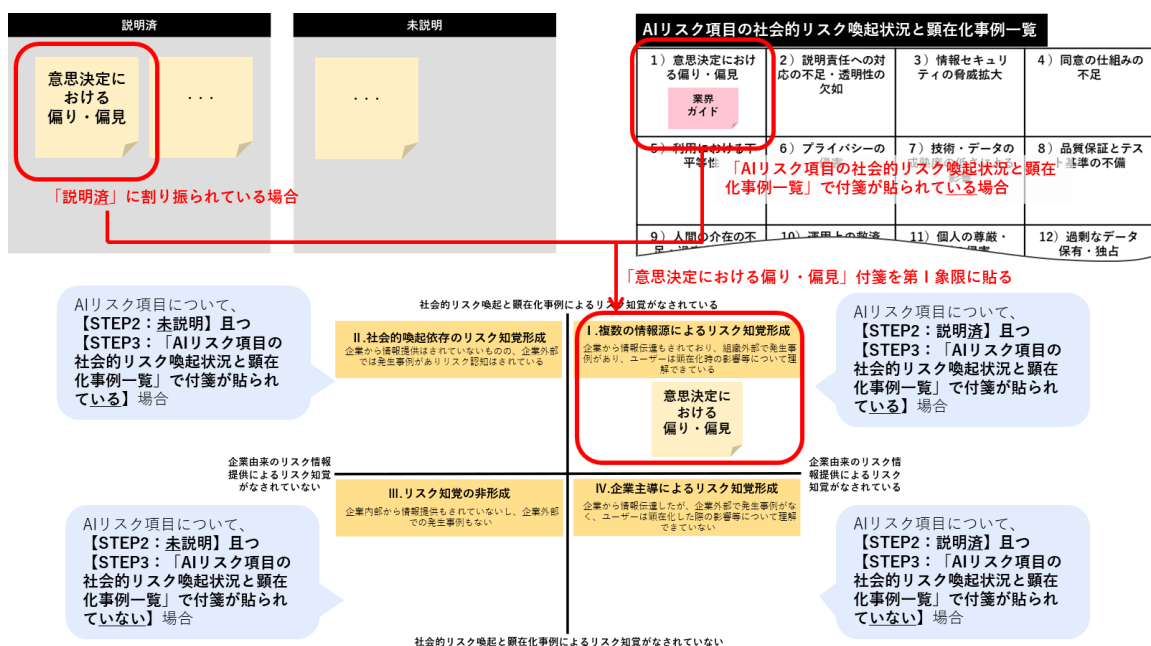


図 3-10 4 象限への振り分け作業イメージ

3.5 リスク性向、リスク知覚状況を踏まえ知覚リスク適正化に向けた情報提供の改善プロセスの設計

3.1 で述べた通り、意思決定に必要な情報の開示と選択肢の多様性の保証は企業にとってはハードルの高い情報提供である。企業が戦略的に信頼関係を構築するためにリスクを隠蔽する行為は、その企業組織事態と利害関係者には従認識されているよりもはるかに有害な損害がその企業自体にももたらされることを示唆している（Robert B Cialdini, 2014）ものの、Dichu Bao ら（2021）によると、企業は状況によって自身の弱点や失敗を競合他社に知られることを避けるためにリスク情報を隠蔽する傾向にあることが課題として指摘されている。

以上のことから、このプロセスの設計においては、知覚リスク適正化に向けた情報提供の改善に対する要求「リスク知覚状況に応じて、企業のリスク情報提供に対する前向きな意識を促進すること」、及び「企業のリスク情報提供に対する改善を支援できること」に応える提案を検討する。

3.5.1 先行研究及びプロトタイプ実施に基づく機能要求定義

知覚リスク適正化に向けた情報提供の改善に対する要求である「リスク知覚状況に応じて、企業のリスク情報提供に対する前向きな意識を促進すること」、及び「企業のリスク情報提供のあり方に対する改善を支援できること」に応える提案を設計するために、筆者の知人から参加者 2 名を募り、それぞれプロトタイプに対するインタビューを実施した。参加者 2 名の一覧を表 3-11 に示す。

表 3-11 プロトタイプ参加者一覧

ID	企業形態・規模	AI サービス・製品の開発における役割	AI サービス・製品の分野
P001	大企業（1000 名以上）	プロジェクトマネージャー	飲料メーカー AI 活用支援
P002	大企業（1000 名以上）	プロジェクトマネージャー	教育支援 AI

プロトタイプに関するインタビューの目的の一つ目は、企業がリスク情報提供に対する前向きな意識を持つのはどのような理由かを特定することとして、二つ目は、企業のリスク情報提供のあり方に対する改善の検討を実効的に行うことが難しい要因を探ることとした。

プロトタイプでは、ケーススタディを用いてリスク情報提供内容・方法の改善のステップを実施してもらったワークの説明を行ったうえで、ワークに関しての所感及びフィードバックについて伺った。AI リスク項目一覧表と各 AI リスク項目についてリスク知覚状況を 4 象限で表現した図（図 3-11）を所与の条件として提示した上で、各象限について改善に向けたアドバイスが記載されたフレームワーク（図 3-12）を参照すれば各 AI リスク項目のリスク情報提供のあり方について改善を検討することができるかというワークとした。

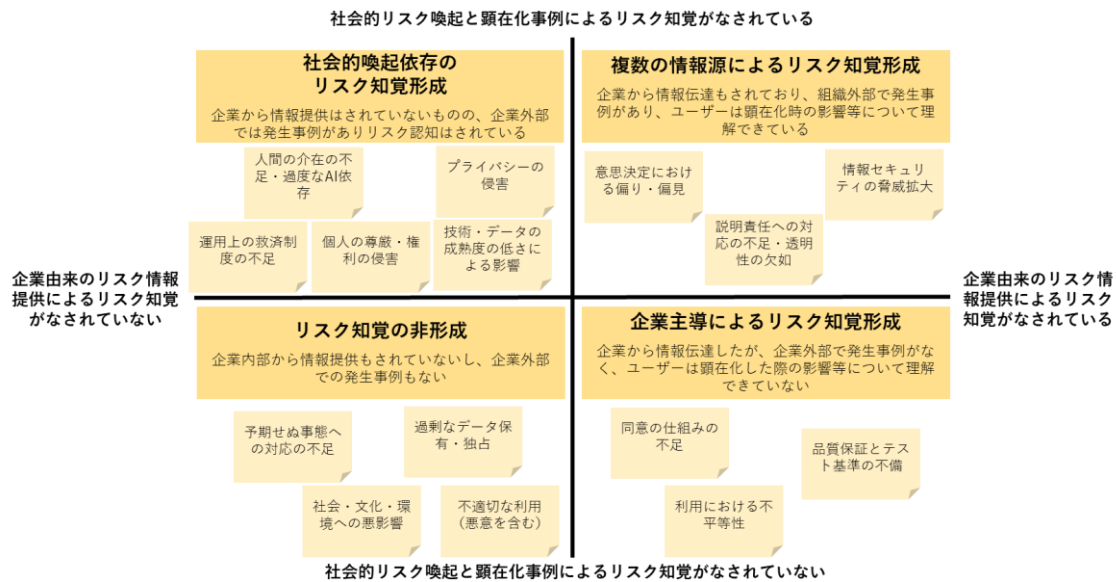


図 3-11 各 AI リスク項目に関するリスク知覚状況（プロトタイプ）

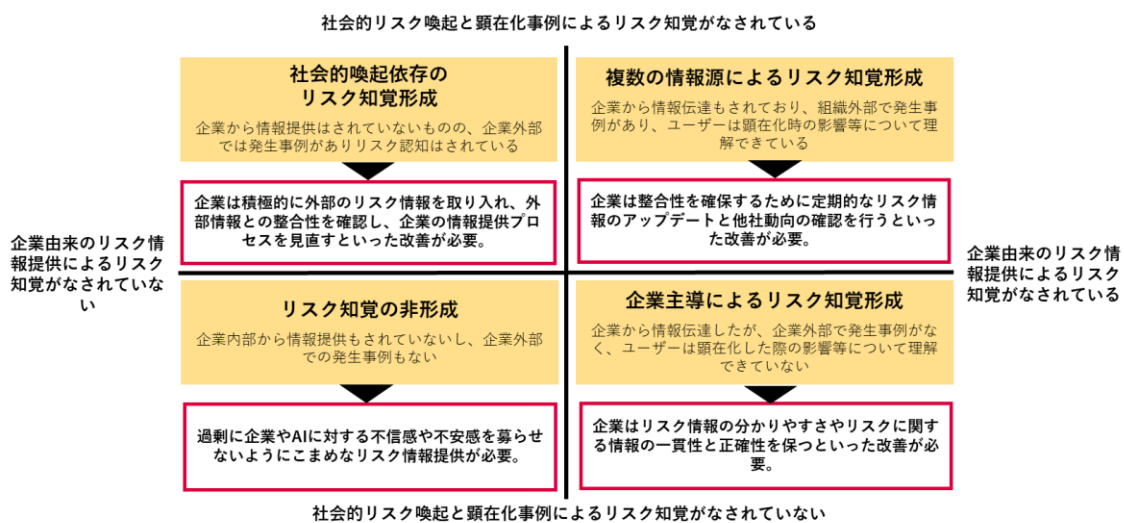


図 3-12 各象限について改善に向けたアドバイスが記載されたフレームワーク（プロトタイプ）

プロトタイプを通して得られたフィードバックは以下の通りであった。

表 3-12 プロトタイプを通し得られたフィードバック

P001 から得られたフィードバック

- ・ 会社がビジネスに AI を活用するという、比較的新しい取組みを進めていたとしても、従来と実施していることは大きく変わっていない。リスク情報提供についてはステークホルダーとの議論上で必要になったことを説明することもあるが、実際には開示項目を決める際には、競合他社のリスク情報提供状況や開示内容を参考にするという、伝統的なアプローチを採用している。
- ・ 現在のアプローチに加え、これまで実施していなかった新たな取組みを検討することが重要である。
- ・ リスクに関する話は、AI システムの運用過程で自然に生じるものである。特に、アルゴリズムやデータ設計を議論する際にリスクの説明が必要になる。従って、リスクについての説明の改善を検討しろ、と言われても考えることが難しいため、AI システムの構成要素を織り交ぜて検討するような流れとした方が実務上でも使いやすいのではないかと。

P002 から得られたフィードバック

- ・ ツールが網羅的になるにつれて、初めの段階では PoC（概念実証）で充分であるが、将来的にどのような内容を記載すべきかが理解できると良い。
- ・ ツールは通常考慮されない項目に気付かせる役割を持ち、新たな気づきを促す。
- ・ ツールによる記載内容を、全社共通の部分、事業部門、AI 開発チームごとに分けることが可能である。
- ・ ツールを使用することが自社の保護につながると認識されると、ツールの使用に対する動機付けになる。

プロトタイプに関するフィードバックとして、上記の通り、P001 より興味深い意見が寄せられた。具体的には、現代の AI 技術を活用しているにもかかわらず、企業がリスク情報を説明・開示する際のプロセスが伝統的な方法論に根差しているという現実が指摘された。これは、企業がリスク情報提供としての情報開示項目を決定する際に、主要な競合他社がどのような情報を公開しているかを調査し、それに倣って情報提供を実施しているという、長年にわたって確立されたアプローチを踏襲していることを意味する。P001 は、このような既存の方法論にとらわれず、これまで実施されてこなかった新たな取組みを検討することの重要性を強調している。さらに、AI システムに関するリスクの扱いに関しても、リスクはそれ自体の話題としてではなく、AI システムの運用中に、例えばアルゴリズムやデータの設計をどのように行うかといった議論の中で自然に取り上げられるべきであると提案している。これは、AI システムの構成要素とリス

ク情報の説明とを結び付けることの重要性を示唆している。また、P002 からのフィードバックでは、ツールの網羅性に関する価値ある視点が提供された。当初は Proof of Concept (PoC) 段階では完全に網羅的である必要はないが、将来的にどのような情報を記載すべきかを理解していることが重要であるとの意見が示された。このツールは、普段は考慮されないような項目に気付かせ、新たな洞察を促す役割を果たすとよいとの指摘があった。さらに、記載すべき内容を全社共通の部分、事業側での部分、AI 開発側での部分といったように分類することで、より効果的な情報管理が可能になると指摘された。これは、自社を守るための戦略としてツールを使用することが、ユーザーのモチベーションとなるという考えに基づいている。

このプロトタイプから得られたフィードバックは、企業の AI 製品・サービスの開発プロジェクト遂行上で実施されるリスク情報提供（コミュニケーション）に関する重要な洞察を提供している。まず、リスク情報提供の具体的な必要性について、企業は競合他社の開示内容を基に自身の情報開示項目を定める従来の方法から一歩進んで、AI 技術の機能と潜在的なリスクを包括的に理解し、これを開示プロセスに反映させる必要がある。リスク情報の提供は、単にリスクの存在を認識すること以上に、それらが企業の運営や AI システムの設計にどのような影響を及ぼす可能性があるかを把握することを含意する。これは、企業が情報開示を通じてリスクに対処し、透明性を高めるための重要なステップである。しかし、従来から踏襲されてきた方法を変えるためには、なぜ企業としてリスク情報提供プロセスを変える必要があるのか、といった点の必要性を十分に理解できないと実行には移すことが困難となる。そこで、リスク情報提供を行うことでリスクマネジメントとして自社を守ることにつながるような記載を追加することが必要だと判断した。以上のことから「リスク情報提供を実施する具体的な必要性を理解できること」を一つ目の機能要求として整理した。

続いて、リスク情報提供内容・方法の改善の検討ステップについての機能要求について検討を行った。AI 技術は、アルゴリズムやデータの設計など、その構成要素において多様なリスクを孕む可能性がある。このため、企業はこれらの構成要素を詳細に分析し、リスクを特定してそれに応じた対策を講じる必要があり、このリスクマネジメントの過程において、AI システム構成要素の設計・運用に伴うリスクを特定することが要求される。そのため、リスク情報提供に関しても AI システム構成要素と関連付けて検討を行うことで、より適切な情報提供のあり方の改善につながると考えた。そこで、「関連する AI システム構成要素を参照しながら改善に向けた検討を行えること」を二つ目の機能要求として整理した。また、今回のプロトタイプではリスク情報提供内容・方法の改善の検討ステップでは改善策を具体化することを求める流れであると説明した。しかし、プロトタイプ参加者からは、すぐに改善策を具体化することは困難であり、実際の業務上はリスク情報提供の改善に向けて、社内の関係者や調査を実施するステップを

挟む必要があるとの意見であった。そのため、具体的な解決策をすぐに決めるのではなく、「改善のためにどのようなアクションを取るかの思考を促進させること」を三つ目の機能要求として整理した。

上記を取りまとめると、このプロセスの設計においては、知覚リスク適正化に向けた情報提供の改善に対する要求「リスク知覚状況に応じて、企業のリスク情報提供に対する前向きな意識を促進すること」、及び「企業のリスク情報提供に対する改善を支援できること」に応える提案について、以下の3つを機能要求として整理した。

- (c-1) リスク情報提供を実施する具体的な必要性を理解できること
- (c-2) 関連する AI システム構成を基に改善に向けた検討を行えること
- (c-3) 改善のためにどのようなアクションを取るかの思考を促進させること

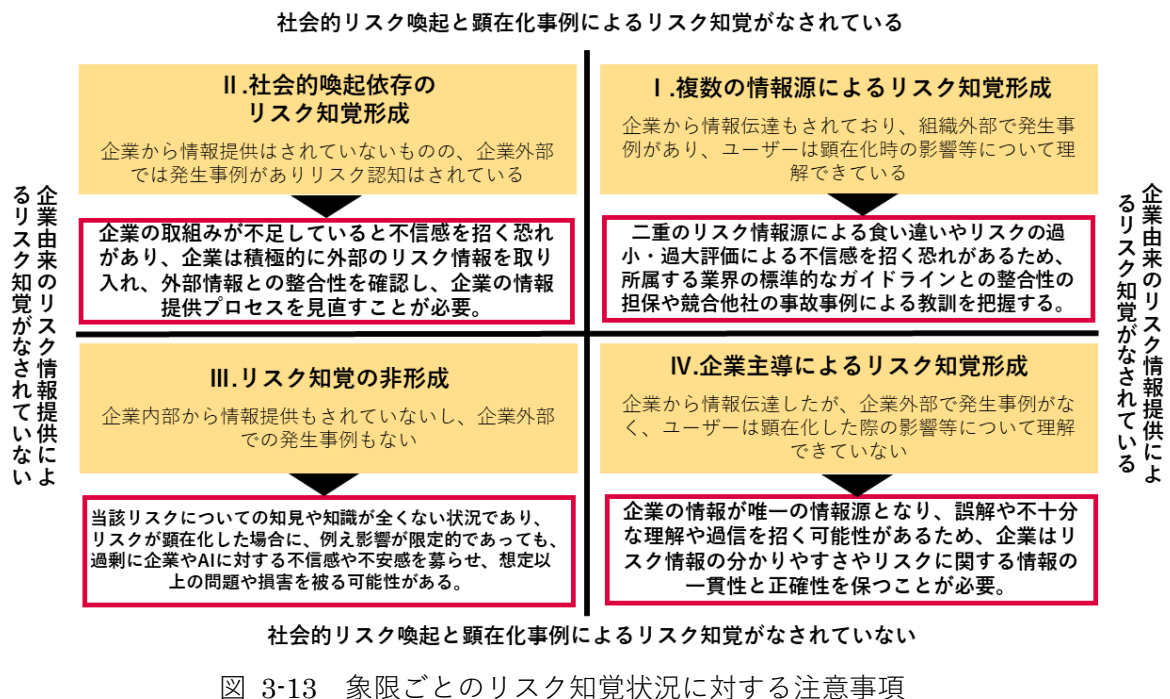
3.5.2 ステップの提案

(1) リスク情報提供のあり方に係る注意事項の把握ステップ

機能要求である「リスク情報提供を実施する具体的な必要性を理解できること」を満たすために、各象限について改善に向けたアドバイスが記載されたフレームワークにおいて、各象限ごとについてリスク情報提供を実施しない場合に発現する企業に対する影響を注意事項として記載する方針とした。

第一象限については、Fischhoff (1995) がリスクコミュニケーションにおいて、正確で一貫した情報提供の重要性が対象者の信頼に強く影響することを指摘している。第一象限に当てはまるリスク項目については、「企業由来のリスク情報提供によるリスク知覚がなされている」、且つ、「社会的リスク喚起と顕在化事例によるリスク知覚がなされている」という複数のリスク情報減によりリスク知覚の形成がなされているため、二重のリスク情報源による食い違いや誤解による不信感を招く恐れがあることを注意事項として整理した。第二象限については、Gambetta (1988) は、信頼は「能力」、「公正さ」、「価値類似性」の3要素が重要であり、その「公正さ」に透明性が含まれると指摘しており、企業からのリスク情報に関する透明性を担保しておかないと信頼には繋がらないという関係を整理している。そのため、第二象限に当てはまるリスク項目については、「企業由来のリスク情報提供によるリスク知覚がなされていない」、且つ、「社会的リスク喚起と顕在化事例によるリスク知覚がなされている」という社会起点のみのリスク知覚の形成がなされているため、AI システム開発企業のリスク情報提供、リスク情報開示に係る取組みが不足していると AI サービスプロバイダー企業に感じられその結

果不信感を持たれる恐れがあることを注意事項として整理した。第三象限については、Slovic (1987) が示唆する、人々のリスク知覚は実際のリスクと異なることが多く、リスク情報が不足していると、人はリスクを適切に理解できず、不必要な不安や不信感を抱える可能性がある点を考慮した。そのため、第三象限に当てはまるリスク項目については、「企業由来のリスク情報提供によるリスク知覚がなされていない」、且つ、「社会的リスク喚起と顕在化事例によるリスク知覚がなされていない」というリスク知覚が非形成である状況については、AI サービスプロバイダー企業がリスクを適切に理解できず、過剰に企業や AI に対する不信感や不安感を募らせることで、予期せぬ問題や損害を被る可能性があることを注意事項として整理した。第四象限については、Akerlof (1970) が市場における情報の非対称性の概念を探究した先行研究において、情報提供者と情報授受者（本先行研究においては消費者）間の情報ギャップが市場にどのような影響を与えるかを考察しており、企業による情報提供が唯一の情報源となると、情報の受け取り手はその情報の正確性を判断する基準がなくなる、という課題が指摘されている。その課題を踏まえ、第四象限に当てはまるリスク項目については、「企業由来のリスク情報提供によるリスク知覚がなされている」、且つ、「社会的リスク喚起と顕在化事例によるリスク知覚がなされていない」とい AI 開発企業の情報が AI サービスプロバイダー企業にとって唯一の情報源である状況について、誤解や不十分な理解や過信を招く可能性があることを注意事項として整理した。これらの注意事項をフレームワークとして取りまとめた。



(2) リスク情報提供内容・方法の改善の検討ステップ

リスク情報提供内容・方法の改善の検討ステップでは、上記のフレームワークを基に改善を検討してもらうが、機能要求である「関連する AI システム構成を基に改善に向けた検討を行えること」、「改善のためにどのようなアクションを取るかの思考を促進させること」を満たすために、改善の検討フレームワークを設計した。象限ごとに、上から AI リスク項目と AI リスク項目に関連する AI システム構成要素を記入した上でリスク情報提供のあり方の改善に向けたアクションを検討する構成とした。関連する AI システム構成要素とは、企業由来のリスク情報提供状況の一覧化ステップで、整理した AI リスク項目一覧表に記載した内容である。つまり、リスク情報提供の方針を検討のアドバイスとして参照しながら各 AI リスク項目について、リスク情報提供のあり方について改善を図るためのアクションを関連する AI システム構成要素を考慮できる作りである。

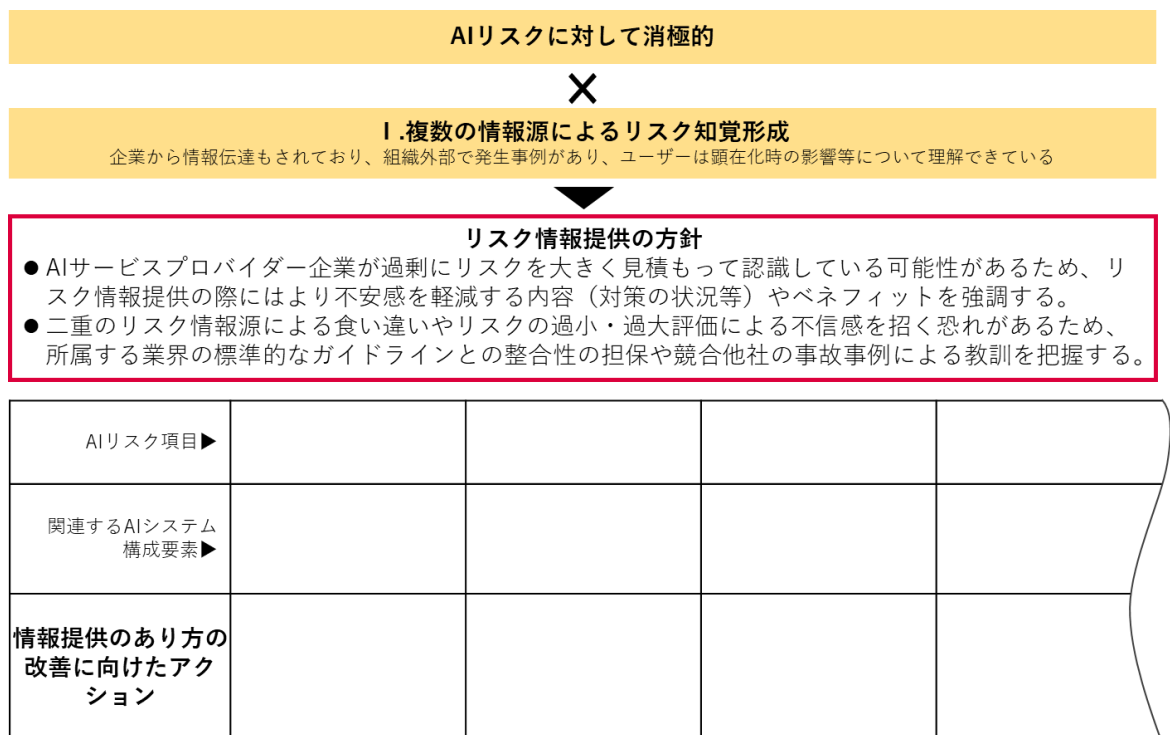


図 3-14 リスク情報提供内容・方法の改善の検討ステップのフレームワーク（第一象限を例に）

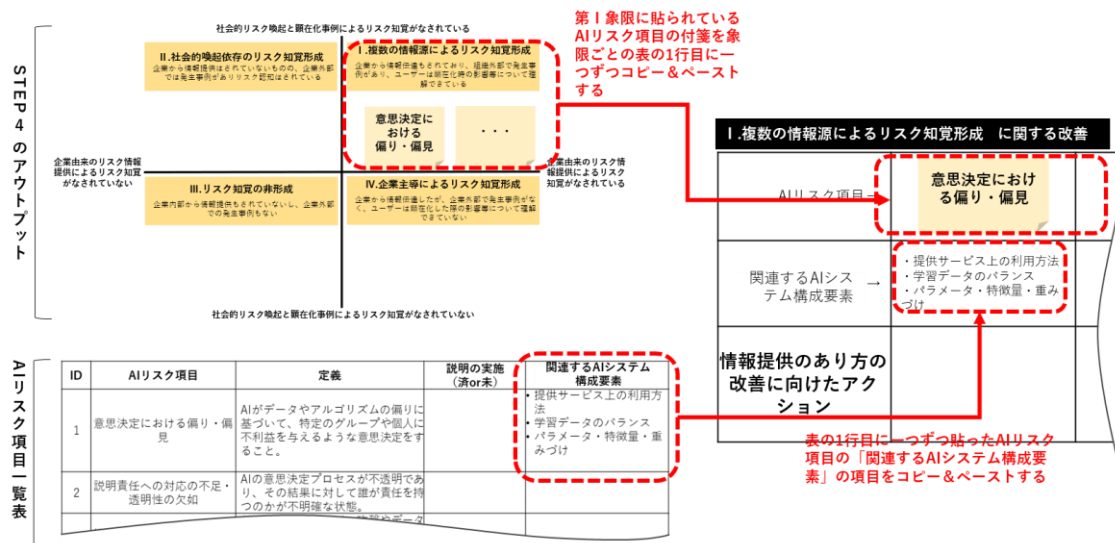


図 3-15 リスク情報提供内容・方法の改善の検討ステップのフレームワークの記入手順

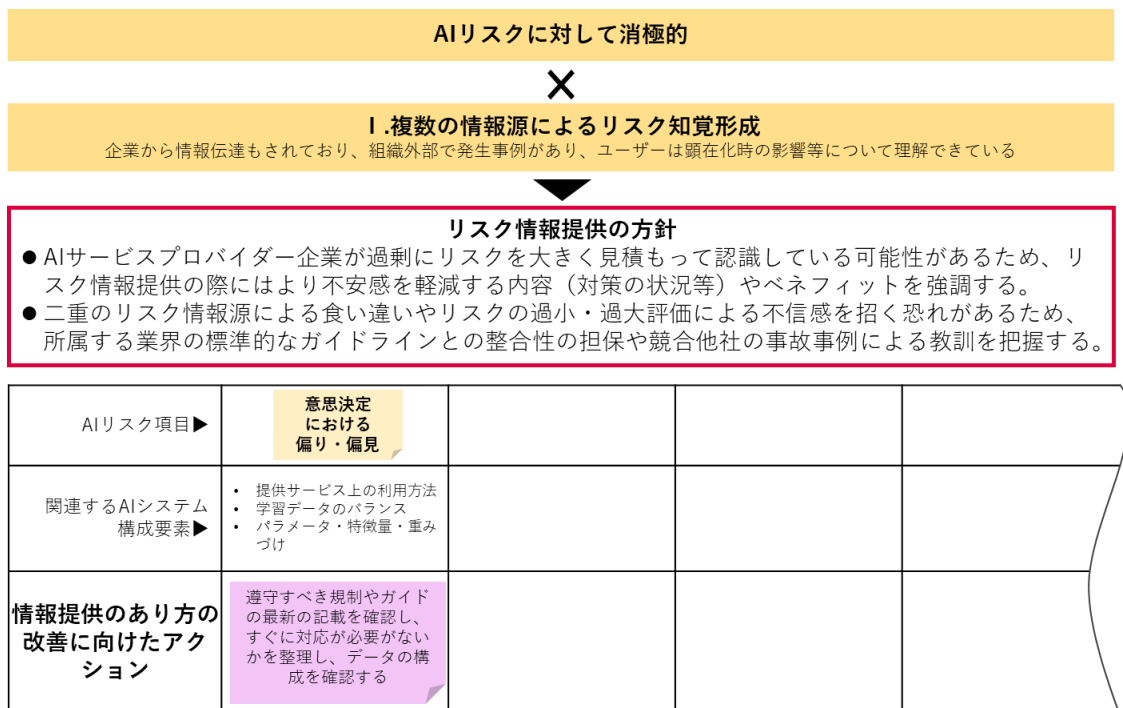


図 3-16 リスク情報提供内容・方法の改善の検討ステップのフレームワークの記載例（第一象限を例に）

3.6 設計したステップの統合

提案ツールは全部で5つのステップに分かれており、AIプロバイダー企業のAIに対するリスク性向を把握したうえで、リスク情報がAI開発企業からどのように伝達さ

れているか、及び、社会的にどのような周知がなされているかを整理し、AI サービスプロバイダー企業のリスク知覚状況を検討し、その整理結果に基づき、リスク情報提供のあり方の改善に向けたアクションを検討する流れである。(図 3-17)

まず、情報として提供されたリスク情報を AI サービスプロバイダー企業がどのように受け止めるかを推測するために、AI サービスプロバイダー企業の AI に対するリスク性向を把握するステップ (STEP1) を提案する。そのうえで、AI 開発企業から AI サービスプロバイダー企業にサービス・製品に係るリスクについてどのように説明されているかを整理するステップ (STEP2) と、AI 開発企業外部で当該 AI サービス・製品に関わるどのようなリスクに関して注意喚起がなされており、リスク顕在化事例が周知されているかを整理するステップ (STEP3) により、AI サービスプロバイダー企業のリスク知覚状況を検討するための情報を整理する流れとしている。それらの情報を踏まえ、リスク性向を踏まえてリスク知覚状況を検討する設計となっている (STEP4)。最後に、その整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善を実施するためにはどのようなアクションが必要かを検討する (STEP5)。

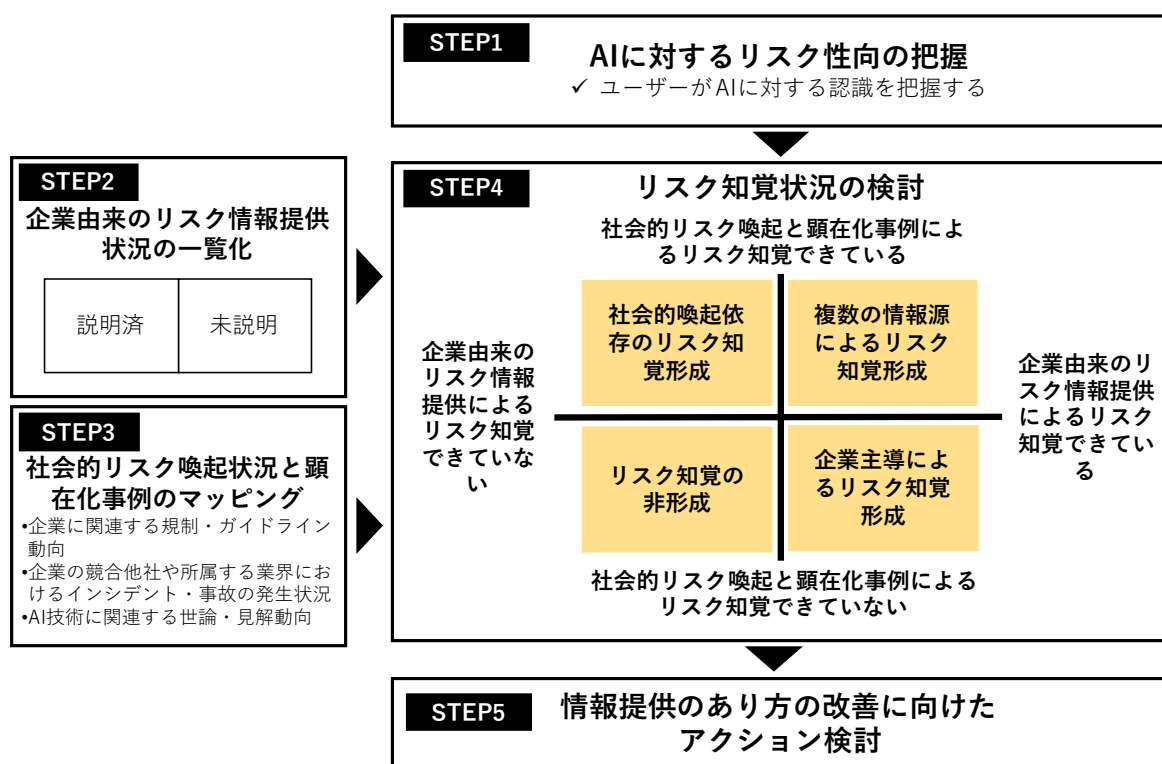


図 3-17 提案ツールのステップ構成

3.7 手順書、提案の使い方

提案に対する要求の一つである「企業の実務上で利用しやすいこと」に応えるべく、提案ツールにはどのようなバックグラウンドを持つ従業員であっても利用できるように手順書を作成した。この手順書があることで、リスク情報提供のあり方やリスクマネジメントの知見が少なくとも、従業員単独でツールを使用することを可能とする。また、検討した結果を実務上のディスカッションや会議の機会を使いやすく、将来的には AI サービス・製品開発に携わる多様なステークホルダー間で共有できるようにするためにオンラインツールを活用した。

3.7.1 手順書

提案ツールには、パワーポイントで視覚的に説明が記載された手順を付属した。この手順書を参照することで、ファシリテーターやガイドなしで個人作業として提案ツールを利用することができる。

手順書には、ツールの目的や準備が必要なもの、全体の流れ、各 STEP での実施内容及び実施イメージを記載した。STEP 3 を例に、各 STEP の説明内容を示す。手順の説明として、各 STEP の実施内容を概観できるように概要及び趣旨をまず提示した上（図 3-18）で、作業の流れ（図 3-19）と実際に作業した場合のアウトプットイメージ（図 3-20）を細かく可視化してある。また、作業に必要な目安となる時間も各ページの右上に提示した。なるべくどのような知見やバックグラウンドであっても適切な理解ができるように実施内容を可視化した図で示した。

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

- AIサービスプロバイダー企業のリスク知覚に影響を与える要因について、以下の観点で調査や発想を実施し、AIリスク項目のそれぞれについて社会的なリスク喚起状況やリスクが顕在化した事例の有無を整理します。具体的なステップは次頁で説明します。

企業に関連する規制・ガイドライン動向

新たな法規制がガイドラインが導入され、それに準拠していないことで、その企業への信頼を損なう要因となる可能性があります。以下について調査・発想してください。

- ・ 新規法令
- ・ 業界ガイドライン等のソフトロー
- ・ 訴訟の結果や判例

企業の競合他社や所属する業界におけるインシデント・事故の発生状況

企業が、競合他社や同業界の企業で発生したリスクや脅威に対して適切な対策を取らないことで、その企業への信頼を損なう要因となる可能性があります。以下について調査・発想してください。

- ・ 他社インシデントの発生状況
- ・ 競合他社や同業界の企業による事故の発生状況

AI技術（そのもの）に関連する世論・見解動向

社会全体の倫理観やプライバシーに対する意識が高まった結果の水準に対して十分な対策を講じていないことで、その企業に対する信頼を損なう要因となる可能性があります。以下について調査・発想してください。

- ・ メディアでのAIやプライバシーの議論
- ・ ユーザーのフィードバックの傾向

図 3-18 概要及び趣旨

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

【STEP3】

- 「自社に係る社会的リスク喚起状況と顕在化事例を構成する要素モデル」について、必要に応じて調査を行いながら、下記の順序で内容やキーワードを、1列目はピンク色の付箋、2列目は青色の付箋、3列目は緑色の付箋に書き出して下さい。
- 列ごとに一番下の行に付箋を貼れたら、記載したキーワード・内容を基に、関連するAIリスク項目一覧リストのリスク項目の枠に付箋をコピー＆ペーストして下さい。

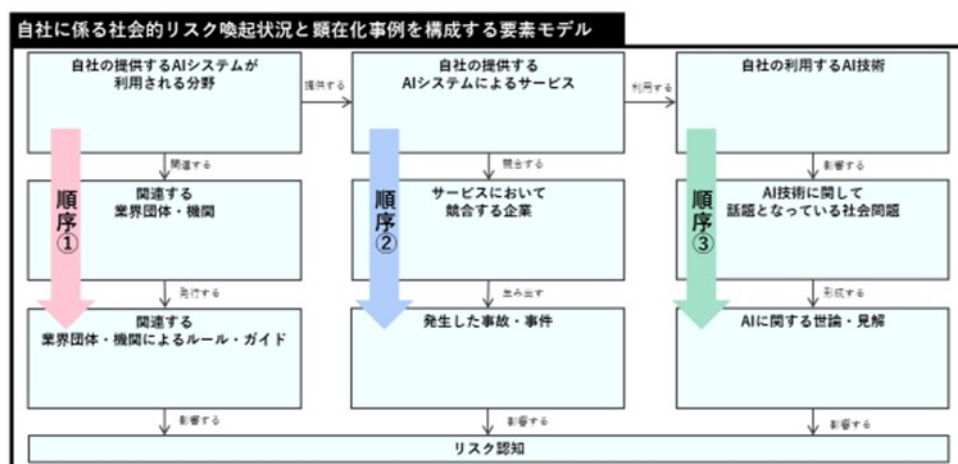


図 3-19 作業の流れ

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

【STEP3】実施イメージ

自社に係る社会的リスク喚起状況と顕在化事例を構成する要素モデル



XXガイドに「プライバシーの侵害」及び「同意の仕組みの不足」について言及がある場合、「関連するAIリスク項目一覧リスト」の当該枠にコピー＆ペーストする

関連するAIリスク項目一覧リスト			
1) 意思決定における偏り・偏見	2) 説明責任への対応の不足・透明性の欠如	3) 情報セキュリティの脅威拡大	4) 同意の仕組みの不足 XXガイド
5) 利用における不平等性	6) プライバシーの侵害 XXガイド	7) 技術・データの成熟度の差による影響	8) 品質保証とテスト基準の不備
9) 人間の介入の不足	10) 誤用・濫用の懸念	11) 個人の尊厳・権利の侵害	12) 適切なデータ保有・抜去

図 3-20 アウトプットイメージ

3.7.2 プラットフォーム

企業間の信頼関係の維持・向上を目的としたツールの設計においてリスク情報の伝達のあり方を検討するため、当初は企業のほとんどが導入している Microsoft アプリケーション（Excel、PowerPoint）のみによるツール提供を考慮していた。しかし、将来的にはサプライチェーン上の様々な企業等のステークホルダーが容易に検討に参加できるよう、創造的な検討や議論が必要なプロセスではオンラインプラットフォームの利用を決定した。これにより、特定の場所に集まる必要なく、どこからでもアクセス可能な形式でリスク情報伝達のあり方を時間や場所に縛られずに検討・議論できるようになる。また、このツールでは、調査を実施しながら発想や調査結果を取りまとめるプロセスを有するため、メモとして使用しやすいよう付箋紙機能や日本語テキスト入力可能なオンラインホワイトボードを使用することとした。このプラットフォームには、要件を満たす限り他のツールに置き換えることも可能だが、本提案では世界で広く利用されている Miro を採用することとした。Miro の使用にはインターネット接続が必要となるが、対象となる企業の担当者は一般的には個人または会社のパソコンを利用できると想定されるため、問題がないと判断した。

3.7.3 ツールの使い方

検討主体は、Miro 上で提供されるツール（図 3-21）を開き、提供された手順書を参照することで知覚リスク適正化に向けた情報提供の改善を実施することができる。手順に従い、順番に進むことで、ステークホルダー間の信頼関係を強化するためのリスク情報提供のあり方の改善に向けたアクションを検討するというアウトプットを最終的に創出が可能となる。

プロトタイプの評価を通じて、各シートの所要時間を確認し、検討主体が計画的に進められるように各シートの左上に所要時間の目安を記載した（表 3-13）。全体のプロセスは約 2 時間で完結するように設計されており、ステークホルダーがリスク情報を効果的に理解し、伝達するための具体的なステップを提供できるよう設計した。

表 3-13 ツールの全体所要時間の目安

内容	目安となる 所要時間（分）	累積所要時間（分）
ご説明	20	20
STEP1	-	20
STEP2	10	30
STEP3	25	55
STEP4	20	75
STEP5	30	105

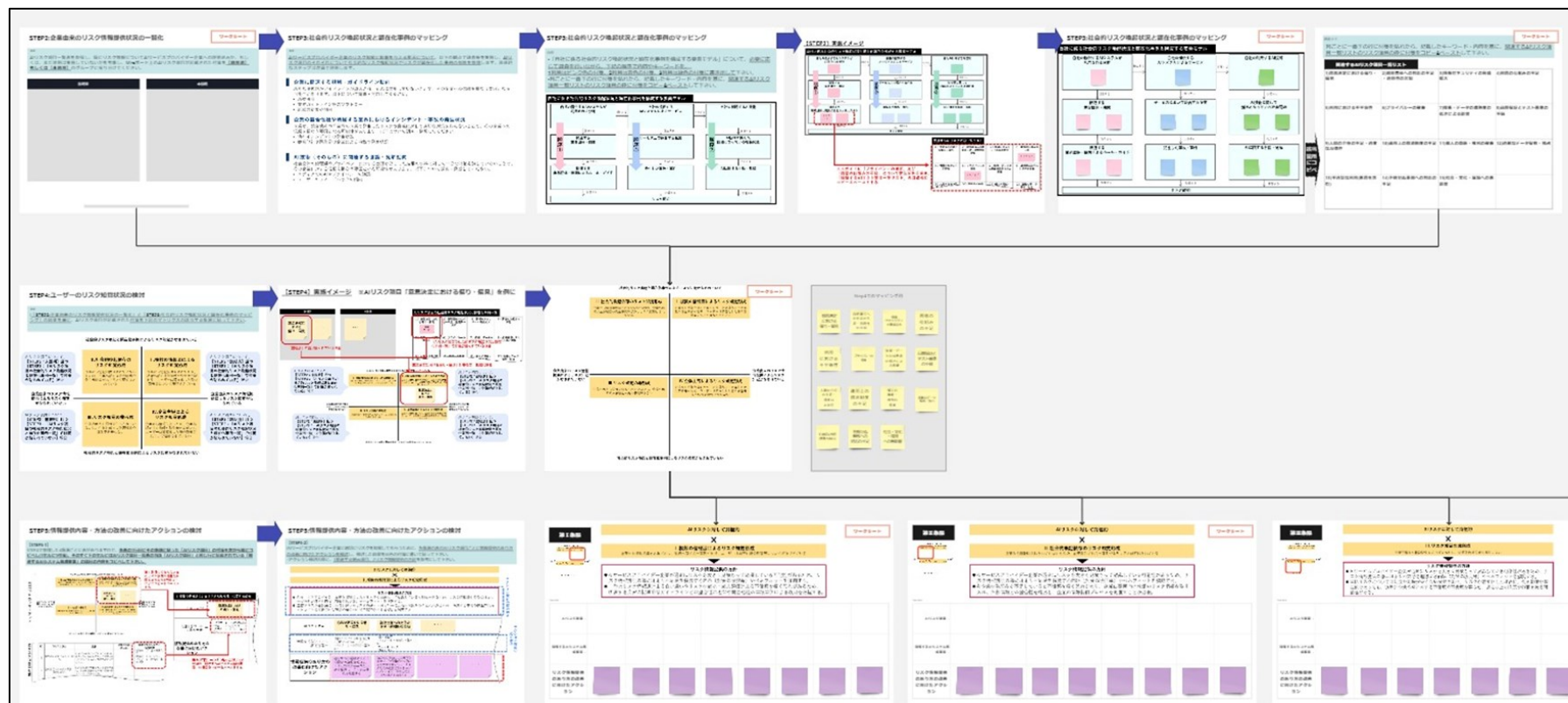


図 3-21 Miro ボード上のツール全体像

第4章 検証と妥当性確認

本章では、第3章で設計した企業間信頼構築のためのリスク情報伝達支援ツールの評価方法、評価結果について記述する。

4.1 検証・妥当性確認の方法

検証・妥当性確認の方法として、企業や公的機関において AI サービス・製品の開発に関与する実務者を対象に、実際にツールを使用してもらうことで、4つの要求（①企業の実務上で利用しやすいこと、②明示的な情報で整理することでリスク知覚状況を客観的に整理できること、③企業のリスク情報提供に対する具体的なアクションを支援できること、④企業のリスク情報提供に対する前向きな意識を促進すること）に対する提案機能が設計通り実行されているのかの検証（Verification）と、リスク情報提供のあり方の改善がなされたのか、という妥当性確認（Validation）を実施した。

提案ツールの検証および妥当性確認の計画を表 4-1 に示す。4つの要求に対する提案機能が設計通り実行されているのかの検証（Verification）について、企業や公的機関において AI サービス・製品の開発に関与する 12 名に検証実験に参加してもらった。事後アンケート及びツールの使用前後での意識や理解の変化量を基に比較による評価を行った。また、ツールの成果物を確認し分析することで、提案の評価を行った。リスク情報提供のあり方の改善がなされたのか、という妥当性確認（Validation）については、ツールの最後のリスク情報提供内容・方法の改善の検討ステップで検討されたアウトプットを対象に客観的評価を実施する。具体的には、成果物に対するリスクコミュニケーションの学術的専門家の2名と、ケーススタディと類似の業務を遂行する実践的専門家からの客観的評価を実施した。

表 4-1 提案ツールの検証・妥当性確認の設計

評価内容	検証対象	検証方法	検証事項	確認事項	
				有効性	理解性・ 利用性
検証	リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと	アンケート (主観的評価)	筆者作成の質問項目	○	○
	明示的な情報で整理することでリスク知覚状況を客観的に整理できるか	アンケート (主観的評価)	筆者作成の質問項目	○	○
		成果物（客観的評価）	STEP3 の記入内容	○	○
	企業のリスク情報提供に対する具体的なアクションを支援できるか	アンケート (主観的評価)	筆者作成の質問項目	○	○
		成果物（客観的評価）	STEP5 の記入内容	○	○
	企業のリスク情報提供に対する前向きな意識を促進するか	アンケート (主観的評価)	筆者作成の質問項目	○	
妥当性確認	企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方を改善できるか	成果物（客観的評価）	筆者作成の評価項目	○	○
		類似業務実施の実践的専門家インタビュー（客観的評価）	オンラインインタビュー内容	○	○

4.2 検証・妥当性確認における評価計画

4.2.1 筆者作成の評価基準による評価

被験者に実施してもらったワークにおいて、その整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善を実施するためにはどのようなアクションが必要かを検討する STEP5 の成果物について評価を実施した。AI システム開発企業、及び AI サービスプロバイダー企業間の信頼構築に向けたリスク情報提供の在り方の改善に貢献するものとなっているかについて筆者作成の評価基準を考慮し、評価を実施した。

評価基準は、文部科学省「リスクコミュニケーション案内」を参考に設計を行った。「リスクコミュニケーション案内」では、リスクコミュニケーションに関するスタンダードな評価方法、指標は存在せず、いまだ発展の途上にあるものの、リスクコミュニケーションの評価基準について一般社団法人リスク研究学会などが中心にまとめた以下の内容が紹介されている。

- ・ 事前に解決したい問題 ・ 目的（ゴール）を設定している。
- ・ リスクを評価した。
- ・ リスク対策を行った。
- ・ 現状の課題を把握している。
- ・ 参加者・関係者（ステークホルダー）の範囲を把握している。
- ・ ステークホルダーのニーズを把握した。
- ・ 対象者の知識レベル、リスクへの意識、リスクリテラシーを把握した。
- ・ 適切な方法の検討と選択を行った。
- ・ 希望者がリスクコミュニケーションを受けるのか、リスクミを受けることがデフォルトで希望者が拒否できるのかを設定する（オプトイン、オプトアウトの設定）。
- ・ 場の設計をした。

上記を踏まえ、STEP5 の成果物に対する評価基準を表 4-2 に示す構造とした。

まず、リスクコミュニケーションをリスク対応としての評価とコミュニケーションとしての評価の2つに類型化した。「(1-1)リスク評価に繋がるアクションか？」という評価基準は、「リスクコミュニケーション案内」の「リスクを評価した」指標と関連している。リスクの特定、分析、評価を行うアクションは、リスクコミュニケーションの成功に不可欠であり、効果的なリスク管理戦略を策定するための基盤を提供するため、

評価基準に含むこととした。「(1-2)リスク対策に繋がるアクションか？」という評価基準は、「リスク対策を行った」という指標に対応する。特定されたリスクに対する対策や解決策の提案は、リスクを最小限に抑え、関係者に信頼感、安心感を提供する上で重要であるとの指摘を参考とした。「(1-3)現状の課題・ニーズ把握に繋がるアクションか？」という評価基準は、「現状の課題を把握している」と「ステークホルダーのニーズを把握した」の両指標と関連する。現在の状況や必要性を理解することは、リスクコミュニケーションを適切に設計し実施するために不可欠である。「(2-1)透明性が担保されているか？」という評価指標は、「プロセスや意思決定の透明性を確保する」という指標に対応する。透明性は信頼性を高め、関係者の不安や誤解を減少させるために重要となる。「(2-2)双方向性があるか？」という評価基準は、「参加者・関係者（ステークホルダー）の範囲を把握している」および「対象者の知識レベル、リスクへの意識、リスクリテラシーを把握した」の指標に対応する。ステークホルダー間の双方向コミュニケーションは、理解と協力を促進し、効果的なリスク管理と情報共有を可能にするという「リスクコミュニケーション案内」の第三章で説明されている通りであり、双方向性という用語はそこから用いることとした。

表 4-2 筆者作成の STEP5 成果物の評価基準

評価基準（大分類）	評価基準（中分類）	内容
(1)リスク対応としての評価	(1-1)リスク評価に繋がるアクションか？	リスクの特定、分析、評価を行うアクションに繋がる検討がなされているか
	(1-2)リスク対策に繋がるアクションか？	特定されたリスクに対する対策や解決策を提案するアクションに繋がる検討がなされているか
	(1-3)現状の課題・ニーズ把握に繋がるアクションか？	現在の状況や必要性を理解し、それに基づいて行動するアクションに繋がる検討がなされているか
(2)コミュニケーションとしての評価	(2-1)透明性が担保されているか？	プロセスや意思決定の透明性を確保するアクションに繋がる検討がなされているか
	(2-2)双方向性があるか？	ステークホルダー間の双方向コミュニケーションを促進するアクションに繋がる検討がなされているか

4.2.2 リスクコミュニケーション専門家による評価

本研究の提案手法について、および本研究の評価結果についての総合的な評価を客観的な視点から行なってもらうことを目的として、専門家インタビューを実施した。専門家インタビューは、オンライン会議ツールによる対面形式のインタビューとした。インタビュー項目は下記に示す通りである。

- ・ 被験者それぞれの成果物についてリスク情報提供の在り方の改善に向けて適切なものになっているか（リスク対応の観点、コミュニケーションの観点それぞれから）
- ・ 全体として、提案手法を使用することでリスク情報提供の在り方が進んでいるといえるか

専門家インタビューの対象者としては、(株)三菱総合研究所で合意形成やリスクコミュニケーションの領域で業務を約 10 年間実施した経験を持つ白井浩介氏を選定した。

4.2.3 ケーススタディの類似事例との比較による評価

提案ツールを評価するため、10 名の被験者にケーススタディを用いてワークを実施してもらった成果物と実際の業務を比較することで評価を実施した。ケーススタディでは、採用プロセスにおける AI の使用を例に取り上げた。これは、M 社と N 社が共同開発した ES を基に採用判定を行うプロジェクトと類似しており、この類似性から、ツールの使用による効果を適切に評価することができると考えた。そこで、被験者の検討した各 AI リスク項目に対するリスク情報伝達の改善に向けたアクション内容に係る成果物を実際の業務に関与する実務者へのインタビューを通じて比較検証し、実業務で実施されていないアクションの特定や新たな気づきや洞察が生まれるかどうかを確認することで、提案ツールの有効性を評価することを目指した。

4.3 実験概要

提案ツールの検証のため、被験者を 12 名集め、2023 年 11 月 30 日から 12 月 15 日の間に検証実験を実施した。AI サービス・製品の開発を実施する企業や公的機関を対象に AI サービス・製品の開発に様々な立場で関与する実務者 10 名から協力を得た。

4.3.1 被験者

被験者一覧を表 4-3 に示す。

AI サービス及び製品開発に携わる様々な背景を持つ被験者を対象に検証を実施した。被験者は、年齢、開発経験年数、関与している分野、役割及び立場によって特徴づけられている。被験者 01 は 35～39 歳の範囲にあり、児童相談業務のリスクアセスメント AI やエンゲージメント分析 AI の開発に 5 年の経験を持つプロジェクトマネージャーである。最も年長の被験者 02 は 50 歳以上で、ファクトリーオートメーション分野において同じくプロジェクトマネージャーとして 5 年の経験を有している。画像類識別を専門とする被験者 03 は 40～44 歳で 7 年の経験があり、プロジェクトマネージャーとして活動している。一方で、30～34 歳の若手層に位置する被験者 04、05、06 は、それぞれ異なる専門分野において 1～3 年の経験を有しており、AI を用いた技術によるアナログ規制見直しの技術実証伴走支援、ゲーム映像のアクション検知、機械制御の画像判定 AI の開発に従事している。特に被験者 04 は品質保証の立場で、05 は個人の取り組みとして、06 はシステムアーキテクトとしての役割を担っている。35～39 歳の中堅層である被験者 07 と 08 は、それぞれユーザーの声分析 AI やニュース解析 AI、児童相談所における AI 活用という分野で 5～6 年の経験を持ち、データサイエンティストとプロジェクトマネージャーの立場からプロジェクトを推進している。最も若い被験者 09 は 25～29 歳であり、LLM を活用したサービス開発に 5 年間従事し、サービス開発の役割を果たしている。経験年数においては被験者 10 が最も長く、10 年の経験を持つ 35～39 歳のプロダクトマネージャーであり、画像認識や物体検出といった技術の開発に関わっている。このように、極端な属性や業務経験の偏りが無いように注意を払い、被験者群を整備した。

表 4-3 被験者一覧

被験者 ID	年齢	AI サービス・製品の開発等経験年数	関与している分野	役割・立場
01	35～39 歳	5	児童相談業務のリスクアセスメント AI、エンゲージメント分析 AI	プロジェクトマネージャー
02	50 歳以上	5	ファクトリーオートメーション	プロジェクトマネージャー
03	40～44 歳	7	画像類識別	プロジェクトマネージャー
04	30～34 歳	1	AI を用いた技術によるアナログ規制見直しに関する技術実証伴走支援	品質保証
05	30～34 歳	3	ゲーム映像のアクション検知	個人の取組
06	30～34 歳	3	機械制御の画像判定 AI	システムアーキテクト
07	35～39 歳	6	ユーザーの声分析 AI、ニュース解析 AI	データサイエンティスト
08	35～39 歳	5	児童相談所における AI 活用	プロジェクトマネージャー
09	25～29 歳	5	LLM を活用したサービス開発	サービス開発
10	35～39 歳	10	画像認識、物体検出	プロダクトマネージャー

4.3.2 実験内容

(1) 実験の流れ

被験者に対して実験概要をメールで説明し、承諾を得た後に事前アンケートを送付した。アンケートの回答を得たのちに、実験に使用する提案ツールを掲載したオンラインホワイトボードの Miro の URL を送付し、ツールへのアクセスの確認を依頼した。また併せてツールの手順説明書を送付した。オンラインホワイトボードの Miro は、被験者それぞれ別のホワイトボードを用意した。

Miro の URL 及びツールの手順説明書の送付後に、オンラインで実験の説明を約 30 分間で実施し、提案ツールの目的や手順書の参照方法、全体の流れについて被験者が理解できたことを確認した。説明後、各被験者に個人ワークを開始してもらい、約 2 週間後にワークの提出締切日として設定した。

(2) 事前アンケート

被験者の属性や提案ツールを使用する前の状態を把握するために表 4-4 に示す項目で事前アンケートを被験者全員に実施した。また、検証では、4 つの要求のうち、②明示的な情報で整理することでリスク知覚状況を客観的に整理できること、③企業のリスク情報提供に対する具体的なアクションを支援できること、④企業のリスク情報提供に対する前向きな意識を促進することに関して、提案ツールの使用前と使用後を比較することで評価を行う。

表 4-4 事前アンケートの質問項目

ID	観点	質問項目	回答方法	選択肢											
1	管理	検証の個人 ID	自由記述												
2	属性	あなたの年齢を選択してください。	選択式												
3	属性	AI サービス・製品の開発等に関わった経験年数を記入してください。（研究やガイドラインの検討も含んでください。）	自由記述												
4	属性	現在関与している AI サービスや製品の分野を具体的に記述してください。（例：クレジットカード不正検知 AI、採用 AI など、概要が分かるようにご記載いただければ幸いです。）	自由記述												
5	属性	現在の AI サービス・製品開発等における役割や職業を選択してください。	選択式	システム アーキテ クト	データサ イエンテ イスト	ソフトウ ェアエン 지니어	データエ ンジンア	AI 研究者	プロジ ェクト マネー ジャー	UX/UI デザイ ナー	品質保 証	ビジネ スアナ リス ト・マ ーケタ ー	営業・ セール ス	その他	
6	属性	5 で回答した役割・職業のほかに、過去に経験した役割や職業があればお答えください。（AI サービス・製品に関係のないものでも構いません。）	自由記述												
7	意識	AI サービス・製品に係るリスク情報提供に関する課題について、あなたの意識の度合いを 5 段階で評価してください。	5 件法	1:小 さい	2	3	4	5:大 きい							

8	意識	AI サービス・製品におけるステークホルダーとの連携に関する課題について、あなたの意識の度合いを5段階で評価してください。	5 件法	1:小さい	2	3	4	5:大きい
9	意識	社外の変化（市場動向、規制、技術進化など）に追従していく必要性に対する、あなたの課題意識の度合いを5段階で評価してください。	5 件法	1:小さい	2	3	4	5:大きい
10	意識	実務上でステークホルダーのリスク性向（リスクに対する考え方や前向きさ等）の把握のやり方が理解できている	5 件法	1:小さい	2	3	4	5:大きい
11	意識	リスク情報を整理するために何を調査すべきかが理解できている	5 件法	1:小さい	2	3	4	5:大きい
12	意識	適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整理ができる	5 件法	1:小さい	2	3	4	5:大きい
13	意識	（ステークホルダーに対する）リスク情報提供の検討に必要なプロセスを理解できている	5 件法	1:小さい	2	3	4	5:大きい
14	意識	（ステークホルダーに対する）リスク情報提供に向けて必要なアクションを実行できる	5 件法	1:小さい	2	3	4	5:大きい
15	意識	リスク情報をステークホルダーに提供することの重要性を理解している	5 件法	1:小さい	2	3	4	5:大きい
16	意識	リスク情報をステークホルダーに提供することのモチベーションがある	5 件法	1:小さい	2	3	4	5:大きい

(3) 事後アンケート

ツールについて STEP すべてを完了した被験者全員に対して事後アンケートを実施した。各 STEP の利用度合いや理解度合いを測るための設問や、4つの要求である、①企業の実務上で利用しやすいこと、②明示的な情報で整理することでリスク知覚状況を客観的に整理できること、③企業のリスク情報提供に対する具体的なアクションを支援できること、④企業のリスク情報提供に対する前向きな意識を促進することに関して、設問を設計した。要求のうち、②明示的な情報で整理することでリスク知覚状況を客観的に整理できること、③企業のリスク情報提供に対する具体的なアクションを支援できること、④企業のリスク情報提供に対する前向きな意識を促進することについては、提案ツールの使用前と使用後を比較することで評価を実施できる構造とした。

ID	観点	質問項目	回答方法	選択肢				
1	理解性	STEP1 の目的や使い方について、どの程度理解できましたか？	5 件法	1:小さい	2	3	4	5:大きい
2	利用性	実務上で STEP1 を実行することは容易だと感じましたか？	5 件法	1:小さい	2	3	4	5:大きい
3	理解性	STEP2 の目的や使い方について、どの程度理解できましたか？	5 件法	1:小さい	2	3	4	5:大きい
4	利用性	指示に従って STEP2 を実行することは容易でしたか？	5 件法	1:小さい	2	3	4	5:大きい
5	理解性	STEP3 の目的や使い方について、どの程度理解できましたか？	5 件法	1:小さい	2	3	4	5:大きい
6	利用性	指示に従って STEP3 を実行することは容易でしたか？	5 件法	1:小さい	2	3	4	5:大きい
7	理解性	STEP4 の目的や使い方について、どの程度理解できましたか？	5 件法	1:小さい	2	3	4	5:大きい
8	利用性	指示に従って STEP4 を実行することは容易でしたか？	5 件法	1:小さい	2	3	4	5:大きい
9	理解性	STEP5 の目的や使い方について、どの程度理解できましたか？	5 件法	1:小さい	2	3	4	5:大きい
10	利用性	指示に従って STEP5 を実行することは容易でしたか？	5 件法	1:小さい	2	3	4	5:大きい
11	有効性	ワークを行う前と比べ、実務上でステークホルダーのリスク性向（リスクに対する考え方や前向きさ等）の把握のやり方ができている。	5 件法	1:小さい	2	3	4	5:大きい
12	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
13	有効性	ワークを行う前と比べ、リスク情報を整理するために何を調査すべきかが理解できている。	5 件法	1:小さい	2	3	4	5:大きい
14	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					

15	有効性	ワーク前と比べて、適切にリスク情報を提供するための客観的な状況の把握・整理ができるようになっている。	5 件法	1:小さい	2	3	4	5:大きい
16	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
17	有効性	ワークを行う前と比べ、リスク情報提供の検討に必要なプロセスを理解できている。	5 件法	1:小さい	2	3	4	5:大きい
18	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
19	有効性	ワーク前と比べて、リスク情報提供に向けて必要なアクションを実行できる（そう）か。	5 件法	1:小さい	2	3	4	5:大きい
20	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
21	有効性	ワークを行う前と比べて、リスク情報をステークホルダーに提供することの重要性を理解している。	5 件法	1:小さい	2	3	4	5:大きい
22	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
23	有効性	ワークを行う前と比べて、リスク情報をステークホルダーに提供することのモチベーションが上がっている。	5 件法	1:小さい	2	3	4	5:大きい
24	有効性	なぜ上記の選択肢を回答したのか理由を教えてください。	自由記述					
25	有効性	あなたの企業の現在の業務プロセスにおいて、このツールを実際に組み込むことは可能だと思いますか？	自由記述					
26	有効性	このツールを日常業務に適用する際に見込まれる主な障害は何だと思いますか？	自由記述					
27	有効性	このツールを使って解決できると考える現実の業務上の課題があれば教えてください。	自由記述					
28	有効性	このツールを使用することにより、あなたの企業にどのようなポジティブな影響が生じると予想しますか？	自由記述					

(4) 実験に使用したケーススタディ

本来は検証・妥当性確認において、実際の企業の実施する業務に係る製品やサービスに基づいて提案ツールを使用して検証・妥当性確認を実施することが望ましい状況ではあったが、実際の企業情報を利用すると非公開情報・機微情報にあたるため、検証においては、架空の AI サービスを想定し、検証・妥当性確認を実施した。

ケーススタディとして設計したテーマは、採用 AI とした。採用 AI とは、エントリーシート書類選考を判断する際の参考情報として使用される AI サービスである。採用 AI のシステムを開発する A 社と採用 AI を使用してサービスを提供する AI サービスプロバイダーとなる B 社間の信頼関係に焦点を当てた。検証・妥当性確認においては、被験者は AI システム開発主体である A 社の立場であるという想定の下でツールを実際に使用して検討を行ってもらった。

ケーススタディでは、A 社は AI システム開発企業であり、AI サービスプロバイダーとなる人材ポータルサイト運営企業の B 社から過去のエントリーシートデータと合否判定（内定の判定）結果を受領し、機械学習（分類モデル）で合否判定をする学習モデルを作成する、という内容とした。また、リスク状況を設定するために、AI システムに関する内容を松本敬史・江間有沙（2020）によるリスクチェーンモデルのケーススタディを参考に表 4-5 の通り設定した。

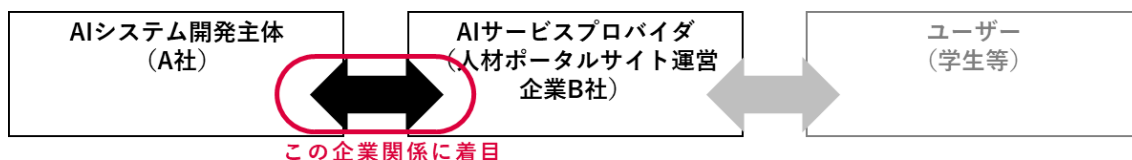


図 4-1 ケーススタディにおける着眼点

表 4-5 ケーススタディの AI システムの内容

適合率 (Precision)	書類選考で合格と判断した中で、面接を経て内定を出した割合で評価し、70%を期待値として設定。 ※蓄積された本番データは採用・不採用の判定結果が入力されたタイミングで適宜学習データとして追加され、日次で学習モデルを更新して本番環境に反映する。ただし、学習プロセスでクロスバリデーションによるテストの結果、正解率が 70%を下回る場合には本番環境の
---------------------------	--

	AI モデルを自動更新しない。AI モデルは過去 1 年分のバージョンを保存している。
使用方法	人材採用部門担当者は AI の判断を参考情報としながら、可否の判断を設定し、人材採用部門長の承認を得て、申込者に可否の通知を行う。
学習データ	過去のエントリーシート（テキスト）＋可否結果を使用。※個人情報あり
本番用データ	最新のエントリーシート（テキスト）を使用。※個人情報あり
AI サービスからの出力内容	可否の判定
出力方法（画面/アクチュエータ等）	人材採用部門担当者の端末上で、申込者のエントリーシートを投入すると、書類選考の可否判定が出力される。
根拠情報の出力	エントリーシートの中で判断に強い影響を与えたキーワードがハイライトされる。
安全性のリスク有無	無
外部 AI への連携	無
外部 AI への連携方法・プロトコル	無

さらに、上記の AI システムやステークホルダーの関係を基に、ケーススタディにおける AI リスク情報提供の状況の詳細も設定し、それを基に被験者にはツールで検討してもらうこととした。AI リスク情報提供の状況の詳細とは、整理された AI リスク一覧表の各リスクの内容と各リスクについて AI 開発企業から AI サービスプロバイダー企業にどのような説明があった状況か、その AI リスク項目に関連する AI システム構成要素を整理したものである。（表 4-6）これらを前提条件として、Excel 及び Miro ボード上の図表で被験者に事前に提供した。

表 4-6 ケーススタディにおける前提条件としての AI リスク一覧表（事前提供資料）

ID	AI リスク項目	どのようなリスクが存在するか	説明の実施（済 or 未）	関連する AI システム構成要素
1	意思決定における偏り・偏見	AI が過去のデータに基づいて意思決定を行うため、過去の採用プロセスにおける偏見が組み込まれるリスク	AI が過去のデータに基づいて判断するため、歴史的な偏見が組み込まれる可能性があり、適合率重視の評価指標が偏見を強化するリスクがあるため、人間の介入を運用プロセス上における工夫や、多様性を考慮したデータセットの作成を行う必要がある。	・提供サービス上の利用方法 ・学習データのバランス ・パラメータ・特徴量・重みづけ
2	説明責任への対応の不足・透明性の欠如	選定理由・基準が不明確である場合、採用の申込者に判断根拠を納得してもらえないリスクがある。	エントリーシートの中で判断に強い影響を与えたキーワードがハイライトされるが、その選定理由・基準が不明確である場合、採用の申込者に判断根拠を納得してもらえない可能性がある。	・説明可能性 ・パラメータ・特徴量・重みづけ ・フィードバック ・問合せ対応
3	情報セキュリティの脅威拡大	個人情報が含まれるデータの漏洩や不正アクセスが発生するリスクがある。	プライベートクラウド環境によるデータ管理において、個人情報が含まれるデータの漏洩や不正アクセスが発生した場合は、社会的な批判を強く受ける可能性がある。	・開発責任者・管理体制 ・実運用データ ・リスクマネジメント・監査体制

4	同意の仕組みの不足	採用の申込者が自分のデータがどのように使われるかを理解せずにサービスを利用するリスクがある。	採用の申込者のデータの利用において、明確な同意プロセスがないと、同意なく個人情報を含めた情報を使用したことで批判を受ける可能性がある。	・実運用データ ・問合せ対応
5	利用における不平等性	一律の評価指標による選考によって、特定のグループに対するサービスの不公平な提供となるリスクがある。	ー ※説明未実施	・ベネフィット・インパクト ・提供サービス上の利用方法 ・学習データのバランス ・パラメータ・特徴量・重みづけ
6	プライバシーの侵害	個人情報が含まれるエントリーシートの使用を行うため、個人情報の不正使用や漏洩のリスクがある。	政府の発行する個人情報保護規則等に準拠し、適切なデータ保護対策を講じない場合、個人情報の不適切な取り扱いだと見なされる可能性がある。	・開発責任者・管理体制 ・実運用データ ・事業者によるモニタリング・メンテナンス
7	技術・データの成熟度の低さによる影響	正解率 70%という期待精度であり、不完全な技術やデータによる誤判断のリスクがある。	ー ※説明未実施	・学習データの概要 ・学習手法 ・パフォーマンス精度・限界
8	品質保証とテスト基準の不備	不適切な品質保証やテスト基準により、予期せぬエラーが発生するリスクがある。	正解率 70%を下回った場合には AI モデルの自動更新を行わないなど、厳格な品質基準を設けること	・変更の適正手続 ・事業者によるモニタリング・メンテナンス

			で、AI システムの品質を担保する必要がある。	・ リスクマネジメント・監査体制
9	人間の介在の不足・過度な AI 依存	AI の判断に対する過度な依存により、人間の判断が排除されるリスクがある。	最終的な合否判断は人間が行い、AI は参考情報として位置付けないと、人間の判断が排除される可能性がある。	・ 利用目的 ・ ベネフィット・インパクト ・ 提供サービス上の利用方法
10	運用上の救済制度の不足	不適切な AI 判断に対する救済措置が不足している場合、採用の申込者に不満が生じるリスクがある。	不適切な判断があった場合のレビュープロセスを設け、救済措置を提供する等の運用に不備が存在すると、採用の申込者に不満が生じる可能性がある。	・ 提供サービス上の利用方法 ・ リスクマネジメント・監査体制 ・ 問合せ対応
11	個人の尊厳・権利の侵害	AI の判断により採用の申込者の尊厳や権利が侵害されるリスクがある。	ー ※説明未実施	・ 利用目的 ・ ベネフィット・インパクト ・ パラメータ・特徴量・重みづけ ・ リスクマネジメント・監査体制
12	過剰なデータ保有・独占	過去 1 年分のバージョンの保存が過剰なデータ保有によるプライバシー侵害につながるリスクがある。	データの最小化原則に従い、必要最低限のデータのみを保有することで、過剰なデータ保有ではない運用とする必要がある。	・ 開発責任者・管理体制 ・ 実運用データ

13	不適切な利用（悪意を含む）	個人情報の不正販売や、虚偽の内容に基づく採用の申込が行われるリスクがある。	システムへのアクセス管理と使用履歴の監視を行うことで、個人情報の不正販売への対策を行う。虚偽の内容に基づく採用の申込は、ES チェック後の面接の中で社員が直接確認することで、対策を行う必要がある。	・事業者によるモニタリング・メンテナンス ・リスクマネジメント・監査体制 ・問合せ対応
14	予期せぬ事態への対応の不足	AI システムの不測の事態に対応できないリスクがある。	緊急時のプロトコルを設定し、障害発生時の迅速な対応を確保することで、AI システムの不測の事態に対応できないリスクに対応できる運用体制・プロセスとする必要がある。	・開発責任者・管理体制 ・変更の適正手続き ・事業者によるモニタリング・メンテナンス ・リスクマネジメント・監査体制
15	社会・文化・環境への悪影響	（未検討）	－ ※説明未実施	・利用目的 ・ベネフィット・インパクト

4.4 検証の評価結果

4.4.1 ツールの理解性・利用性に関する検証結果

提案ツールについてケーススタディを基に使用した被験者 10 名に対して、提案ツールのステップ毎の理解性（理解の度合い）と利用性（利用の容易度合い）を事後アンケートで質問した。理解性とは、各ステップで何をすべきかを理解しているか、であり、利用性とは、その内容（指示内容）を実際に行えたか、である。

提案ツールのステップは、AI サービスプロバイダー企業の AI に対するリスク性向を把握するステップ（STEP1）、AI 開発企業から AI サービスプロバイダー企業にサービス・製品に係るリスクについてどのように説明されているかを一覧化するステップ（STEP2）と、社会的リスク喚起状況と顕在化事例をマッピングするステップ（STEP3）リスク知覚状況を検討するステップ（STEP4）、最後に、その整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善に向けたアクションを検討するステップ（STEP5）で構成される。その各ステップに関する被験者に対するアンケート結果は、表 4-7 の通りである。

被験者 10 名によるアンケート結果に基づいて、提案ツールの理解性と利用性について考察を行った。STEP1（AI に関するリスク性向の把握）は、理解性として理解できたと回答した被験者が 100%であり、また、利用性として 70%の被験者が容易だと回答しているため、STEP1（AI に関するリスク性向の把握）は被験者にとって比較的理解しやすく、使用しやすいことが示唆された。対照的に、STEP3（社会的リスク喚起状況と顕在化事例のマッピング）では、理解性は 90%と高いものの、利用性で困難だと感じる割合が 50%と高く、この部分が最も使いづらいと感じられていることが分かった。STEP3 は、ケーススタディとして設定した企業や AI サービス・製品に関連する社会的動向等を調査するステップであり、ケーススタディに関連する分野の知見を有すると考えられる属性の被験者が少なかったために、提案ツールの STEP3 の実施方法は理解できたものの、実際にタスクを実施することは難しいと受け止められたと考えられる。また、STEP2（企業由来のリスク情報提供状況の一覧化）と STEP4（リスク知覚状況の検討）は、いずれも高い理解性と利用性を示した。特に STEP2 では、全被験者が利用性を高く評価しており、このステップが最も効果的であることが示唆された。一方で、STEP5（情報提供のあり方の改善に向けたアクション検討）では、理解性が 70%、利用性に関しては 30%が困難だと回答しており、このステップにおける改善の余地が示唆された。STEP5 も STEP3 と同様に、ケーススタディとして設定した企業や AI サービス・製品に関連する検討を実施するステップであるために、前提知識等の不足から実際

にタスクを実施することは難しいと受け止められたのではないかと推察される。

表 4-7 提案ツールの理解性・利用性に関するアンケート結果

	STEP1 AI に対するリスク性向の把握		STEP2 企業由来のリスク情報提供状況の一覧化		STEP3 社会的リスク喚起状況と顕在化事例のマッピング		STEP4 リスク知覚状況の検討		STEP5 情報提供のあり方の改善に向けたアクション検討	
	理解性	利用性	理解性	利用性	理解性	利用性	理解性	利用性	理解性	利用性
全く理解できない/非常に困難だと思う	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
あまり理解できない/やや困難だと思う	0%	10%	0%	0%	0%	50%	0%	20%	0%	30%
どちらともいえない	0%	20%	10%	0%	10%	30%	10%	0%	30%	50%
やや理解できた/容易だと思う	40%	50%	10%	30%	50%	20%	40%	30%	60%	10%
非常に理解できた/容易だと思う	60%	20%	80%	70%	40%	0%	50%	50%	10%	10%

4.4.2 「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」に関する検証結果

リスク性向を把握するステップに対する要求である「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」について、前述の実験に際して実施した被験者アンケートに基づき評価を行った。

要求「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」については、事前アンケート及び事後アンケートの質問項目「実務上でステークホルダーのリスク性向（リスクに対する考え方や前向きさ等）の把握の方法を理解できている」の結果を提案ツールの使用前と使用後の比較分析することで考察を行った。事前アンケ

ート及び事後アンケートの質問項目「実務上でステークホルダーのリスク性向（リスクに対する考え方や前向きさ等）の把握の方法を理解ができている」の結果は表 4-8 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-9 に示す通りである。漸近有意確率（両側検定）は 0.034 となり、この値は有意水準（5%）を下回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在することが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化したことが確認されたため、提案ツールを使用することがステークホルダーのリスク性向の把握の方法の理解促進に寄与すると考えられる。

アンケートの自由記述から得られた重要な点を整理する。まず、提案ツールの STEP1（AI に対するリスク性向の把握）において、AI サービスプロバイダー企業に対する AI リスク性向の把握に関するアンケートの実施は、新しいアプローチであるが、実際に行うことに対しては実行の難しさや困難が複数指摘された。これは、実務上でのツールの導入や運用にあたっては、実行可能性や実用性をより一層考慮する必要があることを示唆している。また、提案ツールの STEP1（AI に対するリスク性向の把握）で提示したアンケート構成が普段考慮しない項目も網羅しているという点では、特に AI 開発企業の立場では、リスク性向の把握よりもそもそも AI 技術に対する知識、理解度合いを把握しようという取組みの傾向が強いため、リスク管理の網羅性や新たな視点を提供することで、より深いリスク理解に寄与する可能性があると考えられる。ツールの実用化に向けては、リスクについて考える担当者の明確な位置付けや、ツールの積極的な活用を推奨する組織的なサポートが必要であることも複数の指摘があった。これは、ツールの成功のためには単にツールの開発のみならず、組織文化や運用プロセスの改革も重要であることを示している。

表 4-8 質問項目「実務上でステークホルダーのリスク性向（リスクに対する考え方や前向きさ等）の把握の方法を理解ができている」のアンケート結果

実務上でステークホルダーのリスク性向の把握の方法を理解できている		
ID	事前	事後
01	1	3
02	2	2
03	2	3
04	3	3

05	2	3
06	2	3
07	2	3
08	3	3
09	3	3
10	4	4

表 4-9 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	15
標準誤差	3.536
標準化された検定統計量	2.121
漸近有意確率（両側検定）	0.034

4.4.3 「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」に関する検証結果

AI 開発企業から AI サービスプロバイダー企業にサービス・製品に係るリスクについてどのように説明されているかを一覧化するステップ（STEP2）と、社会的リスク喚起状況と顕在化事例をマッピングするステップ（STEP3）と、リスク知覚状況を検討するステップ（STEP4）に対する要求である「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」について、前述の実験に際して実施した被験者アンケートに基づき評価を行った。

要求「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」については、事前アンケート及び事後アンケートの質問項目「リスク情報を整理するために何を調査すべきかが理解できている」と「適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整理ができる」の結果を提案ツールの使用前と使用後の比較分析することで考察を行った。

事前アンケート及び事後アンケートの質問項目「リスク情報を整理するために何を調査すべきかが理解できている」の結果は表 4-10 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-11 に示す通りである。漸近有意確率（両側検定）は

0.005 となり、この値は有意水準（5%）を下回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在することが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化したことが確認されたため、提案ツールを使用することがリスク情報を整理するための調査方法や内容の理解促進に寄与すると考えられる。

アンケートの自由記述から得られた重要な点を整理する。自由記述では、提案ツールの AI リスク項目一覧表に業務上考慮されない項目を含むことで、リスク管理の網羅性が高まるという指摘がなされた。これは、様々なガイドライン等を集約することで作成された AI リスク項目一覧は、従来のリスクマネジメントのスコープが見落としがちな要素をカバーし、より全面的なリスク評価を可能にするというツールの強みを示していると考えられる。しかし、この網羅性は同時に、提案ツール使用者にとっての情報過多や複雑さを引き起こす可能性もある。また別の自由記述のコメントでは、STEP3（社会的リスク喚起状況と顕在化事例のマッピング）でのルール・ガイドの調査や、AI リスク項目一覧表の使用によって、明示的に情報を整理することができ、被験者がリスク情報をより具体的に、体系的に理解するのに貢献したことが指摘された。自由記述上で複数指摘があったのは、明示的なリスク情報を整理するための多様なアプローチの重要性であった。各象限ごとに異なるアクションが必要であること、また、AI モデルに関するリスク情報だけでなく、適用される業界のガイドラインについても調査する必要があること等の工夫が提案ツールに組み込まれており、ツールを使用することで多様な視点を有する情報源から情報を収集する方法の理解に繋がったと考えられる。また、AI システム開発事業者が当たり前と感じるリスクが、顧客にとっては未知のものである可能性があることを指摘する記述は、リスク情報を客観的に収集し、多角的な視点から評価することの重要性を指摘していると考えられる。これは、提案ツールがリスクマネジメントのプロセスにおいて、異なるステークホルダーの視点を統合する助けとなることが示唆された。

表 4-10 質問項目「リスク情報を整理するために何を調査すべきかが理解できている」の
アンケート結果

リスク情報を整理するために何を調査すべきかが理解できている		
ID	事前	事後
01	3	5
02	2	4
03	2	4
04	4	4
05	2	4
06	2	4
07	3	5
08	3	5
09	2	4
10	2	4

表 4-11 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	45
標準誤差	8.008
標準化された検定統計量	2.81
漸近有意確率（両側検定）	0.005

事前アンケート及び事後アンケートの質問項目「適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整理ができる」の結果は表 4-12 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後を比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-13 に示す通りである。漸近有意確率（両側検定）は 0.03 となり、この値は有意水準（5%）を下回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在することが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化したことが確認されたため、提案ツールを使用することが適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整

理の方法の改善に寄与すると考えられる。

アンケートの自由記述から得られた重要な点を整理する。提案ツールが普段考慮されない項目を含むことで網羅性が高まるとのコメントがある一方でインターネット上の情報収集に一貫したプロセスの妥当性に対する疑問が提起された。提案ツールによって収集される情報の信頼性と品質管理が重要であるため、情報源の選定と検証に係る課題や、現状の調査方法における情報の蓋然性と網羅性に関する不確実性が残ることが示唆された。また、4 象限での分類が情報整理に役立つとの意見は多数あったものの、それが質問項目で示す「客観的な状況の整理」に本当に貢献するかについては、複数の被験者によって曖昧さが指摘されており、提案ツールが提供する整理方法に対する妥当であることの理解のしやすさをさらに高める必要があることが示唆された。

表 4-12 質問項目「適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整理ができる」のアンケート結果

適切にステークホルダーに対してリスク情報を提供するための客観的な状況の把握・整理ができる		
ID	事前	事後
01	3	4
02	4	4
03	3	4
04	3	4
05	4	3
06	2	3
07	4	4
08	2	4
09	2	4
10	3	5

表 4-13 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	33
標準誤差	6.928
標準化された検定統計量	2.165
漸近有意確率（両側検定）	0.03

続いて、提案ツールを評価するため、10名の被験者にケーススタディを用いてワークを実施してもらった成果物と実際の業務を比較することで評価を実施した。ケーススタディでは、採用プロセスにおけるAIの使用を例に取り上げた。これは、M社（AIシステム開発企業）とN社（AIサービスプロバイダー企業）が共同開発したESを基に採用判定を行うプロジェクトと類似しており、この類似性から、ツールの使用による効果を適切に評価することができる考えた。

社会的リスク喚起状況と顕在化事例をマッピングするステップ(STEP3)において、図4-2のフレームを用いてAI開発企業の外でリスクの知覚状況について影響を与える事例を調査するプロセスが、「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」を実現する以下の機能のうち難易度の高い(b-2)の機能を担っている。

- ・ (b-1) リスク情報の共有の有無（対象リスクに関する情報がステークホルダー間で共有されているかどうか）を把握できること
- ・ (b-2) リスク情報の汎用度（対象リスクに関する情報（リスク顕在化による影響や対策等）が一般的な知識としては広まっているかどうか）を把握できること

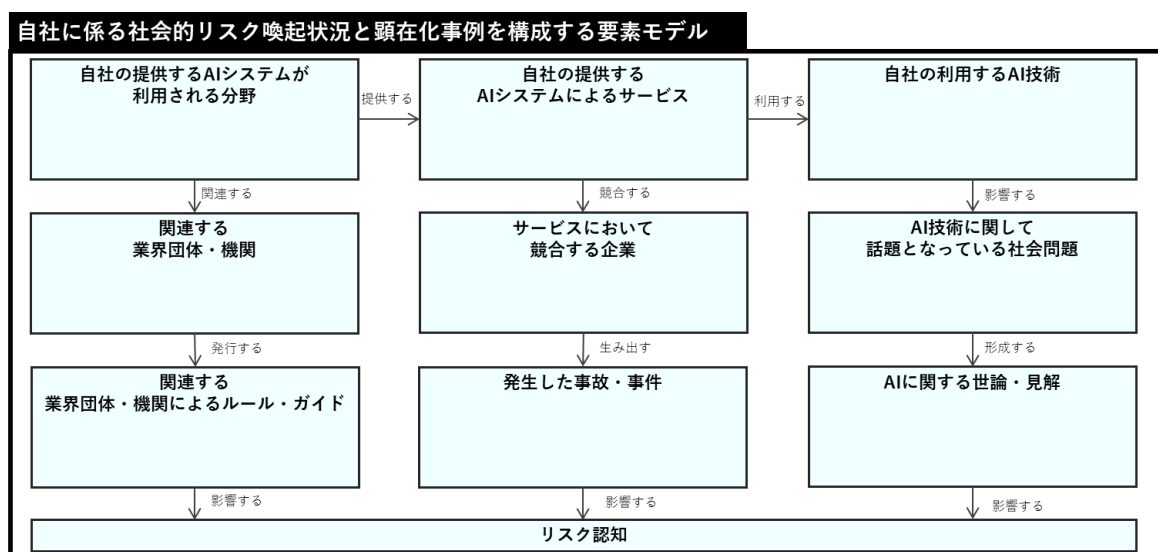


図 4-2 STEP3 における検討フレーム

そこで、M社（AIシステム開発企業）においてESを基に採用判定を行うプロジェクトに参加した中核メンバー2名に合同インタビューを実施することで、提案ツールを利用することが従来の方法を実施するよりも「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」のうち「明示的な情報で整理すること」部分について、貢献しているかどうかを確認した。確認の方法は、10名の被験者が実施したSTEP3

成果物に記載された内容が整理された一覧票（表 4-14）を M 社（AI システム開発企業）のメンバー2 名にインタビューを通して確認してもらったうえで、一覧表に記載されている内容について今まで知らなかった項目やプロジェクトでの検討内容に含まれなかった項目を特定することとした。インタビューを通し、一覧表に記載されている項目のうち、これまで知られていなかったものや、プロジェクトの検討内容に含まれていなかったものを指摘してもらい、提案ツールを使用することで新たな気づきや洞察がどの程度得られるかを評価した。

表 4-14 STEP3 の記載内容の一覧

分類	名称	出所	該当
関連する業界 団体・機関業 界団体・機関 によるルー ル・ガイド	人事データ利活用原則	ピープル・アナリティクス &HR テクノロジー協会	
	人事データ分析の現状と AI ガバナンス・ガイド ラインへの期待	経済産業省	
	匿名データの作成・提供に関するガイドライン	総務省	
	雇用条件を軸としたマッチ群機能の普及に向け た提言	一般社団法人 人材サービ ス産業協議会	
	プロフェッショナル人材活用ガイドブック 2023	内閣府	
	職業安定法	厚生労働省	
	人材協倫理綱領	一般社団法人日本人材紹介 事業協会	
	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	
	採用選考に関する指針（2018 年以降策定して いない）	内閣官房	
	機械学習品質マネジメントガイドライン	産業技術総合研究所（産総 研）	
	AI 利活用ガイドライン	総務省	
	Works156 号 特集 人事の AI 原則	リクルートワークス研究所	
	人事データ利活用原則（peopleanalytics.or.jp）	ピープル・アナリティクス &HR テクノロジー協会	
	元従業員による不正により個人情報漏洩	—	
発生した事 故・事件	サイバー攻撃による派遣の仮登録ユーザーの情 報流出	—	
	リクナビ問題（学生に無許可で辞退率算出）	—	
	「AI による採用面接は違法」人権団体による裁 判	—	1
	内定辞退確率の算出(行政指導)	—	
	AI 判定に対する忌避感の報道	—	
	女性差別※学習データの偏りにより、AI が男性 に高スコアを付与（Amazon）	—	
	人種・性別・経歴による差別	—	
	雇用の減少	—	
AI に対する世 論・見解	責任の所在の明確化の必要性	—	
	AI の透明性・公平性	—	
	定性的な情報の分析に対する懸念	—	
	場競争の不自由さに関する懸念	—	
	過剰にプライバシー情報を集めることについて 懸念	—	
	AI は 100%ではない	—	

M 社（AI システム開発企業）において ES を基に採用判定を行うプロジェクトに参加した中核メンバー2 名に対するインタビューの結果に基づき、表 4-15 に 10 名の被験者によって記載された内容中で何件について新たな気づきや洞察を得られたかを整理した。「関連する業界団体・機関業界団体・機関によるルール・ガイド」については、約 6 割について従来のプロジェクト遂行上では知らなかった、もしくは検討から漏れていたということが示された一方で、「発生した事故・事件」については約 1 割、「AI に対する世論・見解」については 0 という、提案ツールを使用したとしても従来のプロジェクト遂行時と変化が見られない結果となった。この結果から分かることとして、本提案ツールが「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」のうち、「明示的な情報で整理する」部分について従来の方法以上に貢献できる範囲は限定的であることが示された。「関連する業界団体・機関によるルールやガイドライン」については、プロジェクトの初期段階でこれらのルールやガイドラインを調査し参照することが一般的であるものの、AI 業界の急速な変化により新たなガイドラインやルールが頻繁に導入されるため、重要な更新を見落とすリスクが存在するため、提案されたツールの定期的な利用を行うことで、業界の変化や新たに出現したガイドラインとルールを特定するのに有効である可能性が示唆された。また、「発生した事故や事件」や「AI に関する世論や見解」については、これらの情報は主にニュースや SNS を通じて広く共有されるため、不明な情報が少ないことが観察された。しかし、単にこれらの事象を情報として受け止めるのではなく、それらが AI のどのリスク項目に関連しているかを分析し整理することは重要であると指摘があり、提案ツールを用いることが既知の内容を改めて体系的に理解することに貢献することが分かった。

表 4-15 インタビューに基づく STEP3 にて新たな気づきや洞察を得られた件数

分類	新たな気づきや洞察を得られた件数（割合）
関連する業界団体・機関業界団体・機関によるルール・ガイド	13 件中 8 件（61.5%）
発生した事故・事件	8 件中 1 件（12.5%）
AI に対する世論・見解	7 件中 0 件（0%）

4.4.4 「企業のリスク情報提供に対する具体的なアクションを支援できること」に関する検証結果

リスク知覚状況を検討するステップ（STEP4）と、最後に、その整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善に向けたアクションを検討するステップ

(STEP5) に対する要求である「企業のリスク情報提供に対する具体的なアクションを支援できること」について、前述の実験に際して実施した被験者アンケートに基づき評価を行った。

要求「企業のリスク情報提供に対する具体的なアクションを支援できること」については、事前アンケート及び事後アンケートの質問項目「(ステークホルダーに対する) リスク情報提供の検討に必要なプロセスを理解できている」と「(ステークホルダーに対する) リスク情報提供に向けて必要なアクションを実行できる」の結果を提案ツールの使用前と使用後の比較分析することで考察を行った。

事前アンケート及び事後アンケートの質問項目「(ステークホルダーに対する) リスク情報提供の検討に必要なプロセスを理解できている」の結果は表 4-16 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-17 に示す通りである。漸近有意確率(両側検定)は 0.013 となり、この値は有意水準(5%)を下回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在することが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化したことが確認されたため、提案ツールを使用することがリスク情報提供の検討に必要なプロセスの理解促進に寄与すると考えられる。

アンケートの自由記述から得られた重要な点を整理する。ケーススタディやツールの前提として、提案ツールの使用フェーズはサービスリリース直前という設定としていたが、サービスリリース直前に情報提供を行うという現行のプラクティスに対する懸念が示されており、この段階でリスク情報提供を一生懸命に図ったとしても顧客がその時点では既に受け入れられないリスク情報が含まれている可能性が指摘されており、ツールの活用においては、顧客の受容性を考慮したアプローチの改善が必要であることが示唆された。ツールの使用により、従来“なんとなく”で進めていたリスク情報提供が、より系統的な思考に基づくものとなり、全体の流れを把握しやすくなったとの意見や、ツールを使用することで網羅的かつ顧客の知識レベルを考慮しながらのリスク説明が可能となり、リスク情報提供のプロセス全体の理解度が向上したという意見があった。これは、ツールがリスク情報提供のプロセスを可視化し、理解を深めるのに寄与していることを示していると考えられる。具体的には、AI サービスがはらむリスクの全体感への理解、リスク情報を提供する際のスタンスの選択、客観的な情報の位置付け、サービス利用者が現状接しているリスク情報の棚卸しとそれを踏まえた対応検討など、これまでぼんやりとしか理解できていなかったプロセスが明確になったとのことであった。

表 4-16 質問項目「（ステークホルダーに対する）リスク情報提供の検討に必要なプロセスを理解できている」のアンケート結果

（ステークホルダーに対する）リスク情報提供の検討に必要なプロセスを理解できている		
ID	事前	事後
01	4	3
02	3	4
03	2	5
04	4	4
05	2	4
06	3	5
07	3	4
08	1	3
09	2	4
10	3	5

表 4-17 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	43
標準誤差	8.261
標準化された検定統計量	2.481
漸近有意確率（両側検定）	0.013

事前アンケート及び事後アンケートの質問項目「（ステークホルダーに対する）リスク情報提供に向けて必要なアクションを実行できる」の結果は表 4-18 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-19 に示す通りである。漸近有意確率（両側検定）は 0.008 となり、この値は有意水準（5%）を下回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在することが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化したことが確認されたため、提案ツールを使用することが

リスク情報提供の検討に必要なプロセスの理解促進に寄与すると考えられる。

アンケートの自由記述から得られた重要な点を整理する。プロジェクト発足時から顧客との合意形成を行わない場合、プロジェクトリスクが生じる可能性があることが指摘されており、これはリスク情報提供において初期段階からのステークホルダーとのコミュニケーションが重要であると考えられる。また、リスク情報提供に係るアクションの重要性を認識し、その内容が具体的に理解できたという意見もあり、提案ツールがリスク情報提供に必要なアクションを明確にするのに役立つことが示唆された。また、従来は「やれと言われたからやる」という姿勢であったが、ツールの使用により説得力のあるアクションを実行できるようになったとの意見もあり、ツールが理論的な背景と具体的なアクションの組み合わせにより、より効果的なリスク情報提供へと使用者の意識を変容する可能性があることが分かった。一方で、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善に向けたアクションを検討するステップ（STEP5）については、より具体的なアプローチの例示があると有益であるとの意見もあり、提案ツールにおいてアクションの選択肢やアイディアに関しては、さらなる具体性や例示があるとより効果的なツールとなると考えられる。また、リスク情報提供に関する日常的な意識が低いために、必要なアクションの設定に妥当性を判断できないという意見もあり、これはツールがユーザーの基礎知識や経験のレベルに応じてカスタマイズされる必要があることが示唆された。

表 4-18 質問項目「（ステークホルダーに対する）リスク情報提供に向けて必要なアクションを実行できる」のアンケート結果

（ステークホルダーに対する）リスク情報提供に向けて必要なアクションを実行できる		
ID	事前	事後
01	3	3
02	3	5
03	3	4
04	3	4
05	3	4
06	3	3
07	3	4
08	2	4
09	2	3
10	4	5

表 4-19 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	36
標準誤差	6.819
標準化された検定統計量	2.64
漸近有意確率（両側検定）	0.008

続いて、提案ツールを評価するため、10 名の被験者にケーススタディを用いてワークを実施してもらった成果物と実際の業務を比較することで評価を実施した。ケーススタディでは、採用プロセスにおける AI の使用を例に取り上げた。これは、M 社と N 社が共同開発した ES を基に採用判定を行うプロジェクトと類似しており、この類似性から、ツールの使用による効果を適切に評価することができると考えた。

そこで、M 社（AI システム開発企業）において ES を基に採用判定を行うプロジェクトに参加した中核メンバー2 名に合同インタビューを実施することで、提案ツールを利用することが従来の方法を実施するよりも「企業のリスク情報提供に対する具体的なアクションを支援できること」について、貢献しているかどうかを確認した。確認の方法は、10 名の被験者が実施した STEP5 成果物に記載された内容を図 4-3 に示す一覧表に整理し、インタビューを実施することで一覧表に記載されている項目のうち、プロジェクトの実施内容に含まれていなかった具体的なアクションを指摘してもらい、提案ツールを使用することで新たな気づきや洞察がどの程度得られるかを評価した。

ID	AIリスク項目	被験者ID:01	インタビュー	被験者ID:02	インタビュー	被験者ID:03	インタビュー	被験者ID:04	インタビュー	被験者
1	意思決定における偏り・偏見	学習データの偏りによるリスクを説明しつつ活用フェーズでの修正方針を提議する		学習データ・テストデータの多様性が確保されていることを確認する		説明書指図書は確認済みであるが、現実世界がそもそも不平等にできているため、多様性を考慮したデータセットの作成により十分に偏りが出ない可能性も併せて説明するべき。 ※現実の近いシステムはそもそも使われていないし、相手との信頼関係に多大な影響を与える。		偏りの可能性は対応策の説明について、ガイドライン踏まえ確認する。 データセットを既存のガイドライン踏まえ見直す。		偏り等 がある
2	説明責任への対応の不足・透明性の欠如	AI算出ロジック、や信頼性指標との関係について説明する		説明可能なAI（XAI）を導入し、結論に対する透明性を高める		説明の裏指図書に加えて、学習に使われているパラメータが社会通念上許容されることを説明するべき。		説明可能性や申込者とのコミュニケーションに関する説明について、ガイドライン踏まえ確認する。 データセットやアルゴリズムを既存のガイドライン踏まえ見直す。		報告 の可 能性 ある
3	情報セキュリティの脅威拡大	各利用環境ごとの情報漏洩リスクを提供し、対策について提議する		管理責任部門・責任者の所在、最近先の管理体制・方法を明確にし、正しく運用する		個人情報の取り扱いにはリスクを伴うことから、「匿名加工情報」によるデータの取り扱いを推奨し情報セキュリティに関するリスクを管理できる旨を最初に提議するべきであった。				
4	同意の仕組みの不足	他事例を紹介するAI開発制での同意範囲は明確にしておく		明確な同意プロセスの構築と、扱い合わせ窓口を用意する		A社はB社からデータもらう立場にあり、B社が当然考えるべきものであり、A社としての対応は不要 あえてこの場で言うならば、A社は「B社との協力的に繋がる適切にデータを扱います」という態度				
5	利用における不平等性	利用法の一つで生じる個人情報やセキュリティリスクについての情報を提供する		学習データ・テストデータの多様性が確保されていることを確認する		偏りを生んでいるパラメータについて説明をするべき。 例えば「出身大学」のようなパラメータを重視した場合、学習データにない大学出身者は不利となる可能性がある		既存のガイドラインはないものの社会から懸念が示されているため、自社として意思決定における偏りとあわせて説明・対策を検討する。		
6	プライバシーの侵害	個人情報保護等と照らしシステム側からリスク提供を実施するデータ学習の匿名化や特徴量の匿名化を検討する		アウトプットの品質に対するモニタリング・メンテナンス体制（権限・方法・体制）を構築する		個人情報の取り扱いにはリスクを伴うことから、「匿名加工情報」によるデータの取り扱いを推奨しプライバシーの侵害に関するリスクを管理できる旨を最初に提議するべきであった。				注

図 4-3 インタビューに使用した STEP5 成果物一覧表

M 社（AI システム開発企業）において ES を基に採用判定を行うプロジェクトに参加した中核メンバー2 名に対するインタビューの結果に基づき、表 4-20 に 10 名の被験者によって記載された内容に関して、AI リスク項目ごとリスク情報提供のあり方の改善に向けた具体的なアクションに係る新たな気づきや洞察をいくつか得られたかを整理した。表 4-20 がインタビューに基づく、STEP5 成果物に関して類似プロジェクト参加メンバーが新たな気づきや洞察を得られたと回答した件数の結果である。AI リスク項目 15 項目中、11 項目に関して少なくとも 1 件は新たな気づきや洞察を得られたということが分かった。「意思決定における偏り・偏見」や「説明責任への対応の不足・透明性の欠如」については、AI サービスプロバイダー企業に対する説明にとどまらず、社会に対して対外的にどのように説明ができるかという観点での記載が STEP5 成果物にはあり、その点が新たな気づきや洞察に繋がったという指摘があった。

表 4-20 インタビューに基づく STEP5 のインサイト件数

AI リスク項目	インサイト数（件）	割合（％）
意思決定における偏り・偏見	4	40%
説明責任への対応の不足・透明性の欠如	3	30%
情報セキュリティの脅威拡大	1	10%
同意の仕組みの不足	0	0%
利用における不平等性	1	10%
プライバシーの侵害	1	10%
技術・データの成熟度の低さによる影響	0	0%
品質保証とテスト基準の不備	1	10%
人間の介在の不足・過度な AI 依存	0	0%
運用上の救済制度の不足	1	10%
個人の尊厳・権利の侵害	2	20%
過剰なデータ保有・独占	0	0%
不適切な利用（悪意を含む）	2	20%
予期せぬ事態への対応の不足	2	20%
社会・文化・環境への悪影響	1	10%
全体	19	13%

4.4.5 「企業のリスク情報提供に対する前向きな意識を促進

すること」に関する検証結果

提案ツール全体に対する要求である「企業のリスク情報提供に対する前向きな意識を促進すること」について、前述の実験に際して実施した被験者アンケートに基づき評価を行った。

要求「企業のリスク情報提供に対する前向きな意識を促進すること」については、事前アンケート及び事後アンケートの質問項目「リスク情報をステークホルダーに提供することの重要性を理解している」と「リスク情報をステークホルダーに提供することのモチベーションがある」の結果を提案ツールの使用前と使用後の比較分析することで考察を行った。

事前アンケート及び事後アンケートの質問項目「リスク情報をステークホルダーに提供することの重要性を理解している」の結果は表 4-21 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-22 に示す通りである。漸近有意確率（両側検定）は 1 となり、この値は有意水準（5%）を上回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在しないことが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化しておらず、提案ツールを使用することはリスク情報をステークホルダーに提供することの重要性の理解促進に寄与しないと考えられる。

アンケートの自由記述から得られた重要な点を整理する。ワークを通じて、リスク情報を提供しない場合の影響を具体的に想像することができたとの意見があり、これはツールがリスクの現実的な影響を理解する一助となったと考えられるものの、重要性についての理解には従来からの更新がなかったという意見もあり、これは被験者が提案ツールを使用しなくとも従前からリスク情報提供の重要性を理解していたことがアンケートの定量的分析からも自由記述からの定性的分析からも示唆された。他方で、以前からリスク情報提供の重要性を認識していたが、提案ツールを使用しながら客観的な情報を整理する過程で、さまざまな角度からの情報収集を通じたコミュニケーションの必要性を新たに認識できたという指摘もあり、提案ツールがリスクコミュニケーションのモチベーションのさらなる強化に寄与していることが示された。また、4 象限のラベルの直感的な理解の難しさは、ツールのデザインやフレームワークの改善の余地があることについての指摘もあった。提案ツールはリスク情報提供の重要性に対する理解を深める助けとなる可能性があると考えられるが、その影響はすでにステークホルダーに対するリスク情報提供の重要性を理解している参加者には限定的であるとの示唆も得られており、リスク情報提供の重要性を理解している参加者に対する効果を強化するためには改善の余地が残ることが分かった。

表 4-21 質問項目「リスク情報をステークホルダーに提供することの重要性を理解している」のアンケート結果

リスク情報をステークホルダーに提供することの重要性を理解している		
ID	事前	事後
01	5	5
02	5	5
03	4	3
04	4	4
05	5	4
06	4	4
07	4	4
08	4	4
09	4	5
10	4	5

表 4-22 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	5
標準誤差	2.5
標準化された検定統計量	0
漸近有意確率（両側検定）	1

事前アンケート及び事後アンケートの質問項目「リスク情報をステークホルダーに提供することのモチベーションがある」の結果は表 4-16 の通りである。

また、被験者アンケートの結果は、事前事後それぞれ 10 件とデータ数が少ないため、母集団が正規分布していることの判別は困難であり、ノンパラメトリック検定を使用することとした。同一の被験者の回答について、提案ツールの使用前と使用後と比較するため対応のある 2 群間の比較となるため、Wilcoxon の符号付き順位検定にて分析を実施した。検定の結果は、表 4-17 に示す通りである。漸近有意確率（両側検定）は 0.054 となり、この値は有意水準（5%）を上回っているため、提案ツールの使用前と使用後の間に統計的に有意な差異が存在しないことが示唆された。提案ツールの使用前後で被験者のスコアが有意に変化しておらず、提案ツールを使用することはリスク情報をステークホルダーに提供することの重要性の理解促進に寄与しないと考えられる。

アンケートの自由記述から得られた重要な点を整理する。まず、リスク情報を伝え

ないと自社の信用を落とす可能性があることが分かったとの意見があり、提案ツールを使用することで企業の評判や信用に対する懸念の存在を把握でき、リスク情報提供のモチベーションに影響を与えていることが示唆された。一方で、ステークホルダーに対するリスク情報提供そのものへのモチベーションが元々存在していた被験者も多く、そのモチベーションについて大きな変化はないという意見も複数把握されており、これはツールの影響が既にリスクに対して意識が高い個人には限定的である可能性を示している。

表 4-23 質問項目「リスク情報をステークホルダーに提供することのモチベーションがある」のアンケート結果

リスク情報をステークホルダーに提供することのモチベーションがある		
ID	事前	事後
01	5	4
02	5	5
03	2	4
04	5	5
05	3	4
06	4	4
07	4	5
08	1	4
09	3	4
10	4	5

表 4-24 Wilcoxon の符号付き順位検定の結果

合計数	10
検定統計量	25
標準誤差	5.701
標準化された検定統計量	1.93
漸近有意確率（両側検定）	0.054

4.5 妥当性確認の評価結果

4.5.1 筆者作成の評価基準による評価結果

被験者に実施してもらったワークにおいて、整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善を実施するためにはどのようなアクションが必要かを検討する STEP5 の成果物が、AI システム開発企業、及び AI サービスプロバイダー企業間の信頼構築に向けたリスク情報提供の在り方の改善に貢献するものとなっているかについて筆者作成の評価基準に基づき、評価を実施した。

評価基準は、前述の通り、文部科学省「リスクコミュニケーション案内」を参考に設計を行い、表 4-25 に示す設計とした。

表 4-25 筆者作成の STEP5 成果物の評価基準

評価基準（大分類）	評価基準（中分類）	内容
(1) リスク対応としての評価	(1-1) リスク評価に繋がるアクションか？	リスクの特定、分析、評価を行うアクションに繋がる検討がなされているか
	(1-2) リスク対策に繋がるアクションか？	特定されたリスクに対する対策や解決策を提案するアクションに繋がる検討がなされているか
	(1-3) 現状の課題・ニーズ把握に繋がるアクションか？	現在の状況や必要性を理解し、それに基づいて行動するアクションに繋がる検討がなされているか
(2) コミュニケーションとしての評価	(2-1) 透明性が担保されているか？	プロセスや意思決定の透明性を確保するアクションに繋がる検討がなされているか
	(2-2) 双方向性があるか？	ステークホルダー間の双方向コミュニケーションを促進するアクションに繋がる検討がなされているか

表 4-25 に示した評価基準に基づき、すべての被験者の STEP5 の成果物について表 4-26 に示す評価表を作成して評価を実施した。STEP5 では各 AI リスク項目すべてに対してリスク情報提供のあり方の改善に向けたアクションを記載してもらったため、それらの記載内容が評価基準に当てはまるかどうかを確認した。当てはまる場合には、

評価表に1を入れることとした。

10人の被験者に関する評価結果は、表 4-27 の通りである。リスク対応としての評価の合計値（最大3）、コミュニケーションとしての評価の合計値（最大2）を被験者ごとに表にした。

表 4-26 各被験者の STEP5 の成果物に対する評価表（被験者 ID:01 の例）

ID	AI リスク項目	被験者 ID：01 の STEP5 記載内容	リスク対応としての評価			コミュニケーションとしての評価	
			リスク評価に 繋がるアクションか？	リスク対策に 繋がるアクションか？	現状の課題・ ニーズ把握に 繋がるアクションか？	透明性を担保 することに繋 がるアクションか？	双方向性を担 保することに 繋がるアクションか？
1	意思決定における偏り・偏見	学習データの偏りによるリスクを説明しつつ運用フェーズでの是正方針を議論する	1	1	1	1	1
2	説明責任への対応の不足・透明性の欠如	AI 算出ロジック、や客観的指標との関係について説明する	1	1		1	1
3	情報セキュリティの脅威拡大	利用環境ごとの情報漏洩リスクを提供し、対策について提案する	1	1	1	1	1
4	同意の仕組みの不足	他事例を紹介する AI 開発側での責任範囲は明確にしておく				1	
5	利用における不平等性	利便性の一方で生じる個人情報やセキュリティリスクについての情報を提供する	1	1	1	1	1
6	プライバシーの侵害	個人情報保護等と照らしシステム側からリスク提供を実施するデータ学習の匿名化や作業員の関与の範囲についてルールを決める	1	1	1	1	1

7	技術・データの成熟度の低さによる影響	学習データの内容によって職員負荷の増になるため入力データの標準化や作業の効率化に関する検討を実施する		1	1		
8	品質保証とテスト基準の不備	学習データの構築、AI の再学習が重要であることを理解してもらい予算の確保を依頼する			1		1
9	人間の介在の不足・過度な AI 依存	運用設計書で定めた意思決定の流れを定期的に見直す				1	
10	運用上の救済制度の不足	プロバイダー側の運用と AI 開発側の運用とそれぞれの救済制度に関するリスクを明確化しておく	1	1	1	1	1
11	個人の尊厳・権利の侵害	他事例を紹介する AI 開発側での責任範囲は明確にしておく				1	
12	過剰なデータ保有・独占	データ保有に関する承認の実施主体は AI プロバイダーであることを踏まえ、AI 開発側で賄えるリスクについて明確化しておく	1	1	1	1	1
13	不適切な利用（悪意を含む）	完全なセキュリティ対策は存在しないことを理解してもらい、適切な対応、事後ルールについて協議をしておく	1	1	1	1	1
14	予期せぬ事態への対応の不足	運用契約に基づき体制、監査、マネジメント方針を明確化しておく				1	
15	社会・文化・環境への悪影響	全社方針に基づく情報提供程度			1	1	

表 4-27 被験者の STEP5 成果物に対する評価結果

ID	AI リスク項目	被験者 ID:01		被験者 ID:02		被験者 ID:03		被験者 ID:04		被験者 ID:05		被験者 ID:06		被験者 ID:07		被験者 ID:08		被験者 ID:09		被験者 ID:10	
		リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション	リスク 対応	コミュ ニケー ション
1	意思決定における偏り・偏見	3	2	3	1	3	2	3	2	1	0	3	2	3	2	3	2	3	2	3	2
2	説明責任への対応の不足・透明性の欠如	2	2	0	2	3	2	0	2	3	2	0	2	3	2	0	2	3	2	3	2
3	情報セキュリティの脅威拡大	3	2	2	2	3	2	0	1	1	1	3	0	3	2	3	2	3	2	1	0
4	同意の仕組みの不足	0	1	0	2	0	1	0	1	0	2	0	2	3	2	2	1	3	2	2	2
5	利用における不平等性	3	2	3	1	3	2	3	2	1	2	1	2	3	2	2	1	3	2	3	2
6	プライバシーの侵害	3	2	3	2	3	2	0	1	1	1	2	1	3	2	3	2	3	2	2	2
7	技術・データの成熟度の低さによる影響	2	0	1	2	3	2	3	2	2	0	3	2	3	2	1	1	3	2	3	2
8	品質保証とテスト基準の不備	1	1	3	2	3	2	0	1	2	0	2	1	3	2	3	2	3	2	3	1
9	人間の介在の不足・過度な AI 依存	0	1	1	2	0	0	0	1	1	1	1	1	3	2	0	2	3	2	2	2
10	運用上の救済制度の不足	3	2	0	2	0	0	0	1	1	0	3	2	3	2	3	2	3	2	2	2
11	個人の尊厳・権利の侵害	0	1	3	2	3	2	3	2	1	2	2	2	3	2	1	1	3	2	2	2
12	過剰なデータ保有・独占	3	2	1	1	3	2	0	1	1	1	1	1	3	2	3	2	3	2	3	1
13	不適切な利用（悪意を含む）	3	2	3	2	3	2	0	1	1	1	2	1	3	2	3	2	3	2	1	2
14	予期せぬ事態への対応の不足	0	1	2	2	3	2	0	1	1	0	3	1	3	2	3	2	3	2	1	2
15	社会・文化・環境への悪影響	1	1	3	0	2	2	3	0	1	2	3	0	3	2	1	1	3	2	1	2
平均		1.80	1.47	1.87	1.67	2.33	1.67	1.00	1.27	1.20	1.00	1.93	1.33	3.00	2.00	2.07	1.67	3.00	2.00	2.13	1.73

筆者作成の評価基準に基づく 10 人の被験者の実施した STEP5 成果物に関する評価結果の分布を表 4-28 の通り整理を行った。リスク対応としての評価の観点からの 3 つの評価基準である「リスク評価に繋がるアクションか?」、「リスク対策に繋がるアクションか?」、「現状の課題・ニーズ把握に繋がるアクションか?」の 3 つすべてに当てはまる検討結果となっていたものが全体の 54%であり、2 つに当てはまるものが 12%、1 つに当てはまるものが 17.3%で、いずれにも当てはまらないものは 16.7%であった。8 割以上の検討結果がリスク対応の観点からの評価基準に少なくとも 1 つ当てはまっている結果となった。また、コミュニケーションとしての評価の観点からの 2 つの評価基準である「透明性を担保することに繋がるアクションか?」、「双方向性を担保することに繋がるアクションか?」の 2 つすべてに当てはまる検討結果となっていたものが全体の 66.7%であり、1 つに当てはまるものが 24.7%、いずれにも当てはまらないものは 8.7%であった。9 割以上の検討結果がコミュニケーションの観点からの評価基準に少なくとも 1 つ当てはまっている結果となった。リスク対応としての評価の観点からの 3 つの評価基準とコミュニケーションとしての評価の観点からの 2 つの評価基準に基づく評価（採点）結果を足し合わせることによる評価結果の分布を参照すると、1 つも当てはまらないものが全体の 1.3%のみであった。つまり、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善を実施するためにはどのようなアクションが必要かを検討する STEP5 の成果物について、98.7%が部分的にはリスクコミュニケーションの評価基準を満たしていることになり、提案ツールによってリスク情報提供のあり方の改善を支援できることが示唆された。

表 4-28 筆者作成の評価基準に基づく評価結果の分布

	点数	個数	割合
リスク対応	3	81	54.0%
	2	18	12.0%
	1	26	17.3%
	0	25	16.7%
コミュニケーション	2	100	66.7%
	1	37	24.7%
	0	13	8.7%
合計 (リスク対応+コミュニケーション)	5	72	48.0%
	4	15	10.0%
	3	18	12.0%
	2	25	16.7%
	1	18	12.0%
	0	2	1.3%

整理された AI サービスプロバイダー企業のリスク知覚状況のマップを基に、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方の改善を実施するためにはどのようなアクションが必要かを検討する STEP5 において、記述内容の具体性が被験者によって違いがあり、その具体性の違いがリスク情報提供のあり方の改善の出来具合に違いを生み出していると見受けられた。そこで、記述内容の具体性と改善の出来度合いに対する客観的評価を行うために、10 人の被験者の実施した STEP5 成果物の記入文字数と筆者作成の評価基準に基づく 10 人の被験者の実施した STEP5 成果物に関する評価結果に相関関係があるのかを確認した。

10 人の被験者の実施した STEP5 成果物の記入文字数と筆者作成の評価基準に基づく 10 人の被験者の実施した STEP5 成果物に関する評価結果（観点ごとの平均と全体の平均）の一覧は表 4-29 の通りである。STEP5 成果物に関する評価結果のリスク対応、コミュニケーション、その 2 つの観点の合計値、STEP5 成果物の文字数の相関関係について分析した結果を表 4-30 に示す。

相関分析の結果、リスク対応の評価と被験者の検討結果の文字数との間には高い相関（0.782）が見られた。これはリスク対応に対する深い理解や詳細な検討が、より長い検討結果を伴うことを示しており、リスク対応に関する検討の質が高いほど、その説明や分析もより具体的に詳細になる傾向があることを意味している。また、リスク対応とコミュニケーションの評価間には非常に強い相関（0.912）が見られた。一方で、

コミュニケーションの評価と文字数との相関は比較的低いことが分かった。これは、コミュニケーションに関する検討の質が検討内容や分析の具体性に必ずしも影響しないことが示唆された。リスク対応の観点とコミュニケーションの観点の評価結果の合計値と文字数との間には強い相関（.729）があることが分かった。これにより、リスクコミュニケーションの望ましい方向性に即した検討を行っているより高い評価を受けた被験者が、より長い説明や分析を行う傾向にあることが示された。

表 4-29 STEP5 成果物の記入文字数と評価結果の一覧

被験者 ID	該当数			文字数
	リスク対応	コミュニケーション	合計値	
01	1.800	1.467	3.267	36.867
02	1.867	1.667	3.533	40.600
03	2.333	1.667	4.000	75.267
04	1.000	1.267	2.267	37.267
05	1.200	1.000	2.200	48.467
06	1.933	1.333	3.267	58.400
07	3.000	2.000	5.000	83.867
08	2.067	1.667	3.733	39.067
09	3.000	2.000	5.000	94.600
10	2.133	1.733	3.867	37.733

表 4-30 STEP5 成果物に関する評価結果のリスク対応、コミュニケーション、その2つの観点の合計値、STEP5 成果物の文字数の相関関係

		コミュニケーション	リスク対応	合計	文字数
コミュニケーション	Pearson の相関係数	1			
	有意確率（両側）				
	度数	10			
リスク対応	Pearson の相関係数	.912**	1		
	有意確率（両側）	0.000			
	度数	10	10		
合計	Pearson の相関係数	.959**	.991**	1	
	有意確率（両側）	0.000	0.000		
	度数	10	10	10	
文字数	Pearson の相関係数	0.574	.782**	.729*	1
	有意確率（両側）	0.083	0.007	0.017	
	度数	10	10	10	10

* 相関係数は 5% 水準で有意（両側）

** 相関係数は 1% 水準で有意（両側）

4.5.2 リスクコミュニケーション専門家による評価結果

本研究の提案手法について、および本研究の評価結果についての総合的な評価を客観的な視点から行なってもらうことを目的として、専門家評価を実施した。専門家評価の評価項目は下記に示す通りである。

- ・ 被験者それぞれの成果物について企業間の信頼関係構築に向けてリスク情報提供の在り方の改善に向けて適切なものになっているか（リスク対応の観点、コミュニケーションの観点それぞれから）

専門家インタビューの対象者としては、(株)三菱総合研究所で合意形成やリスクコミュニケーションの領域で業務を約 10 年間実施した経験を持つ白井浩介氏を選定した。10 名の被験者が実施した STEP5 成果物を白井氏へ開示し、リスク対応の観点、コミュニケーションの観点それぞれから以下の水準を考慮した 3 段階（○・△・×）で評価してもらったうえで、その評価とした根拠について説明の記述を依頼した。また、全体として、提案手法を使用することでリスク情報提供の在り方が進んでいるといえるかについてもリスク対応の観点、コミュニケーションの観点それぞれから回答を得た。

○：企業従業員の平均的な検討を大きく上回り、理想的な水準

△：企業従業員の平均的な検討を上回る水準

×：企業従業員の平均的な検討と同等の水準

専門家によるリスク対応の観点からの評価結果を表 4-31 に整理した。①自社が現状想定している範囲までか、他事例や将来を含めた範囲か、プロバイダー企業の懸念を含めたか、②対応についても、現状想定している対応のままか、さらなる課題設定とその対処を志向しているか、という二つの観点で、検討するリスクの範囲に差が表れたことにより評価が分かれた。また、専門家によるコミュニケーションの観点からの評価結果を表 4-32 表 4-31 に整理した。情報提供の範囲と内容に差から評価結果に差が生まれた。範囲は、プロバイダー企業への一方向の説明にとどまるか、対外的な公開も含まれるかどうかといった観点であり、内容については現状想定リスクとその対応のみか、プロバイダー企業等の懸念を聴取したうえでのものとするか、という観点である。表 4-33 に示す通り、全体の結果として、リスク対応の観点についてもコミュニケーションの観点についても半数以上の被験者が「○」の評価を得ており、提案ツールを用いることでリスクコミュニケーションとして適切な改善が促進されることが示唆された。

表 4-31 専門家によるリスク対応の観点からの評価結果

	リスク対応の観点からの評価結果
被験者 ID:01	○ 「1.是正方針の協議」や「9.定期的な見直し」によって、プロバイダー企業の持つ懸念・課題に対する適切な対応ができる可能性が高い。
被験者 ID:02	△ 具体的な課題設定がなされているものの、データの確認にとどまるものが多く、改善への志向性が低く感じられる。
被験者 ID:03	○ 技術的な範囲で適切な課題設定と改善可能性（5,7,8）の提案がされている。
被験者 ID:04	△ ガイドラインを踏まえた対応が主であり、課題設定、リスク評価・対策の充分性に疑問が残る。
被験者 ID:05	△ 技術的にはあくまでデータの検討、制度の改善にとどまり、その他の課題設定が不足している印象。
被験者 ID:06	○ 過去の事故やガイドラインを参考と位置づけ、実態に即した課題設定と対応を模索しており、より適切なリスク検討が行われるものと考えられる。
被験者 ID:07	○ 事例の収集や洗い出しといった現段階での想定を超えたリスクについても対応する意思があり、適切なリスク対策が行われる可能性が高い。
被験者 ID:08	△ マニュアルの提示は、あくまで自社内でのリスク認識とその対応にとどまってしまう可能性があり、プロバイダー企業の持つ懸念とそれに即した課題設定と対処がなされるか疑問が残る。
被験者 ID:09	○ プロバイダー企業の不安・懸念が想定されており、ニーズに沿ったリスク対応がなされると思われる。
被験者 ID:10	△ リスク対応についてやや具体性に乏しく、リスク評価、対策が十分か、疑問が残る。

表 4-32 専門家によるコミュニケーションの観点からの評価結果

	コミュニケーションの観点からの評価結果
被験者 ID:01	○ 説明・情報提供が主であるものの、協議や提案、依頼といった協同的な活動が信頼構築に寄与すると思われる。
被験者 ID:02	○ 公開・公表といった対応で一定の透明性が担保されている。また、問い合わせ窓口の設置は、双方向的な対応に資する。透明性、双方向性の観点で信頼構築にもつながると思われる。
被験者 ID:03	△ 説明責任を果たすという点にとどまっている印象。透明性の担保には寄与するが、よりプロバイダー企業、ひいては社会の懸念を収集する仕組みの提案が必要ではないか。
被験者 ID:04	× ガイドラインを踏まえた対応が主であり、その充分性に疑問が残る。補完、説明の方法を工夫しなければプロバイダー企業の懸念に寄り添った誠実な対応とは捉えられず、信頼関係の構築には至れない可能性がある。
被験者 ID:05	× 「5.とりあえず多様性っぽく見える評価指標を設定」、「11.対外的にはサポートの位置づけ」といった対応は不誠実。リスクを大きく見積もっているプロバイダー企業からの信頼にはつながらないのではないか。
被験者 ID:06	○ 基本的にプロバイダー企業との協議・議論による対応としており、懸念の把握のみならず、信頼関係の構築、活用意識の向上につながると考えられる。
被験者 ID:07	○ プロバイダー企業間にとどまらず、対外的な説明を想定しており、透明性の担保を志向していることが読み取れる。
被験者 ID:08	△ あくまでプロバイダー企業への一方向的な提示、説明にとどまる。納得感および信頼感が醸成できるか疑問が残る。
被験者 ID:09	○ プロバイダー企業の不安・懸念が想定されているうえ、ヒアリングでの聴取も含まれており、信頼関係の構築に大きく寄与すると思われる。
被験者 ID:10	△ 強調と説明といった説得的なコミュニケーションを意図していることが読み取れる。そのようなコミュニケーションは必ずしも安心、信頼にはつながらない。

表 4-33 評価結果一覧

観点	評価	件数（割合）
リスク対応の観点	○	5（50%）
	△	5（50%）
	×	0
コミュニケーションの観点	○	5（50%）
	△	3（30%）
	×	2（20%）

4.6 考察

本節では、本研究において提案したツールの検証実験結果を基に、研究の全プロセスを通じて得られた学びを踏まえて考察を行う。

4.6.1 検証結果の考察

本研究では、AI 技術に対する社会的な信頼を高めるために、適切なリスクマネジメントを図ることの主体となる AI システム開発企業、及び AI サービスプロバイダー企業間の信頼構築を本研究の上位目的とし、企業による情報提供のあり方及び社外環境の変化、ユーザーのリスク性向を踏まえ、リスク情報伝達に係る戦略の検討支援ツールを提案した。検証結果を表 4-34 に示す通り、提案ツールを用いることは要求である「リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと」、「明示的な情報で整理することでリスク知覚状況を客観的に整理できること」、「企業のリスク情報提供に対する具体的なアクションを支援できること」を支援することが示唆された。一方で、「企業のリスク情報提供に対する前向きな意識を促進すること」に対する提案ツールの効果は限定的であることが示唆された。

表 4-34 検証結果一覧

評価内容	検証対象	検証方法	検証事項	検証結果
検証	リスクに対する考え方の把握に向けて企業の実務上で利用できるか	アンケート (主観的評価)	筆者作成の質問項目	○
	明示的な情報で整理することでリスク知覚状況を客観的に整理できるか	アンケート (主観的評価)	筆者作成の質問項目	○
		成果物（客観的評価）	STEP3 の記入内容と実際の類似プロジェクトの比較	○
	企業のリスク情報提供に対する具体的なアクションを支援できるか	アンケート (主観的評価)	筆者作成の質問項目	○
		成果物（客観的評価）	STEP5 の記入内容と実際の類似プロジェクトの比較	○
	企業のリスク情報提供に対する前向きな意識を促進するか	アンケート (主観的評価)	筆者作成の質問項目	×

4.6.2 妥当性確認結果の考察

本研究における上位目的である「AI システム開発企業、及び AI サービスプロバイダー企業間の信頼構築に向けてリスク情報提供のあり方を改善できること」に対して提案ツールが貢献するかについての妥当性確認を、STEP5 成果物に対する筆者作成の評価基準に基づく評価、及び、リスクコミュニケーション実践的専門家による評価を実施した。表 4-35 に示す通り、双方の評価結果に基づき有効性が示された。

表 4-35 妥当性確認の一覧

評価内容	検証対象	検証方法	検証事項	検証結果
妥当性確認	リスク情報提供のあり方を改善できるか	成果物（客観的評価）	筆者作成の評価項目	○
		リスクコミュニケーション実践的専門家インタビュー（客観的評価）	筆者作成の質問項目及びオンラインインタビュー内容	○

本研究の妥当性確認の過程で、特に優れた成果物を提出した被験者 2 名が、他の被験者よりも高い評価を受けたことが明らかになった。この顕著な成果の背後にあるプロセスと動機を深く理解するため、この 2 名（被験者 ID07 と ID09）の被験者に対して追加のインタビューを実施し、その内容を分析した。この追加インタビューにより、その成果物が出現するまでの詳細な背景と、それに影響を与えた要因についての考察を行うことができた。優れた成果を示した被験者に対して実施した追加インタビューにおいて、特に以下の 3 つの主要なテーマに焦点を当てた。第一に、STEP5 の検討過程で、被験者が以前から持っていた知識や経験がどのように役立ったかについて質問した。第二に、被験者の意識の高さやモチベーションの源泉を質問することで、その成果に対する動機付けが明らかになった。最後に、STEP5 の検討に際し、提案されたツールのどの部分が特に有用であったと被験者が感じたかについても質問した。

表 4-36 被験者 ID:07 に対するインタビュー結果

属性による知見

- ・ 言語モデルを用いた AI の倫理的リスクを検討するプロジェクトのプロジェクトマネージャーとして経験があり、AI プロダクトを開発している企画側に話して注意を促すような取組みを行っており、その経験を通じて AI に関するリスクの知見を有していた。
- ・ trustworthy AI、GDPR 関連のプライバシー保護技術に関するプロジェクトにも所属した経験があり、データの取扱いに関する知見も有している。これらの経験は、今回の研究における検討過程で役立ったのではないかな。

属性による意識・モチベーション

- ・ 上記の経験を通じて、AI リスク管理の重要性を認識しており、この分野での取り組みに高いモチベーションを有している。リスクマネジメントは開発の速度やモデルの精度に影響を与えるため、所属していた企業でもあまり社員が前向きになりにくい分野であると感じていたが、自分は課題を強く感じていた。そのため個人的にも勉強を実施していた。
- ・ 社内ではリスクに対する意識を高めるためにエンジニアの関心の AI モデルの評価リストに倫理的リスクを組み込むことで、リスクをより理解しやすくするアプローチを採用する等の工夫を図っていた。

ツールのプロセスで役立った部分

- ・ STEP3 の調査が特に有用であった。STEP3 により、ステークホルダーごとのリスクを理解し、影響の範囲を考慮することができた。一方で、STEP3 では特定のルールやガイドについての知見が少なく調査が困難であった。一人でやるよりチームでやるほうが多様な知見を活用でき、より効果を発揮できるのではないかな。

表 4-37 被験者 ID:09 に対するインタビュー結果

属性による知見 ・ 以前、生体認証システムの開発に関わる部署におり、個人情報（顔、指紋など）の取扱いに関して厳格な管理とリスクの考慮が必要であったため、データの扱いにおけるリスクを理解する上で役立った可能性がある。 ・ 今年度からサービス開発部署に移り、マーケティングやブランディング研修を受けたことがあり、顧客のニーズを理解し、負の感情を正のものに変換する方法について学んだため、今回のワークの AI サービスプロバイダー企業のリスクに対する心理を理解し、その不安や懸念の解決がどのように価値創出に繋がるかを考えることができたのではないか。 属性によるモチベーション ・ 後々問題になりやすい点に気づく傾向が強く、さらに普段から問題になりやすい点を見つける努力もしているため、リスクの検討を行うことに向いていたのかもしれない。 ツールのプロセスで役立った部分 ・ STEP3 を実施することで体系的に整理された情報を基にして、様々な角度からバランスよく情報を収集することができた。特に、ルールやガイドを調査することには苦勞したが、情報を偏らせることなく総合的に調査することができた。STEP5 成果物の作成時には、STEP3 で調べた情報を参照して書くことができ、これが非常に参考になった。 ・ STEP4 は、顧客の状況を想像するのに役立った。何もない状態から顧客の状況を想像するのは難しいが、STEP4 によって顧客の認知状況が可視化されるため状態を整理して理解するのに役立った。

まず、両被験者ともに、それぞれの職務経験や学んだ分野から重要な知見を得ておりそれが STEP5 成果物の良い評価に繋がったと示唆された。一人目の被験者は、AI の倫理的リスクに関するプロジェクトでプロジェクトマネージャー（PM）としての経験を持ち、AI 関連のリスク管理に関する深い理解を有していた。また、データプライバシーに関する知見も豊富で、これらの経験が研究の過程で役立ったと考えられる。二人目の被験者は、生体認証システムの開発に携わり、個人情報の取扱いに関する厳格な管理した経験からリスクの知見を得たと考えられる。また、マーケティングやブランディング研修を受け、顧客のニーズや心理を理解する能力を養っていたことが、AI サービスプロバイダー企業のリスクに対する洞察を深める上で役立った可能性がある。次に、意識とモチベーションに関しては、一人目の被験者は AI リスク管理の重要性を強く認識しており、通常企業内では取り組まれにくいこの分野において、積極的に学び、取り組む

姿勢を示していた。二人目の被験者も、普段から問題になりやすい点を見つける努力をしており、リスクの検討に向いていたと言える。これらのことから職務や学びの経験のみならず個人の資質、モチベーションが彼らのワーク成果に直接的な影響を与えたと考えられる。最後に、ツールのプロセスにおいては、両者ともに STEP3 の調査が特に有用であったと感じていた。STEP3 により、多角的な情報をバランス良く収集し、ステークホルダーごとのリスクを理解することが可能となっていた。また、STEP4 の使用によって、顧客の状況をより明確に想像し、理解するのに役立ったことが分かった。

4.6.3 本提案ツールの限界

AI を搭載したサービスや製品の開発を実施する企業に所属し、AI サービス・製品の開発になんらかの立場で関与する従業員が、AI サービスプロバイダー企業のリスク知覚状況を把握し、リスク情報提供戦略を検討するためのツールを提案した。このツールの使用が、①リスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと、②明示的な情報で整理することでリスク知覚状況を客観的に整理できること、③企業のリスク情報提供に対する具体的なアクションを支援できることに対して貢献することが示唆された。一方で、企業のリスク情報提供に対する前向きな意識を醸成することには有効性が出せず、限定的な効果であり、別の方策と組み合わせることが必要であることが分かった。また、企業間の信頼関係維持・向上に向けてリスク情報提供のあり方の改善について有効であることも示唆された。

他方で、被験者インタビュー結果から把握できた通り、ツールの使用者が適切な成果を生み出すためには元々の素養やリスクに関わる知見がある程度必要であることも示唆された。そのため、本提案ツールを最大限に有効活用するためにはリスク教育も併せての実行が必要である。また、本研究では提案ツールの検証についてケーススタディを設定し、それに伴い考慮すべきリスク項目一覧表や対策の状況を筆者が定義した。しかし、リスクマネジメントにおいては、リスク特定のプロセスから始めることで、組織や企業特有のリスクをより明確にすることができる。したがって、提案ツールは、リスク特定及び対応に関するある程度の検討が進んでいる組織において、本研究で検証された有効性を発揮できると考えられる。これらの点を踏まえ、図 4-4 本論文における研究の境界を ISO31000 リスクマネジメントの枠組みにおいて示した。

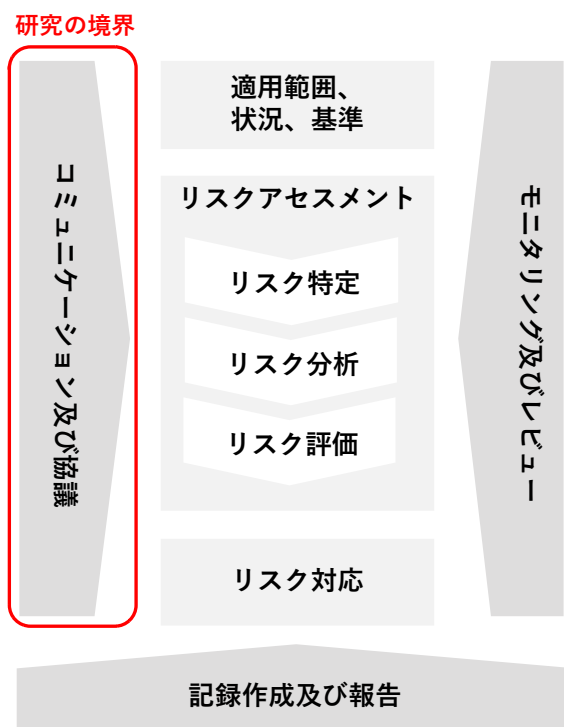


図 4-4 ISO31000 リスクマネジメントに基づく研究の境界の整理

第5章 結論と今後の展望

5.1 結論

本研究では、AIを使用した製品・サービスに対して適切なリスクマネジメントを図ることの主体となるAIシステム開発企業、及びAIサービスプロバイダー企業間の信頼関係の維持・向上を目的に、AIプロバイダー企業のAIに対するリスク性向を把握したうえで、リスク情報がAI開発企業からどのように伝達されているか、及び、社会的にどのような周知がなされているかを整理し、AIサービスプロバイダー企業のリスク知覚状況を検討し、その整理結果に基づき、リスク情報提供のあり方の改善に向けたアクションを検討するためのステップ及びガイドで構成されたツールを提案した。

評価の結果として、提案ツールが設計通りに適切に理解・活用されることで正しく機能し、ツールの使用前と使用後の比較に基づき、第一にリスクに対する考え方の把握に向けて企業の実務上で利用しやすいこと、第二に明示的な情報で整理することでリスク知覚状況を客観的に整理できること、第三に企業のリスク情報提供に対する具体的なアクションを支援できること、に有効であることが確認された。一方で、一方で、企業のリスク情報提供に対する前向きな意識を醸成することには有効性が出せず、限定的な効果であり、別の方策と組み合わせることが必要であることが明らかとなった。

妥当性確認において、筆者作成の評価基準に基づく評価予備リスクコミュニケーション実践的専門家による評価に基づき、AIシステム開発企業、及びAIサービスプロバイダー企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方を適切に改善できることについて有効性が示された。

5.2 今後の展望

本研究の今後の展望を3つ述べる。

第一に、提案ツール使用後の被験者に対するインタビューで得られた意見から、提案ツールにおけるSTEP3(社会的リスク喚起状況と顕在化事例のマッピング)をSTEP5(情報提供のあり方の改善に向けたアクション検討)の検討に際してより活用しやすくする方法について工夫の余地があることが示唆された。具体的には、STEP3(社会的リスク喚起状況と顕在化事例のマッピング)で整理した調査結果の内容からSTEP5(情報提供のあり方の改善に向けたアクション検討)を検討する際に参考となる要素を抽出するステップを設ける必要があると考える。そのようなステップを実装させることで、企業間の信頼関係の維持・向上に向けてリスク情報提供のあり方を適切に改善できることに関してツールによる効果に変化が現れるか検証することが求められる。

第二に、本研究の検証は採用 AI に関するケーススタディを提示することで実施したが、企業での活用を検討するためには本来的には被験者の実際に関与している製品・サービスに関して検討を行ってもらう必要があったと認識している。実際の製品・サービスに基づくケースでの検証を実施することでツールの理解性や利用性に係る課題をより明確に特定し、企業にとってより活用しやすい設計に改善することが可能になると考える。

最後に、4.6.3 で示した本研究の限界の通り、リスクマネジメントプロセスにおけるリスク特定のプロセスは本研究の対象外としたが、リスク特定の別のツールを組み合わせることで対象企業により適した検討が可能となる。そのため、提案ツールと組み合わせるべきツールの特定と適切な組み合わせ方の検討が必要である。

謝辞

指導教員であり本修士論文の主査である慶應義塾大学大学院システムデザイン・マネジメント研究科（SDM）の白坂成功教授には、IPA デジタルアーキテクチャーデザインセンターでの出会いをきっかけに SDM に入学し、研究を進めるにあたり多大な指導を賜りました。白坂先生の下での学びは、この 2 年間にわたり非常に有意義なものでした。

また、副査を務めてくださった神武先生にも、大変お世話になりました。貴重なご意見とアドバイスをいただき、研究を深めることができました。

特任助教の大野嘉子先生には主メンターとして約 1 年間、熱心なご指導を賜りました。また、副メンターとして松浦先生からも多くのアドバイスをいただきました。お二人の先生方が忙しい中で時間を割いてくださったことに、心から感謝しております。

本研究の検証にご協力いただいた 10 名の方々、特に白坂研究室 M1 の山上さん、芳賀さんには、プロトタイプへのご協力や被験者募集にもご協力いただき、深く感謝しております。

最後に、修士論文執筆にあたりご協力いただいた全ての方々に、心より感謝と御礼を申し上げます。皆様の支援があったからこそ、この研究を成し遂げることができました。

参考文献

- 落合浩一郎ら,2019「システムオブシステムズの現状と課題」 ソフトウェア・シンポジウム 2019
- Centre for Data Ethics and Innovation,2020「AI Barometer Report」
- 経済産業省,2019「アジャイル・ガバナンスに向けて」
- 松本敬史・江間有沙,2020「AI サービスのリスク低減を検討するリスクチェーンモデルの提案」
- National Institute of Standards and Technology,2022「Artificial Intelligence Risk Management Framework (AI RMF 1.0)」
- National Institute of Standards and Technology,2022「AI RMF PLAYBOOK」
- Kai Jia & Nan Zhang ,2022「Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines」 Electronic Markets (2021) 32:59–71
- Roy J. Lewicki and Chad Brinsfield,2016「Trust Repair」 Annual Review of Organizational Psychology and Organizational Behavior Vol. 4:287-313 (Volume publication date March 2017)
- Banja,2020「How Might Artificial Intelligence Applications Impact Risk Management?」 AMA journal of ethics
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1)
- Liu, H. W., Lin, C. F., & Chen, Y. J. (2019). Beyond State v Loomis: artificial intelligence, government algorithmization and accountability. International Journal of Law and Information Technology, 27(2), 122–141
- Nelson, G. S. (2019). Bias in Artificial Intelligence. North Carolina Medical Journal, 80(4), 220–222.
- Roger C. Mayer, James H. Davis and F. David Schoorman(1995)「An Integrative Model of Organizational Trust」 The Academy of Management Review , Jul., 1995, Vol. 20, No. 3 (Jul., 1995), pp.709-734
- T. K. Das, Bing-Sheng Teng(2004)「THE RISK-BASED VIEW OF TRUST:A CONCEPTUAL FRAMEWORK」 Journal of Business and Psychology, Vol. 19, No. 1, Fall 2004
- Lewicki RJ, Bunker BB.(1996)「 Developing and maintaining trust in work relationships」 See Kramer & Tyler 1996, pp. 114–39
- Ariño, A., & de la Torre, J. (1998)「Learning from failure: toward an evolutionary model of collaborative ventures」 Organization Science, 9(3): 306-325.

- Robert B Cialdini (2014) 「The Hidden Cost of Organizational Dishonesty」 MIT Sloan Management Review · March 2004
- D. Powell (2005) 「An Introduction to Risk Communication and the Perception of Risk」 Ortwin Renn (1990) 「Risk communication: Towards a rational discourse with the public」 Journal of Hazardous Materials
- Volume 29, Issue 3, February 1992, Pages 465-519
- 木下富雄 (2008) 「リスク・コミュニケーション再考—統合的リスク・コミュニケーションの構築に向けて (1)—」 日本リスク研究学会誌 18 (2) : 3 – 22 (2008)
- Lewenstein, B. V. (2003). 「Models of public communication of science and technology」 Public Understanding of Science, 16(3), 259-274.
- Bucchi, M. (2008). 「Of deficits, deviations and dialogues: Theories of public communication of science.」 In M. Bucchi & B. Trench (Eds.), Handbook of Public Communication of Science and Technology (pp. 57-76). London: Routledge.
- Nisbet, M. C., & Scheufele, D. A. (2009). 「What's next for science communication? Promising directions and lingering distractions.」 American Journal of Botany, 96(10), 1767-1778.
- Simone Grassini (2023) 「Development and validation of the AI attitude scale (AIAS-4): a brief measure of general attitude toward artificial intelligence」 Front. Psychol., 24 July 2023
- Sec. Media Psychology Volume 14 – 202
- Schepman, A., and Rodway, P. (2020) 「Initial validation of the general attitudes towards Artificial Intelligence Scale」 Comp. Hum. Behav. Rep. 1:100014.
- Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., et al. (2021) 「Assessing the attitude towards artificial intelligence: introduction of a short measure in German, Chinese, and English language」 KI-Künstliche Intelligenz 35, 109–118.
- 木下富雄 (2002) 「リスク認知の構造とその国際比較」 安全工学会誌 Vol41 No, 6
- 川端敏朗 (2008) 「消費者契約法における説明義務、情報提供義務」 比較法制研究 (国士館大学) 第 29 号 (2006) 153-176
- 藤垣裕子 (2019) 「科学技術社会論からみたリスクコミュニケーション — 幅のある情報発信と市民の分断, そして RRI をめぐって —」 安全工学 58 (6), 419-425, 2019-12-15
- Dichu Bao, Yongtae Kim, L. Su (2021) 「Do Firms Redact Information from Material Contracts to Conceal Bad News?」
- Davis, F. D. (1989). 「Perceived usefulness, perceived ease of use, and user acceptance of information technology」 MIS Q. 13, 319–340
- Wijayati, D. T., Rahman, Z., Rahman, M. F. W., Arifah, I. D. C., and Kautsar, A. (2022) 「A

study of artificial intelligence on employee performance and work engagement: the moderating role of change leadership] *Int. J. Manpow.* 43, 486–512.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003) 「User acceptance of information technology: toward a unified view」 *MIS Q.* 27, 425–478

Barrett, A. M., and Baum, S. D. (2017) 「A model of pathways to artificial superintelligence catastrophe for risk and decision analysis」 *J. Exp. Theor. Artif. Intell.* 29, 397–414

Frey, C. B., and Osborne, M. A. (2017) 「The future of employment: how susceptible are jobs to computerisation? 」 *Technol. Forecast. Soc. Chang.* 114, 254–280.

Brynjolfsson, E., and McAfee, A. (2014) 「The second machine age: work, progress, and prosperity in a time of brilliant technologies」

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000) 「The affect heuristic in judgments of risks and benefits」 *Journal of Behavioral Decision Making*, 13, 1-17.

Kim, S. D. (2012). Characterizing unknown unknowns. Paper presented at PMI® Global Congress 2012—North America, Vancouver, British Columbia, Canada. Newtown Square, PA: Project Management Institute.

Cleden, D. (2009). 「Managing project uncertainty」 Farnham, UK: Gower.

Baruch Fischhoff (1995) 「Risk Perception and Communication Unplugged: Twenty Years of Process Addressing Agencies' Risk Communication Needs」 A Symposium to Discuss Next Steps, Annapolis, MD, June 6–8, 1994

George A. Akerlof (1970) 「The Market for "Lemons": Quality Uncertainty and the Market Mechanism」 *The Quarterly Journal of Economics* Vol. 84, No. 3 (Aug., 1970), pp. 488-500 (13 pages)

Gambetta, D. (Ed.) (1988). *Trust: Making and breaking cooperative relations.*

Appendix

実験に用いた手順書

ツールの概要

- ツールの目的
 - 企業間の信頼関係の維持・向上を促進することを目的としたツールです。信頼する行為とは「信頼する者が知覚したリスクに対してリスクテイクを行うこと」であるという信頼とリスクの関係性を前提として設計しています。
 - 信頼関係を構築したいステークホルダー（本検証では、AI サービスプロバイダー企業）のリスクの知覚状況を想定して、リスク情報伝達に係る戦略検討を支援できることを目指しています。
- 手元にご準備いただくもの
 - 本手順書
 - Miro ボード（事前にお教えした URL にアクセスしてください。）
 - ※当日までにアンケートにご回答されていること。未回答の場合はご説明前にご回答ください。※
- 全体の流れ

内容	目安となる所要時間（分）	累積所要時間（分）
ご説明	20	20
STEP1	-	20
STEP2	10	30
STEP3	25	55
STEP4	20	75
STEP5	30	105
事後アンケート	15	120

2

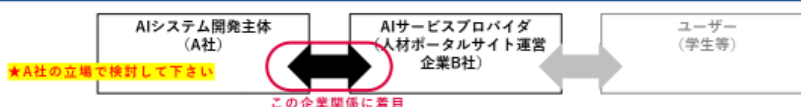
本検証で使用するケーススタディ

- 実際の企業情報を利用すると非公開情報・機微情報にあたるため、検証においては、架空のAI サービスを想定し、検討を行っていただきます。

【ケーススタディ：採用AI】

- エントリーシートの書類選考を判断する際の参考情報として使用されるAIサービス。
- A社はAIシステム開発企業（あなたはA社の社員という立場）であり、AIサービスプロバイダーとなる人材ポータルサイト運営企業のB社から過去のエントリーシートデータと合否判定（内定の判定）結果を受領し、機械学習（分類モデル）で合否判定をする学習モデルを作成。

適合率（Precision）	書類選考で合格と判断した中で、面接を経て内定を出した割合で評価し、70%を期待値として設定。※蓄積された本番データは採用・不採用の判定結果が入力されたタイミングで適宜学習データとして追加され、日次で学習モデルを更新して本番環境に反映する。ただし、学習プロセスでクロスバリデーションによるテストの結果、正解率が70%を下回る場合には本番環境のAIモデルを自動更新しない。AIモデルは過去1年分のバージョンを保存している。
使用方法	人材採用部門担当者はAIの判断を参考情報としながら、合否の判断を設定し、人材採用部門長の承認を得て、申込者に合否の通知を行う。
学習データ	過去のエントリーシート（テキスト）＋合否結果を使用。※個人情報あり
本番用データ	最新のエントリーシート（テキスト）を使用。※個人情報あり
AIサービスからの出力内容	合否の判定
出力方法（画面/アクチュエータ等）	人材採用部門担当者の端末上で、申込者のエントリーシートを投入すると、書類選考の合否判定が出力される。
根拠情報の出力	エントリーシートの中で判断に強い影響を与えたキーワードがハイライトされる
安全性のリスク有無	無
外部AIへの連携	無
外部AIへの連携方法・プロトコル	無

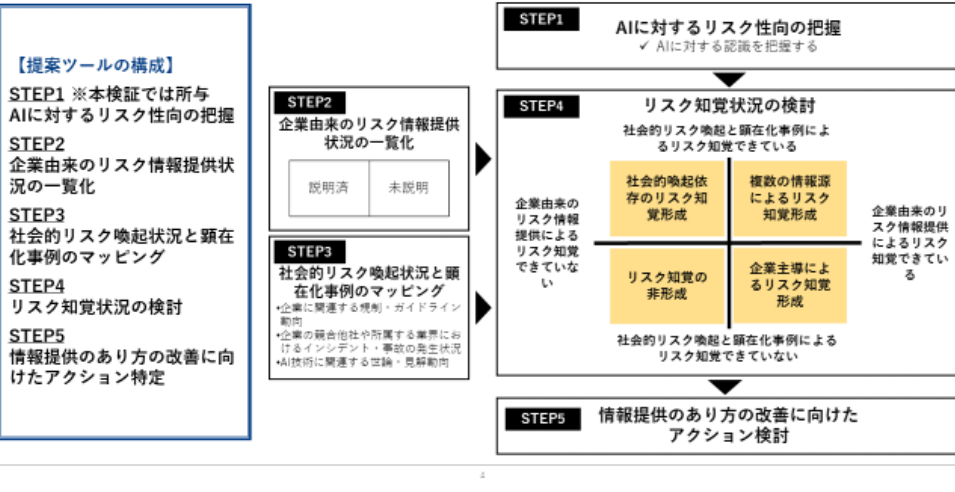


参考)

3

ツールの全体像

- 企業間の信頼関係の維持・向上を促進することを目的としたツールです。**信頼する行為とは「信頼する者が知覚したリスクに対してリスクテイクを行うこと」である**という信頼とリスクの関係性を前提として設計しています。
- 信頼関係を構築したいステークホルダー（本検証では、AIサービスプロバイダー企業）の**リスクの知覚状況を想定して、リスク情報伝達に係る戦略検討を支援できること**を目指しています。



検討の前提

- ケーススタディに基づく、「**AIリスク項目一覧表**」を提供しています。各AIリスク項目について、①どのようなリスクが存在するか（C列）、②そのリスクについてAIサービスプロバイダー企業にどのように説明した状況か（D列）、③そのAIリスク項目と関係するAIシステム構成要素（E列）で整理しています。
- 「**AIリスク項目一覧表**」はSTEP2,STEP5で使用します。Miroボードにも表を掲載していますが、Excel資料でも提供しますので使いやすい方をご使用下さい。

AIリスク項目一覧表				AIシステム構成要素の説明			
項目	AIリスク項目	そのようなリスクがなぜ発生するか	説明の概要（Risk Item）	関連するAIシステム構成要素	項目	説明	関連要素
1	信頼関係の構築・維持	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	信頼関係の構築・維持に寄与するAIシステム構成要素	1-1	プロダクト概要	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。
2	信頼関係の構築・維持	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	信頼関係の構築・維持に寄与するAIシステム構成要素	1-2	プロダクト概要	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。
3	信頼関係の構築・維持	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	信頼関係の構築・維持に寄与するAIシステム構成要素	1-3	プロダクト概要	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。
4	信頼関係の構築・維持	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	信頼関係の構築・維持に寄与するAIシステム構成要素	1-4	プロダクト概要	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。
5	信頼関係の構築・維持	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。	信頼関係の構築・維持に寄与するAIシステム構成要素	1-5	プロダクト概要	AIが信頼の構築・維持に寄与する一方で、信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。信頼関係の構築・維持に寄与しないリスクが存在する可能性がある。

STEP1：AIリスク性向の把握 ※所与条件とするため作業としてはスキップ

- AIサービスプロバイダー企業に対しアンケートを実施し、**AIリスクに対するリスク性向を把握**して下さい。
 - 下記のアンケートを使用することで、認識される利益、潜在的リスク、**AI技術の利用意向など、認識される有用性と技術が社会や人類に与える潜在的影響に焦点を当てることで、個人のAIに対する態度を捉える**ことができる。

AI態度尺度 (AI Attitude Scale; AIAS-5)

- ① 私はAIが私の生活を向上させてくれると信じている。
- ② AIは私の仕事を向上させてくれると信じている。
- ③ 私は将来、AI技術を使うと思う。
- ④ AI技術は人間にとって脅威だと思う（逆項目）。
- ⑤ AI技術は人類にとってプラスだと思う。

▼ 回答の平均値が3以上の場合

AIリスクに対して積極的

- AIサービスプロバイダー企業が過剰にリスクを少なく見積もって認識している可能性があるため、リスク情報提供の際にはよりリスクを強調することが重要となる

OR

▼ 回答の平均値が3より小さい場合

AIリスクに対して消極的

- AIサービスプロバイダー企業が過剰にリスクを大きく見積もって認識している可能性があるため、リスク情報提供の際にはより便益・対策を強調することが重要となる

※本検証ではこちらであったと仮定

5

STEP2：企業由来のリスク情報提供状況の一覧化 10分

【STEP2】

- **AIリスク項目一覧表を参照**し、既にリスク情報についてAIサービスプロバイダー企業への説明済みか、もしくは、まだ説明は実施していないかを考慮し、Miroボード上のAIリスク項目が記載された付箋を【説明済】もしくは【未説明】のグループに振り分けてください。

説明済	未説明
意思決定における偏り・偏見 	利用における不平等性
... ...	

7

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

- AIサービスプロバイダー企業のリスク知覚に影響を与える要因について、以下の観点で調査や発想を実施し、AIリスク項目のそれぞれについて社会的なリスク喚起状況やリスクが顕在化した事例の有無を整理します。具体的なステップは次頁で説明します。

企業に関連する規制・ガイドライン動向

新たな法規制がガイドラインが導入され、それに準拠していないことで、その企業への信頼を損なう要因となる可能性があります。以下について調査・発想してください。

- ・ 新規法令
- ・ 業界ガイドライン等のソフトロー
- ・ 訴訟の結果や判例

企業の競合他社や所属する業界におけるインシデント・事故の発生状況

企業が、競合他社や同業界の企業で発生したリスクや脅威に対して適切な対策を取らないことで、その企業への信頼を損なう要因となる可能性があります。以下について調査・発想してください。

- ・ 他社インシデントの発生状況
- ・ 競合他社や同業界の企業による事故の発生状況

AI技術（そのもの）に関連する世論・見解動向

社会全体の倫理観やプライバシーに対する意識が高まった結果の水準に対して十分な対策を講じていないことで、その企業に対する信頼を損なう要因となる可能性があります。以下について調査・発想してください。

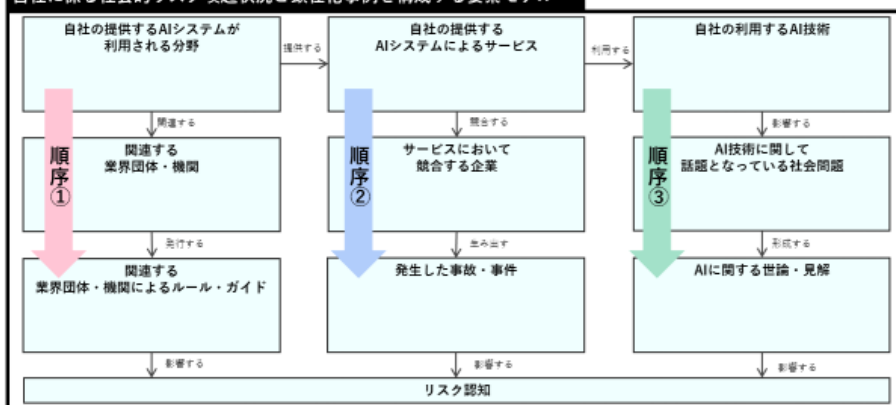
- ・ メディアでのAIやプライバシーの議論
- ・ ユーザーのフィードバックの傾向

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

【STEP3】

- 「自社に係る社会的リスク喚起状況と顕在化事例を構成する要素モデル」について、必要に応じて調査を行いながら、下記の順序で内容やキーワードを、1列目はピンク色の付箋、2列目は青色の付箋、3列目は緑色の付箋に書き出して下さい。
- 列ごとに一番下の行に付箋を貼れたら、記載したキーワード・内容を基に、関連するAIリスク項目一覧リストのリスク項目の枠に付箋をコピー＆ペーストして下さい。

自社に係る社会的リスク喚起状況と顕在化事例を構成する要素モデル

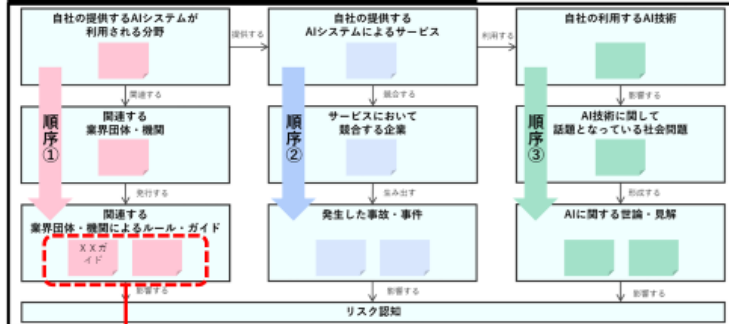


25分

STEP3：社会的リスク喚起状況と顕在化事例のマッピング

【STEP3】実施イメージ

自社に係る社会的リスク喚起状況と顕在化事例を構成する要素モデル



関連するAIリスク項目一覧リスト

1) 意思決定における偏り・偏見	2) 説明責任への対応の不足・透明性の欠如	3) 情報セキュリティの脅威拡大	4) 同意の仕組みの不足 XXガイド
5) 利用における不平等性	6) プライバシーの侵害 XXガイド	7) 技術・データの成熟度の低さによる影響	8) 品質保証とテスト基準の不備
9) 人間の介入の不	10) 運用上の脆弱性	11) 個人の尊厳・	12) 過剰なデータ保有・独占

XXガイドに「プライバシーの侵害」及び「同意の仕組みの不足」について言及がある場合、「関連するAIリスク項目一覧リスト」の当該枠にコピー＆ペーストする

10

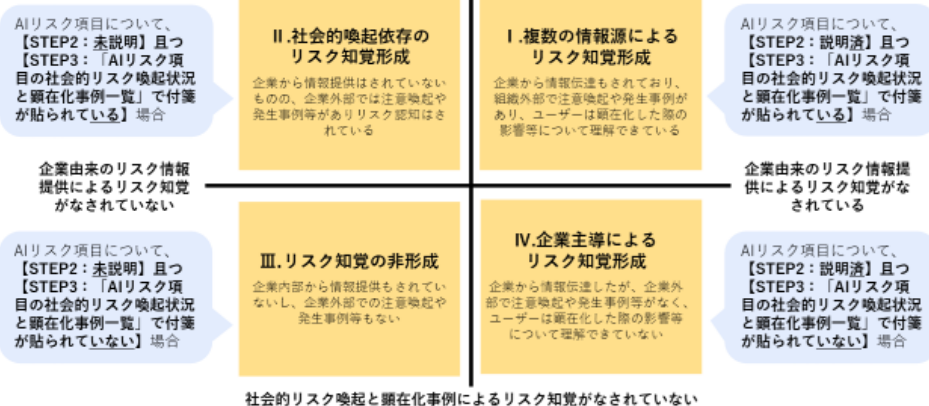
10分

STEP4：リスク知覚状況の検討

【STEP4】

- 「STEP2：企業由来のリスク情報提供状況の一覧化」と「STEP3：社会的リスク喚起状況と顕在化事例のマッピング」の結果を基に、AIリスク項目が記載された付箋を下記のマトリクスの該当する象限に貼って下さい。

社会的リスク喚起と顕在化事例によるリスク知覚がなされている

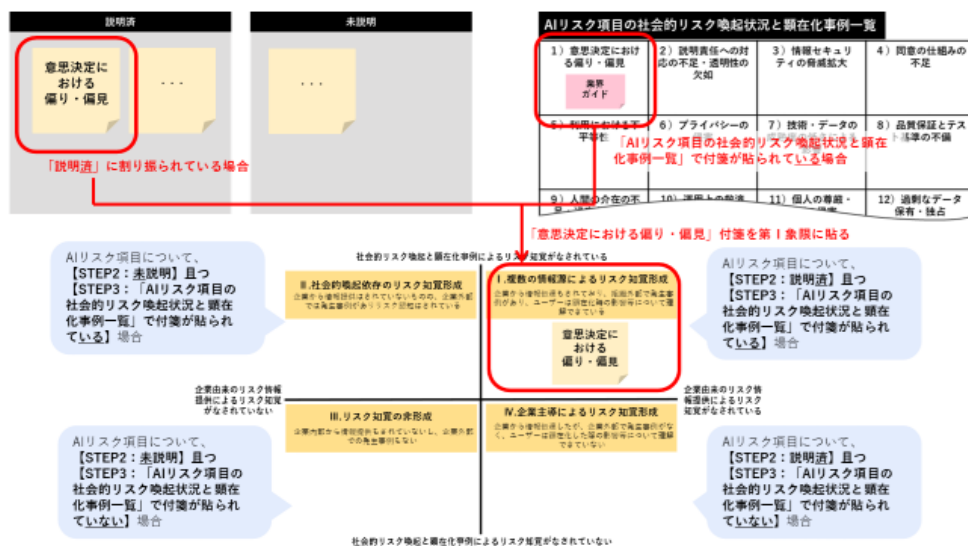


社会的リスク喚起と顕在化事例によるリスク知覚がなされていない

11

10分

AIリスク項目「意思決定における偏り・偏見」を例に



30分

- STEP4で整理した4象限ごとに表がありますので、各表の1行目にその象限に貼った「AIリスク項目」の付箋を左から順に貼り(1セルに1付箋)、そのすぐ下のセルにはAIリスク項目一覧表の当該「AIリスク項目」と同じ行に記載されている「関連するAIシステム構成要素」の項目の内容をコピーして下さい。

[illegible]

STEP5：情報提供のあり方の改善に向けたアクション検討

【STEP5-2】

- AIサービスプロバイダー企業に適切にリスクを知覚してもらうために、**各象限の表のAIリスク項目ごとに情報提供のあり方の改善に向けたアクションを検討**し、検討した結果を灰色の付箋に書いて貼って下さい。
- アクション検討の際に、①関連する開示項目、②リスク情報提供の方針を参考にして下さい。

【STEP5-2】実施イメージ



STEP5：情報提供のあり方の改善に向けたアクション検討

【STEP5-2】参考資料

- 情報提供のあり方の改善に向けたアクション検討の際には、以下の方針を参考として下さい。

