

生成 AI に関する知識と認識に関する多国間調査

吉川 諒¹ 方 詩涛¹ 徐 安然¹ 矢谷 浩司¹

概要: 生成 AI が急速に普及し、多くの用途で使われ始めている一方で、利用者が生成 AI に関する基礎知識をどの程度有しているのか、そして生成 AI にどういった感情を抱き、評価しているのかは十分に明らかになっていない。そこで本研究では、クラウドソーシングサービスを用いて、4 カ国の居住者 (N=400) の生成 AI に関する知識や、生成 AI に対する主観的評価を尋ねる定量的調査を実施した。調査を通じて、生成 AI に関する基礎知識の認知度が、居住国、年代、性別によって異なることが明らかになった。また、主観的評価を問う設問の因子分析を行い、生成 AI に対する評価基準として「生成 AI の脅威」「生成 AI の利便性」「生成 AI の出力内容」の 3 項目が見出された。

1. はじめに

与えられたプロンプトに対してデータや情報を生成する人工知能技術である生成 AI は、文書生成や画像生成など多くの用途で利用されている。イギリスの政府機関、Ofcom [11] の 2023 年 9 月の調査によれば、イギリスの 16 歳以上のインターネット利用者のうち 3 割が生成 AI を使ったことがあり、さらに 16-24 歳ではその割合が 74% にのぼるとされており、若年層を中心に広く用いられていることがわかる。特に、ChatGPT^{*1}、Bard(Gemini)^{*2}等をはじめとするテキスト生成 AI は今後更に、教育現場や業務などの様々な局面で、多くの利用者に用いられるようになると考えられる。

高い利便性の反面、不正確な内容の出力や、機密情報の入力による情報漏洩の可能性など様々な課題も存在する。しかし、生成 AI が急速に普及したことから、十分な知識を持たずに生成 AI を利用する人もいると考えられ、生成 AI が悪影響をもたらすことが懸念される。生成 AI の適切な利用を推進するためには、基礎的な知識を啓発することが重要であるが、そのためには現時点でどのような知識がどの程度広まっているか認識することが欠かせない。しかし、生成 AI の利用経験等に関する調査はなされてきた一方で、生成 AI に関する基礎的知識がどの程度知られてい

るか調査した研究は少ない。

また、人々が生成 AI をどのように捉え、どのような感情を抱いているのかを調査する意義も大きい。生成 AI の普及以前から、人工知能技術に対する感情や評価について研究されてきた。例えば Kelley ら [8] は、人工知能に対する感情を Exciting, Useful, Worrying, Futuristic の 4 つの軸に沿って分析した。

生成 AI に対する感情についても研究されている。例えば HynBaek と Kim [6] によれば、生成 AI に対して信頼や不気味さといった様々な感情が存在し、一見相反するこれらの感情は時に両立することも明らかになっている。

一方で、生成 AI に対する評価軸にどのようなものがあるかについては、さらなる探究の余地がある。生成 AI に対する人々の態度を理解することは、生成 AI の普及が進む中で意義が大きいと言える。そこで本研究では、クラウドソーシングを用いた多国間調査によって、大規模言語モデルを用いた生成 AI に関する知識や利用習慣等について定量的調査を行うこととした。

本研究では以下のことを行った。

- 生成 AI の性能の限界や問題点等に関するトピックの探索。一般利用者向けの生成 AI ガイドラインの内容から、生成 AI を適切に利用する上で重要となる基礎知識を挙げ、4 つの大分類とそれを細分化した 9 つの小分類にまとめた。
- 生成 AI に関する認知度調査。上述の分類結果に基づいて生成 AI に関する 10 種類 (19 問) の知識問題を作成し、幅広い国や年代の参加者を募集して正答率を

¹ 東京大学 Interactive Intelligent Systems Laboratory
Interactive Intelligent Systems Laboratory, The University of Tokyo

^{*1} <https://chat.openai.com/auth/login>

^{*2} <https://bard.google.com/chat>

測定した。さらに、正答率と参加者の属性（年代、性別、居住国）の関連を分析した。

- 生成 AI の利用習慣や、生成 AI に対する感情・捉え方の調査。生成 AI の利用習慣、そして生成 AI に対する主観的評価を問うアンケートを作成し、クラウドソーシング調査を通じて実態を明らかにした。さらに、アンケート設問の因子分析を通じて、生成 AI に対するさまざまな感情や評価に通底する評価軸を見出した。

2. 関連研究

2.1 生成 AI に関する総説論文

生成 AI 技術について、その仕組みや活用方法、課題点等を網羅的に把握する調査研究がなされてきた。Wu ら [16] の研究では ChatGPT の利点や課題点について指摘した。この研究では、倫理的側面に加え、ChatGPT の演算が地球環境に及ぼす影響など、将来生じうる課題についても言及している。Ray [12] も ChatGPT についての調査研究を行い、ChatGPT をはじめとした大規模言語モデルの比較や、ChatGPT の利点・課題点について明らかにした。この論文では、いくつかの試験において、ChatGPT が人間の平均値を上回るなど、ChatGPT の高い性能を示す事例が紹介されている一方で、生成 AI の性能の限界や倫理的な課題、生成 AI 出力の偏りなどについて、多くの観点で指摘されている。

生成 AI の特徴や性能上の限界に関して理解することは、生成 AI を適切に利用する上で欠かせないといえる。

2.2 AI に関する知識やリテラシー

生成 AI が普及する前から、従来の人工知能について、利用者の知識や捉え方についての研究がなされてきた。Long と Magerko [9] は、過去の研究の調査を通じて「AI リテラシー」という語の定義を明確にし、AI を批評的に評価するために必要な能力として 17 個の項目を提示した。また、Ng ら [10] は、AI リテラシーに関する自己評価を問う質問項目、AILQ を開発した。この指標では、人工知能に関するリテラシーの習得を Affective learning, Behavioural learning, Cognitive learning, Ethical learning の 4 分類 (ABCE) に分類し、ABCE の枠に基づいた質問項目を作成した。

このように、人工知能技術に関する知識や能力を体系立てるとともに、それらを測定するための質問項目が設計されてきた。一方で、生成 AI について具体的にどのような知識が知られているのかについては十分に明らかになっていない。また、生成 AI の普及が進む中で、定期的に生成 AI に関する適切な知識の広まりを調査することの意義も大きい。

2.3 AI に対する感情や捉え方

生成 AI の普及以前から、AI という技術を利用者がどのように捉え、どのような感情を抱いているか研究されてきた。Kelley ら [8] は、8 カ国の調査参加者を対象に、AI に対する意見を調査し、Exciting, Useful, Worrying, Futuristic の 4 つの感情群に沿って各国の参加者の傾向を明らかにした。Kelley ら [7] はさらに、10 カ国の調査参加者を対象に、AI がプライバシーに及ぼす影響をどのように捉えているか調査した。調査で得られた回答の分類を通じて、Data at Risk, Highly Personal, Without Consent, State and Surveillance の 4 つの大きな懸念があることが明らかとなった。

生成 AI に対する感情についても研究が進められている。Hyn Baek と Kim [6] は、ChatGPT の利用者が、どのような動機で ChatGPT を利用しているのかを調査した。この調査では、ChatGPT が仕事効率化に役立つという評価と、ChatGPT に不気味さを感じるという感情の間に正の相関があることなどがわかった。このように、生成 AI に対する捉え方や感情は複雑に関連していることがわかる。また、Baldassarre ら [1] は、ChatGPT が与える善悪両面での影響、新たな動向、改善の余地のある点の 4 側面に関して、2023 年 5 月までに公開されていた論文やコラム、ブログ等のテキスト分析を通じて頻出の概念などを明らかにした。

AI に対する態度を測定する指標の開発も試みられてきた。Schepman と Rodway [13] は General Attitudes towards Artificial Intelligence Scale (GAAIS) という指標を開発し、AI に対する態度を測定することを目指した。この指標では、ポジティブな設問とネガティブな設問を 16 問ずつ作り、リッカート尺度による評価を収集した。Schepman と Rodway [14] はさらに、確認的因子分析を行い、GAAIS の内部一貫性を検証した他、Big Five との相関についても調査した。

このように、人工知能技術に対する感情や捉え方について研究が行われ、感情を測定する指標も開発されてきた。また、生成 AI に対する感情についても研究が行われている。本研究ではさらに、調査で尋ねる設問の因子分析を行い、生成 AI に対する感情の背景にある評価軸などを明らかにすることを試みる。

3. 設問の作成

3.1 生成 AI に関する基礎知識の分類

本研究では生成 AI の機能上の制約や不適切な利用方法に関する認知度調査を行う。そのためには、現在提起されている生成 AI の性能の限界や社会的な影響を把握し、その内容に基づいた設問を作成する必要がある。そこで、政府や自治体など、公的機関が発表している生成 AI ガイド

ラインの内容を分類することで、主要な知識を明らかにすることを目指した。公的機関の発表するガイドラインは一般利用者が知っておくべき事柄が網羅的にまとまっており、調査対象として適切だと考えられる。本研究では特に業務や教育への利用に関するガイドラインを対象にすることとした。

これらの条件に基づき、ガイドラインを選定した。対象のガイドラインは以下の通りであり、いずれも 2023 年の 7-9 月に公開された。

- (1) カナダ政府 [5] が公表した、一般利用者向けのもの
 - (2) イギリス政府 [4] が公表した、公務員向けのもの
 - (3) UNESCO[15] が公表した、各国政府や行政向けのもの
 - (4) 文部科学省 [19] が公表した、学校関係者向けのもの
 - (5) 東京都 [17] が公表した、一般利用者向けのもの
 - (6) 福岡県 [18] が公表した、県庁職員向けのもの
- 続いて、選定したガイドラインをもとに、生成 AI の利用上の注意点等に関する項目の探索を行った。手順は以下の通りである。

- (1) ガイドラインを読み、生成 AI を利用する上での倫理的な課題や注意点に関する項目を書き出した。
- (2) 書き出した項目を統合し、9 つの小分類を作成した。ここでは、各ガイドラインの子見出しとなるような内容を中心にトピックを抽出した。
- (3) 内容の類似した小分類をまとめ、大分類 4 つを作成した。
- (4) ガイドラインを再度読み、上記で作成した小分類に関する事項が含まれているかを確認した。

分類の名称については、次節の表 1 に示す。

3.2 知識問題の作成

前節で作成したカテゴリに基づき、生成 AI に関する知識問題を作成した。問題形式は全て選択式である。知識問題の題材は、生成 AI に関する知識、特に生成 AI の機能上の制約や適切な取り扱いについてである。全ての知識問題は、前章で作成した 9 つの小分類のいずれかに関連する内容とした。

知識問題の作成にあたっては、主に 2 点に注意した。まず、生成 AI サービスによって機能や仕様が異なることを踏まえ、一部のサービスにしか当てはまらない事項は、「～という特徴を持つ生成 AI サービスも存在する。」などの記述とした。また、細かな文言の差異が正答率に影響を及ぼす可能性を考慮した。特に、同じ内容であっても、肯定的な質問方法と否定的な質問方法で正答率に差が出る可能性がある。そこで、正誤問題については、同じ内容について、正解が“True”となる問題（問題 a）と“False”となる問題（問題 b）を 1 問ずつ作成することとした。

最終的に、正誤問題を 9 セット 18 問（Q1a-Q9b）と選

択式問題を 1 問（Q10）、合計 19 問を作成した。表 1 に、前節で作成した分類と、その内容に対応するクイズを示した。

表 1: 生成 AI に関する知識の分類と作成した知識問題の内容の対応関係を示した表。太字は大分類、その下の細字は小分類を指す。

分類	クイズ
入力する内容 個人情報・機密情報を入力してはならない	Q1,Q2
情報の正確性 数学的な計算には適していない 最新の情報を学習していないモデルも存在する 誤った内容や無関係な内容を出力することがある	Q3 Q7 Q10
出力の不安定さ 同じプロンプトに対して違う返答をすることがある 偏見や、攻撃的な内容を出力する可能性がある	Q4 Q8
著作権・出典 出力内容の根拠が不明な場合がある 生成 AI の出力物は、その旨を明記して公表すべきである 生成物が著作権を侵害する可能性がある	Q5 Q6 Q9

3.3 アンケートの作成

続いてアンケート形式の設問を作成した。まず、利用頻度と利用用途について、選択肢形式で問う設問を作成した（H1, H2）。

次に、生成 AI の利用方法を問う設問を 2 問（S1, S2）、生成 AI に対する感情・捉え方について尋ねる設問を 8 問（S3-S10）作成した。これらの 10 題はすべて、5 段階リッカート尺度で回答する形式である。生成 AI の利用動機には、信頼や不気味さといった感情が関連するとされている [6]。また、AI 技術がプライバシーを侵害するといった懸念を持つ人もいる [7]。これらの内容を参考にして、生成 AI に対して抱きうる評価・感情に関して問う設問とした。

設問作成にあたっては、設問の文言が回答内容に影響する可能性を考慮した。例えば、生成 AI に対する感情や捉え方については、ポジティブな内容とネガティブな内容が双方ある。そのため、両者が均等な設問数となるようにした。また、奇数番目の設問をポジティブ、偶数番目の設問をネガティブな内容の設問とした。

作成した設問の内容一覧を表 2 に示した。

4. 定量的調査

第 3 章にて作成した知識問題・アンケートについて、ク

表 2: 作成したアンケートの設問の趣旨. 生成 AI の利用習慣や生成 AI に対する捉え方について尋ねるもので, 全 12 問ある.

ID	設問の趣旨
H1	テキスト生成 AI の利用頻度
H2	生成 AI の利用用途
S1	私は生成 AI の出力内容の根拠を確認する.
S2	私は生成 AI に個人情報などを入力する.
S3	生成 AI は生活の利便性を向上させると思う.
S4	生成 AI はプライバシーを侵害すると思う.
S5	生成 AI の出力性能は質・量とも高いと思う.
S6	生成 AI の出力は不正確だと思う.
S7	生成 AI の出力は人間みたいだと思う.
S8	生成 AI は人間の能力を低下させると思う.
S9	生成 AI が役立つ場面は多いと思う.
S10	生成 AI の出力を良くする工夫が要ると思う.

クラウドソーシングを用いて回答を収集した. 調査はオンラインで完結し, ウェブ上のアンケートフォームによって質問に答えてもらった. この調査は, 事前に所属機関の倫理審査委員会からの審査・承認を受けた上で実施された.

4.1 参加者の募集

参加者は, クラウドソーシングプラットフォームである, Prolific^{*3}にて募集した. 参加者の条件は以下の通りとし, 募集要項に明記した.

- 20 歳以上であること.
- 英語でのアンケートを読み, 回答できること.
- アメリカ, カナダ, イギリス, ドイツの居住者であること. 生成 AI の利用者で, かつ英語の設問に回答できる参加者が多いと考えられるなどの条件から, この 4 カ国を対象とした.

各参加者は 1 回のみ調査に参加できることとした. アメリカ, イギリス, カナダ, ドイツの居住者から 100 名ずつ, 合わせて 400 名を募集し, 適切に回答を完了した参加者に報酬として 1 ポンドを支払った.

4.2 質問フォームの構成

質問フォームは, 以下の 4 部構成となっている.

- 参加者の情報. 年代, 性別, 居住国について尋ねた.
- 知識問題. 作成した正誤問題 9 問, 用語を問う選択肢問題 1 問を出題した. さらに, 後述するクオリティコントロール用の設問 1 問も含めた. 正誤問題とクオ

リティコントロールの設問は全て, “True”, “False”, “I cannot tell” の 3 択で回答する形式とした.

- アンケート. 表 2 に記した, 生成 AI の利用習慣や生成 AI に対する意識に関する説問を出題した. H1, H2 は利用習慣を問う設問で選択肢式, S1-S10 は利用の実態や生成 AI に対して抱く感情や捉え方を問う設問で, 5 段階リッカート尺度による回答形式となっている.
- **SeBIS [3]** の設問. 本調査では, SeBIS への回答も調査内容に含めた. SeBIS は, 情報セキュリティに関する習慣を尋ねる設問群で, 4 つのサブカテゴリに関する習慣の自己評価を定量化することができる.

3.2 節にて述べた通り, 正誤問題の作成時には, 1 つのテーマにつき正解が “True” になる問題 (問題 a), “False” になる問題 (問題 b) を 1 問ずつ作成した. これらの問題の問うている内容は実質的に同一であり, 両方とも出題すると双方の回答内容に影響を及ぼす可能性がある. そのため, フォームを 2 つ作成し, 問題 a, b を別のフォームで出題することとした. それぞれを Form I, Form II と呼ぶ. 設問内容の偏りを防ぐために, 両方のフォームに問題 a, 問題 b がほぼ同数含まれるように作成した. また, Form I, II には同人数 (各国 50 人ずつ, 合計 200 人) が回答するようにした.

また, 知識問題の出題部分では, “Please answer ‘True’ to this question.” という設問 1 題を含めた. この問題は, 知識問題の途中に出題されるクオリティコントロール用の設問である. 知識問題の出題部分は本設問を含め 11 問となった.

4.3 回答者数と参加者の属性

集まった回答のうち, 対象国以外の居住国を記述した回答 (計 2 件) を除外した. また, 同じフォームに複数回回答した, もしくは複数のフォームに回答した 5 名について, 2 件目以降の回答 (計 5 件) を除外した. また, タスクを完了したものの回答が確認されなかった 1 名のデータが欠如している. なお, クオリティコントロールの質問に誤答した参加者はいなかった.

結果として, 393 件の有効回答が集まった. 性別の内訳は男性, 女性, その他が各々 187, 198, 6 件, 居住国の内訳はアメリカ, イギリス, カナダ, ドイツが各々 96, 100, 99, 98 件, 年代の内訳は 20,30,40,50,60 代, 70 代以上が各々 167, 125, 51, 31, 15, 2 件であった. なお, 性別と年代はそれぞれ 2 名が回答を望まなかった. また, 参加者が回答したフォームは Form I, II で 198, 195 件だった.

次章以降で, 得られた回答のうち, 知識問題への正答率, リッカート尺度で回答する形式のアンケート設問への回答について詳しく分析を行う.

*3 <https://prolific.com>

5. 知識問題の正答率の分析

5.1 正答率の結果

表 3 に、全ての知識問題の正答率を示した。なお、Q9b において出題文言に誤りがあることが明らかとなった。本来の問題の正答は“False”だが、出題時の文言を踏まえ、ここでは“True”を正答とした際の正答率を記した。

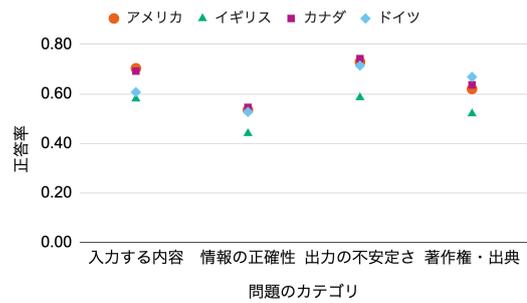
Form I と Form II の対応する問題の正答率の差をカイ二乗検定で分析したところ、平均正答率には有意な差は見られなかった。また、出題の誤りのあった Q9 を除く Q1–Q8 について、7 組において問題 a の方が有意に正答率が高かった。

表 3: 知識問題の正答率のフォーム別一覧と、正答率のカイ二乗検定の結果。Q9b は問題の不備があったが、正解を“True”とした場合の正答率を示し、分析も行なった。太字は比較ペアのうち有意に正答率の高かったものを示す。

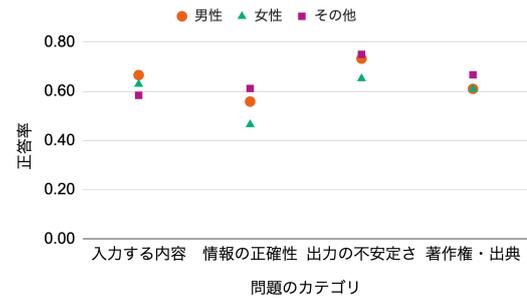
	出題した問題と正答率		統計量		
	Form I	Form II	$\chi^2(1,393)$	p	ϕ
Q1	a / 0.80	b / 0.63	13.42	<.001	0.19
Q2	b / 0.35	a / 0.81	83.98	<.001	0.47
Q3	a / 0.47	b / 0.09	70.91	<.001	0.43
Q4	b / 0.60	a / 0.92	54.24	<.001	0.38
Q5	a / 0.65	b / 0.45	14.34	<.001	0.20
Q6	b / 0.61	a / 0.73	6.11	.01	0.13
Q7	a / 0.76	b / 0.71	1.25	.26	0.06
Q8	a / 0.70	b / 0.55	8.58	<.01	0.15
Q9	b / 0.57	a / 0.77	16.61	<.001	0.21
Q10	a / 0.53	b / 0.51	0.06	.80	0.02
総合	0.61	0.60	0.19	.66	0.01

さらに、Q1-8, Q10 の問題をカテゴリ別に分け、参加者の属性別に正答率の散布図を作成した。問題が属するカテゴリは、表 1 に記した通りである。問題 a, 問題 b のいずれに回答したか区別をせずに正答率を集計した。また、性別や年代を回答しなかった参加者（各 2 名）は、数が少ないことから、散布図には含めなかった。年代が 70 代以上の参加者も少数であったため、60 代の参加者とまとめて「60 代以上」として取り扱った。

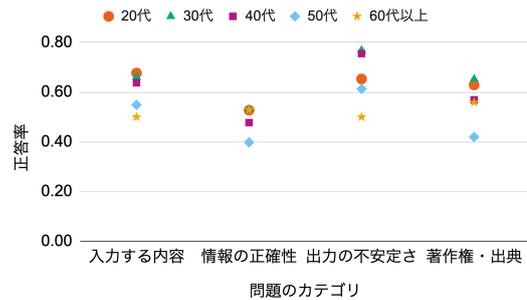
図 1 は、各カテゴリ内の問題の平均正答率を、参加者の居住国・性別・年代ごとに散布図に示したものである。散布図からは、年代や居住国などの属性によって、正答率に大きな違いがあることが見て取れる。



(a) 回答者の居住国ごとの正答率平均。



(b) 回答者の性別ごとの正答率平均。



(c) 回答者の年代ごとの正答率平均。

図 1: カテゴリ別の知識問題の正答率の平均を、参加者の属性別にプロットした散布図。

5.2 回答者の属性による正答数の差

前節の図 1 で見られた属性別の正答率の差が有意なものであるかを明らかにするため、参加者間の一次元分散分析 (One-way ANOVA) によって属性別の正答率の差を定量的に検証した。

各大分類 (表 1 で示したもの) ごとの正答数を変数として扱い、正答数に有意な差があるかを検証した。正答率ではなく正答数を変数としたのは、後の post-hoc テストで示される統計量の解釈が明確になるためである。カテゴリ別 (4 種) に、居住国・性別・年代 (3 種) による差を検証するため、合計 12 組の分散分析を行う。なお、年齢や性別を回答しなかった参加者のデータは分析に含めなかった。

まず変数の等分散性を検証するために、Bartlett テストを行ったが、全ての組み合わせで等分散性の仮定は棄却されなかった。そのため、続いて一次元分散分析を行なったところ、12 組のうち 8 組において有意差が確認された。

続いて、具体的にどのグループ間（例えば 20 代と 30 代など）に差があるか検証するため、一次元分散分析で有意な差が見られた 8 組について Tukey-Kramer 法を用いた post-hoc テストを行い、どの属性間に有意な正答率の差が見られるかを検証することとした。結果として、問題のカテゴリ別に、下記の通りの有意な差が見られた。

- 「入力する内容」について：有意差のあるペアはなし
 - 「情報の正確性」について：男性参加者は女性参加者より有意に正答数が多い（差は $0.28(p = .01)$ ）
 - 「出力の不安定さ」について：
 - アメリカ、カナダ、ドイツ在住の参加者はイギリス在住の参加者より有意に正答数が多い（差は $0.32(p = .00)$, $0.27(p = .02)$, $0.30(p = .01)$ ）
 - 男性参加者は女性参加者より有意に正答数が多い（差は $0.16(p = .04)$ ）
 - 30 代の参加者は 20 代、60 代以上の参加者より有意に正答数が多い（差は $0.23(p = .02)$, （差は $0.54(p = .01)$ ）, 40 代の参加者は 60 代以上の参加者より有意に正答数が多い（差は $0.51(p = .04)$ ）
 - 「著作権・出典」について：
 - ドイツ在住の参加者はイギリス在住の参加者より有意に正答数が多い（差は $0.30(p = .02)$ ）
 - 20 代、30 代の参加者は 50 代の参加者より有意に正答数が多い（差は $0.42(p = .03)$, （差は $0.47(p = .01)$ ））
- 一次元分散分析では、カテゴリと属性の組み合わせ 12 組のうち 8 組で有意差が確認されていたが、Tukey-Kramer 法で 1 つ以上のグループ間で有意差が確認されたのは 12 組中 6 組であった。この違いは、多重比較の手法の検出力によるものと考えられる。

5.2.1 結果の考察

一次元分散分析とその post-hoc テストにより、居住国・性別・年代ごとに正答率が異なる傾向が見られた。代表的な傾向として、男性や若年層で正答率が高く、イギリスの参加者の正答率が低いといった結果が確認された。

また、問題のカテゴリ別に見ると、「出力の不安定さ」に関する問題で属性ごとの正答数の差が最も多く確認された。一方で、このカテゴリの平均正答率は他の 3 つのカテゴリより高い。このことから、いわば常識的と思われる知識であっても認知度に差があることが示唆され、適切な啓発を行うことの重要性がわかる。

6. 生成 AI に対する感情・捉え方に関する設問の分析

6.1 設問への回答結果

生成 AI の利用実態や、生成 AI への感情や捉え方について尋ねた設問（表 2 の S1–S10）への回答結果を示す。これらの設問は、全てリッカート尺度で回答する形式である。

表 4 に各設問への回答の平均と標準偏差を示した。

表 4: 生成 AI に対する感情や捉え方を尋ねる設問への回答の平均と標準偏差。

(a) ポジティブな内容の設問 (b) ネガティブな内容の設問
(奇数番目) の回答結果. (偶数番目) の回答結果.

	μ	σ		μ	σ
S1	2.82	1.25	S2	1.64	0.91
S3	3.41	0.97	S4	2.90	1.00
S5	3.46	0.86	S6	3.09	0.94
S7	2.37	1.08	S8	3.03	1.09
S9	4.06	0.86	S10	3.68	0.87

6.2 生成 AI に対する主観的評価の因子分析

続いて、生成 AI に対する感情や捉え方に関する設問への回答結果から、生成 AI に対する主観的評価が、具体的にどのような評価軸に基づいているのかを探索する。生成 AI に対する主観的評価に通底する要因が明らかになれば、その要因を定量的に評価する指標を作成し、測定したスコアの文化間比較などを行うことも可能となる。本研究では、生成 AI に対する主観的評価の背景となる要因の解明に向けて、設問の因子分析を行う。

6.2.1 設問の一貫性

評価尺度では、各設問の内容が同じ概念を問うものであること、すなわち一貫性があることが求められる。一貫性の評価の中で相関係数の計算を行うが、その際には逆転項目のスコアを反転させ、正しく相関関係があらわれるようにする必要がある。本研究では、偶数番目の設問（生成 AI に対する否定的な捉え方を問う設問）への回答スコアを逆転する処理を行った。

続いて、設問の一貫性の指標であるクロンバックの α [2] を計算した。 $0 \leq \alpha \leq 1$ であり、 α が大きいほど設問群の一貫性があることを示す。本調査で得られた回答について α を計算したところ、 $\alpha = 0.43$ となった。続いて各設問の Item-Total Correlation(項目テスト相関) を計算した。これはある質問項目への回答と、全ての設問項目の得点合計の和の相関係数を指す。結果として項目テスト相関が低い項目として S1(0.31), S2(0.00), S10(0.07) が明らかとなった。これらの設問は、生成 AI に対する評価より、生成 AI の利用習慣に関する内容であるともいえることから、設問の内容と項目テスト相関に基づいて S1,S2,S10 を因子分析の対象外とした。結果として α は 0.61 に向上した。

6.2.2 因子分析

続いて、この 7 問に対して因子分析を行った。まず、Bartlett の球面性検定を行ったところ、 $\chi^2(45) =$

620.93($p < .001$)となり、設問への回答間の相関が有意にある、すなわち背後に因子が存在しうることがわかった。そのため、これらの設問に通底する因子を検証する。図2にスクリープロットと平行分析の結果を示した。スクリープロットの傾き（急峻になるところ）と平行分析の結果（図2の赤線と青線が交わる点）を踏まえ、本研究では適切な因子数を3とした。

続いて、因子負荷の計算を行った。因子負荷は、3つの因子が各設問にどのように影響するかを示すものである。ここでは最小残差法を用い、因子軸の回転方法にオブリン回転を指定した。0.25以上因子負荷の一覧を表5に示した。これは先行研究のSeBIS [3]と同じ基準である。

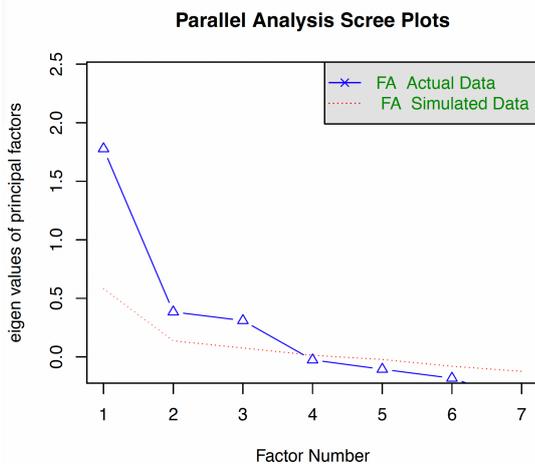


図2: S3-S9の7設問のスクリープロット。縦軸が因子の固有値を表し、横軸が因子の数を表す。青実線が実際の固有値を表し、赤点線が乱数生成されたデータの固有値を表している。

表5: 3つの因子を仮定し、設問毎に因子負荷を記した表。最小残差法を用いて計算し、0.25を上回ったもののみを記載している。

	S3	S4	S5	S6	S7	S8	S9
因子1		0.97		0.38		0.29	
因子2	0.31						0.99
因子3	0.38		0.78	0.26	0.53		

6.2.3 因子の表す内容の解釈

ここまでの解析により、生成AIに対する感情や捉え方を問う設問の背景に、3つの因子があることが推定された。それぞれ以下の通りに対応が見られた。

- 因子1...S4, S6, S8
- 因子2...S3, S9
- 因子3...S3, S5, S6, S7

因子に対応する設問の内容を踏まえ、因子1-3の内容をそ

れぞれ「生成AIの脅威」「生成AIの利便性」「生成AIの出力内容」と決定した。この結果から、調査の参加者は、生成AIがもたらす利便性と弊害、現時点での出力の性能を加味しながら生成AIに対する評価を行っているとし唆される。

一方で、因子に関連する設問の内容を見ると、因子1（「生成AIの脅威」）、因子2（「生成AIの利便性」）はそれぞれ否定的な内容、肯定的な内容の設問のみにあらわれている。さらに因子3（「生成AIの出力内容」）はS6を除き肯定的な内容の設問にあらわれている上、S6の因子負荷の値も比較的低い。このことから、設問が問うている内容より、設問の内容が肯定的か否定的かが強く影響している可能性も考えられる。

7. 結論と今後の展望

7.1 結論

本研究では、生成AIに関する知識や利用習慣、生成AIに対する感情や捉え方などについて、幅広い居住国、年代の参加者を対象とした調査を行った。

まず、生成AIの利用ガイドラインの分類を通じて、生成AIに関する基礎知識を4つの大分類と、それを細分化した9つの小分類に分類した。そして、分類の内容に基づき作成した知識問題の正答率調査を通じて、生成AIに関する基礎知識の認知度が明らかになった。また、年代・性別・居住国によって正答率に差があることも判明した。本調査で得られた知見は、今後生成AIの利用者が一層増加すると考えられる中で、知識を啓発し、適切な生成AI利用を推進する上で役立つと期待される。

本研究では、生成AIに対して利用者が抱く感情や捉え方についても調査した。さらにそれらの判断基準を明らかにすることを旨とし、因子分析を行った。本研究では10の設問を作成して参加者に回答を依頼した。因子分析を通じて、7つの設問に通底する3つの因子が見出された。それぞれを「生成AIの脅威」「生成AIの利便性」「生成AIの出力内容」と決定した。因子分析により、生成AIがもたらす利便性、脅威、そして出力の性能が生成AIに対する評価の基準になっていることが示唆された。

7.2 本研究の制約と今後の展望

本研究には、複数の制約・限界も存在する。その一つが参加者の偏りである。募集条件や、そもそもクラウドソーシングで参加者を募集したことなどから、参加者の属性や生成AIの利用経験などに偏りがあった可能性がある。今後はクラウドソーシング以外の参加者募集方法を用いることや、より多様な居住国から参加を募ることで、普遍的な参加者からの回答結果が得られると考えられる。

また、知識問題や設問の個数、内容に関しても今後拡張

の余地があるといえる。今回作成した設問は、2023年時点での生成AIの現状を元に作成したものであり、題材も大規模言語モデルに基づく生成AIを対象としていた。今後は画像生成AIに関する問題なども追加することで、生成AIに関する認知度をより様々な面から測定できると期待される。さらに、生成AI技術は進歩を続けており、それに伴い社会的な影響も変容している。設問の内容も最新の状況に合わせて改変する必要がある。

また、生成AIに対する主観的評価について、設問の追加や改善、再度のアンケート調査などを通じて、体系的に明らかにすることも大きな意義がある。生成AIに対する考え方を決定する要因が明らかになれば、その要因を定量化する指標を作り、スコアの差について統計分析を行うことができる。統一的な指標が開発されれば、多数の研究・調査間で利用され、広範かつ長期間にわたる結果の比較も可能になる。

これらの研究が進むにつれ、今後生成AIが幅広く、適切に利用されるようになるために大きな役割を果たすと期待される。本研究はその実現に向けた端緒として位置付けられる。

謝辞 研究を進めるにあたり多くの助言を頂いた研究室のメンバーと、調査参加者の皆様に深くお礼申し上げます。

参考文献

- [1] Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D. and Ragone, A.: The Social Impact of Generative AI: An Analysis on ChatGPT, *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, GoodIT '23, New York, NY, USA, Association for Computing Machinery, p. 363–373 (2023).
- [2] Cronbach, L. J.: Coefficient alpha and the internal structure of tests, *Psychometrika*, Vol. 16, No. 3, pp. 297–334 (online), available from (<https://doi.org/10.1007/BF02310555>) (1951).
- [3] Egelman, S. and Peer, E.: Scaling the Security Wall: Developing a Security Behavior Intentions Scale (Se-BIS), *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, Association for Computing Machinery, p. 2873–2882 (2015).
- [4] Gov. UK: Guidance to civil servants on use of generative AI (2023). <https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai>.
- [5] Government of Canada: Guide on the use of Generative AI (2023). <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html>.
- [6] Hyun Baek, T. and Kim, M.: Is ChatGPT scary good? How user motivations affect creepiness and trust in generative artificial intelligence, *Telematics and Informatics*, Vol. 83, p. 102030 (2023).
- [7] Kelley, P. G., Cornejo, C., Hayes, L., Jin, E. S., Sedley, A., Thomas, K., Yang, Y. and Woodruff, A.: "There will be less privacy, of course": How and why people in 10 countries expect AI will affect privacy in the future, *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*, Anaheim, CA, USENIX Association, pp. 579–603 (2023).
- [8] Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T. and Woodruff, A.: Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries, *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, New York, NY, USA, Association for Computing Machinery, p. 627–637 (online), available from (<https://doi.org/10.1145/3461702.3462605>) (2021).
- [9] Long, D. and Magerko, B.: What is AI Literacy? Competencies and Design Considerations, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA, Association for Computing Machinery, p. 1–16 (2020).
- [10] Ng, T. K., Wu, W., Leung, J., Chiu, T. K. and Chu, S.: Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach, *British Journal of Educational Technology*, pp. 1–23 (2023).
- [11] Ofcom: Online Nation – 2023 report (2023). https://www.ofcom.org.uk/_data/assets/pdf_file/0029/272288/online-nation-2023-report.pdf.
- [12] Ray, P. P.: ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, *Internet of Things and Cyber-Physical Systems*, Vol. 3, pp. 121–154 (2023).
- [13] Schepman, A. and Rodway, P.: Initial validation of the general attitudes towards Artificial Intelligence Scale, *Computers in Human Behavior Reports*, Vol. 1, p. 100014 (online), DOI: <https://doi.org/10.1016/j.chbr.2020.100014> (2020).
- [14] Schepman, A. and Rodway, P.: The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust, *International Journal of Human-Computer Interaction*, Vol. 39, No. 13, pp. 2724–2741 (2023).
- [15] UNESCO: Guidance for generative AI in education and research (2023). <https://unesdoc.unesco.org/ark:/48223/pf0000386693>.
- [16] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L. and Tang, Y.: A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development, *IEEE/CAA Journal of Automatica Sinica*, Vol. 10, No. 5, pp. 1122–1136 (2023).
- [17] 東京都デジタルサービス局：文章生成AI利活用ガイドライン version 1.2 (2023). https://www.digitalservice.metro.tokyo.lg.jp/ict/pdf/ai_guideline.pdf.
- [18] 福岡県生成AI検討プロジェクトチーム：生成AI庁内利活用ガイドライン(2023). <https://www.pref.fukuoka.lg.jp/uploaded/attachment/209591.pdf>.
- [19] 文部科学省：初等中等教育段階における生成AIの利用に関する暫定的なガイドライン ver 1.0 (2023). https://www.mext.go.jp/content/20230710-mxt_shuukyoo2-000030823_003.pdf.