

研究経過報告書

2025年 4月 15日

研究員 (留学生)	所属 政経学部 職 准教授 氏名 加藤 将貴
派遣期間	2024年 4月 1日 ~ 2025年 3月 31日
研究主題等	情報の信用評価における認知バイアス AIの先端研究と技術動向・環境変容
報告事項	<p>(研究活動の概要、内容、成果等、添付書類の見出し等)</p> <p>1. 研究活動の概要</p> <p>派遣期間中は東京大学大学院情報学環・学際情報学府に受け入れていただき、フェイクに対する認知的・情動的要因について研究調査や分析を行った。加えて、近年では生成AIをはじめとする情報処理技術の発展と、それに伴う社会的・倫理的課題の多様化など、研究領域周辺の環境変化が急速に生じており、最新の研究や技術動向を把握する必要性を強く感じた。そのため、専門的な知見の深化と多角的な視点の獲得を目的として、情報処理学会や情報ネットワーク法学会といった関連学会に参加し、技術発展の趨勢やAI活用・倫理、制度整備などに関する研究発表、議論に触れる機会を得た。</p> <p>2. 研究の背景</p> <p>本研究の背景は、フェイク情報の社会的影響力増大に対する懸念である。かつて、画像や動画の加工には高度な専門知識と高性能な機器・ソフトウェアが求められ、写真や動画の真正性は高いものとされていた。しかし、ハードウェアの性能向上とソフトウェアの発展により、データの加工が容易に行えるようになり、SNSや動画共有サイト利用者数の急増と相まって、インパクトの強い情報が、真偽や正確性が十分に検証されないまま拡散される事態が増加している。インターネットを介して広がったフェイクが、人々のプライバシー侵害や人権侵害を惹起したケースは枚挙にいとまがなく、冤罪による殺人事件の引き金となった事例も存在する。情報メディアが大きな影響力を持つ現代において、悪意のあるフェイクが拡散した場合、国家の安全保障すら脅かしかねない。</p> <p>3. 研究内容</p>

3. 1 情報と認知バイアス

フェイク情報の信憑性に対する認知には、複数の学術分野において多角的な検証が進められており、既存の知見を整理すると、概ね以下の5つの要因に分類される。

(1) Cognitive Biases

人間の判断は、しばしば認知バイアスに左右されることが知られており、こうした認知バイアスはフェイク情報の受容過程において重要な役割を果たす。代表的なものとして以下が挙げられる。

・ Confirmation Bias

自身の信念や価値観と一致する情報を選好し、反証的な情報を排除する傾向。特に政治的・宗教的主張と結びついたフェイク情報に対して顕著に見られる。

・ Anchoring Bias

最初に接触した情報が後続の判断基準として機能し、情報評価に過大な影響を及ぼす傾向。

・ Availability Heuristic

印象的で想起しやすい情報が、実際よりも頻度が高く、信頼性が高いと判断されやすい傾向。

・ Representativeness Heuristic

特定の情報がステレオタイプや既存の枠組みに合致する場合に、その情報を真実と見なしてしまう傾向。

・ Illusory Truth Effect

同一の情報に繰り返し接触することで、それが真実であるかのように錯覚される現象。

(2) Emotional Factors

「恐怖」「不安」「怒り」「正義感・義憤」「共感」「驚き」といった情動は、合理的な情報処理を妨げ、直感的・感情的な受容行動を引き起こす要因となりうる。

フェイク情報の多くは、これらの情動を意図的に刺激するよう設計されており、受容者の批判的思考を回避させることで、真偽を問わず拡散されやすい傾向にある。特に、感情的に反応しやすい内容を含むフェイク情報は、事実性や論理的整合性よりも印象や共感を通じて受け入れられ、SNS等における情報拡散においても強い影響力を持つことが、先行研究により示されている。

(3) Lack of Information Literacy

情報リテラシーの不足は、しばしばフェイク情報を批判的に検討する力を損ない、受容・拡散のリスクを高める要因となる。具体的には、以下のような認知的資源の欠如が指摘される：

- ・ 情報源の信頼性評価に対する不十分な認識
- ・ デジタル情報の加工や改変を見抜く知見や経験の不足

・複数の情報源を比較検証する習慣や能力の欠如
特に若年層においては、表層的に整った情報や検索結果の上位表示に過度な信頼を置く傾向が強く、事実確認や出典検証を怠ることが多いことが報告されている。

(4) Social Factors

フェイク情報の信頼性評価や拡散行動には、個人の属する社会的環境や集団的影響を受けることが多く、これらはSNSにおける情報の選択と拡散とも関連性が深い。

- ・ **Conformity Pressure**

周囲の意見に迎合しようとする心理的傾向。

- ・ **Social Identity**

自らの所属する集団に都合の良い情報を信じやすくなる傾向。

- ・ **Echo Chamber**

意見の異なる情報が遮断され、同質的な意見のみが強化される情報環境。

- ・ **Bandwagon Effect**

多くの人が信じている情報であるという事実そのものが、その情報の信頼性を保証するかのよう機能する現象。

(5) Authority Bias

情報の信頼性評価において、内容の正確性、整合性よりも、肩書きや社会的地位といった「外的信頼性指標」に基づいて判断が行われる傾向があり、これらは、専門家や著名人、政府機関などから発せられる情報に対して、内容の精査を経ずに受容してしまうという行動傾向として現れる。

以上のように、フェイク情報の信憑性認知に影響を及ぼす要因は多岐にわたり、これらはしばしば複合的に作用することから、単一の視点からでは十分に説明できないことが多い。したがって、認知的知見にとどまらず多層的視点からの分析が求められる。

3. 2 先端研究・技術動向の把握

更に、研究を深化させるにあたり、近年急速に進展している生成AIなどの技術発展の趨勢やAI活用の実情、倫理や制度整備に関する知見を取り込む必要があると判断した。技術の発展はフェイク情報の精緻化と可視的信頼性の擬似的強化をもたらしており、従来の枠組みでは対処困難な局面が増加しているためである。

このような観点から、派遣期間中に情報処理学会ならびに情報ネットワーク法学会といった関連学会に参加し、技術的進展や法的規制の動向、ならびにAI倫理を巡る議論等に触れる機会を得た。

情報ネットワーク法学会（第24回研究大会）においては、基調講演であった「偽・誤情報対策とニューストラスト」が印象深く、偽・誤情報への対策の現状と課題について具体的な指摘がなされた。特

に「健全性」や「偽・誤情報」の定義が不明確であること、表現の自由やメディア規制との関係において慎重な議論が求められることなど、対策にあたって制度設計の難しさが強調されていた。その他、AI技術の進展に伴う新たな法的課題へのアプローチも紹介され、生成AIと法制度の接点について多くの示唆を得るものであった。

また、情報処理学会（第87回研究大会）では、情報技術分野における最前線の研究成果が幅広く発表されており、学術的および技術的側面の両面から多角的な視座に触れることができた。とりわけ「万有知能化の世紀へ -実世界に行きわたる人工知能-」と題されたセッションでは、AIによって知識の遍在化が進みつつあるとして、実世界におけるAIの現在地について理解を深めることができた。

4. 今後の教育・研究への活用

本派遣期間中に得られた知見及び示唆は、今後の教育・研究において活用する。教育面においては、生成AIやフェイク情報に関連する諸問題への対応が重要性を増す中、学生に対して情報の真偽を評価する能力や、情報の発信・受容に関する倫理的判断力を涵養することを重視したい。研究面においては、今後もフェイク情報の認知に関する理論的探究と研究成果の公表を進めていきたい。

5. 謝辞

本研究活動を遂行するにあたり貴重な1年間の研究期間を賜り、研究に専念する機会を与えてくださった本学ならびに関係各位、併せて、私の不在中に業務をご担当くださった教職員の皆様のご理解とご協力に、この場をお借りし心より感謝申し上げます。