AI 誤判断による価値損失の定量的評価

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するために，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．本論文では，架空の AI サービスの誤判断に関するアンケート調査を行い，コンジョイント分析を用いて AI 誤判断による価値損失（AI 誤判断による不具合の影響）の定量的な評価を試み，それらの結果について考察する．

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，利用時品質

Quantitative Measurement of the Decline in Quality in Use due to AI Misjudgment

SHIGEYOSHI SHIMA[1,2]　RYUICHI OGAWA[2]
YOUICHI SAGAWA[2]　TOSHIHIKO TAKEMURA[3]

Abstract: There is a risk of AI misjudgment in AI systems (IT systems equipped with an AI learning model) because the judgement accuracy of AI learning models usually does not reach 100%. The risk can adversely affect AI system users in building trust with the systems by reducing the systems' quality in use, which AI system users expect when they use the systems. Our research goals are to analyze the process of building the trust with AI systems and to develop a measurement method of the AI misjudgment risk. This paper describes the result of questionnaire regarding drawbacks of AI systems due to misjudgment and its conjoint analysis.

Keywords: AI System, Misjudgment Risk, Risk Perception, Risk Acceptance, Building Trust, Quality in Use

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%にならず，誤判断や誤認識してしまうことが知られている[1]．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 システムの導入では，利用者が AI 誤判断リスクの受容の可否を判断できず，導入に至らないケースが見られる．筆者らは，AI システムの導入を容易にするために，AI 誤判断リスクに関して「リスクを許容して AI システムを利用してもよい」という，利用者の AI システム導入に対するトラスト構築の重要性に注目している．これまでに，トラスト構築を包括的に表すトラスト階層モデルと，トラスト構築プロセスを提唱している[2][3][4]．

これらのトラスト階層モデルとトラスト構築プロセスでは，AI 誤判断リスクへの対応を製品品質の一部として捉えている．また，トラスト構築の主体となる「利用者」を AI 誤判断リスクの捉え方や許容の考え方が異なる個人，組織，社会の 3 つの類型に分類している．個人，組織，社会の順に，利害を異にする関係者や考慮すべき項目・制度等が増えることでトラスト構築のプロセスは複雑になり，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している．

現在筆者らは主体に関してトラスト構築プロセスが一番容易であると考える個人をケースに取り組んでいる．まず，個人の AI 誤判断リスクの認知や受容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．このアンケート調査は，基礎的な AI 誤判断リスクの認知や受容に関する質問項目や，AI システムの利用意図へのリスク認知や受容の影響を分析するための TAM（Technology

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

Acceptance Model）に基づく質問項目，AI 誤判断による価値損失を定量的に評価するコンジョイント分析の質問項目で構成される．このうち，リスクの認知や受容に関するアンケート調査を分析した結果は報告済みである[5]．本論文では，上記のアンケート調査のうち，コンジョイント分析の質問項目を分析して，AI 誤判断の損失価値を定量化した結果とその考察について述べる．

2. トラスト階層モデル

筆者らは，AI システム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提唱している[4]．

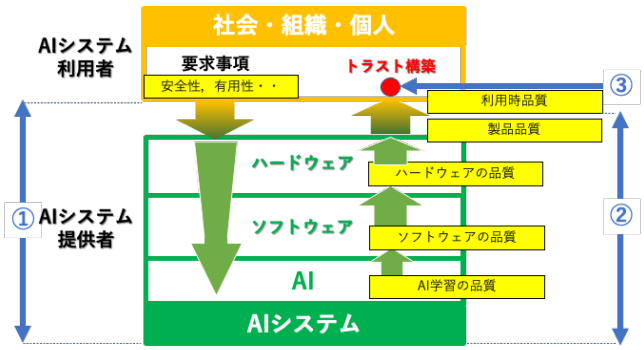

図 1：トラスト階層モデル

Figure 1：Trust Hierarchy Model

トラスト階層モデルでは，トラスト構築プロセスを図 1 の①～③の 3 つのプロセスに分けている．

① 利用時品質の製品品質への落としこみ（図 1 の左の下向きの矢印）

② 製品品質の確保（図 1 の右の上向きの矢印の流れ）

③ トラスト構築（図 1 の利用者の赤丸）

本トラスト階層モデルの特徴は，AI システムに要求される製品品質や利用時品質が図 1 の「AI 学習の品質」で満たせなくても，「ソフトウェアの品質」「ハードウェアの品質」によって満たせるよう対応可能であるとしたところである．このため，利用者の要求事項は，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．

トラスト構築プロセスでは，AI システム利用者が自身の要求事項に基づき，AI システム提供者の提供する利用時品質が満たされていると判断して利用する場合，トラストが構築される．トラスト構築の主体となる利用者は，利用におけるリスクに対応する単位の大きさ（係る人数や利用範囲等）の違いから，「個人」「組織」「社会」の 3 つの類型に分類される．

筆者らは，トラスト構築プロセスが比較的シンプルであると考える個人のケースから取り組んでいる．まず以下のようなプロセス，環境を想定する．

- リスクの認知と受容：個人が AI 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 AI 誤判断リスクの把握は一般に難しく，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る．
- 適用環境：個人が AI システムを利用する環境は専有，共有に分類できる．本論文では検討が容易な専有をターゲットとする．

3. AI 誤判断によるサービス価値損失の分析

筆者らは，トラスト階層モデルの利用時品質について，ソフトウェア品質の国際標準である SQuaRE（IS0/IEC25010, JIS X25010）を参照している．SQuaRE は，利用時品質として，5 つの項目（「リスク回避性」「有効性」「効率性」「満足性」「利用状況網羅性」）を定義している．筆者らは，これら 5 つの項目のうち，リスク回避性が AI 誤判断リスクに関係する項目と捉えている．

リスク回避性では，以下の「経済リスク緩和性」「健康・安全リスク緩和性」「環境リスク緩和性」の 3 つ項目が挙げられている[6]．

- 経済リスク緩和性
  意図した利用状況において，財政状況，効率的運用操作，商業資産，評判又は他の資源に対する潜在的なリスクを，製品又はシステムが緩和する度合い．例えば，動作の停止や精度の低減等による影響を防止する配慮を行うことである．
- 健康・安全リスク緩和性
  意図した利用状況において，製品又はシステムが人々の安全等に対する潜在的なリスクを緩和する度合い．例えば，人命や健康等に係る被害を低減する配慮を行うことである．
- 環境リスク緩和性
  意図した利用状況において，環境に対する潜在的なリスクを製品又はシステムが軽減する度合い．例えば，環境汚染，騒音等の防止や，省エネルギー化等を配慮することである．

リスク回避性では，AI システムや AI システムを用いたサービスで顕在化したリスクに対して，これら 3 項目の観点から緩和策を検討する．このためには，どのようなリスクがあるかの洗い出し，それらのリスクを評価して受容可能なレベルになるようにリスク緩和策を講じる必要がある．

4. 関連研究

顕在化したリスクは，リスクの影響や損失を定量的に評価する際に金額として表すことが一般的である．セキュリティリスクについては，特に個人情報漏えいの経済損失の定量化に関して以下の研究がある．

日本ネットワークセキュリティ協会（JNSA）のセキュリティ被害調査ワーキンググループでは，個人情報の漏洩に

関して，顧客一人当たりへの想定損害賠償額を算出する JO モデル（JNSA Damage Operation Model for Individual Information Leak）を提唱している[7]. この JO モデルでは，漏洩個人情報価値の算出式を式 1 で示している．

（式 1）

漏洩個人情報価値
＝基礎情報価値×機微情報度×本人特定容易度

式 1 は，「入力項目決定」「入力値定量化」「専門家の助言」「算定式の策定」の手順を経て作成されている．「専門家の助言」では，弁護士などの専門家の意見を取り入れて，算出精度を高めている．

右辺の 3 項目の入力値は，以下のように定義されている．

- 基礎情報価値：基礎値として一律 500 ポイントを設定．
- 機微情報度：個人情報を 3 段階のレベルに設定し，その値からセンシティブの度合いを算定し，機微情報を算出．
- 本人特定容易度：定義した本人特定容易度判定基準に基づき決定．

本論文にて，JO モデルを参考にする場合は，AI 誤判断による損失を算出する入力項目を決定し，入力項目の入力値の定量化を行い，算出式を作成する必要がある．式 1 の適用する際には，「AI 誤判断の損失において情報漏洩は部分的なもの（他の損失もありうる）」であるため，そのまま適用することは難しい．また，JO モデルに倣って新たな定式化を行うとした場合，SQuaRE の利用時品質を要素とする定式化が必要であると想定されるが，妥当性検証に時間を要すると思われることから今回は採用しないこととする．

個人情報の漏洩被害額の算出の研究としては，過去の損失額に基づいた統計的な算出手法がある．Romanosky らは，2005 年から 2014 年の間に米国の企業で発生した 11,705 件のインシデント情報に基づいて，企業が毎年に被った総コストを算出する算出式（式 2）を提案している[8].

（式 2）

$$\log(cost_{i,t}) = \beta_0 + \beta_1 \cdot \log(revenue_{i,t}) + \beta_2 \cdot \log(records_{i,t}) + \beta_3 \cdot repeat_{i,t} + \beta_4 \cdot malicious_{i,t} + \beta_5 \cdot lausuit_{i,t} + \alpha \cdot FirmType_{i,t} + \lambda_t + \rho_{ind} + \mu_{i,t}$$

また，山田らの研究では，企業が発表している決算の情報から，インシデント発生企業で生じた損害額の情報を収集し，114 件のインシデント情報に対して重回帰分析のモデルを提案している[9]. この重回帰分析モデルでは，「決算短信」に記載されている「特別損失額」に計上されている「セキュリティ対策費」から特別損失額との関係を分析した漏洩損失額を目的変数として設定している．

本論文にて，これらの統計的手法を適用する場合には，過去に発生した AI 誤判断によって発生した損失事例を集めなければならない．しかし，AI 誤判断による損失事例のデータは集まっておらず，これらの研究を参考にすることは現実的でない．

市場では取引されない漏えいしたプライバシー情報（基本的な情報）の価値に関して，金銭的な実被害額を全額補償した上で支払われる慰謝料をどの程度要求するかというシナリオ等に基づいてアンケート調査を行い，精神的被害や実被害等の金銭的価値を評価する研究がある[10][11]. これらの研究で用いられているコンジョイント分析は，市場に出る前の（様々な属性をもつ）製品やサービスなど，消費者の評価がわからないものについて，架空のサービスシナリオを想定してそのサービスの属性の定量的評価をする手法の 1 つであり，マーケティングの分野でよく用いられている．論文[10][11]は，利用者が自らのプライバシー情報の価値や情報漏洩した企業の事後対応などの価値について測定を試みている．コンジョイント分析は利用者の評価であることから，上述した研究（JNSA のモデル，Romanosky らや山田ら手法）とはアプローチが大きく異なっており，上述した研究で示した課題もクリアする．このため，本論文にて，このコンジョイント分析を用いる手法を採用することにした．

5. 調査設計と分析方法

本調査では，AI システムを用いた架空サービスを想定し，そのサービス中で AI 誤判断によって損失が起こるシナリオに基づき，AI 誤判断の頻度・影響・受容できる対応コストに関するアンケート調査を実施する．そして，コンジョイント分析の結果から AI 誤判断リスクを含むサービスの限界支払意思額（支払ってよいと思う対価の最大値）を求め，金額による定量的な評価を行う．

5.1 コンジョイント分析に関わる調査内容

本調査では，AI 誤判断による価値低下に関して以下の項目を試算する．

（1） AI 誤判断による価値低下の試算

AI 誤判断の評価に，深刻度と頻度を設定する．深刻度は異なる 2 種類の AI 誤判断のケースを想定し，頻度は 3 段階想定する．コンジョイント分析を行なった結果から，限界支払意思額を算出して，金額的な観点から定量的な影響の大きさや，AI 誤判断を減らすことでのリスク低減効果について考察する．

（2） AI 誤判断への補償の価値の客観的な試算

（1）の AI 誤判断による価値低下に対する緩和策として，金銭的な損失を補償する保険を設定し，コンジョイント分析を行った結果から限界支払意思額を算出し，リスク移転の効果について考察する．

（3） AI 誤判断リスクの認知と受容の程度の異なる被験者グループの試算結果の比較

筆者らは，論文[5]において，被験者をその属性に基づいて 2 群に分けて分析し，リスクの認知や受容に関して，U 検定において統計的な有意な差が見られるグルーピングを調査した．本論文では，有意差が見られた以下の二

種類のグルーピングに関して（1）と（2）の分析を行って比較する．

(3-1) デジタルネイティブと非デジタルネイティブ

論文[5]と同様に，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ートフォン，インターネット利用が当たり前にあった世代（デジタルネイティブ）とそうでない世代（非デジタル非ネイティブ）の2群に分ける．デジタルネイティブは 40 歳未満とし，非デジタルネイティブは40歳以上とした．

(3-2) AIに関する知識レベル

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[1]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，導入の阻害になると述べている．この影響を評価するため，AI に関する知識のある人とない人の 2 群に分けた．AIに関する知識レベルは，アンケート調査で出題したクイズの正解数によって判定した．

5.2 コンジョイント分析の流れ

コンジョイント分析は，以下の手順で行った．なお，手順の詳細は参考にした文献[12]を参照されたい．

① サービスシナリオの作成
② 属性と水準の設定
③ 部分実施要因計画の作成
④ アンケート調査の作成
⑤ アンケート調査結果の分析

5.2.1 サービスシナリオの作成

本調査では，架空のサービスとして，お片付け掃除ロボットサービスを想定し，月額利用料や，2 種類の AI 誤判断のケース, 保険の有無に関する以下のシナリオを作成した．

（サービスシナリオ）

以下のシナリオのお片付け掃除ロボットは、１つ前のシナリオのお掃除ロボットに、片付け機能がついたものです。お片付け掃除ロボットは、床に置かれた衣類や箱、本、ゲーム機等のモノ（物）を片付けたり、紙屑や空き缶等のゴミを捨てたりしてから掃除をします。モノの置き場所を学習させておけば、床に置かれているモノを置き場所に戻してくれます。また、置き場所を学習させていないモノは指定した場所に置く他、紙屑、空き缶等のゴミと判断されたモノは指定のゴミ箱に入れます。その後、掃除を開始します。しかし、モノの置き場所を間違えて戻してしまうことや、モノをゴミと誤って捨ててしまうことが稀にあります。このお片付けロボットは、月額サービスで貸し出され利用する形態です。またオプションとして、お片付けロボットが誤ってものを捨てたり、片付けの際にモノを破損させてしまったりした際の損害を補償する保険を付けることができます。

本サービスシナリオは，5.1 節の（1）に基づいて深刻度の異なる 2 種類の誤判断ケースとして，「置き場所の間違い」（深刻度低）と「ゴミとして捨てる」（深刻度高）を含めており，（2）に基づいて「保険」を含めている．

5.2.2 属性と水準

コンジョイント分析で用いるプロファイルの属性及び水準を表 1 のようにした．

表 1：属性と水準

Table 1：Attribute and Level

属性	水準	内容
サービス価格	500 円	シナリオのサービスが 500 円の場合
	1,000 円	シナリオのサービスが 1.000 円の場合
	1,500 円	シナリオのサービスが 1.500 円の場合
	2,000 円	シナリオのサービスが 2.000 円の場合
場所の戻し間違い	週 1 回	お片付けロボットが間違う頻度
	月 1 回	お片付けロボットが間違う頻度
	半年 1 回	お片付けロボットが間違う頻度
ゴミと間違え捨てる	週 1 回	お片付けロボットが間違う頻度
	月 1 回	お片付けロボットが間違う頻度
	半年 1 回	お片付けロボットが間違う頻度
保険	なし	間違いによる損失を補償する保険なし
	あり	間違いによる損失を補償する保険あり

「場所の戻し間違い」「ゴミと間違え捨てる」の深刻度の異なる 2 種類の AI 誤判断ケースの水準に関するコンジョイント分析では，「1 週間に 1 度」「1 ヶ月に 1 度」をダミー変数化することで，「半年に 1 度」も表現することができるため，「場所の戻し間違い」「ゴミと間違え捨てる」はそれぞれ 2 つのダミー変数で表現することにした. これにより，3 段階の頻度（水準）の変化をわかりやすくすることにした．

また、金銭的な属性として AI システムを用いたサービスの「サービス価格」を含める．これは，コンジョイント分析において，「サービス価格」の属性の限界効用の比から各属性の評価額を求めるためである．この評価額は，個人が支払ってもよいと考える最大金額（限界支払意思額）として捉えることができる．

5.2.3 アンケート調査の作成

アンケート質問に用いる完全実施要因計画は，属性と水準の数から 4×3×3×2=72 通りであり，アンケートでの回答数が多くなってしまう. 一般にコンジョイント分析では，一定の統計的信頼性を保ちつつ回答者に提示するプロファイルの数を絞り込む部分実施計画が用いられる．本調査では，部分実施要因計画を完全実施要因計画の 72 通りの半分以下の 30 通りに縮減した．

本調査では，回答者に示すプロファイルデザイン図（図2）のように，「サービス 1 を利用したい」と「サービス 2

を利用したい」の選択肢を作成に加えて,「どちらも利用したくない」という選択外オプションを設定した.

アンケート調査では，回答者が以下の（問題文）を読んだ後に，図 2 に示すような 30 問の質問に対してサービス 1 とサービス 2 の内容を見比べ，3 つの選択肢から 1 つを選択する.

（問題文）

月額利用料，2 種類の誤り（置き場所の間違い，ゴミとして捨てる）の頻度，保険の有無のオプション項目があり，これらの組み合わせが異なるサービス 1 とサービス 2 があります．これらを比較して利用したいサービスを選択するか，どちらも利用したくない場合は「どちらも利用したくない」を選択してください.

サービス1	サービス1を利用したい	どちらも利用したくない	サービス2を利用したい	サービス2
金額：月額1000円 場所の戻し間違い：週1回程度 ゴミと間違え捨てる：週1回程度 保険：なし	□	□	□	金額：月額1500円 場所の戻し間違い：月1回程度 ゴミと間違え捨てる：月1回程度 保険：あり

図 2：プロファイルデザイン

Figure 2：Profile Design

5.2.4 アンケート調査結果の分析

サービス価格に関しては, 水準として「500 円」「1,000 円」「1,500 円」「2,000 円」の 4 つを設定し，サービス価格が選択の効用に与えている影響について分析する.

AI 誤判断に関しては,「場所の戻し間違い」「ゴミと間違え捨てる」の水準において,「半年に 1 度」を基準に「1 週間に 1 度」「1 ヶ月に 1 度」に変化した際の損失価値の変化と，選択肢の効用に与えている影響を分析する.

保険に関しては,「あり」「なし」を設定して，保険が選択肢の効用に影響を与えているかを分析する.

また，本論文では非価格属性の変数の限界支払意思額を算出することで，定量的に評価する．非価格属性は「場所の戻し間違い」「ゴミと間違え捨てる」「保険」である．非価格属性の変数の限界支払意思額は,「サービス価格」である価格属性の変数に基づいて，非価格属性の変数が 1 単位変化したときの評価額を表す．また，1 単位とは,「場所の戻し間違い」を例に挙げると「半年に 1 度」から「1 週間に 1 度」もしくは「1 か月に 1 度」からの変化となる.

5.3 アンケート調査

アンケート調査では，調査対象者を，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 2：回答者の年齢・性別の構成

Table 2：Age and Gender Composition of Respondents

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

6. 分析結果

本節では，データ解析環境 R を用いて，コンジョイント分析と限界支払意思額を求めた結果について述べる.

6.1 AI 誤判断による価値低下

6.1.1 回答者全体

表 3 は回答者全体のコンジョイント分析の結果を示している．なお，表 3 の ASC は選択肢固有定数であり，例えば「サービス 1 を利用したい」「サービス 2 を利用したい」のどちらかを選択していれば 1,「どちらも利用したくない」を選択していると 0 が付与されるダミー変数である.

表 3：コンジョイント分析結果 I

Table 3：Conjoint Analysis Result I

	coef	exp(coef)	se(coef)	z	p
ASC	2.175E-01	1.243E+00	4.156E-02	5.235	1.65e-07 ***
サービス価格	-1.096E-03	9.989E-01	2.442E-05	-44.893	< 2e-16 ***
場所の戻し間違え1週間1回	-1.157E-01	8.907E-01	2.95E-02	-3.922	8.79e-05 ***
場所の戻し間違え1か月1回	-1.057E-01	8.997E-01	3.13E-02	-3.381	0.000722 ***
ゴミと間違え捨てる1週間1回	-3.322E-01	7.174E-01	3.08E-02	-10.798	< 2e-16 ***
ゴミと間違え捨てる1か月1回	-2.266E-01	7.972E-01	3.173E-02	-7.142	9.17e-13 ***
保険あり	7.514E-01	2.120E+00	2.62E-02	28.722	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

表 3 の p 値（p）を見ると，いずれの変数も 1%水準で統計的に有意になっていることが確認できる．また,「保険あり」の係数は正, それ以外の項目の変数は負となっている．理論的に考えると,「保険あり」の係数は正，それ以外の変数の係数は負となることが想定されており，実際に表 3 の結果もその想定と一致している．例えば,「場所の戻し間違

い」の場合，（半年に 1 回から）その頻度が増えると，選択肢の効用に負の影響を与えることを意味している．

表 4 は表 3 の結果を用いて算出した限界支払意思額（非価格属性の変数の係数を価格で除したもの）となる．

回答者全体の限界支払意思額の代表値である．「場所の戻し間違い」は，半年に 1 回と比較して，1 週間 1 回はサービス価値が 106 円低下し，1 か月に 1 回はサービス価値が 96 円低下することが確認できる．また，「ゴミと間違え捨てる」は，半年に 1 回と比較して，1 週間 1 回はサービス価値が 303 円低下し, 1 か月に 1 回はサービス価値が 207 円低下することがわかる．一方で，「保険あり」の場合，サービスの価値が 685 円向上することがわかる．

この結果から，「場所の戻し間違い」や「ゴミと間違え捨てる」という AI 誤判断の結果はサービスの価値を下げること，また，その誤りの程度によっても利用者の評価が異なることが確認できる．

表 4：限界支払意思額 I

Table 4：Calculated Willing to Pay I

	限界支払額
場所の戻し間違え1週間1回	-105.5587
場所の戻し間違え1か月1回	-96.44355
ゴミと間違え捨てる1週間1回	-302.9973
ゴミと間違え捨てる1か月1回	-206.7313
保険のあり	685.3844
	単位 円

6.2 異なる 2 群での金銭的な価値の違いの把握

6.2.1 ・デジタルネイティブと非デジタルネイティブ

表 5 と表 6 はデジタルネイティブ群，表 7 と表 8 は非デジタルネイティブ群のコンジョイント分析と限界支払意思額の結果を示している．

表 5 と表 7 の「サービス価格」「ゴミと間違え捨てる 1 週間 1 回」「ゴミと間違え捨てる 1 か月 1 回」「保険あり」の係数はいずれも 1%水準，「場所の戻し間違い 1 か月 1 回」の係数はいずれも 5%水準となっている．また，表 5 の「場所の戻し間違い 1 週間 1 回」の係数は 10%水準, 表 7 の「場所の戻し間違い 1 週間に 1 回」の係数は 5%水準となっている．これらのことから，いずれの変数の係数も統計的に有意となっていることが確認できる．

表 6 と表 8 に示した限界支払意思額を比較すると，いずれの変数に関してもデジタルネイティブ群と非デジタルネイティブ群の評価が異なっていることが確認できる．この結果から，デジタルネイティブ群は非デジタルネイティブ群よりも AI の誤判断に対して厳しいと判断している（よりサービスの価値を低下させるものと考える）ということが示唆される．また，保険に対する評価は前者の方が後者よりも高いことが確認できる．

表 5：コンジョイント分析結果 II（デジタルネイティブ）

Table 5：Conjoint Analysis Result II (Digital-native Group)

	coef	exp(coef)	se(coef)	z	p
ASC	2.732E-01	1.314E+00	5.715E-02	4.78	1.76e-06 ***
サービス価格	-9.813E-04	9.990E-01	3.255E-05	-30.147	< 2e-16 ***
場所の戻し間違え1週間1回	-7.833E-02	9.247E-01	4.010E-02	-1.953	0.0508 .
場所の戻し間違え1か月1回	-9.597E-02	9.085E-01	4.289E-02	-2.238	0.0253 *
ゴミと間違え捨てる1週間1回	-3.524E-01	7.030E-01	4.204E-02	-8.384	< 2e-16 ***
ゴミと間違え捨てる1か月1回	-2.130E-01	8.082E-01	4.366E-02	-4.879	1.07e-06 ***
保険あり	7.892E-01	2.202E+00	3.569E-02	22.112	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

表 6：限界支払意思額 II（デジタルネイティブ）

Table 6：Calculated Willing to Pay II (Digital-native Group)

	限界支払額
場所の戻し間違え1週間1回	-79.81946
場所の戻し間違え1か月1回	-97.79529
ゴミと間違え捨てる1週間1回	-359.1423
ゴミと間違え捨てる1か月1回	-217.0569
保険あり	804.2658
	単位 円

表 7:コンジョイント分析結果 III（非デジタルネイティブ）

Table 7：Conjoint Analysis Result III (Non-digital-native Group)

	coef	exp(coef)	se(coef)	z	p
ASC	1.948E-01	1.215E+00	6.107E-02	3.19	0.001424 **
サービス価格	-1.247E-03	9.988E-01	3.722E-05	-33.501	< 2e-16 ***
場所の戻し間違え1週間1回	-1.602E-01	8.520E-01	4.383E-02	-3.655	0.000257 ***
場所の戻し間違え1か月1回	-1.171E-01	8.895E-01	4.599E-02	-2.547	0.010856 *
ゴミと間違え捨てる1週間1回	-3.152E-01	7.296E-01	4.547E-02	-6.933	4.13e-12 ***
ゴミと間違え捨てる1か月1回	-2.430E-01	7.842E-01	4.658E-02	-5.217	1.81e-07 ***
保険あり	7.131E-01	2.040E+00	3.872E-02	18.417	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

表 8：限界支払意思額 III（非デジタルネイティブ）

Table 8 : Calculated Willing to Pay III (Non-digital-native Group)

	限界支払額
場所の戻し間違え1週間1回	-128.4695
場所の戻し間違え1か月1回	-93.93252
ゴミと間違え捨てる1週間1回	-252.7627
ゴミと間違え捨てる1か月1回	-194.8987
保険あり	571.797
	単位 円

6.2.2 AI に関する知識レベルの高／低

表 9 と表 10 は AI に関する知識レベル高い群，表 11 と表 12 は AI に関する知識レベルが低い群のコンジョイント分析と限界支払意思額の結果を示している．

表 9：コンジョイント分析結果 IV（知識高）

Table 9：Conjoint Analysis Result IV (High-knowledge)

	coef	exp(coef)	se(coef)	z	p
ASC	3.497E-01	1.419E+00	5.554E-02	6.297	3.04e-10 ***
サービス価格	-1.076E-03	9.989E-01	3.214E-05	-33.485	< 2e-16 ***
場所の戻し間違え1週間1回	-9.952E-02	9.053E-01	3.917E-02	-2.541	0.0111 *
場所の戻し間違え1か月1回	-8.304E-02	9.203E-01	4.156E-02	-1.998	0.0457 *
ゴミと間違え捨てる1週間1回	-3.456E-01	7.078E-01	4.087E-02	-8.457	< 2e-16 ***
ゴミと間違え捨てる1か月1回	-2.360E-01	7.898E-01	4.246E-02	-5.557	2.74e-08 ***
保険あり	7.763E-01	2.173E+00	3.482E-02	22.291	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

表 10：限界支払意思額 IV（知識高）

Table 10：Calculated Willing to Pay IV (High-knowledge)

	限界支払額
場所の戻し間違え1週間1回	-92.48323
場所の戻し間違え1か月1回	-77.16947
ゴミと間違え捨てる1週間1回	-321.2101
ゴミと間違え捨てる1か月1回	-219.2858
保険あり	721.4133
	単位 円

表 11：コンジョイント分析結果 V（知識低）

Table 11：Conjoint Analysis Result V　(High-knowledge)

	coef	exp(coef)	se(coef)	z	p
ASC	7.485E-02	1.078E+00	6.302E-02	1.188	2.349E-01
サービス価格	-1.130E-03	9.989E-01	3.774E-05	-29.945	< 2e-16 ***
場所の戻し間違え1週間1回	-1.375E-01	8.716E-01	4.506E-02	-3.051	0.00228 **
場所の戻し間違え1か月1回	-1.364E-01	8.725E-01	4.769E-02	-2.86	0.00423 **
ゴミと間違え捨てる1週間1回	-3.188E-01	7.271E-01	4.694E-02	-6.79	1.12e-11 ***
ゴミと間違え捨てる1か月1回	-2.177E-01	8.044E-01	4.805E-02	-4.53	5.89e-06 ***
保険あり	7.234E-01	2.061E+00	3.984E-02	18.157	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1					

表 9 と表 11 の結果から，表 9 の「サービス価格」「ゴミと間違え捨てる 1 週間 1 回」「ゴミと間違え捨てる 1 か月 1 回」「保険あり」，表 11 のいずれの変数の係数も 1%水準で統計的に有意となっており，表 9 の「場所の戻し間違い 1 週間 1 回」「場所の戻し間違い 1 か月 1 回」の係数は 5%水準となっていることから，いずれの変数の係数も統計的に有意となっていることが確認できる．

続いて，表 10 と表 12 に示した限界支払意思額を比較すると，「場所の戻し間違い」に対しては AI に関する知識レベルが低い群の方が高い群よりもよりサービスの価値を低下させるものと考えるものの，「ゴミと間違え捨てる」に対しては AI に関する知識レベルが高い群の方が低い群よりもサービスの価値を低下させるものと考える傾向にあることが伺える．また，保険に対する評価は前者の方が後者よりも高いことが確認できる．

表 12：限界支払意思額 V（知識低）

Table 12：Calculated Willing to Pay V (Low-knowledge)

	限界支払額
場所の戻し間違え1週間1回	-121.6427
場所の戻し間違え1か月1回	-120.6847
ゴミと間違え捨てる1週間1回	-282.031
ゴミと間違え捨てる1か月1回	-192.6153
保険あり	640.0263
	単位 円

7. 考察

本節は，6 節の結果に基づいて考察する．

7.1 利用者個人の AI 誤判断リスクに関する金銭的評価

表 4 の限界支払意思額を見ると，「場所の戻し間違い」「ゴミと間違え捨てる」のどちらも，半年に 1 回と比較して，1 か月に 1 回，1 週間に 1 回と，間違える頻度が増加するにつれてサービス価値が低下する．また，「場所の戻し間違い」と「ゴミと間違え捨てる」の半年に 1 回と比べて 1 か月に 1 回，1 週間に 1 回のサービス価値の落ち方を見ると，「場所の戻し間違い」が-96 円→-106 円，「ゴミと間違えて捨てる」が-207 円→-303 円であり，「ゴミと間違え捨てる」の頻度を減らす方が，サービスの価値が高まることが示唆される．一方で、「場所の戻し間違い」の 1 週間 1 回が-106 円， 1 か月 1 回が-96 円と差が小さいことから，「場所の戻し間違い」については半年 1 回まで AI 誤判断の頻度を下げると，サービス価値が向上することが示唆される．

また，表 3 の「保険」の係数が他の属性よりも絶対値として大きいことから，選択肢の効用に大きな影響を与えていることから，利用時品質のリスク回避性の観点から，本サービスにおいて保険の意味があることが示唆される．

7.2 デジタルネイティブと非デジタル非ネイティブ

表 5 の分析結果では，デジタルネイティブの「場所の戻し間違い 1 週間 1 回」が 5%水準で有意でなく、1 週間に 1 回は選択肢の効用に影響していない．デジタルネイティブにとって，1 週間に 1 回間違うようであれば，選択肢としての評価の対象として捉えていないと言うことができる．デジタルネイティブよりも非デジタルネイティブでは，「場所の戻し間違い 1 か月 1 回」の限界支払意思額の差は 4 円

程度デジタルネイティブがマイナスである．このため，場所の戻し間違いが半年に 1 回から 1 か月に 1 回に変化した場合，デジタルネイティブの方がサービスの価値が低下するが，4 円程度であり同程度の効果と捉えることができる．一方，デジタルネイティブの方が「ゴミと間違え捨てる 1 週間 1 回」の限界支払意思額の差は 106 円マイナスであり，「ゴミと間違え捨てる 1 か月 1 回」は 22 円マイナスであり，デジタルネイティブの方が AI 誤判断の頻度を減らすと，サービス価値の向上の効果が高い．

「保険あり」は，デジタルネイティブの方が非デジタルネイティブよりも限界支払意思額が 232 円高いことから，デジタルネイティブの方は保険があれば非デジタルネイティブよりもサービス価値が高くなる．

これらのことから，非デジタルネイティブよりも，デジタルネイティブの方が，本サービスの AI 誤判断リスクを受容してもらうために，頻度を少なくし，保険ありにすることが効果的であると示唆される．

7.3 AI に関する知識レベルの高と低

AI に関する知識レベルの高い群と低い群の限界支払意思額を比較してみると，AI に関する知識レベルの高い群の方が「ゴミと間違え捨てる」の限界支払意思額のマイナスの金額が大きく，一方 AI に関する知識レベルの低い群の方が「場所の置き間違え」の限界支払意思額のマイナスの金額が大きいことから，2 群でサービスの価値評価が低下する AI 誤判断が異なっている．このことから，サービスの受容には，サービスのターゲットが重視するリスクを事前に把握することが効果的であると考える．

7.4 本論文の分析のアプローチについて

本論文では，AI システムを用いた架空サービスとしてお片付け掃除ロボットサービスを想定して，2 種類の AI 誤判断による損失価値の金額的な評価を試みた．今回は特に誤判定頻度に関する損失価値の定量的な差を見ることができ，2 種類の AI 誤判断で，それぞれどこまで頻度を低減することが有効かの検討比較ができた．しかし，2 種類の AI 誤判断が同じ頻度の時の損失価値の比較（いいかえれば誤判断の深刻度の直接的な比較）はできていない．2 種類の AI 誤判断の損失価値を比較するには，ダミー変数を用いずに各 AI 誤判断を属性としての限界支払意思額を比較すれば可能である．

本論文の結果は仮説シナリオに基づいた AI 誤判断の損失価値の評価であり，実際のサービスを利用した際の結果と比較しなければ，妥当性は評価できない．しかし，4 節に述べたように，現段階では本手法は AI 誤判断による損失価値を把握するために役立つと考える．

また，分析上の注意として，AI システムのサービスの利用者を想定したアンケート調査を実施しないと，適切な結果が得られないことに気をつける必要がある．

8. まとめ

本論文では，利用者として個人をターゲットとし，AI 誤判断によるサービスの価値損失の定量的な評価を行った．架空の AI システムを用いたシナリオに対してアンケート調査を実施し，コンジョイント分析と限界支払意思額の計算を行い，誤判断リスク受容における発生頻度低減，および保険の効果を定量的に検証した．

今後はリスク定量化手法の検討を深めるとともに、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）のための質問項目のアンケート結果を分析した結果を報告する予定である．

利用者の主体として個人のトラスト構築に関してまとまってきたのちに，主体を個人以外の組織や社会にも広げていく．

また，コンジョイント分析の結果に基づいて，マーケティングの観点から分析を進めることで，AI 誤判断の頻度に基づいた購入確率(普及率)についての関係を明らかにすることも行いたい．

参考文献

[1] 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 2022-07-26).

[2] 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.

[3] 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.

[4] 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）.

[5] 島成佳，小川隆一，佐川陽一，竹村敏彦. “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”．情報処理学会 2019-SPT-47 巻 9 号 1-8.

[6] 日本規格協会. “JIS X 25010（システム及びソフトウェア製品の品質要求および評価（SQuaRE）-システム及びソフトウェア品質モデル）”, 2013.

[7] NPO 日本ネットワークセキュリティ協会(JNSA)セキュリティ被害調査ワーキンググループ, “情報セキュリティインシデントに関する調査報告書別紙 第 1.0 版”, 2018.

[8] Sasha Romanosky. “Examining the costs and causes of cyber incidents”, Journal of Cybersecurity, 2(2), pp. 121-135(2016).

[9] 山田道洋，菊池浩明，松山直樹，乾孝治. “個人情報漏洩の損害額の新しい数量モデルの提案”, 情報処理学会, 2018-DPS-174 巻 18 号 P1-7.

[10] 竹村敏彦，片山佳則，鳥居悟，古川和枝. “プライバシー情報の価値の測定”. 2022 年暗号と情報セキュリティシンポジウム（SCIS2019）.

[11] 竹腰智，小川隆一，竹村敏彦. “コンジョイント分析による SNS アカウント情報の価値の測定”. 2022 年暗号と情報セキュリティシンポジウム（SCIS2019）.

[12] 合崎英男，西村和志. “データ解析環境 R による選択型コンジョイント分析入門”． 農工研技報 151-173, 2007.