
「人を対象とした」データ利活用におけるELSIのリスク・マネジメント・プロセスの開発

Development of Risk Management Process for Ethical, Legal, and Social Issues in Human Data Utilization

大場 義浩*
Yoshihiro OBA

草薨 真人*
Masato KUSANAGI

荒海 麻由佳*
Mayuka ARAUMI

富上 良樹*
Ryoju TOGAMI

要 旨

近年のAI技術の進化と共に、人を対象としたデータの利活用が急速に進んでいる。その一方で、倫理的観点で社会からの批判を受け、サービスの提供を終了するケースが少なくない。そのため、これまでに様々な機関からAI技術についてのガイドライン等が公表されてきた。しかしながら、抽象的な原則を示すだけでは、開発の現場におけるリスク低減の活動には直接つながりにくい。そこで、本研究では、リスク・マネジメントの国際標準を参考にしつつ、個別の開発案件における具体的な倫理的・法的・社会的課題（Ethical, Legal and Social Issues: ELSI）を発見する方法を開発した。そして、ケーススタディを通じて、リスク低減に関する有効性を示すことができた。本稿では、導入としてAI倫理やELSIについて解説しつつ、開発した方法と今後の課題について報告する。

ABSTRACT

With the evolution of AI technology, the use of human data has increased rapidly, and there are many cases where services are discontinued upon receiving criticism for ethical reasons. While various organizations have published guidelines and frameworks on AI technology, merely outlining abstract principles often fails to directly contribute to risk reduction activities in the actual development process. Therefore, in this study, we developed a practical method for discovering specific Ethical, Legal, and Social Issues (ELSI) in individual development projects, drawing inspiration from international standards in risk management. Through case studies, we were able to demonstrate the effectiveness of risk reduction. In this paper, we provide an introduction to AI ethics and ELSI, present the developed method, and report on the challenges ahead.

* 先端技術研究所 HDT研究センター
Human Digital Twin Institute, Advanced Technology R&D Division

1. はじめに

1-1 人を対象としたデータ利活用の状況

2010年代のユビキタスネットワーク環境の整備を基盤として、IoTやAI技術の進化と共に、近年は様々なデータの利活用が急速に進んでいる。また2020年に始まった新型コロナウイルス感染症の流行は、社会のデジタル化を促進する機会ともなり、仕事の環境も大きく変化した。そのような中、人が創造力を発揮し、イノベーションを創出することや働きがいを実感できることがより重要視されるようになってきている。

創造力の発揮に寄与する技術を開発するためには、人を対象としたデータを利活用することが必要になる。近年では、人を対象としたデータを収集、解析、そして情報を提供する新たなサービスが増えており、個人情報などパーソナルなデータが利活用される機会も増えている。

その一方で、公平さの欠如やプライバシーへの配慮不足など倫理的な観点で社会から批判を受けるケースやソーシャルメディアを介してネット炎上に発展するケースも目立つようになっている。さらには、AI技術やビッグデータの利活用は社会を破壊すると警鐘をならす科学者らもおり¹⁾、人を対象としたデータの利活用が必ずしも善い影響だけをもたらすわけではないことに留意が必要である。

1-2 AI倫理のガイドライン化

前節のような状況を受けて、国内外の様々な機関からAI倫理に関するガイドライン等が公表されてきている。例えば、2017年に米国で行われたアシロマAI会議では、“人工知能の研究課題”、“倫理と価値”、“長期的な課題”の3領域で23の原則が採択されている²⁾。日本においても2017年に制定された人工知能学会の倫理指針³⁾をはじめ、2019年3月に内閣府による『人間中心のAI社会原則』⁴⁾、2019年8月に総務省による『AI利活用ガイドライン』⁵⁾が公表されている。また、2021年4月に欧州委員会

で提出された『人工知能に関する調和の取れたルールを定める規則の提案』⁶⁾のように罰則付きの規制化の流れもある。これらでは、透明性 (Transparency)、公正さ (Justice & fairness)、無危害 (Non-maleficence)、責任 (Responsibility)、プライバシー (Privacy) など、人間中心の社会を実現するための概念が重視されている⁷⁾。

1-3 リスク・マネジメント・プロセスの開発目的

ガイドライン等が多く公表される一方で、抽象的な原則を示すだけでは意味がないとの指摘もある⁸⁾。そのため、今後必要なことは、個別の開発案件に対してAI倫理に基づく分析・評価や対策を現場で実践するための具体的な方法を構築することと考えられる。

これまでもAI倫理を先導する立場の機関や企業がガイドライン等に基づいた評価ツールを開発し社会に提供するなど、より一歩具体的な実践に取り組んでいる。例として、IBM社で開発が始められた機械学習における学習バイアス対策ツール⁹⁾、一般社団法人日本データマネジメント・コンソーシアムによるAI・データ活用のためのチェックシート¹⁰⁾、FUJITSU社のAIの倫理影響評価方法¹¹⁾などが挙げられる。これらの多くは開発した技術やアルゴリズム、具体的なデータフローについてセルフチェックできるツールとなっている。

しかしながら、評価ツールを使っても開発者の自己評価のみでは評価結果の妥当性の確認については課題が残ってしまう。また、評価の対象が開発済みの技術に限定されてしまうと、後戻りを避けたい開発者心理が働くことから評価結果を反映した修正が難しくなってしまう課題もある。

そこで本研究では、開発された技術を主観的に評価するのではなく、研究の初期段階における開発者の認識を客観的に評価するリスク・マネジメント・プロセス（以下、RMPと記載）を開発することを目的とした。

2. リスク・マネジメント・プロセスの解説

2-1 開発したRMPの特徴

本章では、開発したRMPの詳細を説明する。本研究ではケーススタディを通じて実験を重ね、常に内容を改良しながらRMPを開発する。ここでは現状での方法を示す。RMPの特徴は下記の3点である。以下、節ごとにこれらの特徴について説明する。

- ・ ELSIの概念を反映した評価方法
- ・ ガイドライン等に基づくアセスメント項目
- ・ 開発者の認識に対する客観的な評価方法

2-2 ELSIの概念を反映した評価方法

近年、まだ社会実装されていない、あるいは実装されて間もない新しい技術を分析・評価するための考え方として“倫理的・法的・社会的課題 (Ethical, Legal and Social Issues: ELSI)”という概念が改めて注目されている。2021年に閣議決定された第6期科学技術・イノベーション基本計画でも、ELSI対応を促進する必要性が言及されている¹²⁾。そこで本研究では、RMPにおける評価方法にELSIの概念の反映を試みた。はじめにELSIについて説明する。参考情報¹³⁾を基に、Fig. 1にELSIを構成する課題の由来と特徴を整理した。

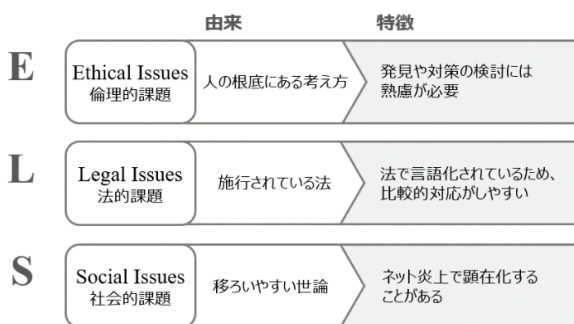


Fig. 1 Origin and characteristics of E, L, and S in ELSI.

倫理的課題は、人間の尊厳や自由といった社会的規範の根本にある考え方に由来している。その考え方は普遍的だが、社会からの意見は必ずしも明解で

はなく、倫理的課題の発見や対策の検討には熟慮が必要となる。

法的課題は、施行されている法に由来している。法的課題は法で言語化されているため、比較的对応がしやすい。しかしながら、新しい技術に対しては、法の施行が後追いとなってしまう場合が多い。

社会的課題は、世論に由来している。世論はその時々々の情報に対する主観的な反応であることが多く、状況によって移ろいやすい。そのため、社会的課題はソーシャルメディアを介したネット炎上で顕在化することがある。

この考え方を基本として、RMPが対象とするリスクは人を対象としたデータ利活用におけるELSIとした。また、RMPではFig. 2に示すフローとELSI発見のポイントを基本の考えとした。以下ではそれらについて説明する。Fig. 2にAIシステムに関する一般的なデータ利活用のフローを示す。このフローには、開発と利用のプロセスがある。

ここでは開発プロセスを研究開発時にAIシステムを構築する作業と定義する。開発プロセスでは、データの取得に始まり、データの加工（例えばデータのクレンジングやタグ付けなど）、そしてアルゴリズムの設計の手順で作業が行われる。場合によっては、第三者にデータを提供するなど二次利用が行われるケースもある。また、利用プロセスはAIシステムをサービスとして実際に使用する際の作業と定義する。利用プロセスでは、データの提供者である利用者から生データを取得することに始まり、学習済みモデルでの解析、そしてそのアウトプットを利用する手順で進む。

このデータ利活用のフローにおいてELSIを発見するには、データは適切に入手したか、アルゴリズムは公正か、サービスは利用者にとって役に立つものかを問うことがポイントとされる。AI技術の利活用で生じる潜在的なリスクは、主に「データ」「アルゴリズム」「入出力」「目的とその影響」で混入しやすいとも説かれる¹⁴⁾。データ利活用のフローには人が担う作業が含まれている。その作業において、AIシステムの開発者の個人的な見方や先

- ELSI発見のポイント**
- データは適切に入手したか
 - アルゴリズムは公正か
 - 利用者にとって役に立つものか

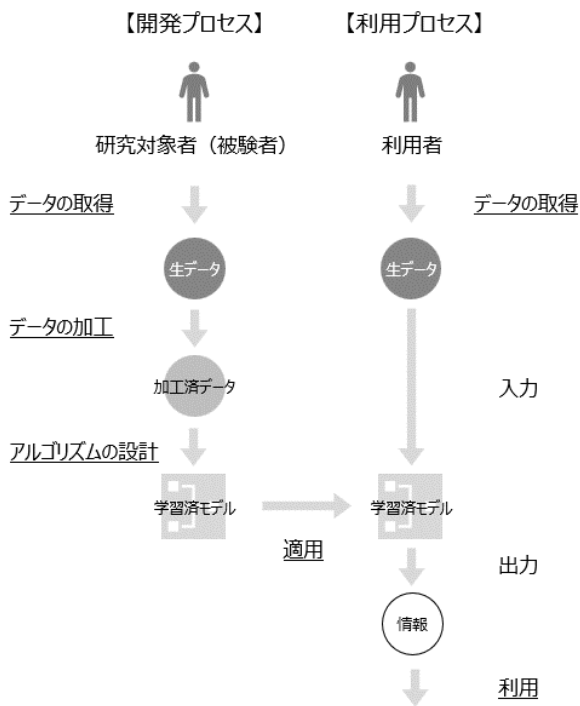


Fig. 2 Flow of data utilization.

入観はシステムやデータ仕様の決定に対して無意識にも影響を与える可能性がある。これによってELSIが生じる場合がある。RMPではELSIの概念を反映するものとして以上に留意して評価を行う。

2-3 ガイドライン等に基づくアセスメント項目

本研究では、ガイドライン等で用いられる諸原則を参考にRMPにおけるアセスメントの項目をFig. 3に示す32項目に設定した。

また、各項目を「自律」「安全」「有益」「公正」「理解可能」「リコーウェイ¹⁵⁾」「適切な利用」の7つのカテゴリに分類した。「自律」「安全」「有益」「公正」は、生命医学倫理の諸原則にある人に対する敬意（respect for persons）、無危害（nonmaleficence）、慈恵（beneficence）、正義（justice）の原則を、RMPのアセスメント原則とし

- 1. 自律**
 - データ取得の適正さ
 - 選択肢を排除しない
 - Human in the loop
 - 学習し成長するシステム
- 2. 安全**
 - プライバシー保護
 - データミヌム
 - 合理的なセキュリティ
 - 信頼性
- 3. 有益**
 - メリットとデメリットのバランス
 - 不正の防止
 - 持続可能
 - 装置間の連携
- 4. 公平**
 - 認知バイアス
 - 差別の禁止
 - 公共性
 - Diversity & Inclusion
- 5. 理解可能**
 - トレーサビリティ
 - 解釈可能性
 - 説明可能性
 - 透明性
- 6. リコーウェイ**
 - CUSTOMER-CENTRIC
 - ETHICS AND INTEGRITY等
- 7. 適切な利用**
 - ターゲットの適正
 - 通知方法
 - 自由意思の確認
 - 合意の方法
 - 保守管理
 - 目的外利用の規制
 - 利己的利用の対処
 - 責任ある運用体制
 - 安全な運用体制
 - 信頼ある運用体制

Fig. 3 List of assessment perspectives.

て、その解釈を読み替えたものである。以下では、各カテゴリの説明を行う。

「自律」は、自己決定権の尊重に関わる項目を収めたカテゴリである。ここではデータ提供者である研究対象者（被験者）やサービスの利用者との手続きの適正さや、人間の役割や判断を組み込んだ人間参加型のシステムであるか、データ提供者からフィードバックされる意見の活用方法などをアセスメントの項目としている。

「安全」は、プライバシー保護を含め、データ提供者の安全確保に関わる項目を収めたカテゴリである。ここではプライバシー情報の特定やセキュリティ対策の合理性、取得するデータは目的に対して適切で必要最小限に留めているかなどをアセスメントの項目としている。

「有益」は、技術がもたらす恩恵に関わる項目を収めたカテゴリである。ここでは技術を社会実装することによるメリットとデメリットのバランスや、技術が悪用される可能性、他のシステムと組み合わせる場合の影響などをアセスメントの項目としている。

「公平」は、偏りがなく全てを同等に扱うこと、広く社会の価値観に調和することに関わる項目を収めたカテゴリである。ここでは取得したデータ等が

潜在的に備える各種バイアスの有無や、差別につながる可能性、誰もが使える公共性があるかなどをアセスメントの項目としている。

「理解可能」は、近年の多くのAI倫理のガイドライン等でも言及される説明可能なAI (Explainable AI) の概念を反映させたカテゴリである。ここではAIシステムの影響を人間が予測可能にするために、アルゴリズムの解釈の容易さや情報開示の計画、データの保管方法などをアセスメントの項目としている。

「リコーウェイ」は、弊社の企業理念を反映させたカテゴリである。弊社ではリコーグループの日々の判断や活動の基礎となる普遍的な理念をリコーウェイと呼び、社員が使命を果たし目指す姿を実現するための行動指針として以下の7つの価値観を設定している。

- ・ CUSTOMER-CENTRIC：お客様の立場で考え、行動する
- ・ PASSION：何事も前向きに、情熱を持って取り組む
- ・ GEMBA：現場・現物・現実から学び改善する
- ・ INNOVATION：制約を設けず、柔軟に発想し、価値を生み出す
- ・ TEAMWORK：お互いを認め合い、すべての人と共創する
- ・ WINNING SPIRIT：失敗をおそれず、まずチャレンジし、成功を勝ち取る
- ・ ETHICS AND INTEGRITY：誠実に、正直に、責任を持って行動する

RMPでは上記の価値観のうち2つをアセスメントの項目としている。一方は、他の価値観にもつながる基本的な価値観であると考え、CUSTOMER-CENTRICとする。もう一方は、依頼人が自身のアイデアに対する認識に合わせて他の価値観から自由に選択する。

「適切な利用」は、サービスの提供の際に利用者に対して確認すべき項目を収めたカテゴリである。

原則に従って綿密に設計されたシステムも、使い方を誤れば想定外のトラブルを招く可能性がある。そのため、開発者はサービスを提供するまでではなく、適切な利用を促すところまでを検討しておくことが望ましい。ここでは利用者への通知方法や内容の適切さや、サービスの利用が強制的にならないか、目的外の利用に対する対処方法、運用体制の構築は適切かなどをアセスメントの項目としている。

2-4 開発者の認識に対する客観的な評価方法

本節では、RMPの手順を示すとともに特徴である客観的な評価方法について説明する。RMPは、新しい技術やサービスのアイデアを持った開発者（以下、依頼人と記載）を利用の対象者としている。RMPを実施する目的は、依頼人のアイデアにおけるELSIを発見し対策の実施を促すことである。そこで、RMPによって依頼人とステークホルダーとの認識の乖離がコミュニケーションを通して可視化されることを目指した。

RMPはリスク・マネジメントの国際標準であるISO31000（対応する日本工業規格、JIS Q 31000:2019）¹⁶⁾のフレームワークを参考にしている。そのためRMPはFig. 4に示す3つのパートで構成される。以下では各パートで実施する内容について説明する。なお、RMPでは円滑な進行のために依頼の問い合わせ先（以下、事務局と記載）が必要に応じて支援を行う。

パート1では、主に依頼人と事務局とで事前相談を行う。依頼人からの説明を基に、事務局はFig. 2に示したELSI発見の3つのポイントに留意しつつ、AI倫理のガイドライン等を参考に当該技術のリスクレベルを大まかに把握し、パート2への移行、留保、不要の判断などを依頼人に提案する。

パート2では、依頼人が、依頼人とは異なる知識や経験あるいは感覚を有することが見込まれる人物（以下、評価者と記載）とのコミュニケーションを通して依頼人だけでは見過ごしがちなリスクの発見

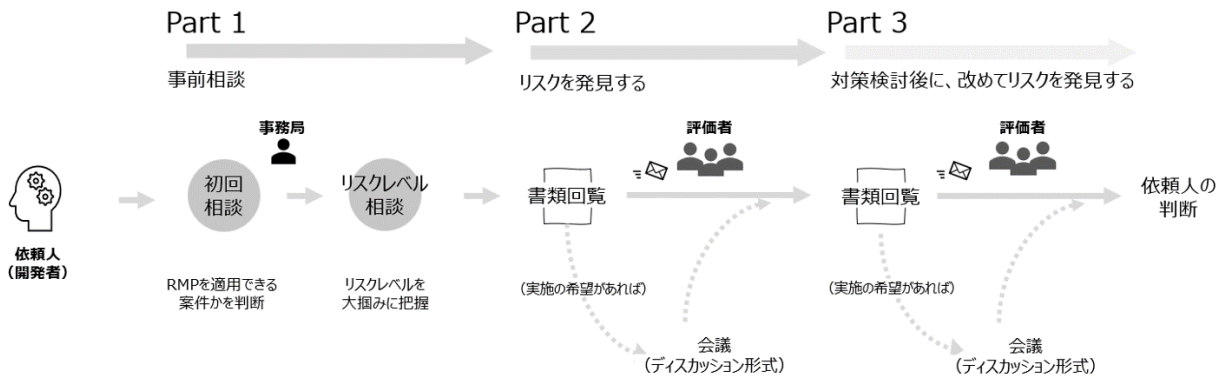


Fig. 4 Overview of the risk management process in this study.

と分析を行う。なお、評価者は少なくとも依頼人が行う技術・サービス開発に直接関わっていない人物から選ばれる。

パート2における依頼人と評価者のコミュニケーション手段は、依頼人のアイデアに関する情報を記述した書類を回覧する方法と、ディスカッション形式の会議を行う方法を併用する。ただし、ディスカッション形式の会議を行う方法は依頼人が希望する場合に実施する。

RMPでは、書類を回覧する方法をコミュニケーション手段の基本としている。それは過去、ディスカッション形式の会議において様々な課題に直面してきたためである。ディスカッション形式の会議では、話題が頻繁に流動したり、同じ議論が繰り返されたりすることが起きうる。また、進行時の評価者の心理的安全性の確保や発話者の意見に流され聴くことに専念してしまう懸念など配慮すべきことが多い。そのためRMPでは、あらかじめ論点を整理した上で、複数の評価者が独立して自身のペースで思考できるよう、書類を回覧する方法を開発した。

書類回覧による方法は、以下の①から④の作業からなる。依頼人の希望があれば、③から⑤の手順を反復する。

- ① 事務局は、評価者に対しパート1で依頼人から受けた説明を基に依頼人のアイデアを紹介する。この際、書類回覧による方法や手順を

確認するとともに、回答は匿名で扱われることを伝える。

- ② 事務局は、依頼人に対し2-3に記載したアセスメント項目について自身の技術に対する認識を書類に回答するよう依頼する。
- ③ 依頼人は、自身のアイデアに対する認識を書類に回答し事務局に提出する。その後、事務局は、内容確認ののち書類を評価者に転送する。
- ④ 評価者は、受け取った書類に記載された依頼人の認識に対して質問や所感などを自由に追記し事務局に提出する。その後、事務局は、内容確認ののち依頼人に書類を転送する。
- ⑤ 依頼人は、受け取った書類に記載された評価者の回答を読み、自身の認識を改めつつリスクについて分析する。

また、書類の回答は文章で自由記載する形式とした。これは評価者が感じたELSIを論理的に説明することはできなくとも、その率直な反応をすくい取られることを狙いとしたためである。「よくわからないけど、何か嫌な感じがする」のように素朴で純粋な反応にこそ、潜在的なELSIが含まれていると考えたためである。

一方、本研究ではELSIの分析と対策の効率化のためにリスクの定量化にも取り組んでいる。定量化することで、依頼人がリスクの大きさに応じて対策を検討できるようにするためである。RMPでは、

依頼人と評価者の認識に関する状態をリスクの評価指標としている。この状態は両者の認識が一致しているのか、あるいは乖離しているのかを表すものである。状態のスコアをAgreement Score（以下、ASと記載）と定義し、主観アンケートに対する両者の回答から定量化する。

主観アンケートは、Fig. 5 (a)に示すように依頼人には技術開発の検討熟度について、評価者には技術を社会実装することに対する許容度について、10段階で回答してもらう内容になっている。Fig. 5 (b)に両者の回答とASの対応を示す。ASは依頼人の検討熟度が高くかつ評価者の許容度も高いほど高くなる。ASが高ければ両者の認識が一致しておりリスクは小さい。一方、ASは依頼人の検討熟度が高くかつ評価者の許容度が低いほど低くなる。ASが低ければ両者の認識が乖離しておりリスクは大きい。また、ASは依頼人の検討熟度が低いほどゼロに近づく。

これは依頼人が今後検討してみれば両者の認識は合う可能性があり、中間の状態であることを表している。なお、依頼人が対象外と回答した場合のASには評価者の回答が反映されるようにした。逆に、評価者がよくわからないと回答した場合のASには依頼人がある程度検討している場合に限り依頼人の検討熟度が反映されるようにした。

また、主観アンケートで許容度10%と回答した評価者の割合をRejection Level（以下、RLと記載）と定義する。RLは“!”一つが2割を指し、“!!!”であれば評価者の6割が「拒否感がある」と回答していることを意味する。RLは少数ながらも強い反対の声をすくい上げるための指標として用いる。

RMPでは、評価者の記述回答、AS、ASの平均値（以下、AS_AVEと記載）、RLを参考に依頼人がリスクについて分析する。AS_AVEは集団としての許容を捉えるための指標として用いる。

RMPでは2-3に記載したアセスメント項目で評価を実施する。評価者の記述回答、AS、AS_AVE、RLは、アセスメントの項目それぞれで得られるため分析は項目ごとに行う。なお、RMPではAS_AVEが15以下かつRLが1（“!”）の場合を注意、

(a) 主観アンケートの内容

依頼人	評価者
検討熟度について ** : この質問は、対象外だと思う 10% : まったく考えてこなかった 30% : 少しは考えてみた 50% : そこそこ考えたつもり 60% : 自分では十分考えたと思える 70% : 80% : 90% : 100% : これ以上ないほど考え尽くした	許容度について ** : よくわからない 10% : いやまずいだろと拒否感がある 20% : なんとなく違和感を感じる 40% : 50% : 60% : 70% : 80% : 90% : 100% : 十分責任を果たせている

(b) 両者の回答とASとの対応

依頼人の回答	評価者の回答										
	**	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
**	0	-40	-30	-20	-10	0	10	20	30	40	40
10%	0	0	0	0	0	0	0	0	0	0	0
20%	0	-10	0	0	0	0	0	0	0	0	10
30%	0	-20	-10	0	0	0	0	0	10	20	20
40%	0	-30	-20	-10	0	0	0	10	20	30	30
50%	0	-40	-30	-20	-10	0	10	20	30	40	40
60%	10	-50	-40	-30	-20	10	20	30	40	50	50
70%	20	-60	-50	-40	-30	10	30	40	50	60	60
80%	30	-70	-60	-50	-40	20	40	50	60	70	70
90%	40	-80	-70	-60	-50	20	50	60	70	80	80
100%	50	-90	-80	-70	-60	20	60	70	80	90	90

Fig. 5 Correspondence between survey and AS.

AS_AVEが30以下またはRLが2（“!!”）以上の場合を警告に値するリスクの大きさであるとして、これらに該当する項目をその中に重大なELSIが認められる状態として表す。

パート2の後、依頼人にはリスクに対する対策を検討してもらう。ただし、32項目全てについて対策を検討することは必須ではなく、項目の選択は依頼人の判断に任せることとした。そしてパート3では、パート2と同様のコミュニケーション手段で、依頼人が対策を検討した上での認識と評価者の認識との乖離を改めて確認する。

RMPを用いることで最終的なリスクを低くコントロールすることが可能になる。そして、RMPに関する記録は透明性確保とナレッジの共有のために開示されることが望ましい。依頼人はリスクレベルが許容可能であることを自ら説明し、また、リスクに対する自らの判断責任や開発に関わる責務は負うものと、認識しておく必要がある。

3. ケーススタディ

本章では、RMPを弊社で開発している画像解析による従業員の健康管理サービス¹⁷⁾に適用した例を示す。

はじめにサービスの概要を説明する。このサービスは、主にオンラインでの朝礼に参加した営業職の従業員を対象としている。その内容は、従業員の顔表情や会話をカメラ画像や音声データとして取得し、画像・音声解析等の技術を用いて従業員の状態の変化（例えば、笑顔が減ったなど）や会議の状態をスコア化し、従業員の上司へその結果を通知するものである。

次に評価者について説明する。本ケースでは6名の弊社社員に評価を依頼した。評価者は広く一般から選ばれることが望ましいが、本ケースがRMPの有効性を示す探索的な試みであったことから、弊社社員から選ばれた。人選では、性別、年齢層、組織、

職種が分散するように努めた。本ケースでは、希望のあった3名には事務局が個別に書類に関する補足説明を行いながら聴取する形式（以下、インタビュー形式と記載）で実施した。

本ケースのパート2、パート3の結果をFig. 6に示す。ここではアセスメントの項目それぞれについてAS_AVEとRLを示した。Fig. 6に示すように、RMPを導入したことにより開発案件における潜在的なELSIを俯瞰することが可能になった。本ケースでは、パート2で『メリットとデメリットのバランス』『公共性』『Diversity & Inclusion』『自由意思の確認』『合意の方法』のアセスメント項目の中に重大なELSIが認められた。

RMPでは、はじめにELSIの見当を付けたのち、評価者の記述回答を精査し、その問題を分析していく。本ケースにおいては、『メリットとデメリットのバランス』『公共性』の項目では「スコア化されるのが従業員だけなのは不信感を感じる。上司のよ

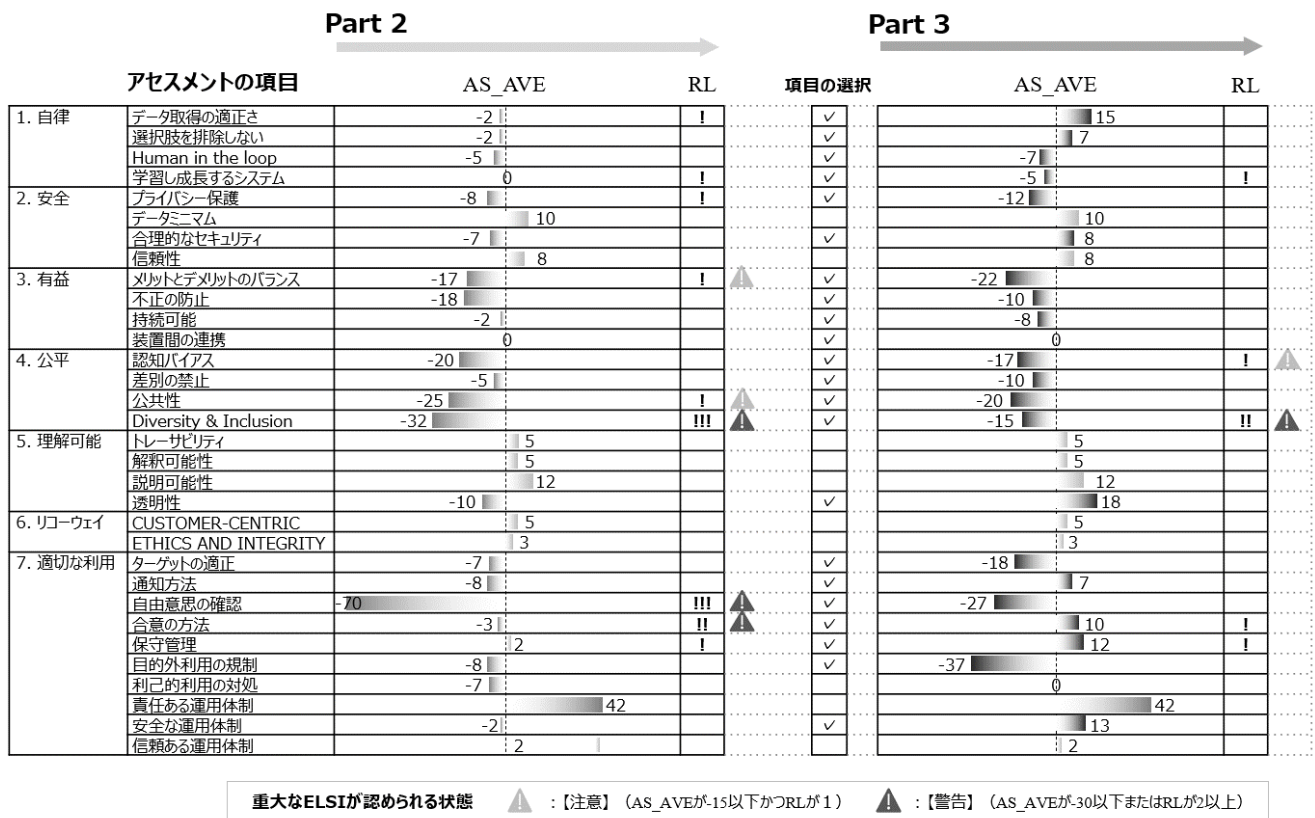


Fig. 6 Results of this process attempt.

うな管理者も同様にスコア化されるべき。」など従業員の立場に立った検討が不足していることに関する回答があった。また『Diversity & Inclusion』の項目では「朝から笑顔であることを無条件に善いこととして価値観を強制している気がする。」など感情労働につながる可能性を示唆する回答があった。

『自由意思の確認』『合意の方法』の項目では「チームで利用するサービスなので利用に関しては同調圧力が働き、ほぼ辞退できない。」など個人の希望によらず実質的には強制参加になる可能性に関する回答があった。

本ケースでは、依頼人はアセスメントの項目から22項目を選択し対策の検討を行った。Fig. 6においてチェックマークの付いた項目が選択された。検討の結果、依頼人はスコア化される対象に上司のような管理的立場の者も含める仕様の変更や、利用辞退の通知やカメラ録画のオンオフを従業員自身がコントロールできる機能をソフトウェアに実装するなどの追加開発を行った。

この対策を受けて実施したパート3では、『自由意思の確認』『合意の方法』の項目などでAS_AVEやRLの改善がみられた。今回のケーススタディを通じてリスクが低減する結果が得られ、我々が開発したRMPの有効性を示すことができた。

4. ケーススタディを通じて発見された課題

筆者らが所属する弊社研究センターでは、第3章で示したケースを含め複数のケースで試用したのち、2023年5月よりRMPの運用を開始している。しかしながら、RMPにもケーススタディを通じて幾つかの課題を発見した。発見した課題は、RMP自体に関する課題とRMPを運用する側のオペレーションに関する課題とに分けられる。以下では発見した課題について述べる。

はじめに、RMP自体に関する課題について述べる。一つは、評価者の記述回答の解釈方法に課題がある。評価者の記述回答にはELSIに関する多くの

示唆が含まれる。しかしながら、記述から潜在する意味を解釈するには時間と経験を要する。本ケースでは依頼人の主観で解釈してもらったが、今後は自然言語処理に関するAI技術の応用として解釈を機械が支援することも有効と考えられる。ただしその一方で、人間の判断が介在する仕組みを残すことも必要になる。

また、リスクに対する対策の評価にも課題がある。リスク・マネジメントは研究段階から社会実装後の使われ方までも含む長期にわたる取り組みである。本研究では、RMPを社会実装前に実施するものとして開発したが、今後はリスクに対して対策した技術やサービスが社会実装後にどのように受け入れられたのか、市場からの意見を取り入れる仕組みも必要になる。

次に、RMPを運用する側のオペレーションに関する課題について述べる。一つは、人材に関する課題がある。RMPにおける評価者はアセスメントする技術やAI倫理、ELSIに関する前提の知識を備えていない場合が多い。したがって、本ケースでは回覧する書類のみでの説明では理解が十分追従せず、追加で説明するなど個別のフォローが必要であった。またインタビュー形式で回答してもらった場合、事務局は評価者から考えを引き出すために問いかけをする必要があるが、評価者の考えを事務局の考えに誘導してしまうことは避けなければならない。事務局を務めるにも一定の知識や訓練が必要である。RMPを円滑に運用するには、作業の効率化のほか、ELSIに関する教育や啓発の活動を通して人材を育成することが必要になる。

また、人材の育成に向けてはRMPを運用する組織におけるELSIやリスク・マネジメントに対する意識の醸成に課題がある。その解決に向けては、組織のトップマネジメントが方針やコミットメントを明示し、運用の体制を確立していくことが必要になる。

5. おわりに

本研究では、ISO31000とELSIの概念を参考に、人を対象としたデータを利活用するAI技術に関するリスク・マネジメントの方法を開発した。そしてケーススタディを通じてそれが実践可能であることを確認するとともに、今後の課題を明らかにした。

第4章で説明したようにRMPにはまだまだ課題が多い。しかしながら、このRMPは机上で完成度を高めるものではなく、実践を通じた絶えざる維持改善を行うものである。弊社では今後もこの実践を継続していく。

謝辞

本研究は、大阪大学社会技術共創研究センターと弊社との共創研究プロジェクト「新規技術およびソリューションの社会実装プロセスの研究」での議論を基に行われた。

参考文献

- 1) キャシー・オニール, 久保尚子: あなたを支配し、社会を破壊する、AI・ビッグデータの罠, pp. 6-25, インターシフト (2018).
- 2) Future of Life Institute: AI Principles Japanese, <https://futureoflife.org/open-letter/ai-principles-japanese/> (参照2023-12-22).
- 3) 人工知能学会倫理委員会: 「人工知能学会 倫理指針」について, <https://www.ai-gakkai.or.jp/ai-elsi/archives/471> (参照2023-12-22).
- 4) 内閣府: 人間中心のAI社会原則 (統合イノベーション戦略推進会議決定), 人間中心のAI社会原則会議, <https://www8.cao.go.jp/cstp/ai/ningen/ningen.html> (参照2023-12-22).
- 5) 総務省: AI利活用ガイドライン～AI利活用のためのプラクティカルリファレンス～, AIネットワーク社会推進会議, <https://www.soumu.go.jp/iicp/research/results/ai-network.html> (参照2023-12-22).
- 6) European Commission: Proposal for a Regulation laying down harmonised rules on artificial intelligence, <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence> (参照2023-12-22).
- 7) A. Jobin, M. Ienca, E. Vayena: The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, Vol. 1, pp. 389-399 (2019).
- 8) The Linux Foundation: AI Fairness 360 Understand and mitigate bias in ML models, <https://ai-fairness-360.org> (参照2023-12-22).
- 9) 古川直裕ほか: Q&A AIの法務と倫理, p. 130, 中央経済社 (2021).
- 10) JDMC: https://japan-dmc.org/?page_id=21982 (参照2023-12-22).
- 11) FUJITSU: AI倫理ガイドラインに基づきAIシステムの倫理上の影響を評価する方式を開発、手順書や適用例とともに無償公開, <https://pr.fujitsu.com/jp/news/2022/02/21.html> (参照2023-12-22).
- 12) 内閣府: 第6期科学技術・イノベーション基本計画 本文, 第6期科学技術・イノベーション基本計画, <https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html> (参照2023-12-22).
- 13) 電通: ELSI対応なくして、データビジネスなし?! 話題のELSIとは, <https://dentsu-ho.com/articles/7123> (参照2023-12-22).
- 14) 保科学世, 鈴木博和: 責任あるAI-「AI倫理」戦略ハンドブック, p. 81, 東洋経済新報社 (2021).
- 15) RICOH: 企業理念「リコーウェイ」, リコーについて, <https://jp.ricoh.com/about/ricoh-way> (参照2023-12-22).
- 16) JIS Q 31000: 2019 リスクマネジメント—指針.
- 17) 佐々木俊英ほか: 人情報を活用した組織・個人活性化ソリューション, *Ricoh Technical Report*, No. 46 (2024).