

人工知能の開発・利用に関するリスク認知調査

谷 辺 哲 史

抄 録

人工知能 (AI) に関するリスクに対する一般市民の認知を調査した。AI の発展に対して、全体的にはマイナス面よりもプラス面が多いという回答が優勢で、AI の社会的受容が進んでいると考えられるが、リスクもあると認知されていた。特に、想定していないリスクが生じるという懸念を示す回答が最も多く、先端技術ゆえの未知性がリスク認知につながっていると考えられる。また、AI の学習のためのデータ利用に関わるリスクについては、収集した情報が流出することへの懸念が強く、学習のためにデータが利用されることへの懸念は比較的弱かった。ただし、情報の種類 (SNS への投稿、購買履歴など) によってリスク認知には差があったため、ユーザーが抱く懸念を踏まえてデータの取り扱い方法を考える必要がある。さらに、リスク認知に影響を与える要因を分析したところ、AI は規則に従って判断するというイメージが想定外のリスクへの懸念につながる、AI を身近に感じる人はデータの利用に対して容認的な態度を示すといった効果が見られた。

問 題

人工知能の普及と市民のリスク認知

本研究の目的は、人工知能 (artificial intelligence; AI) に関するリスクについて、一般市民のリスク認知とその背後にある要因を明らかにすることである。

2010年代の第3次 AI ブーム以降、AI は医療、自動車の運転、外国語の翻訳、文章や画像の生成など幅広い分野で利用されるようになり、2020年代半ばの現在も発展を続けている。利用場面が拡大するのに伴って、AI を安全に利用するための制度設計などの倫理的・法的・社会的課題 (ELSI) についても議論が行われている (松尾, 2015; 西垣・河島, 2019)。

AI をめぐる ELSI にはさまざまな論点があるが、その中で本研究が注目するのは、AI が普及することによって生じるリスクに対する一般市民の認知である。AI を利用した製品やサービスが普及する際に、エンドユーザーとなる一般市民が新技術を安心して使えることは、技術が実際に社会の中で利用されるために必要な条件である。市民の安心は技術的な観点での安全性と完全

に対応するとは限らないため、技術的な発展だけでなく市民がどのようなリスクを認識しているかを明らかにすることで、AIが社会的に受容されるための方策を検討しやすくなると期待できる。

加えて、本研究では特に、AIの開発のためのデータ利用に関わるリスクについて、市民の認知を検討する。第3次AIブーム以降のAIでは機械学習という技術が重要な役割を果たしているが、機械学習には現実世界から得られるデータが必要であり、買い物の内容のような私的な情報も収集の対象となりうる。そのため、プライバシーに関する倫理的な課題⁽¹⁾や、収集した大量の情報をAI開発企業が適切に管理しているのかという安全性への懸念があり、こうした問題が解消されなければAI技術そのものへの不信感につながりかねない。そこで本研究では、情報の種類によって学習に利用されることへの許容度が変わるのではないかと考え、さまざまな形でユーザーが発信する情報（SNSへの投稿、通販サイトでの購買行動など）がAIの学習に利用されることをリスク事象と捉えて、市民のリスク認知を検討することとした。

リスク認知の2つの側面

本研究では、リスク認知の測定において、発生可能性の認知と問題の大きさの認知を区別することを試みた。リスク認知の研究では、事故や自然災害、パンデミックのように明らかに被害が生じる事象に対する認知を調査することもあるが（e.g., Slovic, 1987; 中谷内・島田, 2010; 小森他, 2023）、AI開発におけるデータ利用については、それが悪いことだと考えるかどうかも議論の対象になる。というのも、機械学習によって適切な判断を行えるAIを開発するには多くのデータを収集する必要がある、データを学習して判断の精度が向上することはユーザーの利益につながるからである。たとえば通販サイトのおすすめ商品は、多くの消費者の購買行動の情報を収集することで適切に機能する。そのため、ユーザーはAIの学習によって自分自身が大きな利益を得られるならば、自分自身の行動などの情報が利用されることを問題だとは考えないかもしれない。

このような場面では、リスク認知という言葉の多義性に注意する必要がある。ある事象について「リスクがあると思うか」「不安があるか」といった質問によってリスク認知の高低を定量的に測定しても、リスク認知が低いという測定結果が「その事象が起きることはない」という判断を反映しているのか、「その事象が起きて構わない、問題ではない」という判断を反映しているのかを区別できないからである。そこで本研究では、AIの開発におけるデータの利用や管理に関して、「あるリスク事象が起きるだろうという可能性の認知」と「その事象が社会にとって悪いことであるという価値判断」を区別し、さまざまなリスク事象に対する一般市民の認知を2つの観点で調査した。

(1) AIがユーザーの情報を収集することとプライバシー権をめぐる問題については久木田他（2025）の第8章を参照。また、収集した情報の管理は個人情報保護に関する一般的な問題でありAIの開発に特有の問題ではないが、学習のためにユーザーから情報を収集することに伴って生じる問題であるため本論文の議論に含めた。

調査の概要

本研究では AI に関するリスク認知を幅広く調査することを目指したため、明確な仮説は設定せず、さまざまなリスクに対する認知を調査し、変数間の関連を探索的に分析した。ただし基本的な方針として、AI に対して一般市民が抱いているイメージが（その認識が AI の技術の実態に合っているかどうかとは関わりなく）リスク認知を形成する要因になると想定して分析を行った。AI に対するイメージの内容としては、コンピュータープログラムとして決められた通りに動く機械的な AI のイメージと、その対極として臨機応変に判断を変えられる柔軟なシステムのイメージがあると仮定し、こうしたイメージの個人差を測定した。

方法

参加者と手続き

2024年12月にウェブ調査を実施した。調査会社の回答者パネルの登録者に参加を依頼し、日本国内に在住の1000名（男性498名、女性496名、その他・無回答6名）が調査に参加した。参加者の年齢は20歳から69歳（ $M=44.88$, $SD=13.79$ ）だった。生成 AI を仕事または趣味や日常生活で使用したことがある参加者は354名⁽²⁾、そのうち少なくとも月1回程度使用している参加者は185名だった。

なお、本研究は早稲田大学「人を対象とする研究に関する倫理審査委員会」の承認を受けて実施された（承認番号2024-417）。

調査票の構成⁽³⁾

科学技術への関心 調査参加への同意を取得し、性別、年齢等の個人属性情報を尋ねた後、科学技術に関するニュースへの関心の程度を尋ねた（「1 = 関心がない」から「5 = 関心がある」の5件法）。

科学技術の功罪 科学技術の発展を全体的に考えた場合に、功罪のどちらが多いかの評価を求めた（「1 = マイナス面が多い」から「5 = プラス面が多い」の5件法）。最初に科学技術全般についての評価を求めた後、6つの具体的な科学技術について同様に評価を求めた。質問項目として用いたのは人工知能（AI）、原子力発電、遺伝子組み換え作物、代替肉（大豆などの植物から作られる、肉のような食材）、培養肉（牛などの細胞を人工的に培養して作る食肉）、宇宙開発（月や惑星の探査、有人宇宙飛行など）の6つだった⁽⁴⁾。この質問項目は、安全性や倫理的課題に

(2) 情報通信白書（総務省、2025）では2024年における日本の生成 AI 利用経験率は26.7%と報告されており、本研究のサンプルは総務省の調査よりも利用経験率が高い。

(3) 科学技術への関心、科学技術の功罪のうち科学技術全般に対する評価、科学的基礎知識についての質問項目は、文部科学省の「科学技術に関する国民意識調査」（細坪ほか、2020）の項目を利用した。

ついて議論が行われている主な技術について評価を求めることで、本研究の題材である AI が他のさまざまな科学技術と比較してどの程度社会的に受容されているかを把握することを目的としていた。

AI のイメージ AI に対して参加者が持っているイメージを、「人間よりも AI のほうが公平な判断をする」などの 6 項目に対してどの程度同意するかという形で質問した（「1 = 当てはまらない」から「5 = 当てはまる」の 5 件法）。質問項目は図 2 に示した 6 項目で、いずれも「人間よりも AI のほうが」に続けてそれぞれの文言を提示した。順序効果を避けるために項目の提示順序をランダム化した⁽⁵⁾。

AI の開発方針 将来の AI の開発の方向性について、どのような AI の開発を進めるべきかを尋ねた。「人間が決めた規則に従って判断するだけでなく、倫理的な善悪を判断する AI」、「判断の結果だけでなく、判断の理由を人間にわかりやすい形で説明する AI」、「翻訳や自動車の運転といった個別の作業だけではなく、1 つのシステムでさまざまな知的作業を行う AI」⁽⁶⁾の 3 つについて、それぞれ「1 = 技術開発を進めるべきではない」から「5 = 技術開発を進めるべきだ」の 5 件法で評価を求めた。これら 3 つの質問項目は提示順序をランダム化した。

AI に関するリスク認知 AI が普及することに伴って生じるリスクに対する認識を尋ねた。「AI が普及することで、」に続けて「今までに全く想定していなかったリスクが生じる」など 6 つのリスクを提示し、それぞれのリスクがあると思う程度を尋ねた（「1 = 当てはまらない」から「5 = 当てはまる」の 5 件法。質問項目の文言は図 3 に示した）。質問項目の提示順序はランダム化した。

AI の開発における情報管理に関するリスク認知 AI の学習のためのデータ収集や、収集したデータの管理に関わるリスクについて、さまざまなリスク事象の発生可能性と、その事象が起きることの問題の大きさの評価を求めた。

リスク事象の内容は以下の 9 項目だった：「AI の学習のために集められた情報が、外部からのサイバー攻撃によって外部に流出する」、「AI の学習のために集められた情報が、AI 開発者の不正行為によって外部に流出する」、「AI の学習のために集められた情報が、不注意や事故によって外部に流出する」、「特定の相手に送信した内容（電子メール、LINE など）が、AI の学習に使用される」、「インターネット上のサービスを利用するために登録した個人情報、AI の学習に使用される」、「インターネット上で公開する範囲を制限して投稿された内容（Instagram や X（Twitter）の非公開アカウントでの投稿など）が、AI の学習に使用される」、「インターネット上で不特定の相手に公開されるように投稿された内容（Instagram や X（Twitter）の投稿など）が、AI の学習に使用される」、「インターネット上のサービスを利用したときの行動履歴（買い物、

(4) 6 つの科学技術は括弧内の文言も含めて、本文中に示した文言の通りに参加者に提示した。

動画や音楽の視聴)が、AIの学習に使用される]、「ChatGPTなどの生成AIを利用する際に利用者が入力した内容が、AIの学習に使用される」。

これらの事象が起きる可能性を尋ね(「1 = 可能性が低い」から「5 = 可能性が高い」の5件法)、次に、これらの事象が起きるのは社会にとって悪いことだと思うかを尋ねた(「1 = 悪いことではない」から「5 = 悪いことである」の5件法)。項目の提示順序は回答者ごとにランダム化した。回答やすさを考慮し、回答者内では発生可能性についての質問と問題の大きさについての質問で同じ順序とした。

AIに関する意識、利用状況 参加者にとってのAIの身近さの認識(「1 = 身近に感じない」から「5 = 身近に感じる」の5件法)、AIに関する主観的な知識量(「1 = 全く知らない」、「2 = あまり知らない」、「3 = 人並みには知っている」、「4 = よく知っている」の4件法)、生成AI(文章や画像・映像を作成するAI)の利用状況(業務での利用、趣味や日常生活での利用の2つについて、「1 = 使用したことがない」、「2 = 使用したことはあるが、日常的には使用していない」、「3 = 月に1回程度」、「4 = 月に数回～週に1回未満」、「5 = 週に1回～数回程度」、「6 = ほとんど毎日」の6件法)をそれぞれ質問した。

科学的基礎知識 最後に科学全般に関する知識量を測定した。「地球の中心部は非常に高温である」など11項目の正誤問題に「正しい」、「誤っている」、「わからない」の3択で回答し、正答数を知識量の指標とした。質問項目の提示順序はランダム化した。

結 果

科学技術の功罪に対する評価

まずAIに対する全般的な評価を確認するため、AIを含む科学技術の発展に対する評価を集計した(図1)。AIに対して肯定的な回答(プラス面が多い、どちらかというとならプラス面が多い)が46.0%、否定的な態度(マイナス面が多い、どちらかというとならマイナス面が多い)は16.2%となり、肯定的な反応が優勢だった。

他の科学技術と比較しても、AIは宇宙開発に次いで肯定的な回答が多かった。平均値⁽⁷⁾を分

(5) AIのイメージを尋ねると同じページに、不注意回答者の確認のための質問項目を設けた。「読み取り確認のため、この質問には「どちらかというとなら当てはまる」と答えてください」という質問を行い、指示通りに回答しなかった参加者は調査票を注意深く読んで回答していないと判断し分析対象から除外した。本研究ではデータのスクリーニングも含めて調査会社に委託したため、調査会社から納品された1000名分のデータはすべて不注意回答チェックを通過したものである。

(6) 第2項目は説明可能AI、第3項目は汎用AI(AGI)を想定した質問項目である。

(7) これ以降の分析も含め、選択式の質問項目への回答を分析する際には、回答の選択肢(当てはまる、当てはまらないなど)に割り振られた数字を便宜的に間隔尺度と見なして分析に用いた。なお、参加者に提示したウェブ調査の画面では、選択肢の文言のみを表示し、数字は表示されていない。

散分析によって比較すると、科学技術の種類によって有意な差があった ($F(5, 4995) = 131.30, p < .001$)。Shaffer 法による多重比較の結果、AI への評価は宇宙開発よりも有意に低かったが ($t(999) = 5.81, p < .001$)⁽⁸⁾、遺伝子組み換え作物 ($t(999) = 15.17, p < .001$)、培養肉 ($t(999) = 10.50, p < .001$)、原子力発電 ($t(999) = 10.61, p < .001$) よりも有意に高かった。AI と代替肉では有意な差がなかった ($t(999) = 2.03, p = .085$)。

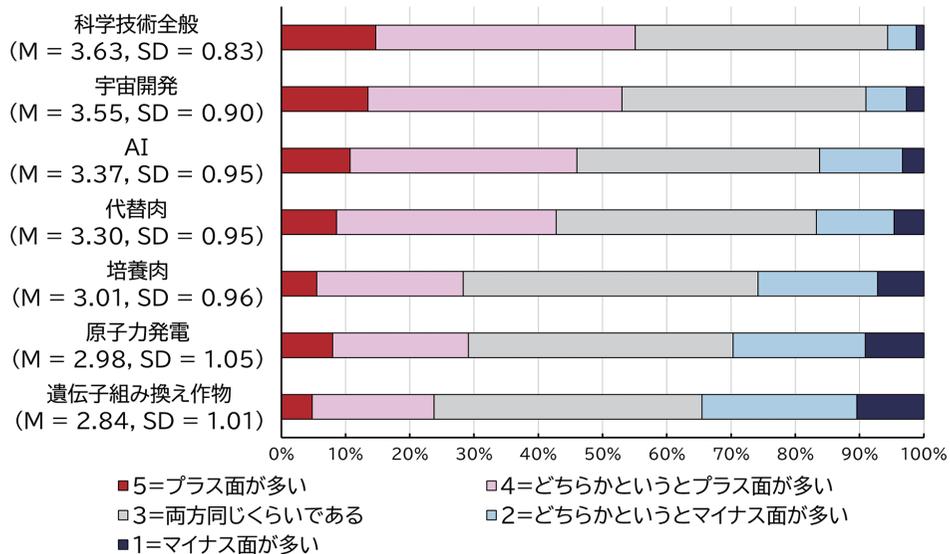


図1 科学技術の功罪に対する評価の回答分布

AI に対するイメージ

AI に対するイメージの回答分布は図2のようになった。各項目の平均値を分散分析によって比較すると、項目間で有意な差があった ($F(5, 4995) = 225.78, p < .001$)。Shaffer 法による多重比較の結果、「臨機応変に判断できる」と「予想外の事態に対処できる」の間には有意な差がなかったが ($t(999) = 0.71, p = .480$)、それ以外の対比較ではすべて有意な差があった ($t_s(999) > 5.6, p_s < .001$)。AI に対するイメージは、規則に従ってミスなく判断するという定型的な動きをするイメージが強く、柔軟な判断をするイメージはあまり持たれていなかった。

回答の度数分布を見ると、中間の「どちらともいえない」の回答率が最も少ない「規則を守る」でも30.8%で、他の項目ではいずれも40%を超えていた。

(8) 多重比較の結果は Shaffer 法による補正後の調整 p 値を報告する。

人工知能の開発・利用に関するリスク認知調査

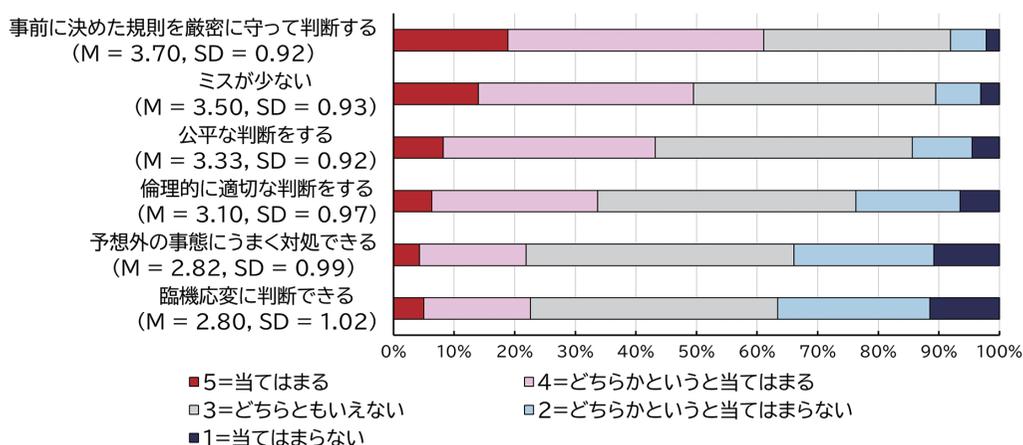


図2 AIに対するイメージの回答分布

AIの開発方針

将来のAIの開発を進めるべきかの判断について、説明可能AI ($M=3.70$, $SD=0.88$)、汎用AI ($M=3.58$, $SD=0.90$)、倫理的AI ($M=3.29$, $SD=0.94$)の順に肯定的な反応が強く、この3つの平均値は全て互いに有意な差があった ($F(2, 1998) = 115.72$, $p < .001$; 多重比較 (Shaffer法) は全ての対比較で $t_s(999) > 5.0$, $p_s < .001$)。

AIに関するリスク認知

AIが普及することによるリスクの評価は図3のようになった。各項目の平均値を分散分析によって比較すると、項目間で有意な差があった ($F(5, 4995) = 126.01$, $p < .001$)。Shaffer法による多重比較の結果、AIへの過信と失業の間には有意な差がなかったが ($t(999) = 1.37$, $p = .171$)、それ以外の対比較ではすべて有意な差があった ($t_s(999) > 2.8$, $p_s < .01$)。

想定していないリスクが生じると考える参加者が65.8%を占め、回答の平均値も最も高かった。つまり、AIに関するリスク認知は具体的なリスク事象に対する認知よりも、何が起きるかが分からないという漠然としたリスクが最も強く認知されていた。

上記のAIに関するイメージがリスク認知に与える影響を検討するために重回帰分析を行った(表1)。説明変数は前述のAIのイメージに加えて、科学技術全般に対する関心と知識量、AIに対する意識(主観的知識量、身近さの認知)だった。さらに統制変数として年齢と性別⁽⁹⁾を加えた。リスク認知の各項目を目的変数として、同一のモデルでそれぞれ重回帰分析を行った。

(9) 性別は男性を0、女性を1とするダミー変数。性別がその他または無回答だった参加者はこの分析には含まなかった。

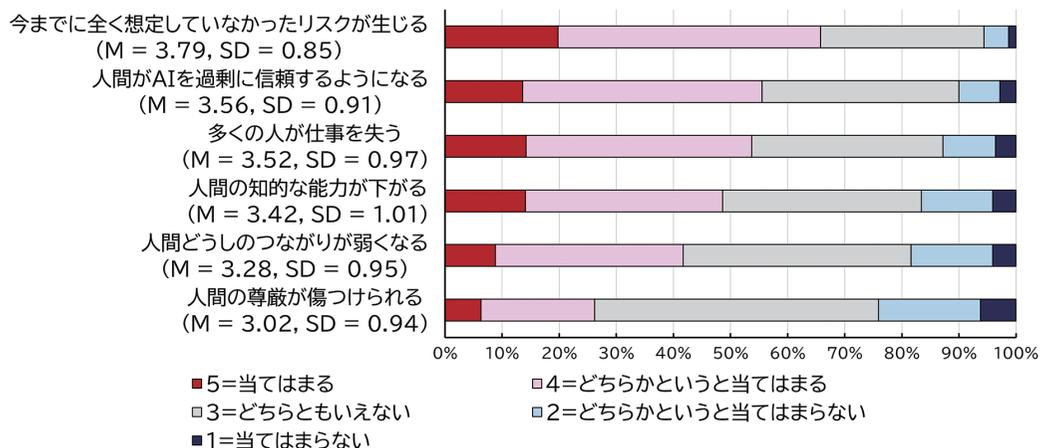


図3 AIに関するリスク認知の回答分布

表1 AIに関するリスク認知を説明する重回帰分析

	想定外の リスク	AIへの 過信	失業	知的能力 の低下	人間関係 の希薄化	人間の 尊厳の毀損
性別 (0 = 男性, 1 = 女性)	.084**	.023	.132***	.097**	.137***	.070*
年齢	.037	.077*	-.014	.008	.049	.049
科学への関心	.083*	.047	-.033	.039	.036	.023
科学的基礎知識	.123***	.052	-.017	.001	-.034	-.052
AIに関する主観的知識量	.012	.020	-.027	-.036	-.090*	-.068†
AIの身近さ	-.110**	-.048	-.051	-.058	-.020	-.032
AI: 規則を守る	.229***	.147***	.033	.111**	.049	-.041
AI: ミスが少ない	-.093*	.007	.110**	.015	.014	.027
AI: 公平	-.024	.036	-.049	-.063	-.048	-.131**
AI: 倫理的	.011	.056	-.005	.002	.046	-.063
AI: 予想外に対処する	-.074†	-.048	-.064	-.089*	-.004	.017
AI: 臨機応変	-.051	-.018	.092*	.037	.006	.052
R^2	.110***	.057***	.044***	.032**	.041***	.054***
Adj. R^2	.099	.046	.033	.020	.029	.043

(注) 表中の数値は標準化偏回帰係数。*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

AIの開発における情報管理に関するリスク認知

情報管理に関するリスクについて、発生可能性と問題の大きさそれぞれについて、回答の分布は図4、図5のようになった。発生可能性、問題の大きさともにリスク事象による差があったが（それぞれ $F(8, 7992) = 23.59, p < .001$; $F(8, 7992) = 182.12, p < .001$ ）、発生可能性は平均値が最大のもの（生成AIへの入力学習に利用される： $M = 3.86, SD = 0.85$ ）と最小のもの（電子メール

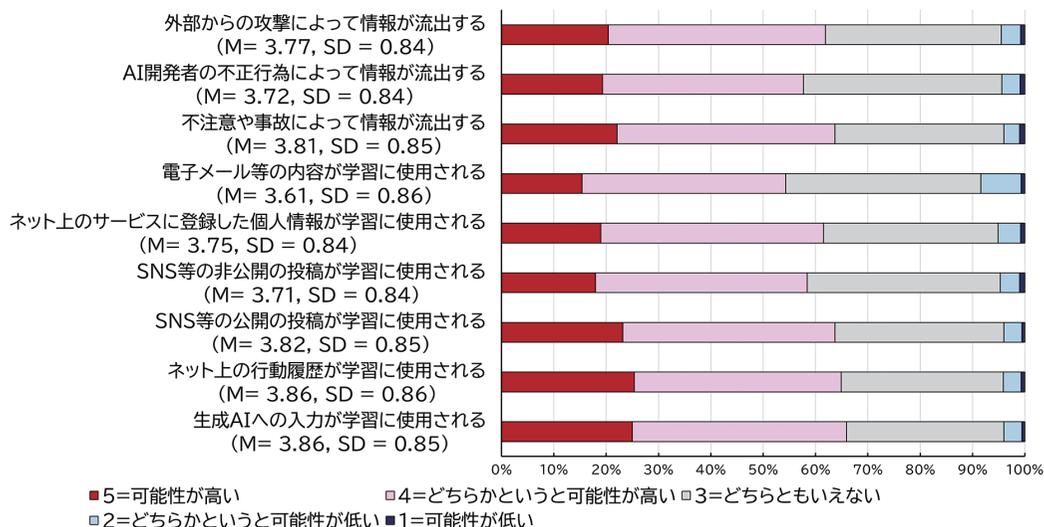


図4 情報管理に関するリスクの発生可能性の回答分布

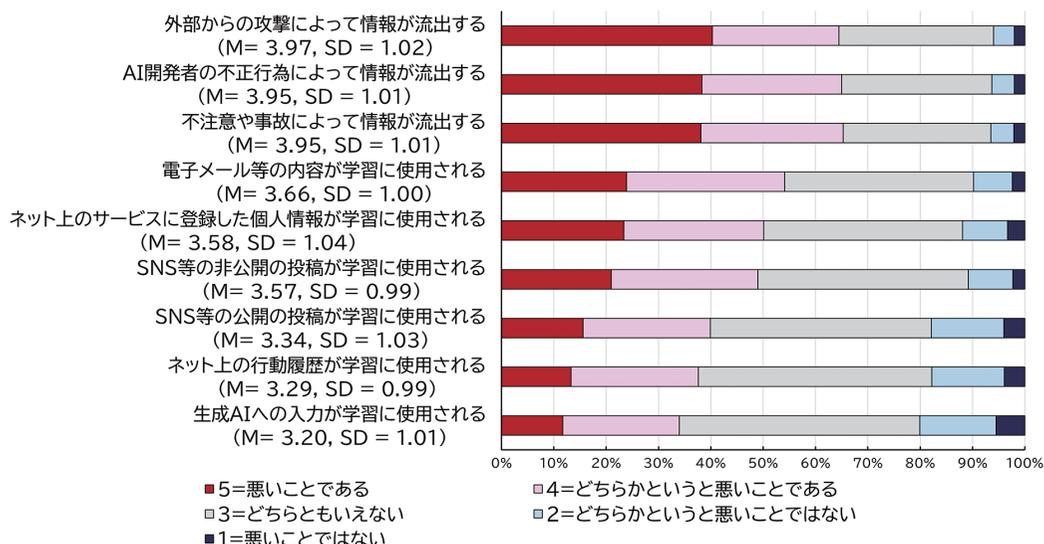


図5 情報管理に関するリスクの問題の大きさの回答分布

ル等の内容が学習に利用される： $M=3.61$, $SD=0.86$)でも0.25程度しか差がなかった。一方で問題の大きさは最大のもの（外部からの攻撃によって情報が流出する： $M=3.97$, $SD=1.02$)と最小のもの（生成AIへの入力学習に使用される： $M=3.20$, $SD=1.01$)で0.77の差があった。つまり、リスク事象の問題の大きさは事象ごとにある程度区別して認知されていたのに対して、それぞれの事象が実際に起きるかどうかはさほど明確に区別して認知されていなかった。

リスク事象間での問題の大きさを比較するために多重比較（Shaffer法）を行ったところ、平均値の上位を占めた情報の流出に関する3項目は、いずれも他の6項目のすべてと有意な差があった一方で（ $t_s(999) > 10.0$, $p_s < .001$ ）、これら3項目の間では互いに有意な差はなかった（ $t_s(999) < 1.1$, $p_s > .99$ ）。

発生可能性の認知と問題の大きさの認知との関連については、発生可能性の上位3項目（生成AIへの入力学習に使用される、ネット上の行動履歴が学習に使用される、SNS等の公開の投稿が学習に使用される）と問題の大きさの下位3項目が一致していたが、その他の6項目では両者の順位に明確な関連は見られなかった。

前項のリスク認知についての分析と同様に、情報管理に関するリスク事象の発生可能性、問題の大きさについても、AIのイメージとの関連を検討するために重回帰分析を行った（表2、表3）。

表2 情報管理に関するリスクの発生可能性を説明する重回帰分析

	攻撃による情報流出	不正行為による情報流出	事故による情報流出	電子メール等を学習に利用	個人情報を学習に利用	非公開の投稿を学習に利用	公開の投稿を学習に利用	行動履歴を学習に利用	生成AIへの入力を学習に利用
性別（0 = 男性, 1 = 女性）	.079*	.075*	.090**	.034	.071*	.052	.063*	.042	.044
年齢	.036	.019	.017	.024	.025	.052	.005	.018	-.031
科学への関心	.091**	.052	.048	.110**	.097**	.084*	.111**	.066†	.064†
科学的基礎知識	.124***	.117***	.155***	.030	.101**	.112**	.186***	.140***	.172***
AIに関する主観的知識量	.066†	.046	.045	.002	.055	.019	.053	.100**	.107**
AIの身近さ	-.081*	-.071*	-.073*	-.005	-.019	.004	-.029	-.020	-.036
AI：規則を守る	.198***	.222***	.258***	.154***	.187***	.194***	.219***	.249***	.262***
AI：ミスが少ない	-.029	-.040	-.039	-.025	.017	-.061†	.000	.014	.019
AI：公平	.018	-.037	.000	-.010	-.031	-.038	-.004	.010	-.023
AI：倫理的	.004	.040	-.059	.004	-.011	.038	-.039	.000	-.093*
AI：予想外に対処する	-.046	-.108**	-.022	-.036	-.005	-.033	-.101**	-.073†	.009
AI：臨機応変	-.066†	-.022	-.040	.000	-.013	-.003	-.001	-.061	-.039
R^2	.104***	.094***	.119***	.042***	.079***	.074***	.145***	.147***	.151***
Adj. R^2	.093	.083	.108	.030	.068	.063	.134	.136	.141

（注）表中の数値は標準化偏回帰係数。*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

表3 情報管理に関するリスクの問題の大きさを説明する重回帰分析

	攻撃による情報流出	不正行為による情報流出	事故による情報流出	電子メール等を学習に利用	個人情報を学習に利用	非公開の投稿を学習に利用	公開の投稿を学習に利用	行動履歴を学習に利用	生成AIへの入力を学習に利用
性別 (0 = 男性, 1 = 女性)	.125***	.160***	.143***	.106***	.158***	.112***	.118***	.094**	.111***
年齢	-.012	.007	.002	.022	.038	-.029	.016	.037	.065*
科学への関心	.013	.009	.001	-.008	.007	-.001	.018	.022	-.010
科学的基礎知識	.186***	.178***	.171***	.104**	.101**	.078*	.001	.025	.028
AIに関する主観的知識量	.138***	.132***	.141***	.112**	.087*	.146***	.112**	.098**	.041
AIの身近さ	-.154***	-.116***	-.124***	-.120***	-.081*	-.086*	-.096**	-.151***	-.108**
AI: 規則を守る	.155***	.144***	.163***	.143***	.088*	.131***	-.008	.025	-.052
AI: ミスが少ない	.040	.004	.001	-.037	-.053	-.037	-.056	-.033	-.012
AI: 公平	.019	.030	.021	-.009	.021	.001	-.024	-.086*	-.076†
AI: 倫理的	-.117**	-.115**	-.119**	-.139***	-.187***	-.136***	-.102**	-.079*	-.066†
AI: 予想外に対処する	-.032	-.062	-.059	-.034	-.047	-.054	.029	-.006	.023
AI: 臨機応変	-.137***	-.103**	-.122**	-.035	.044	-.025	-.010	.067†	.038
R ²	.159***	.145***	.154***	.088***	.083***	.080***	.047***	.053***	.054***
Adj. R ²	.149	.135	.144	.076	.072	.069	.035	.042	.042

(注) 表中の数値は標準化偏回帰係数。*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

考 察

AIに対するイメージ、評価の概要

本研究ではAIに対する全般的なポジティブ、ネガティブな評価やAIのイメージといった基本的な認識、将来のAIの開発に対する期待を調査した。その結果、AIは他の科学技術と比べて比較的ポジティブな評価を受けており、AIに対する社会的受容が一定程度進んでいることが確認できた。

AIのイメージは規則を守って判断する、ミスが少ないといった定型的な判断をする機械というイメージが強く、予想外の事態に対処したり臨機応変に判断したりといった柔軟性があるというイメージは弱かった。AIはプラス面が多いというポジティブな評価も、こうした定型的な判断を行うAIに対するものだと考えられる。また、将来のAIの開発に対しても、倫理的な判断を行えるAIや汎用AIよりも、説明可能AIの開発を進めるべきだという回答が多かった。説明可能AIは現在すでに普及が進んでいる機械学習を用いたAIについて判断の過程を分かりやすく示すものである。人々が将来のAIに期待しているのは判断規則の透明性を高めるということであり、現在のAIのイメージの延長線上にあるものといえる。倫理的な判断を行うAIへの

期待はさほど高くないという結果は、人は AI に倫理的判断を任せることを望んでいないという先行研究の結果 (Bigman & Gray, 2018) とも整合的である。

しかし、近年はプロンプト (ユーザーが入力した指示) に応じて文章や画像を作成する生成 AI が急速に普及し、話し相手など人間の他者のような役割を AI に求める人もいるため (e.g., 日本経済新聞, 2025)、今後は柔軟で非定型的な判断を行うものという AI のイメージがより強く意識されるようになるかもしれない。そうすると、AI に対する人々の態度もあらためて検討する必要が出てくるだろう。

AI に関するリスク認知

AI の普及によってどのようなリスクが生じると思うかを尋ねると、最も強くリスクが認知されていたのは、失業などの具体的な問題ではなく、全く想定していなかったリスクが生じるという不確実なリスクへの認知だった。リスク認知を未知性と恐ろしさの 2 つの因子 (Slovic, 1987) に分けて考えると、急速に進歩している AI は未知の技術と認知されており、それゆえ具体的にどのような問題が生じるかということ以上に、何が起きるかが分からないという不安が生じているということだろう。AI を身近に感じる人ほど想定外のリスクを小さく認知するという結果を踏まえると、AI を利用した製品やサービスが普及するにつれて、「何が起きるか分からない」という漠然としたリスク認知は低下していくかもしれない。

しかし一方で、科学技術への関心の高さや科学的基礎知識は、想定外のリスクの認知を促進する効果があった。未知性という観点では、科学技術への関心や知識があればリスク認知は抑制されても不思議ではないが、調査結果はその逆となった。この 2 変数が効果を示したのは想定外のリスクの認知のみで、他の個別のリスクとは関連していなかった点も興味深い結果である。AI は急速な進歩が日々のニュースで伝えられ続けているため、科学技術への関心や知識が高い人々ほどそうした最新の情報に触れており、AI が今後どのように変化していくか予測できないものだという認識を持っているのかもしれない。

AI のイメージとリスク認知の関連については、リスクの種類によって影響のあるイメージの内容はまちまちだったが、その中で「AI は規則を守る」というイメージは想定外のリスク、AI への過信、人間の知的能力の低下という 3 つのリスクの認知を高めていた。規則に正確に従って判断できることは AI の性能の高さとも言えるが、同時に、柔軟性がなく定型的な作業しかできないと捉えることもできるため両面的な評価につながりうる。そのため、AI を規則に従うものと見なす人にとっては、人間の行っている判断を AI に置き換えるのは AI への信頼が過剰であるとか、人間が本来行うべき判断をできなくなっていくという懸念を生じさせるのだろう。また、規則に従うというイメージは想定外のリスクへの認知も高めていた。規則に忠実であることは予測可能性を高めると考えることもできそうだが、この分析結果に基づくと、むしろ予測不可能性

を感じさせていたと考えられる。AIはミスが少ないというイメージは想定外のリスクの認知を低下させていることも踏まえると、規則に従うことと適切な判断を行うことは人々のイメージの中で区別されているようである。現実世界における問題は事前に明示的な規則を示せるとは限らないため、想定外の状況に直面した際に、AIは規則に従うがゆえに不適切な結果をもたらしてしまうという懸念につながると解釈できる。

AIの開発における情報管理に関するリスク認知

リスク認知の中でもAIの開発、特に機械学習のためのデータ利用に関わるリスクについて、リスク事象の発生可能性と問題の大きさの認知を区別して調査した。その結果、事象ごとの発生可能性の認知と、問題の大きさの認知との間に明確な関連は見られなかった。つまり、リスク事象が起きる可能性を認識していることと、その事象が起きるのは悪いことだと考えることは、人々の認知において実際に区別して判断されていた。AIの学習のためのデータ利用については、データを収集・利用するのが悪いことなのかということ自体が明確ではないため、リスクへの対応について議論する際には、リスク認知という言葉に複数の側面があることを明確にしておくことで混乱を避けることができるだろう。

問題の大きさの認知をリスク事象間で比較すると、情報流出に関する項目と、学習のためのデータ利用の項目で明確に差があった。情報流出は企業等が当然求められる情報管理を適切に行っていないという事象であり、学習のためのデータ利用とは異なり、明確に悪い出来事と見なされるのは自然なことといえる。一方で学習のためのデータ利用については、情報の種類によって問題の大きさの判断が異なっていた。情報収集の手段や、学習に利用されることでユーザーが得る利益など要因はさまざまに考えられるが、データの利用に対するユーザーの受容度は変化するということを踏まえて、ユーザーに受容される利用のあり方を検討する必要があるだろう。

問題の大きさの認知が低かった事象に注目すると、生成AIの利用時にユーザーが入力した内容、買い物などの行動履歴、SNSで公開した投稿といった情報がAIの学習のために利用されることについては、人々はそのようなデータ利用が実際に行われるだろうと考えているが、それが悪いことであるとは必ずしも考えていないことが示された。AIの学習のために開発者がユーザーの情報を収集することについてはプライバシーの観点から問題が指摘されることもあるが、情報の種類によってはユーザーもさほど強い抵抗を感じていないと考えられる。これらの事象の問題の大きさが低く認知された理由について本研究のデータから明確な結論を得ることはできないが、SNS等の公開での投稿はユーザー自身が発信した時点で不特定の相手に公開されることを了承しているため、その情報が他者に利用されることにも抵抗は生じにくいだろうと考えられる。生成AIへの入力と行動履歴の情報は公開情報ではないが、AIから出力される文章などの内容や通販サイトのレコメンドなどユーザー自身の利便性につながる情報が個人に最適化されるという

入出力の関係が理解しやすいため、ユーザー自身の利益の大きさゆえに容認されるのかもしれない。ただし問題の大きさの平均値は最も低いものでも3.20と、5件法の尺度の中点を超えており、データの利用を全く問題のないものと判断しているわけではないことには注意が必要である。

AIのイメージやその他の変数がリスク認知に与える影響を分析すると、AIが規則に従うというイメージは全てのリスク事象に関して発生可能性の認知を高めていた。AIが何らかの規則に従って判断するものだと考えている人は、その判断の基準としてユーザーから収集した情報が利用されているという認識も持ちやすいと考えられる。ただし比較的問題が小さいと判断されていた3項目では、規則に従うAIのイメージは発生可能性の認知を高める一方で問題の大きさの認知には影響を与えていなかった。したがってデータ利用の方法によっては、データが利用されていると認識しつつ、それが悪いことだとは考えないという受容的な態度につながっていると考えられる。

AIが倫理的に適切な判断をするというイメージは、リスク事象の発生可能性の認知にはほとんど影響しない一方で、問題の大きさは多くの項目で抑制する効果があった。AIが倫理的に適切な判断をできると考える人にとっては、そうした判断の材料としてユーザーから収集した情報が利用されても問題ないと許容できるのだろう。

ただし、人々のリスク認知は必ずしも合理的な推論に基づくものではないかもしれないことには注意が必要である。AIが規則に従うというイメージは、情報流出に関するリスク事象の発生可能性、問題の大きさの両方を大きく認知することにつながり、AIが倫理的に適切な判断をするというイメージはこれらのリスクの問題を小さく認知することにつながっていた。しかし情報流出は企業などの情報管理体制の問題であり、開発するAIの性質によってリスクの発生可能性が変わるとは考えにくいし、情報流出の悪さも変わらないはずである。AIのイメージがこれらのリスクに対する認知を変えるということは、具体的なリスクの内容ではなく、AIに対する全般的なポジティブまたはネガティブなイメージがリスク認知も変化させていると考えられる。これは、ポジティブなイメージを持っている技術に関してリスクのようなネガティブな情報に触れることが認知的不協和 (Festinger & Carlsmith, 1959) を生じさせるため、不協和を解消するためにリスク認知を変容させるという認知的な過程が存在すると考えられる。

AIのイメージ以外の要因の効果で興味深いのは、AIを身近に感じる人ほど、どのリスク事象に対しても問題が小さいと判断するという関連があったことである。個々の参加者がどのような要因によって身近さを感じていたかは分からないが、身近であるという感覚は個別のリスクの内容に対する認識を超えて、AIの開発におけるデータ利用に対する全般的な態度を肯定的なものにする効果があると考えられる。今後、AIを利用した製品やサービスに触れる機会が増えていけば、多くの人々が今よりもAIを身近に感じるようになり、それによってデータ利用への受容的な態度も強くなっていくだろうと予想される。AIの社会的受容が進むことはさまざまな分野で

社会に利益をもたらすと期待されるが、一方で AI が身近になることによってデータの利用や管理の問題が小さくなるわけではないため、このような要因による受容的態度はリスクの見過ごしや AI 開発者への過信を引き起こすおそれもある。

反対に AI に関する主観的知識量や科学的基礎知識は、いくつかのリスク事象に関して問題が大きいという判断を強めていた。科学技術について知ることは、科学技術の利用に伴う問題についても意識することにつながっているようである。具体的にどのような知識が AI に関するリスク認知に影響を与えるのかは明確ではないものの、科学技術についての正しい理解を促進するような情報提供をすることで、リスクの見ごしを抑制できる可能性がある。

本研究の限界と今後の展望

最後に本研究の限界と今後の展望を議論する。

第一に、リスクの種類やリスク認知に関わる要因を拡大して、実証的な知見を蓄積していく必要がある。本研究ではさまざまなリスクに対する認知や AI に対するイメージを幅広く調査し、変数間の関連を探索的に検討した。その結果、どのようなリスクが高いと認知されているか、リスク認知は AI のイメージとどのように関連するかといった観点で人々の認知を明らかにできた。しかしイメージやリスク認知といった心理変数どうしの関連の検討に留まるため、AI に対するさまざまなイメージがどのように形成されるのかは解明できていない。AI の利用経験など、AI に対するイメージの形成に関わる要因を明らかにできれば、AI の社会的受容を促進するための方策など、より実践的な議論も可能になる。また、検討すべきリスクの種類も今後はさらに増えていこう。現在、特に注目が集まっているのは生成 AI であるが、事実と反する情報を出力するハルシネーションの問題が指摘されている。また、話し相手となる対話エージェントに精神的に依存する人も増えていこう。誤情報の出力やユーザーの依存といったリスクは本研究では扱わなかったが、生成 AI の普及に伴って大きな問題になっていくと考えられるため、ユーザーがそれらのリスクを認識できているかを確認し、リスクを抑制しながら生成 AI を活用できる仕組みを検討していくことが求められる。

第二に、調査対象者の代表性の問題がある。AI の開発や利用に関する問題はすべての人に関わるものであるため、本研究でも日本に住むすべての人を母集団として AI に対する認知を検討することが理想的である。本研究ではウェブ調査会社の回答者パネルを利用して調査参加者を募集しており、母集団からの無作為抽出ではないため、回答の分布や平均値は母集団の情報を推測するには不適切である。ウェブ調査に自発的にアクセスする人という偏りは、インターネットを利用しない人を排除しており、特に AI などの情報技術に関しては親和的な態度を持つ人に偏るおそれがあるため、本研究で得られた結果を解釈するには注意が必要である。今後は郵送調査などの手法も組み合わせて、一般市民の認識をより広く調査するべきである。

最後に、本研究で得られた経験的な証拠と、データの収集や利用、管理をどのように行うべきかという規範的な議論とを結びつけることによって、AIを社会に有益な形で開発、利用するための仕組みづくりにつなげていく必要がある。一般市民の認知に関する情報は、規範的な議論に対して直接的に回答を与えるわけではない⁽¹⁰⁾。しかし、データの取り扱いに関する制度を市民にとって安心できるものにするためには、市民がどのようなリスクを重視しているかという情報があることで、より適切な制度設計に向けた議論が可能になるだろう。

謝辞

本研究の実施にあたり、公益財団法人セコム科学技術振興財団「特定領域研究助成：情報セキュリティ分野」、科研費24K21346の助成を受けた。

文献

- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58(2), 203–210.
- 細坪護拳・角田英之・加納圭・岡村麻子・星野利彦 (2020). 科学技術に関する国民意識調査——新技術の社会受容性—— *NISTEP RESEARCH MATERIAL*, No. 296. 文部科学省科学技術・学術政策研究所
- 小森めぐみ・武田美亜・高木彩 (2023). COVID-19第1波におけるリスク認知と予防行動——知識および社会間題リスク認知との関連—— *実験社会心理学研究*, 62(2), 195–207.
- 久木田水生・神崎宣次・佐々木拓・本田康二郎 (2025). AI・ロボットからの倫理学入門 名古屋大学出版会
- 松尾豊 (2015). 人工知能は人間を超えるか——ディープラーニングの先にあるもの—— KADOKAWA
- 中谷内一也・島田貴仁 (2010). 日本人のハザードへの不安とその低減 *日本リスク研究学会誌*, 20(2), 125–133.
- 日本経済新聞 (2025). 対話型 AI は「救世主」か 9月15日朝刊, 17.
- 西垣通・河島茂生 (2019). AI 倫理——人工知能は「責任」をとれるのか—— 中央公論新社
- Slovic, P. (1987). Perception of risk. *Science*, 236(4799), 280–285.
- 総務省 (2025). 情報通信白書令和7年版 <https://www.soumu.go.jp/johotsusintokei/whitepaper/r07.html> (2025年9月19日閲覧)
- 鈴木貴之 (編) (2020). 実験哲学入門 勁草書房

(10) 規範的な問いと経験的証拠の関係については鈴木 (2020) を参照。