

奨励金No.1613

検索エンジン・生成 AI のサービス利用がアンコンシャス・バイアスの拡散に与える影響に関する心理学研究

山田 歩

滋賀県立大学人間文化学研究院 准教授

谷口 友梨

滋賀県立大学人間文化学研究院 准教授

Psychological research on the impact of search engine/generative AI service use on the spread of unconscious bias

Ayumi Yamada

The University of Shiga Prefecture, Associate Professor

Yuri Taniguchi

The University of Shiga Prefecture, Associate Professor

本研究は、検索エンジンや生成 AI のサービス利用がアンコンシャス・バイアスの拡散に与える影響を検証した。実験参加者は「手描き」、「Google 検索」、「いらすとや」、「Copilot」のいずれかの手段を用いて、「小学生に向けた職業紹介に使用するイラスト」を用意する課題に取り組んだ。実験の結果、サービス上でアクセスできる画像に表現上の偏りがあるほど、実験参加者はその偏った表現の画像を選択しやすくなることが確認された。

This study examined the impact of using search engines and generative AI services on the spread of unconscious bias. Participants were tasked with creating an illustration for introducing occupations to elementary school students using one of four methods: hand-drawing, Google Search, Irasuto-ya, or Copilot. The experiment revealed that the more biased the images accessible on a given service, the more likely participants were to select images reflecting that bias.

1. 研究内容

1.1 研究背景・目的

IT や AI の情報技術は生活に大きな恩恵をもたらす一方で、その副作用も懸念されている。本研究は、検索エンジンや生成 AI のサービス利用が、アンコンシャス・バイアスの拡散に与える影響を検討する。アンコンシャス・バイアスとは「無自覚の偏見・思い込み」を指す。意図的に表出される偏見と異なり、特定の社会集団に対して無自覚的に適用される偏見である。画像検索や画像生成の既存サービスがアウトプットする「画像」には、「男は仕事、女は家庭」といったように男女の役割

に固定的な観念をあてはめる「ジェンダー・ステレオタイプ」をはじめ、表現の偏りが広範にみられる。

本研究は、検索エンジンや生成 AI のサービス利用がアンコンシャス・バイアスを拡散する原因になっている可能性があることを行動実験によって検証する。実験では、画像検索および画像生成の既存サービスを実際に使用して特定の資料を作成する課題を用いて、これらのサービス利用がユーザーの行動に与える影響が検討される。サービス上でアクセスできる画像に表現上の偏りがあるほど、多様性に欠ける画像が使用されやすくなると

予想する。さらに、ジェンダー意識などの規範意識も実験参加者から測定し、これらの規範意識が、アンコンシャス・バイアスの拡散・抑制に果たす役割に関する知見も得る。

1.2 方法

1.2.1 実験参加者

226名の大学生。PCトラブルなどが確認された回答を除いた、191名の回答を最終分析対象とした。

1.2.2 手続き

参加者は4つの条件に無作為に割り振られ、実験者が指定した方法で「小学生に向けた職業紹介に使うイラスト」を用意する課題に取り組んだ。職業は「研究者」と「警察官」の2つが指定された。本報告書では紙面の都合から「研究者」の結果を報告する。

「手描き条件」に割り振られた参加者（N=41）は手描きで課題のイラストを作成した。「Google検索条件」に割り振られた参加者（N=48）はGoogle検索を用いて課題のイラストを検索し、最も適したものを1つ選ぶように求められた。検索語は「研究者 イラスト」に固定した。「いらすとや検索条件」に割り振られた参加者（N=43）はフリー素材サイト「いらすとや」で課題のイラストを検索し、もっとも適したものを1つ選ぶように求められた。検索語は「研究者」に固定した。「Copilot生成条件」に割り振られた参加者（N=59）は、生成AIサービスのCopilotを用いて課題のイラストを計10枚の画像を生成し、最も適したものを1つ選ぶように求められた。指示文は「研究者 イラスト」に固定した。当条件の参加者は課題の中で生成された10枚の画像も全て提出した。

上記の手続きの後、「男女共同参画社会に関する意識」（内閣府、2019）を短縮・一部改変したものについて回答を求めた。

1.2.3 情報環境の把握

本実験で定めた検索方法・生成方法で得られる画像のジェンダー表現にどの程度のバイアスがみられるかを確認した。「Google検索」と「いらすとや」については、実験参加者とは独立した環境で、同じ実験日、同じ検索語を用いて同サービス利用を通して得られた検索結果、「Copilot」については、実験参加者が実際に得た生成結果について、その内容を分析した。各サービスから出力された画像について、そのジェンダー表現に従って、①男性メイン、②女性メイン、③中立のいずれかに分類された。Google検索といらすとやを比較すると、Google検索のほうが男性メインに描かれた研究者イラストが多かった。また、その傾向は上位表示される画像ほど強かった。Copilotについては、実験参加者全体でみると、ジェンダー表現に大きな偏りはなかったが、参加者個人でみると特定のジェンダー表現の画像が多く出現するなど、ジェンダー表現に偏りが生じるケースが多く確認された。

1.3 結果

1.3.1 提出イラストの分類とカウント

各条件の参加者が「職業紹介に使用するイラスト」として提出したイラストを、そのジェンダー表現に従って、①男性メイン、②女性メイン、③中立のいずれかに分類し、カウントした（図1）。

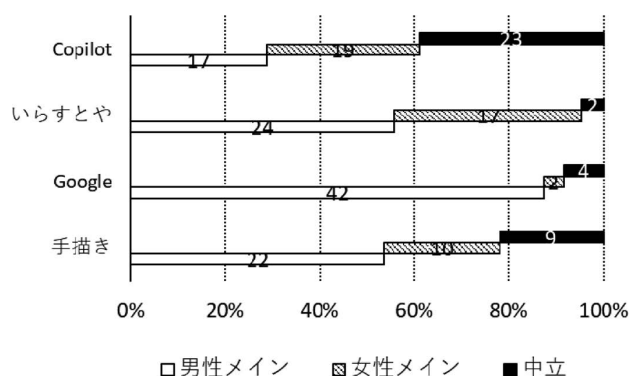


図1 職業紹介用として提出された研究者イラストのジェンダー表現

1.3.2 提出イラストにおけるジェンダー・バイアス

上記分類データから構成された「実験条件」と「提出イラストのジェンダー表現」の分割表を対象にカイ二乗検定をおこなった。その結果、職業紹介用イラストを用意する際に利用するサービスによって、参加者が最終的に提出するイラスト画像のジェンダー表現が異なることが確認された ($\chi^2(6, 191) = 48.87, p < .01$)。

先述したように、Copilot 条件では参加者によって生成画像のジェンダー表現には偏りがあった。そこで、特定のジェンダー表現のイラストの出現頻度ごとに、そのジェンダー表現のイラストがどの程度選択されたかについて集計をおこなった。この集計結果からは、特定のジェンダー表現の出現頻度が高いほど、そのジェンダー表現のイラストが選択されやすいことが確認された ($r = 0.93$)。

1.3.3 規範意識とジェンダー・バイアス

実験のなかで測定された「男女共同参画社会に関する意識」のうち「今後女性がもっと増える方がよいと思う職業」として「研究者」を選んだ参加者と選ばなかった参加者に分け、両グループの提出イラストを比較した。しかし、男女共同参画に関する意識は、参加者の行動に影響を与えていなかった。

1.4 考察

著者が実験参加者とは独立しておこなった調査、実験参加者が実際に各種サービスを利用するなかで得られた出力を内容分析した結果からは、男性メインの研究者イラストがアウトプットされた割合は「Google 検索」が最も高く、次いで「いらすとや」、そして「Copilot」の順であることが確認された。また「Copilot」では参加者個人レベルでは偏りが生じていることも確認された。

実験の結果、サービス上でアクセスできる画像に表現上の偏りがあるほど、実験参加者はその偏った表現の画像を選択しやすくなることが確認

された。たとえば、「Google 検索条件」では実験参加者の 88 パーセントが研究者の職業紹介用イラストとして男性メインのイラストを提出した一方で、「いらすとや検索条件」におけるその割合は 56 パーセント、「Copilot 生成条件」では 30 パーセントであった。この順位は、各情報技術サービスの情報環境で見られたジェンダー表現の偏りの順位と一致している。さらに、Copilot 生成条件から得られた結果について、実験参加者ごとに生じた表現の偏りを加味した分析をすると、利用可能性の高いジェンダー表現のイラストが選ばれやすくなる傾向があることが確認された。これらの知見をあわせると、情報技術サービスは、それぞれが構成する情報環境の性質に応じて、ユーザーから異なる行動を引き出していることが示唆される。

また本研究では、「男女共同参画社会に関する意識」についても実験参加者から回答を得た。「男女共同参画社会に関する意識」から測定された規範意識は実験参加者の行動を調整することはなかった。インターネットサービスを利用している最中は、ユーザーは比較的無自覚なたちでその情報環境から影響を受けることが示唆される。アンコンシャス・バイアスの低減を狙った様々な啓発運動や教育が取組まれているが、ユーザーの行動に対してより直接的な影響を与えることができる手法の開発に力点を置くなど、規範意識の醸成以外の取組みに力をいれる必要があることが示唆される。

2. 発表（研究成果の発表）

山田歩・谷口友梨（2024）. 検索エンジンの使用がジェンダー・バイアスを拡散させる 日本心理学会第 88 回大会（熊本城ホール、2024 年 9 月）

山田歩・谷口友梨（2024）. 生成 AI の利用はジェンダー・バイアスを拡散させるか 日本心理学会第 89 回大会（東北学院大学、2025 年 9 月）