

生成 AI は人間のように信頼されるべきか？

—生成 AI によるフィードバックに対する大学生の見解—

Should artificial intelligence be trusted as humans?

—Perspectives towards trust issues of generative AI feedback by university students—

大滝 文一[†], 大島 律子[†], 大島 純[†]
Bunichi Otaki, Ritsuko Oshima, Jun Oshima

[†]静岡大学

Shizuoka University

otaki.bunichi.23@shizuoka.ac.jp

概要

本研究では、AI と人間の教員によるフィードバック (FB) に対する学生の認識を、信頼と学びのメタファの観点から議論することを目的とした。データ収集として、スウェーデンの大学で学部生と修士学生を対象にフォーカス・グループ・インタビューを実施し、再帰的テーマティック分析を行った。結果から、生成 AI は獲得メタファに基づく文法やスペル修正における能力信頼を獲得し得るが、人間の教員は参加メタファに紐づく、関係性や時間軸を伴うコミュニケーション信頼を得ている観点を考察した。

キーワード：人工知能, 学習, メタファ

1. はじめに

2022 年以降、生成 AI 技術の基盤となる大規模言語モデルが急速に発展し、高等教育における学習機会への活用が積極的に模索されている。UNESCO の報告書 [1] では、高等教育場面における生成 AI の様々な役割を擬人的なメタファで表現することで、人間の教師や学生同士が担ってきた役割を AI に付与するなど、AI と人間との協調的な学びが期待されている。具体的には、議論を活性化させる「ソクラテス的な対話相手 (Socratic opponent)」,あるいは共同作業における課題解決を促進する「協働コーチ (collaborative coach)」など、従来は人間の教師や学生同士が担ってきた役割を、AI が擬人的なメタファを付与された形で担うことが期待されているという点が注目に値する。

一方で、人工知能と脳・認知科学間における学問領域間の言語交差において、conceptual borrowing のプロセスで生まれる悪影響を懸念する考えもある。例えば Floridi and Nobre (2024) は「機械学習」という言葉が脳科学や認知科学で用いられる際、人間や動物がある環境における経験の結果として新しい態度や精神を習得

したり、既存の知識を修正したりするような学習概念を意味するものではないと指摘している [2]。

また、人間の学習については Paavola and Hakkarainen (2005) が 3 つの学習メタファ (獲得, 参加, 知識創造) を用いてその詳細を論じている [3] が、人同士あるいは人と物との対話プロセスの中で頻繁に学習が起きる参加や知識創造の学習メタファの特徴を見る限りでは、このような学習に AI を導入しても AI と人間との対話的な学びが人間同士と同じように行われるとは捉え難い。このような学習メタファを通して人間の AI への認識を調査することにより、AI との学習プロセスにおける潜在的な限界や課題を明確にし、より効果的な教育・学習設計や実践的な応用に役立てることができると考える。

これらの背景をもとに、本研究では人間と AI との対話的な学びを検討する際に信頼と FB という側面から議論する。信頼は学習者の主体性や学習プロセスにおいて重要であり [4], また対話の中で頻繁に起こる相手からの FB に深い関連があるとされている。関連する研究として信頼と FB の関係を「能力信頼」と「コミュニケーション信頼」として論じたもの [5] や、生成 AI が能力信頼だけでなく人間のようにコミュニケーション信頼も獲得すべきか、という問いへの批判的検討の必要性を提起したもの [6] などがある。

今回は上記の問題意識と学びのメタファ及び信頼に関する概念的枠組みを踏まえ、対話型 AI による FB を人間の教員によるそれと比較することで、学生がどのような学びのメタファの元にそれらの差異を理解し、信頼しているかを明らかにすることを目的とした。

2. 方法

2023 年 3 月から 4 月にかけてスウェーデンの 2 つの国立大学にて、17 名の大学生・大学院生を対象にフォ

ーカス・グループ・インタビューを実施した。フォーカス・グループは、研究トピックに焦点を当てて、被験者にインタラクティブな議論をさせることによってデータを生成するように設計されている (Wilkinson, 2004)。

2.1 被験者

被験者 17 名はスケジュールの都合を見ながら 5 つに分け (1 グループ 3~4 名), グループ 1 と 2 は修士課程の学生, グループ 3 から 5 は修士課程と学士課程の学生の混合グループで編成した。被験者の全員が勉強のために生成 AI を使用した経験を持っていた。国際性に富むスウェーデンの大学の性質上, 被験者の国籍はスウェーデン, ポルトガル, イタリア, ラトビア, イラン, パキスタン, 中国, 日本から, また所属学部は社会科学, 自然科学, IT, 教育, 人文系学部からの参加があった。

2.2 インタビュー

質問テーマは生成 AI と人間の対話的 FB について大きく次の 4 つのカテゴリーの質問を中心に構成した: (1)人間と AI が生成する FB の比較, (2)FB に対する感情的反応, (3)倫理的考慮, (4)AI と人間による FB の統合。被験者は質問に基づき, 自身が卒業論文やグループ課題に取り組むといった大学での学びの場面における, 自身と生成 AI とのやり取り, また教員とのやり取りを想定しながら, 互いに意見を述べ議論した。インタビューは英語で, 各グループにつき約 1 時間ずつ行った。

2.3 分析手法

分析については被験者の見解における生成 AI と人間の FB に対する認識に関する発言を同定するために, 収集されたインタビューデータについて再帰的テーマティック分析[7]を行った。分析は筆頭著者を中心に行った。分析段階として初めに録音データの書き起こしを行った後に, 最初のコーディングを行った。関連するデータの抽出が特定され, 暫定的なコードで分類した。この段階で, 研究課題に関連する潜在的なテーマを生成した。続いて, テーマに基づきコード化された抜粋とデータセット全体をレビューした。次に, 残った各テーマを定義し, 各テーマが表すデータの具体的な側面を理解できるようにした上で, 最終的に分析結果をナラティブに統合した。分析のプロセスは反復的で, 分析が進むにつれてテーマやパターンを再検討し, 修正した。

3. 結果

本論文の目的に基づきインタビューデータを再帰的テーマティック分析した結果, (1) 道具的

(instrumental) な FB と包括的 (holistic) な FB, (2) 感情的な受け止めやすさのダイナミクス, (3) 関係性に基づく時間軸の 3 つのテーマが抽出された。

3.1 道具的(instrumental)な FB と包括的(holistic)な FB

学生はインタビューの中で生成 AI の FB における利点と人間による FB との差異を見出した。何人かの学生は生成 AI の FB における利点を文法やスペル修正などで, 人間より利用しやすく優れていると認識していた。修士学生の一人は, 教員に文法ミスなどで負担をかけることを恐れ, 生成 AI を他の英文校正 AI のように効率的に自身の文章を修正してもらうための効果的なツールだと捉えた。一方で, 学生は AI による FB を受けた後, 教員との「face-to-face」による活発な相互コミュニケーションを通じた FB を求めた。

AI feedback is best to use as a first step. For example... grammar mistakes. (If) my supervisor found tons of grammar mistakes, they're going to be so mad. So yeah, AI feedback is really good as a first step. But after that, I really want to receive that active, vibrant feedback... That is much better than AI feedback, I would say. But like you said, it's the balance, you know."

上記のように, 生成 AI は学生が求めている FB のある側面を補助することはできるが, バランスが重要であろうと話した。ある学生は, 人間の教員は生徒の返答に基づいて FB を適応させることができるが, 生成 AI は「子供」のようなもので, 何でも簡単な言葉で伝えないといけなかった。またある学生は, 教員に「経験に基づく, 具体的で専門的なフィードバック」を期待していた。同様に, 質問に対する教師の返答は, 授業全体の状況を包括的に把握しているため, よりインパクトのあるものになる傾向があると指摘する学生もいた。

3.2 感情的な受け止めやすさのダイナミクス

一方で, 人間による FB の感情的な側面は, ポジティブな特徴としてだけで捉えられているわけではない。ある学生は, AI との気軽なやり取りのしやすさを評価し, 指導教授から (厳しい) FB を受けるのは心理的な負担があるため, 下記のように ChatGPT のような生成 AI の FB が好ましいと話した。"I don't speak much in front of the supervisor, so ChatGPT is really good for me. You can ask a stupid question and [...] you don't need to think about the other person."

別の学部生は、ChatGPT に気軽に何度でも質問できることが、人間とのやり取りと比べて「more relaxed (リラックスできる)」と指摘した。この意見に同意する学生は、「AI には (同じ内容について) 10 回質問しても大丈夫だけれど、人間 (の先生) には 10 回質問したら、きっと大丈夫ではないだろう」と話した。学生は自分を緊張する場にさらすことなく、感情的に中立に生成 AI からの FB を得ることができると認識していた。

しかし、この感情的な中立さという面でも AI の FB が人間の FB より優れているという単純な結論には帰着しない。ある学部生は、同じ批評を生成 AI と人間の教員から受ける時の心境について次のように話した。

I wonder if I would be like, if the same thing was said by a teacher, the exact same like critique, I wonder if I would feel like, what the difference would be. I think I would take the critique by the teacher more to heart and the AI sort of, maybe more in line in a sort of way as an like online quiz or something where I get a score like a percentage and you're just like, oh yeah, I have to do it again. But you don't take it so much to heart.

上記のように、学生は生成 AI の FB を「オンラインクイズのようなもの」と見なし、人間の教員による FB をより心に留めると言及した。また別の修士学生も、AI からの FB を「真剣に受け止めない」とした上で、次のように述べた。

If I receive negative feedback (from a human teacher), I will be very sad, but I will also trust them. So I will feel like, okay, this needs to change. If I respect that person, then I think, okay, they are right. This hurts me, but it is true...And I think that respect, almost like admiration or looking up to whoever is giving you feedback is so important.

上記のように学生は人間の教員からの FB は尊重し、「前向きなものであれば喜びを感じ、ネガティブなのは痛みを感じるが、真実であり、成長のために受け取め、信頼する」と話した。

3.3 関係性に基づく時間軸

教員と学生との関係性に基づく時間軸についてある修士学生は次のように言及した。「私と先生との時間

軸は連続している。私が大学に入って、先生が 2 年間のプログラムのリーダーで……徐々に時間軸 (timeline) ができあがっていく。(略)感情的に、他のものよりも、連続的なタイムラインによって、人間により深く関わることができる」と感じるとした。話を受け、別の修士学生は次のように話した。"Teachers, while giving feedback they also put some subjective um...not judgment but comments. Subjective comments but also like drawn from personal experience." 学生にとって人間の先生は、FB をする時に、ジャッジするだけではなく個人的な経験に基づくコメントをしてくれる。そして、生徒はそれを自分の経験に結びつけることができ、自分は間違っていないと思うことができると考え、そのような経験は AI には持ち得ないのではないかと述べた。

4. 考察

結果から、生成 AI または教員からの対話的 FB についての学生による認識差が顕になった。生成 AI は言語や文法を修正するといった能力信頼を得る傾向にあり、これは獲得メタファに関連していると考えられる。一方教員については、認知的徒弟制に紐づくであろうコミュニケーション信頼が見られたことから参加メタファに関連しているものと考えられる。

コミュニケーション信頼は、FB の受容性と効果に大きな影響を与える。学生は、信頼できる教員からの FB を真剣に受け取り、尊重する傾向がある。例えば、学生は教員からのネガティブな FB を受け取った際、自己の成長のために受け入れると述べている。これは、教員が学生との間に築いた信頼関係が、FB の受容に直接影響していることを示している。一方、生成 AI からの FB は感情的に中立であり、利用することが容易であるが、その分、FB の重みや影響力が薄いと感じられている。学生は、生成 AI の FB を「オンラインクイズ」のように受け止め、信頼ある人間の教員からの FB をより深く心に留める傾向がある。しかし、教員が学生と信頼関係を築くことができなかつたり、また教員の FB を厳しく、感情的に受け止められない経験を学生が積み重ねた場合、教員との FB のやり取りを避け、より中立的に受け止めやすい生成 AI の FB を求めたりするかもしれない。

学びのメタファの違いに基づく FB の認識差を検討することも重要である。学習メタファは、教育の設計や実践において、特定の行動や認識を引き出すフレーム

ワークとして機能し得る。学生は生成 AI と獲得メタファに基づく修正的なやり取りをしながらも、人間の教員とはより固有の関係性に基づく参加的な学びを期待している。つまり、学生は学びに対するメタファ及び認識を複数持ち、AI と人間とで別のメタファに基づいて認識を切り替えている傾向が示唆された。知識創造メタファに結びついた言及は今回見られなかったが、学生、もしくは教員がこれらの学習メタファを認識することで、AI との対話コミュニケーション及び認識に変化が起きるかどうかは検討の余地がある。

生成 AI は今後の教育や学びにおいて革新的な変化をもたらす可能性がある。一方で、生成 AI と人間の違いを捉えることなく、AI との対話を人間の代替かのように捉えて学びを設計することには注意が必要である。これは、AI がより性能を高め、更なる能力信頼を高めることで代替し得るものではないと考えられる。学習者及び教員は、AI と人間を本質的に異なるものとして理解することで、それぞれを信頼することが可能ではないかと思われる。学習メタファを元に、教育者や研究者は AI と人間に対する異なる視点やアプローチを理解し、より包括的な教育実践を検討することができる。また、学習者自身がどのようなメタファに基づいて学びを捉えているのかを理解することで、より適切な指導方法を見つける手がかりとなる。今後は場面に応じて異なる学習メタファを検討し、使い分けることで、AI と人間の協調的な学びの実現を検討し、捉えていくような実証的研究が求められる。

文献

- [1] Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: Quick startguide. UNESCO Publishing.
- [2] Floridi, L., & Nobre, A. C. (2024). Anthropomorphising machines and computerising minds: The crosswiring of languages between Artificial Intelligence and Brain & Cognitive Sciences.
- [3] Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor: An emergent epistemological approach to learning. *Science & education*, 14, 535-557.
- [4] Bryson, C., & Hand, L. (2007). The role of engagement in inspiring teaching and learning. *Innovations in Education and Teaching International*, 44(4), 349-362.
- [5] Carless, D. (2012) Trust and its role in facilitating dialogic feedback. *Feedback in Higher and Professional Education*, Routledge, pp.90-103.
- [6] 大滝文一, 神拓実, 長沼祥太郎, Lindwall O., 大島純. (2024). 生成 AI によるフィードバックの信頼に対する大学生と教員の見解. 日本教育工学会第 44 回全国大会講演論文集, 499-500.
- [7] Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, 9(1), 3.