

AI 誤判断に関する利用時品質のトラスト構築についての考察^a

島成佳^{1,2} 小川隆一² 佐川陽一² 竹村敏彦³

概要: AI システム (AI 学習モデルを搭載する IT システム) には, AI 学習モデルの判定精度が 100%にならないことから, AI の誤判断によって処理や動作を誤ってしまう AI 誤判断リスクが存在する. AI 誤判断リスクは, 利用者が AI システム導入の際に, その利用における品質 (利用時品質) が受容可能であるかに影響する. 筆者らは AI 誤判断リスクが AI システムと利用者とのトラスト構築 (AI システム受容) にどう関わるかに注目しており, 利用者のトラスト構築のプロセスを明らかにし, 受容の可否判断に有効な, AI 誤判断リスクに関わる利用時品質の評価手法・指標を創出することを目標としている. 筆者らはまず, AI システムを利用する個人をトラスト構築の対象者に選び, AI 誤判断リスクの認知と受容に関するアンケート調査を実施した. 本論文ではアンケート調査の一部を速報という位置付けで報告する.

キーワード: AI システム, 誤判断リスク, リスク認知, リスク受容, トラスト構築

Consideration of Building Trust for Quality in Use about AI Misjudgment

SHIGEYOSHI SHIMA^{†1} RYUICHI OGAWA^{†2}
YOUICHI SAGAWA^{†2} TOSHIHIKO TAKEMURA^{†3}

Abstract: There is a risk of AI misjudgment in AI system (IT systems equipped with an AI learning model) because the judgement accuracy of the AI learning model does not reach 100%. The AI misjudgment risk which is adversely affected to an AI system user is acceptable in building trust from the view point quality in use when the AI system user uses the AI system. In the building trust, our research goals are to analyze the process of building trust of the AI system user and to create measurement method about the AI misjudgment risk. In this paper, we conduct a questionnaire survey on recognition and acceptance about AI misjudgment risks and describe about results of part of the questionnaire survey.

Keywords: AI System, Misjudgment Risk, Risk Perception, Risk Acceptance, Building Trust

1. はじめに

ディープラーニング (深層学習) を代表とする機械学習等の AI(Artificial Intelligence: 人工知能) 技術では, 学習データに基づいて, 探索的・反復的に AI 学習モデルを構築する. この AI 学習モデルは, 学習を網羅的に行うことが困難であり, 判定精度が 100%にならないことから誤判断や誤認識をする可能性がある. このため, AI 学習モデルを搭載する IT システム (以降, AI システム) には, 誤判断や誤認識によって処理や動作を誤ってしまうリスク (以降, AI 誤判断リスク) が存在する.

AI システムは既に製品やサービスとして広がり始めており, 社会経済の発展や社会問題の解決等に重要な役割を果たすと期待されているが, 利用者が AI 誤判断リスクの受容可否を判断できず AI システムの導入に至らないケースが見受けられる. 筆者らは, AI 誤判断リスクに関して AI システムの導入を進めるために「リスクを許容して AI システムを利用してもよい」という, AI システム導入に際し

て利用者観点でのトラスト構築の重要性に注目している [1][2]. これまでに, このトラスト構築を包括的に表すトラスト階層モデル, 及びトラスト構築プロセスを定義している [3]. 本トラスト構築プロセスでは, AI 誤判断リスクを製品の品質の一部として捉え, 利用者が AI システムに求める要求事項の提示から AI システムを導入するまでを 3 つのプロセス (「利用時品質と製品品質からの落とし込み」「製品品質の確保」「トラスト構築」) で定義している. 利用者は AI システムの利用時品質 (利用者の要求事項) に基づいて品質を評価し, 要求した品質を満たしていると判断できるときにトラストが構築され, AI システムの導入・利用に至る, とするものである.

本モデルでは, トラスト構築の主体となる「利用者」を個人, 組織, 社会の 3 つの類型に分類する. 各主体で AI 誤判断リスクの捉え方や許容の考え方が異なるため, トラスト構築プロセスも異なる部分もある. トラスト構築のプロセスは, 個人, 組織, 社会の順に, 誤判断リスクの許容を合意する関係者や考慮すべき項目等が増えることで複雑に

^a 本研究の意見は, 著者たち個人に帰属し, 所属機関の公式見解を示すものではないことをことわっておく.

¹ 長崎県立大学
University of Nagasaki

² 独立行政法人情報処理推進機構
Information-technology Promotion Agency, Japan (IPA)

³ 城西大学
Josai University

なり、AI 誤判断リスクの捉え方や許容可能かの判断も難しくなると想定される。

本論文では、トラスト構築の主体を、トラスト構築プロセスが一番容易であると思われる個人とする。筆者らはまず、個人の AI 誤判断リスクの捉え方や許容についての状況を明らかにするため、AI システムを実際に利用している個人（以降、AI システム利用者）を対象としてアンケート調査を実施した。これは実際に AI システム利用者は、何らかのトラスト構築プロセスを経験したという仮定に基づいてである。本論文では、速報という位置付けで AI 誤判断リスクの認知や受容に関する質問項目の集計結果や分析結果について述べる。

2. 関連研究

2.1 トラスト階層モデル

筆者らは、AI システムの利用者のトラスト構築に関して、図 1 に示すトラスト階層モデルとプロセスを定義している[3]。

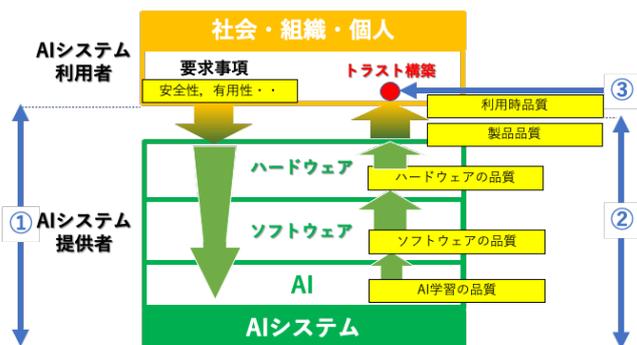


図 1：トラスト階層モデル

本トラスト階層モデルは、AI システムの実装に関連する専門家との議論を通して得た知見から、AI システムに要求される製品品質や利用時品質が AI の品質のみで満たせなくても、システムを構成するソフトウェアやハードウェアによって満たすことが可能であり、要求事項をハードウェア、ソフトウェア、AI の 3 つのレイヤの品質指標に落とし込めるとして 3 層に分けているのが特徴である。そして、本トラスト階層モデルでは、トラスト構築のプロセスを「利用時品質と製品品質からの落としこみ」「製品品質の確保」「トラスト構築」の 3 つに分けて、以下のように定義している。

- ① 利用時品質の製品品質への落としこみ（図 1 の左側の下矢印の流れ）
 利用者が要求事項（利用時品質）を提示することを起点として、提供者がこの要求事項を受けて求められるハードウェアやソフトウェア、AI の要件（製品品質）を規定する。
- ② 製品品質の確保（図 1 の右の上矢印の流れ）

AI の品質に基づいて、ソフトウェアやハードウェアを含む製品品質を確保する。これまでの AI の品質管理活動では AI の品質が利用時品質に直結し、ハードウェアやソフトウェアの品質も包括した利用時品質を考えるとという視点が十分でなかった。

- ③ トラスト構築（図 1 の利用者の赤丸）
 利用者が要求事項（利用時品質）に基づいて、提供される製品品質の総体を評価し、提供された製品品質で利用時品質が満たされた場合に導入を受け入れる。ここで、利用時品質について利用者・提供者が合意できていることが重要である。

AI システム提供者が保証する品質（製品品質）は、図 1 の緑色（AI システム）の部分において、AI システムの提供者や開発者の視点から検討されてきており、AI プロダクト品質保証コンソーシアム（QA4AI）から「AI プロダクト品質ガイドライン」[4]や産業技術総合研究所から「機械学習品質マネジメントガイドライン」[5]等のドキュメントが公開されている。

AI システム利用者から求められる品質（利用時品質）は、図 1 のオレンジ色（AI システム利用者）の部分において、「契約締結における AI 品質ハンドブック」[5]において、AI ソフトウェアを提供する提供者と自社のビジネスに導入する利用者の間で契約する場合にコミュニケーションが大切であることが述べられている。しかし、利用者が AI システムを利用してもよいというトラスト構築（図 1 の赤い点）の具体的な手法やプロセスについては記載されていない。このため、本論文では、「③トラスト構築」について、利用者が AI 誤判断リスクの受容の可否を判断するトラスト構築のプロセスを明らかにすることを試みる。

2.2 利用時品質

本論文では、利用時品質に関してはシステム及びソフトウェア製品の品質要求及び評価に関する国際規格である ISO/IEC 25000 シリーズ(SQuaRE : System and software Quality Requirement and Evaluation)を参照している。SQuaRE の利用時品質モデルを図 2 に示す。



図 2：利用時品質の図

利用時品質の要素としては、「有効さ」「効率」「満足性」

「利用状況網羅性」「リスク回避性」の5つが挙げられている。これら利用時品質に基づいて、利用者はシステムやソフトウェアの製品品質を評価して受け入れる。

利用時品質においてAIの判定精度が高くなれば、5つの要素の品質を一度に高めることになるため、利用者が100%の判定精度を求めることは十分考えられる。しかし、一般にAIの判定精度は100%にならず、どの程度の判定精度であれば利用時品質を満たしているかの評価手法や指標が必要になる。

利用時品質では、AI誤判断リスクをリスク回避性に関わる項目と位置付けることができる。リスク回避性の項目には、「経済リスク緩和性」「健康・緩和リスク緩和性」「環境リスク緩和性」の3つがあり、AIシステムの適用環境によって検討する項目や内容が異なる。例えば、エンドユーザー向けのAIスピーカーで、環境に影響がない利用範囲を想定するものであれば、環境リスクの緩和を考慮・検討しなくても良い。

AI誤判断リスクは、AIの判定精度に関わるため、他の4つの要素との関係も考慮する必要があると考える。例えば、SOC（セキュリティオペレーションセンター）では、分析者が見切れないIDSアラートが発生しており、AIシステムによってIDSアラートを分析者が見ることのできる範囲に絞り込んでくれば、誤判断（重要でないアラートを重要と判定）してしまう比較的低い精度であっても導入効果がある[2]。このようなAI誤判断リスクに係るリスク回避性に関する調査研究は知る限り見当たらない。

情報セキュリティにおいて利用時品質に係る研究としては、ユーザブルセキュリティがあり、セキュリティの担保やプライバシーの保護に関し、高いユーザビリティを持つ技術やプロセスが研究されている。現行のユーザブルセキュリティ研究は、利用時品質における満足性、効率性、実用性に注目しており^b、リスク回避性を網羅していないが、セキュリティシステムにも誤判断（例えば、IDSの誤検知）が存在するため、ユーザビリティという観点で誤判断リスクの評価が包含される可能性があると考えている。

3. トラスト構築

前述のとおり、トラスト構築の主体となる利用者に関して、リスクの捉え方や受容の考え方の違いから、個人、組織、社会の3つの類型に分類している。各類型のリスク受容のプロセスは以下のように仮定できる。

- ・ 個人：個人が誤判断リスクを把握して、リスクを受容できるかどうかを判断する。事前の誤判断リスクの把握は難しく、個人のAIに対する知見等で判断される場合が多いと考えられる。
- ・ 組織：リスクアセスメントに基づいて、組織としての

誤判断リスクを把握・受容することになる。リスクによる業務上の悪影響よりも有用性やコストダウン等のベネフィットが上回れば、利用すると判断に至る。誤判断に伴うコストをどう評価するか（例えば人手によるリカバリコスト）が焦点となる。

- ・ 社会：社会基盤サービスとしてAIシステムを導入する場合、誤判断リスクの社会的な影響把握や受容の合意が必要となる。一般的に社会的な合意に基づいて法律等のルールを整備し、それらルールの運用を強制することによって、リスク受容することになるため、合意形成に時間を要する。

また、筆者らはAIシステムの利用環境としてトラスト構築の主体が専有の環境（例えば、個人の家の敷地内等）とトラスト構築の主体以外も存在する共有の環境（例えば、道路等）の2つに分類している。主体が専有する環境であれば、AIシステムの利用にあたり、利用者のみがトラスト構築をすればよい。しかし、他者との共有する環境では、AIシステムのAI誤判断リスクのトラスト構築の際に、利用環境を共有する他者も考慮する必要がある。

筆者らは、AI誤判断リスクに関する利用者のトラスト構築プロセスについて、単純な個人、組織、社会の順番で取り組んでいこうと考えている。また、同様の理由から適用環境についても専有、共有の順番で取り組もうと考えている。

4. 本論文での主体

本論文では、主体を個人、適用環境を専有の場合のAI誤判断リスクのトラスト構築プロセスを対象とする。主体が個人である場合は、リスクの捉え方や許容に関して、個人がAI誤判断リスクを認知して許容できるかどうかを判断すると仮定している。そこで本論文ではまず、個人の心理的なAI誤判断リスクの認知・受容のプロセスについて明らかにするため、AIの誤判断リスクに関する認知と受容についてのアンケート調査を実施した。

本論文では、本アンケート調査の集計結果と基本的な分析結果の一部について速報の形で報告する。さらに、今後アンケート調査の分析を進める、リスク回避性品質との関係や、リスク評価手法と指標についての検討を進めていく。

5. 調査仮説

本論文では、AIシステム利用者を対象として、AI誤判断リスクの認知や受容の状況把握や、誤判断リスクの認知や受容に関係すると考える要因に関しての仮説を述べる。

5.1 誤判断リスクの認知と受容

AI誤判断リスクの認知や受容に関する調査研究に関する文献を調べたが、現時点で見つけることができていない。

^b この考えが正しいかはユーザブルセキュリティの有識者との間で精査が必要

このため、本論文では、AI 誤判断リスクの認知や受容の状況や検証すべき仮説を独自に設定し、すでにトラスト構築を経験したと思われる AI システム利用者を対象にアンケート調査を実施することとした。

今回の検証仮説としては、AI 誤判断リスクの認知や受容の状況に関して、個人の AI に対するイメージ、AI に関する知識や年齢がどう関わるかを 2 群比較により行うこととした。すなわち、5.2 節から 5.4 節の項目の回答者を 2 群に分けて、Mann-Whitney U 検定を行い、その 2 群における統計的有意差を確認して、誤判断リスクの認知や受容との関係を検証する。

5.2 AI を使った製品やサービスについての基本的なイメージ

AI に対するイメージの調査としては、消費者庁の「第 1 回消費者意識調査結果 (AI に対するイメージについて)」[7]が公開されている。消費者庁の調査では、サンプルに AI システムの利用経験あり/なしの両者が含まれた調査である。ただし、スクリーニングにおいて「人工知能という言葉を知らない」「AI を利用した製品・サービスの開発に従事している」と回答した回答者を本調査実施対象から除外。AI に対して「暮らしを豊かにする」「生活に良い影響を与える」というポジティブなイメージをもつ人が多く、また半数近くの回答者は「不安」「何となくこわい」というネガティブなイメージを持っているという結果であった。この調査から、AI に対してポジティブなイメージを持つ人と、ネガティブなイメージを持つ人でリスクの捉え方が違うのではないかと考えた。

そこで、AI システムの利用者 (トラスト構築の経験者) を対象として、AI を使った製品やサービスに関するイメージを調査し、イメージのポジティブな利用者とはネガティブな利用者で AI 誤判断リスクの認知や受容の状況が異なるとの仮説を立てた。

5.3 AI に関する知識

AI プロダクト品質保証コンソーシアム (QA4AI) が発行する AI プロダクト品質保証ガイドライン[4]では、AI の導入にあたり、AI の学習精度が 100%にならないことを導入する際に利用者に理解してもらわなければ、導入の障害になることを述べている。これを参考にして、筆者らは AI に関する知識の有無が AI システムの誤判断リスクに関しての認知や受容に影響すると仮説を立てた。

5.4 デジタルネイティブと非デジタルネイティブ世代

コンピュータやインターネット等の IT (情報技術) を利用しており、物心ついた頃から既にパソコンやスマホ、インターネットが当たり前であった世代 (デジタルネイティブ) とそうでない世代 (非デジタルネイティブ) では、AI の受け入れ方が異なるのではないかと仮説を立てた。

実際、第三次 AI ブームの前後の時期に出てきた SNS サービスである Instagram や TikTok は、若いほど利用者の割

合が多いことや、デジタルネイティブ世代以降かどうかで割合が異なるように見える。

デジタルネイティブ (digital native) には、いくつかの定義がある。本論文では デジタルネイティブは学生時代からインターネットやパソコンのある生活環境の中で育ってきた世代であるとし、40 歳未満をデジタルネイティブ、40 歳以上を非デジタルネイティブとした。

6. アンケート調査

今回のアンケート調査では、AI システム利用者を対象とした AI 誤判断リスクに関する状況を把握・分析のため、基本的な質問、AI 誤判断リスクが AI システムへの利用意図に与える影響を分析するための技術受容モデル (TAM) の質問、AI 誤判断の価値を測るためのコンジョイント分析の質問等を実施した。

調査対象者は、AI を利用した製品やサービスを「いつも使っている」「たまに使っている」「使ったことがある」と回答した AI システム利用者を対象とした。AI を利用とした製品やサービスとしては、「スマートフォンの音声アシスタント」「スマートフォンの手書き入力」「お掃除ロボット」「スマートスピーカー」としている。

アンケート調査は、調査会社が保有する Web アンケートシステムおよびモニタ会員を用いて実施した。今回 Web アンケート調査を用いた理由としては、AI を利用する製品やサービスの利用者を一定数確保するためである。

アンケート調査では、調査対象者であるかのスクリーニング調査を実施したのちに、その中から条件を満たす 1,000 人を抽出して本調査に回答してもらうという 2 段階の方式を採用している。調査期間は 2022 年 3 月 15 日から 3 月 17 日である。回答者の構成は表 1 に示すように、性別と年齢に偏りがないように等サンプルを取るようにした。また、デジタルネイティブ、非デジタルネイティブの割合が同じになるように、40 歳未満と 40 歳以上で等サンプルとなるようにしている。

表 1: 回答者の年齢・性別の構成

| | 10・20代 | 30代 | 40代 | 50代以上 |
|----|--------|-----|-----|-------|
| 男性 | 125 | 125 | 125 | 125 |
| 女性 | 125 | 125 | 125 | 125 |

7. 集計・分析結果

7.1 アンケートの質問項目と集計結果

7.1.1 AI 誤判断リスクの認知

AI 誤判断リスクの認知の状況について、図 3 に示す 5 つの質問項目に AI システム利用者がリッカード尺度 5 段階で回答した結果に示す。

図 3 の 5 つの質問項目のうち、前半の 2 問は AI 誤判断リスクに関する全般的な設問である。1 問目は、一般的に

AIは学習しながら精度を高めていくことから、その過程でAI誤判断が発生するため、このAI誤判断リスクをどう捉えるかに関する設問である。2問目は、AIの判定精度は100%にならないため、このAI誤判断リスクに対して不安を感じているかに関する設問である。そして、後半の3問目は、AIスピーカー、お掃除ロボット、自動運転といったAIシステムにおいて、AIが判断を誤った場合の影響を想像できるとリスクを特定できることから、AIシステムでAIが判断を誤った場合に自分自身への影響を想像できるかに関する設問である。これらの設問は筆者ら独自に作成したものである。

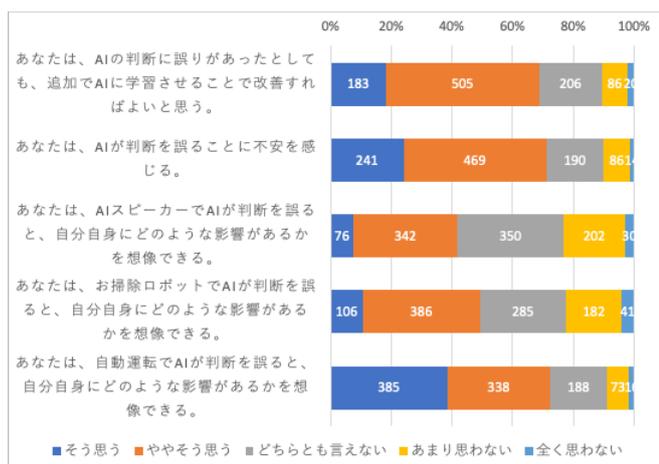


図 3 : AI 誤判断リスクの認知に関するアンケート結果

「AIの判断に誤りがあったとしても、追加でAIに学習させて改善すればよいと思う」は、「そう思う」「ややそう思う」が68.8%、一方「全く思わない」「あまり思わない」が10.6%であった。「AIが判断を誤ることに不安を感じる」は、「そう思う」「ややそう思う」が71%、一方「全く思わない」「あまり思わない」が10%であった。

AIスピーカー、お掃除ロボット、自動運転に関して、AIが判断を誤ったときに、自分自身にどのような影響があるか想像できるかに関して以下の結果となった。AIスピーカーは、「そう思う」「ややそう思う」が41.8%、一方「全く思わない」「あまり思わない」が23.2%であった。お掃除ロボットは、「そう思う」「ややそう思う」が49.2%、一方「全く思わない」「あまり思わない」が22.3%であった。自動運転は、「そう思う」「ややそう思う」が72.3%、一方「全く思わない」「あまり思わない」が8.9%であった。

7.1.2 AI 誤判断リスクの受容

AI誤判断リスクの受容状況について、具体的な間違いのケースを5つ挙げて受容できるかどうかについて、「下記のAIのサービスをあなたが利用していることを前提として、AIの判断の誤りの受け入れに関してあなたの考えに最も近いものを1つずつ選んでください。」とのリード文に続いて、図4に示す5つの質問項目にAIシステム利用者がリッカート尺度5段階で回答した結果を示す。

図4の5つの質問項目のうち、前半の4つの質問項目は、音声認識、画像認識、お掃除ロボット、自動運転といったAIシステムの具体的な誤りを示し、この誤りをAIシステム利用者が許容できるかに関する設問である。5問目は、AIの精度の期待として、AIが人よりも判定精度が低い場合に、それを許容できるかの設問である。これらの設問は筆者らが独自に作成したものである。

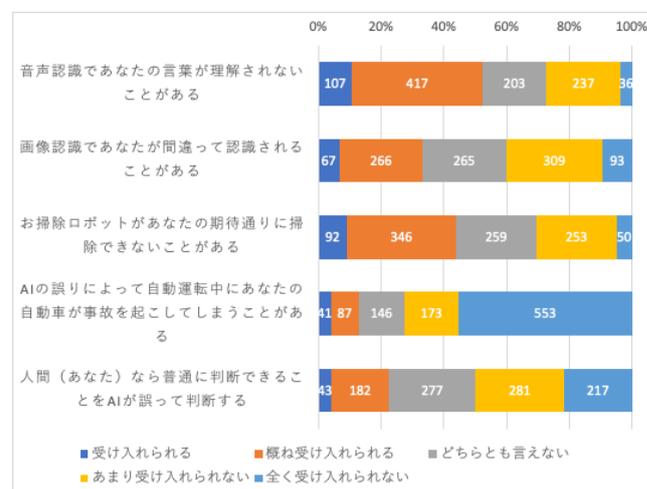


図 4 : AI 誤判断リスクの受容に関するアンケート結果

音声認識の誤りは、「受けられる」「概ね受け入れられる」が52.4%、一方「全く受け入れられない」「あまり受け入れられない」が27.3%であった。画像認識の誤りは、「受けられる」「概ね受け入れられる」が33.3%、一方「全く受け入れられない」「あまり受け入れられない」が40.2%であった。お掃除ロボットの誤りは、「受けられる」「概ね受け入れられる」が43.8%、一方「全く受け入れられない」「あまり受け入れられない」が30.3%であった。自動運転の誤りは、「受けられる」「概ね受け入れられる」が12.8%、一方「全く受け入れられない」「あまり受け入れられない」が72.6%であった。

「人間なら普通に判断できることをAIが誤って判断する」は、「受けられる」「概ね受け入れられる」が22.5%、一方「全く受け入れられない」「あまり受け入れられない」が49.8%であった。

7.1.3 AI の製品やサービスの利用に関するイメージ

AIを使った製品やサービスの利用についてどのようなイメージで捉えているかに関して、「下記のAIの利用に関する意見に対して、あなたの考えに最も近いものを1つずつ選んでください。」とのリード文に続いて、図5に示す5つの質問項目にシステム利用者がリッカート尺度5段階でAI回答した結果を示す。図5の質問項目は、AIを使った製品やサービスの利用について、有用性、不安、補償、心地よさ、利便性に関するイメージを問うもので、消費者庁のAIに対するイメージについての調査[7]や、Web上に調査会社等の公開している調査の質問内容を参考に、筆者ら

が独自に作成した。

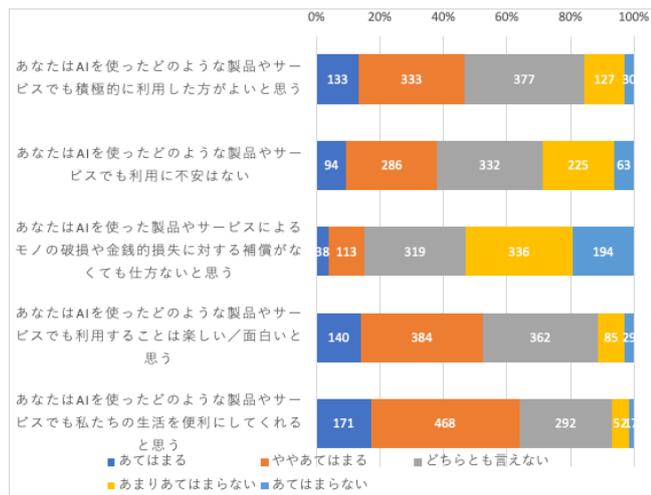


図 5 : AI の製品やサービスの利用に関するイメージ

「積極的に利用した方がよいと思う」(有用性)は、「そう思う」「ややそう思う」(ポジティブなイメージ)が46.6%、一方「全く思わない」「あまり思わないが」(ネガティブなイメージ)が15.7%であった。

「利用に不安がない」(不安)は、「そう思う」「ややそう思う」が38%、一方「全く思わない」「あまり思わないが」が28.8%であった。

「モノの破損や金銭的損失に対する補償がなくても仕方ない」は、「そう思う」「ややそう思う」が15.1%、一方「全く思わない」「あまり思わないが」が53.0%であった。

「利用することは楽しい／面白いと思う」(心地よさ)は、「そう思う」「ややそう思う」が52.4%、一方「全く思わない」「あまり思わないが」が10.4%であった。

「生活を便利にしてくれると思う」(利便性)は、「そう思う」「ややそう思う」が63.9%、一方「全く思わない」「あまり思わないが」が6.9%であった。

今回、5つの設問のうち、リスクの認知や受容においては、一番ネガティブなイメージのものが関係すると考え、「モノの破損や金銭的損失に対する補償がなくても仕方ない」についてポジティブとネガティブな回答者の群で比較することにした。

7.1.4 AIに関する知識レベル

AIに関する知識レベルについて、AIの技術的な観点からのクイズをAIシステム利用者に5問出題し、1問の正解を1点として0点から5点までの人数の分布を図6に示す。クイズはインターネットからAIの技術に関するものを参考にして作成した(付録A.1)。「正しい」「間違っている」「わからない」の選択から回答するクイズであり、「わからない」は不正解とした。

AIクイズの得点の結果は、平均点は2.561点であり、最頻値は3点であった。この結果からAIの知識レベルが高い(3点以上:538人)と知識低い(2点以下:462人)の

2群に分けることにした。

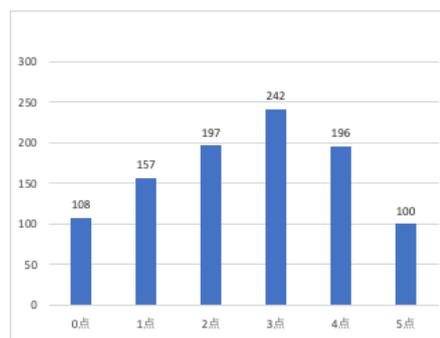


図 6 : AI クイズの得点の結果

7.2 アンケートの分析結果

7.2.1 AI 誤判断のリスク認知

AIシステム利用者のリスク認知に係る質問項目に関して、AI誤判断の損害への補償、AIに関する知識レベル、デジタルネイティブの項目における2群にMann-WhitneyのU検定を行った結果を表2から表4に示す。

- ・ AI誤判断の損害への補償の必要/不要

「モノの破損や金銭的損失に対する補償がなくても仕方ない」に関して、補償が不要(「そう思う」「ややそう思う」と補償が必要「全く思わない」「あまり思わないが」)の2群について比較した。

表 2 : 補償の必要の必/不の2群のU検定結果

| | | | | | |
|-----------|---|-----|----------|--------|----------------------------|
| AI 誤判断は追加 | 必 | 288 | 78294 | 96336 | $z = -7.897$ |
| 学習すれば良い | 不 | 380 | 145152 | 127110 | $\text{Prob}> z = 0.0000$ |
| AI 誤判断に不安 | 必 | 288 | 111825 | 96336 | $z = 6.694$ |
| を感じる | 不 | 380 | 111621 | 127110 | $\text{Prob}> z = 0.0000$ |
| AI スピーカーで | 必 | 288 | 84869 | 96336 | $z = -4.857$ |
| 影響を想像可能 | 不 | 380 | 138577 | 127110 | $\text{Prob}> z = 0.0000$ |
| お掃除ロボットで | 必 | 288 | 88537 | 96336 | $z = -3.309$ |
| 影響を想像可能 | 不 | 380 | 134909 | 127110 | $\text{Prob}> z = 0.0009$ |
| 自動運転で影響 | 必 | 288 | 94647.5 | 96336 | $z = -0.730$ |
| を想像可能 | 不 | 380 | 128798.5 | 127110 | $\text{Prob}> z = 0.4654$ |

5つの設問の全てで、5%水準で統計的に有意となっているため、5つの設問において、補償が必要な回答者と不要な回答者で差があることがわかる。

- ・ AIに関する知識レベル

7.1.4節に記載した知識の低い利用者と知識の高い利用者の2群について比較した。

表 3 : AIの知識レベルの低/高の2群のU検定結果

| | | | | | |
|-----------|---|-----|--------|--------|----------------------------|
| AI 誤判断は追加 | 低 | 462 | 216985 | 231231 | $z = -3.382$ |
| 学習すれば良い | 高 | 538 | 283515 | 269269 | $\text{Prob}> z = 0.0007$ |
| AI 誤判断に不安 | 低 | 462 | 221782 | 231231 | $z = -2.218$ |
| を感じる | 高 | 538 | 278718 | 269269 | $\text{Prob}> z = 0.0266$ |
| AI スピーカーで | 低 | 462 | 201188 | 231231 | $z = -6.922$ |
| 影響を想像可能 | 高 | 538 | 299312 | 269269 | $\text{Prob}> z = 0.0000$ |

| | | | | | |
|-----------------|---|-----|----------|--------|----------------------------|
| お掃除ロボットで影響を想像可能 | 低 | 462 | 195000.5 | 231231 | $z = -8.332$ |
| | 高 | 538 | 305499.5 | 269269 | $\text{Prob}> z = 0.0000$ |
| 自動運転で影響を想像可能 | 低 | 462 | 197824.5 | 231231 | $z = -7.745$ |
| | 高 | 538 | 302675.5 | 269269 | $\text{Prob}> z = 0.0000$ |

5つの設問の全てで、5%水準で統計的に有意となっているため、5つの設問において、AIに関する知識レベルが高い回答者と低い回答者と差があることがわかる。

- デジタルネイティブ・非デジタルネイティブ

5.4節に記載したデジタルネイティブ（低年齢（40歳未満））と非デジタルネイティブ（高年齢（40歳以上））2群について比較した。

表4：年齢の低/高の2群のU検定結果

| | | | | | |
|-----------------|---|-----|----------|--------|----------------------------|
| AI誤判断は追加学習すれば良い | 低 | 500 | 248229.5 | 250250 | $z = -0.478$ |
| | 高 | 500 | 252270.5 | 250250 | $\text{Prob}> z = 0.6324$ |
| AI誤判断に不安を感じる | 低 | 500 | 245365.5 | 250250 | $z = -1.143$ |
| | 高 | 500 | 255134.5 | 250250 | $\text{Prob}> z = 0.2529$ |
| AIスピーカーで影響を想像可能 | 低 | 500 | 249169 | 250250 | $z = -0.248$ |
| | 高 | 500 | 251331 | 250250 | $\text{Prob}> z = 0.8039$ |
| お掃除ロボットで影響を想像可能 | 低 | 500 | 252023.5 | 250250 | $z = 0.407$ |
| | 高 | 500 | 248476.5 | 250250 | $\text{Prob}> z = 0.6843$ |
| 自動運転で影響を想像可能 | 低 | 500 | 248698 | 250250 | $z = -0.359$ |
| | 高 | 500 | 251802 | 250250 | $\text{Prob}> z = 0.7198$ |

5つの設問の全てで、5%水準で統計的に有意となっていないため、5つの設問において、年齢が高い回答者と低い回答者と差がないことがわかる。

7.2.2 AI誤認識リスクの受容

リスクの受容に係る質問項目に関して、AI誤判断の損害への補償、AIに関する知識レベル、デジタルネイティブの項目における2群にMann-WhitneyのU検定の結果を表5から表7に示す。

- AIシステムの損害への補償の有/無

「モノの破損や金銭的損失に対する補償がなくても仕方がない」に関して、補償が不要（「そう思う」「ややそう思う」と補償が必要「全く思わない」「あまり思わないが」）の2群について比較した。

表5：補償の必要の必/不の2群のU検定結果

| | | | | | |
|-----------------|---|-----|----------|--------|----------------------------|
| 音声認識のAI誤判断を許容 | 必 | 288 | 84896 | 96336 | $z = -4.901$ |
| | 不 | 380 | 138550 | 127110 | $\text{Prob}> z = 0.0000$ |
| 画像認識AI誤判断を許容 | 必 | 288 | 83861.5 | 96336 | $z = -5.220$ |
| | 不 | 380 | 139584.5 | 127110 | $\text{Prob}> z = 0.0000$ |
| お掃除ロボットAI誤判断を許容 | 必 | 288 | 88301 | 96336 | $z = -3.378$ |
| | 不 | 380 | 135145 | 127110 | $\text{Prob}> z = 0.0007$ |
| 自動運転AI誤判断を許容 | 必 | 288 | 87411 | 96336 | $z = -4.027$ |
| | 不 | 380 | 136035 | 127110 | $\text{Prob}> z = 0.0001$ |
| 人が判断できることを誤るを許容 | 必 | 288 | 83414 | 96336 | $z = -5.382$ |
| | 不 | 380 | 140032 | 127110 | $\text{Prob}> z = 0.0000$ |

5つの設問の全てで、5%水準で統計的に有意となっている

ため、5つの設問において、補償が必要な回答者と不要な回答者と差があることがわかる。

- AIの知識の低い/高いによるU検定結果

7.1.4節に記載した知識の低い利用者と知識の高い利用者の2群について比較した。

表6：AIの知識レベルの低/高の2群のU検定結果

| | | | | | |
|-----------------|---|-----|----------|--------|----------------------------|
| 音声認識のAI誤判断を許容 | 低 | 462 | 217212 | 231231 | $z = -3.237$ |
| | 高 | 538 | 283288 | 269269 | $\text{Prob}> z = 0.0012$ |
| 画像認識AI誤判断を許容 | 低 | 462 | 222109 | 231231 | $z = -2.075$ |
| | 高 | 538 | 278391 | 269269 | $\text{Prob}> z = 0.0380$ |
| お掃除ロボットAI誤判断を許容 | 低 | 462 | 213507.5 | 231231 | $z = -4.049$ |
| | 高 | 538 | 286992.5 | 269269 | $\text{Prob}> z = 0.0001$ |
| 自動運転AI誤判断を許容 | 低 | 462 | 236662.5 | 231231 | $z = 1.316$ |
| | 高 | 538 | 263837.5 | 269269 | $\text{Prob}> z = 0.1883$ |
| 人が判断できることを誤るの許容 | 低 | 462 | 220162.5 | 231231 | $z = -2.507$ |
| | 高 | 538 | 280337.5 | 269269 | $\text{Prob}> z = 0.0122$ |

4つの設問（自動運転を除く）で、5%水準で統計的に有意となっているため、4つの設問において、AIに関する知識レベルが高い回答者と低い回答者と差があることがわかる。

- デジタルネイティブ・非デジタルネイティブ

5.4節に記載したデジタルネイティブ（低年齢（40歳未満））と非デジタルネイティブ（高年齢（40歳以上））2群について比較した。

表7：年齢の低/高の2群のU検定結果

| | | | | | |
|-----------------|---|-----|----------|--------|----------------------------|
| 音声認識のAI誤判断を許容 | 低 | 500 | 266328 | 250250 | $z = 3.702$ |
| | 高 | 500 | 234172 | 250250 | $\text{Prob}> z = 0.0002$ |
| 画像認識AI誤判断を許容 | 低 | 500 | 276664 | 250250 | $z = 5.992$ |
| | 高 | 500 | 223836 | 250250 | $\text{Prob}> z = 0.0000$ |
| お掃除ロボットAI誤判断を許容 | 低 | 500 | 257065.5 | 250250 | $z = 1.553$ |
| | 高 | 500 | 243434.5 | 250250 | $\text{Prob}> z = 0.1205$ |
| 自動運転AI誤判断を許容 | 低 | 500 | 262310 | 250250 | $z = 2.913$ |
| | 高 | 500 | 238190 | 250250 | $\text{Prob}> z = 0.0036$ |
| 人が判断できることを誤るを許容 | 低 | 500 | 272483 | 250250 | $z = 5.021$ |
| | 高 | 500 | 228017 | 250250 | $\text{Prob}> z = 0.0000$ |

4つの設問（お掃除ロボットを除く）で、5%水準で統計的に有意となっているため、4つの設問において、年齢が低い回答者（デジタルネイティブ）と高い回答者（非デジタルネイティブ）で差があることがわかる。

8. 考察

8.1 AIシステム利用者のAI誤認識リスクの認知

図3の結果に基づいて、AI誤認識リスクの認知に係る状況や仮説の分析結果を考察する。

AIの判断に誤りがあったとしても、追加でAIに学習させて改善すればよいと思っている回答者が68.8%であり、誤りを学習で改善することは、多くのAIシステム利用者

に受け入れられると考える。しかし、図4の5問目の「人間なら普通に判断できることをAIが誤って判断する」において、受容可能な回答者は22.5%であったことから、AIシステム提供時には人間よりも高い精度であることが求められると考える。

AIシステムでAIが判断を誤った場合に自分自身への影響を想像できるかに関しては、AIスピーカー、お掃除ロボット、自動運転の順で影響を想像できる割合が高く、自動運転、お掃除ロボット、AIスピーカーの順でリスクを特定できる割合が高いと考える。

AIが判断を間違えることを不安に思う回答者は71%と高い割合である。しかし、図5の2問目の「AIを使ったどのような製品やサービスでも利用に不安がない」で不安に思う回答者は28.8%低いことから、今後AI誤判断リスクへの不安が、AIシステム利用の不安にどのような影響があるのかを分析したいと考えている。

8.2 AIシステム利用者のAI誤認識リスクの受容

図4の結果に基づいて、AI誤認識リスクの受容に係る状況や仮説の分析結果について考察する。

音声認識、画像認識、お掃除ロボット、自動運転における誤判断に関して、受容できる（「受けられる」「概ね受け入れられる」と受容できない（「全く受け入れられない」「あまり受け入れられない」）の割合それぞれ異なるため、AIシステムによって受容状況が異なると思われる。また、自動運転は、他の3つと比較して受容できない割合が非常に高い。この結果はAIシステムの普及状況からAIスピーカーと比べて自動運転のような人命に係るリスクの方が受容されにくいとの推測[2]を裏付けるものとなった。

AI誤認識リスクの受容に関して、仮説の結果は、AIが誤った際の補償の不要/必要では全て、AIに関する知識の有/無では4つ（自動運転を除く）、デジタルネイティブ/非デジタルネイティブ（お掃除ロボットを除く）の2群において差が見られたため、これらに関してAI誤判断リスクの受容への影響についての分析を進めていきたいと考えている。また、デジタルネイティブと非デジタルネイティブでは、図3の認知の質問項目で差がなかったが、図4の受容の質問項目で差があったことから、今後図3の認知と図4の受容の関係性を分析したいと考えている。

9. おわりに

本論文では、AIシステムの利用者を対象として実施したアンケート調査の初期の報告として、AIを使った製品やサービスに関してAIの誤判断リスクの認知と受容に関しての集計結果や仮説に基づく分析結果について述べた。

今後、AI誤判断リスクがAIシステムへの利用意図に与える影響を分析するための技術受容モデル(TAM)の質問項目、AI誤判断の価値を測るためのコンジョイント分析の質問項目等に関しての分析結果を報告する予定である。

また、現在、利用時品質の標準化作業でも、「利用者」の定義の精緻化のための検討が行われている[9]。利用時品質では、これまで利用者をAIシステムの直接的な操作者をイメージしてきた。さらに運用者や間接的な利用者等のステークホルダーも含めて利用者として利用時品質を論じる必要があると言われている。このため、利用者としての主体を個人以外の組織や社会にも広げるにあたり、今後利用時品質の標準化作業も参考にして研究を進める。

参考文献

- [1] 小川 隆一, 島 成佳. “機械学習システムのトラスト構築に関する課題分析”. 情報処理学会 2019-SPT-32 巻 21 号 P.1-P.6.
- [2] 島 成佳, 小川 隆一. “機械学習システムのトラスト構築に関する課題分析 (2) ～期待性能に関する考察～”. コンピュータセキュリティシンポジウム 2019 論文集 P.1100-P.1107.
- [3] 島成佳, 小川隆一, 佐川陽一. “AIシステムの利用者視点からのトラスト構築の考察”. 2022年暗号と情報セキュリティシンポジウム (SCIS2022).
- [4] AIプロダクト品質保証コンソーシアム (QA4AI). “AIプロダクト品質保証ガイドライン 2020.08 版”. 2020年8月1日. <http://www.qa4ai.jp/QA4AI.Guideline.202008>, (参照 2022-04-10).
- [5] 国立研究開発法人産業技術総合研究所. “機械学習品質マネジメントガイドライン第2版”. 2021年7月5日. <https://www.digiarc.aist.go.jp/publication/aiqm/AIQM-Guideline-2.1.0.pdf>, (参照 2022-04-10).
- [6] 一般社団法人日本ディーラーニング協会「契約締結におけるAI品質保証の在り方」研究会. “契約締結におけるAI品質ハンドブック”. 2021年7月 https://www.jdla.org/wp-content/uploads/2021/08/jdla2021_handboo, (参照 2022-04-10).
- [7] 消費者庁. “第1回消費者意識調査結果 (AIに対するイメージについて)”. https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/consumer_policy_cms101_20316_03.pdf(参照 2022-04-10).
- [8] 木村忠正. “デジタルネイティブの時代 なぜメールをせずに「つぶやく」のか”. 平凡社(平凡社新書). 2012.
- [9] 福住 伸一. “利用時品質規格の動向”. 日本人間工学会 2021.

付録 A.1 アンケート調査の質問項目

(1) AIの知識を問うクイズ

以下に、アンケート回答者に出題したクイズを示す。

- Q4-1. AIは、近年ディーラーニング等の機械学習という技術で注目されている。(1つ選択)
選択肢：正しい 間違っている わからない
- Q4-2. AIは、誤判断することがある。(1つ選択)
選択肢：正しい 間違っている わからない
- Q4-3. AIを賢くするには、学習データが必要であり、学習データは量より質が重要である。(1つ選択)
選択肢：正しい 間違っている わからない
- Q4-4. AIでは、用途に応じて最適な技術(アルゴリズム等)が異なる。(1つ選択)
選択肢：正しい 間違っている わからない
- Q4-5. AIを賢くする際には、偏った学習データで学習させても自動的に補正される。(1つ選択)
選択肢：正しい 間違っている わからない