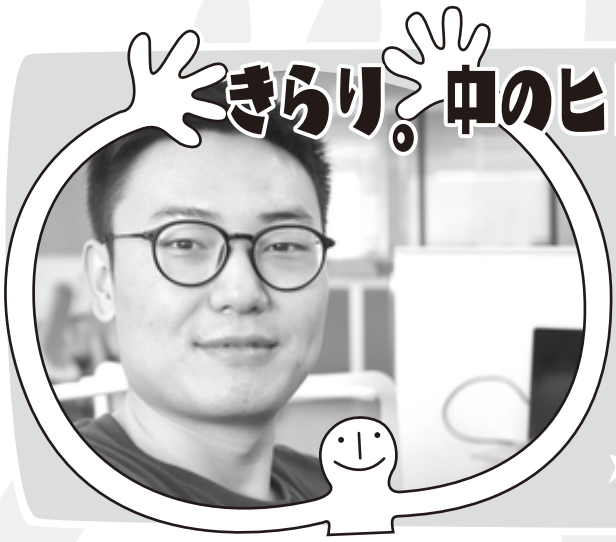

第26回

人間とAIの協同進化へ

王　博文†

まえがき

大阪大学産業科学研究所で助教を務めております，王博文と申します．これまで私は，大規模言語モデルをはじめとする人工知能技術を背景に，AIの信頼性や説明可能性といったテーマを中心に研究を進めてきました．今回，このような形で随筆の執筆機会をいただき，改めて「今，自分は何を考えているのか」，「どこに向かおうとしているのか」を振り返るよい機会だと感じています．

当初は，これまでの研究成果を整理し，体系的に紹介することも考えました．しかし，それ以上に本稿では，私自身の中で近年強まってきた問題意識や，研究関心の重心がどのように変化しつつあるのかについて，率直に書いてみたいと思うようになりました．そこで本稿では，大規模言語モデルをめぐる研究経験を踏まえつつ，近年私が特に関心を寄せている「人間とAIの協同進化」という考え方について，自身の研究や日常の実感と結びつけながら紹介したいと思います．

研究内容について

近年，大規模言語モデルは，自然言語理解や生成の分野において急速な発展を遂げてきました．質問応答や要約，翻訳といった従来のタスクにとどまらず，医療，都市分析，知識サービスなど，より複雑で文脈依存性の高い実世界の課題にも応用され始めています．こうした技術的進歩は，人工知能が人間の知的活動に深く関与する段階に入りつつあることを示しているように思います．

私の最近の研究では（**図1**），大規模言語モデルを単なる「正解を返すシステム」や「作業を効率化する道具」として扱うのではなく，推論や判断，さらには責任が伴う状況に関与する存在として捉えることを重視しています．現実世界の問題，とりわけ医療や専門的意思決定の場面では，入力される情報は必ずしも完全ではなく，不確実性や曖昧さを多く含んでいます．そのような環境において，モデルがいかに高い精度を示したとしても，その判断の根拠が不透明であれば，人は安心してそれを受け入れることができません．例えば，医療テキスト[5]の理解や診断推論の場面では，記述の不足や表現の揺れ，さらには専門家間の見解の違いが日常的に存在します．こうした状況下で，大規模言語モデルが一見もっともらしい結論を提示したとしても，「なぜその結論に至ったのか」，「どの情報が決定的だったのか」，「判断に迷いはなかったのか」といった点が明らかでなければ，その出力を人間が適切に評価し，修正することは困難です．

しかし，研究や開発の現場では，しばしば性能指標の向上が最優先され，その背後にある判断過程や前提条件への目配りが後回しにされがちです．私自身も，こうした流れの中で研究を進める一方で，「このモデルの判断は，どこまで信頼してよいのだろうか」，「人はこの判断にどう関与できるのだろうか」といった問いを，次第に強く意識するようになりました．こうした研究経験を通じて，私は次第に，AIの性能向上そのものだけでは，現実世界での長期的な活用や社会的な信頼の確立には充分ではないと感じるようになりました．

AIを人間の判断や価値観から切り離された存在として設計するのではなく，人間がその振る舞いを理解し，対話し，ときに立ち止まらせることのできる存在として位置づける必要があるのではないか．そのような問題意識が，私の研究の中心に据えられるようになっていったのです．現在の研究では，大規模言語モデルに対して，知識の構造化や推論過程の可視化といった要素を組み合わせることで，モデルが「何を答えたか」だけでなく，「どのように考えたか」を人間と共有できる仕組みを模索しています．これは，AIにすべてを任せるための技術ではありません．むしろ，人間がAIの判断に関与し続けるための余地を残し，人とAIが同じ問題空間を見渡しながら考えるための基盤を整える試みだと考えています．

AI研究に惹かれたきっかけと，これまでのキャリア

振り返ってみると，私が現在取り組んでいる大規模言語モデルを中心とした研究は，決して最初から明確な目標

† 大阪大学 産業科学研究所
"Towards the Collaborative Evolution of Humans and AI" by Bowen Wang (SANKEN, the University of Osaka, Osaka)

映像情報メディア学会誌 Vol. 80, No. 3, pp. 337～339（2026）

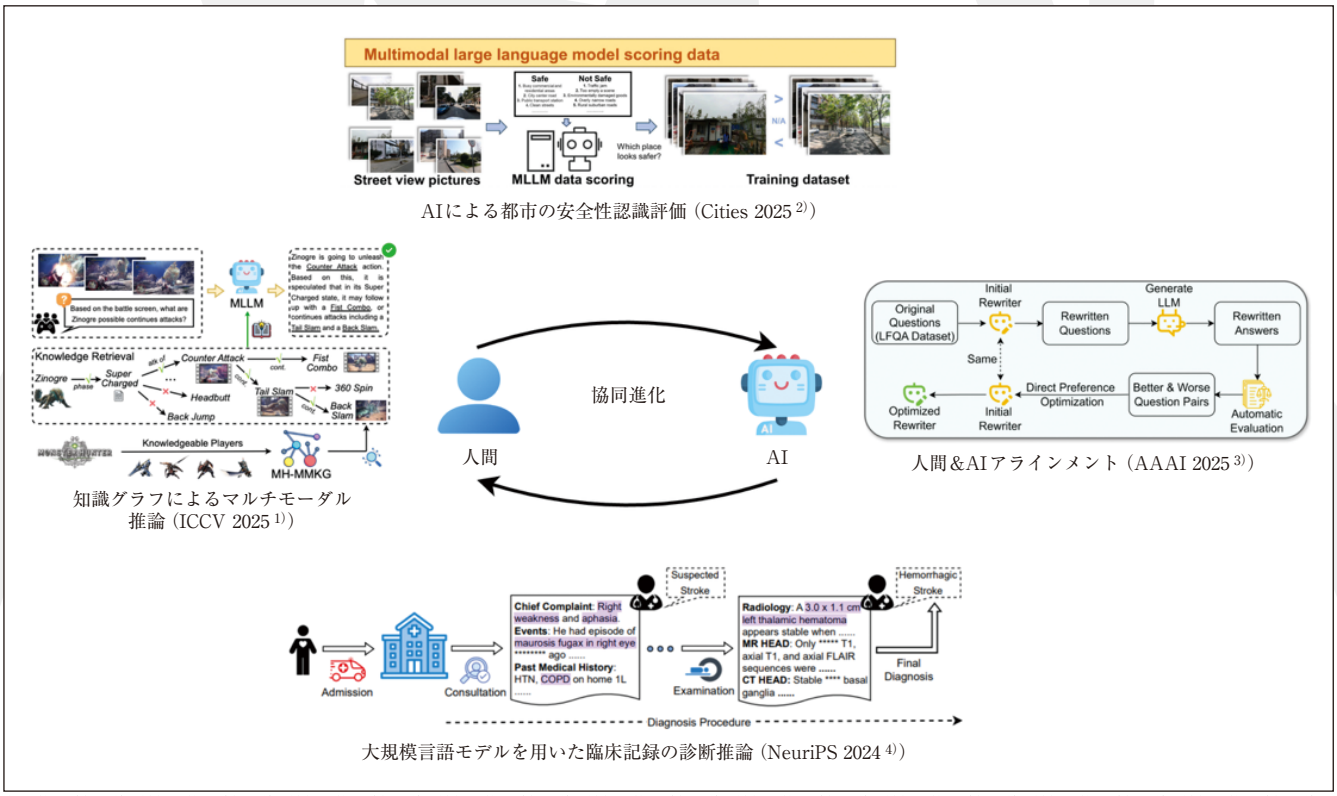

図1　私の研究テーマ

として定まっていたわけではありません．むしろ，その時々に直面した課題に導かれる形で，少しずつ研究の関心が移り変わってきたように思います．研究の初期段階では，コンピュータビジョンや医用画像解析，説明可能人工知能といった分野に取り組み，モデルが下した判断を人間がどのように理解し，納得できる形で提示できるのかという問題に向き合ってきました．当時の関心は，「精度を上げること」そのものよりも，「その判断はどのような根拠に基づいているのか」，「人はそれをどう受け取るのか」といった点にありました．

特に印象に残っているのは，医師や医療研究者との共同研究を通じて得た経験です．医療の現場では，人工知能が高い性能を示したとしても，その判断理由が説明できなければ，実際の意思決定に用いられることはほとんどありません．「なぜこの結果になったのかを説明できないのであれば，参考意見としても使いにくい」―― そうした率直な言葉に触れる中で，私は，人工知能にとって説明可能性は付加的な機能ではなく，信頼を成立させるための前提条件なのだと強く実感しました．こうした経験は，後に推論型の大規模言語モデルへと研究関心を広げていく上で，重要な土台となっています．

大規模言語モデルの能力が急速に向上し，流暢で自然な文章を生成できるようになるにつれ，私は次第に，別の種類の違和感を覚えるようになりました．それは，「あまりにも自然に答えすぎてしまうこと」への違和感です．モデルがもっともらしい答えを即座に提示できる一方で，その背後にある推論の過程や，不確実性に対する態度は必ずしも明示されません．そのとき私は，「流暢に答えられること」以上に，「なぜその答えに至ったのかを説明できること」，そして場合によっては「今は答えるべきではないと判断できること」が，実世界で AIが人と協働するためには不可欠なのではないかと考えるようになりました．

この問題意識は，AIを完全に自律した存在として設計するのではなく，人間の関与を前提とした知能として設計する方向へと，私の研究を導いています．AIがすべてを決めるのではなく，人間が問いを立て，判断の責任を担い，その過程を AIと共有する．そのような関係性の中でこそ，人工知能は初めて現実の社会に根付いていくのではないかー― そうした考えが，現在の研究活動の根底にあります．

研究者としての日々

研究者としての日常は，論文執筆や実験だけで成り立っているわけではありません．私の一日は，学生や共同研究者との議論から始まり，研究テーマの方向性や仮説の妥当性について考えることに多くの時間を費やしています．「この問いは本質的なものか」，「現実の問題とどのようにつながっているのか」，「人間が最終的に判断すべき部分はどこに残すべきか」といった点を，日々問い直しています．

また，国内外の学会や研究会に参加し，さまざまな分野の研究者と意見を交わすことも，研究活動の重要な一部です．AIに関する議論は急速に進んでおり，異なる立場や分野の研究者との対話を通じて，自分自身の問題意識が整理されたり，新たな視点を得たりすることが少なくありません．発表や質疑応答を通じて，自身の研究の位置づけを客観的に見直す貴重な機会にもなっています．研究室内では，学生との日常的なミーティングを通じて，研究の進め方そのものについても一緒に考えています．結果だけでなく，どのような思考過程を経て結論に至ったのかを共有することで，研究を「一人で進めるもの」ではなく，「対話を通じて深めていくもの」として捉えることを大切にしています．

こうした日々の活動を通じて，私はAIを単なる自動化の道具としてではなく，人間の思考や議論を豊かにする協働的な存在として位置づけたいと考えています．人と人との議論にAIが加わることで，研究の探索空間が広がり，これまで見過ごされてきた問いに気づくきっかけが生まれる．そのような研究環境を実現することが，私自身の目標の一つです．

むすび

技術が高度化するほど，その使われ方や向き合い方が問われるようになります．AI研究に携わる中で，私自身もまた，答えを出すこと以上に，考え続ける姿勢の重要性を意識するようになりました．人とAIが同じ問いに向き合い，異なる立場から補い合う関係を築くことができれば，研究の営みそのものも，より開かれたものになっていくはずです．今後もそうした関係性を丁寧に探りながら，静かに研究を積み重ねていきたいと思います．

（2026年1月17日受付）

〔文　献〕

1) B. Wang, Z. Jiang, Y. Susumu, S. Miwa, T. Chen, Y. Nakashima: "Taming the Untamed: Graph-Based Knowledge Retrieval and Reasoning for MLLMs to Conquer the Unknown", IEEE/CVF International Conference on Computer Vision (ICCV), 2025
2) J. Zhang, Y. Li, T. Fukuda, B. Wang: "Urban Safety Perception Assessments via Integrating Multimodal Large Language Models with Street View Images", Cities (2025)
3) J. Chen, B. Wang, Z. Jiang, Y. Nakashima: "Putting People in LLMs' Shoes: Generating Better Answers via Question Rewriter", Association for the Advancement of Artificial Intelligence (AAAI) (2025)
4) B. Wang, J. Chang, Y. Qian, G. Chen, J. Chen, Z. Jiang, J. Zhang, Y. Nakashima, H. Nagahara: "DiReCT: Diagnostic Reasoning for Clinical Notes via Large Language Models", Advances in Neural Information Processing Systems 2024 (NeurIPS) (2024)
5) A. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. Pollard, S. Hao, B. Moody, B. Gow, L. Lehman, L. Celi and R. Mark. MIMIC-IV, a freely accessible electronic health record dataset. Scientific data, 10 (1) :1 (2023)

新刊図書

現代社会を生きるための AI×哲学

谷口忠大・鈴木貴之・丸山隆一　著

日常生活からビジネスシーンに至るまで，AIは驚くべきスピードで社会に浸透しています．AIは多くの利便性をもたらす一方で，フェイク画像や動画の流出，著作権侵害，プライバシー侵害，偏見の助長など，さまざまなリスクや社会問題も引き起こしています．私たち一人ひとりが，今後AIとどのように向き合い，社会生活をより豊かなものにしていくかが問われています．

本書は，「技術としてのAI」，「心の哲学としてのAI」，「社会の中のAI」という三つの観点から構成されています．技術解説にとどまらず，AIがもたらすリスクや，AIと人間・社会との関係，さらには未来社会における在り方までを幅広く扱っており，多角的で立体的にAIを捉えられる点が大きな魅力です．

0～1章の導入では，生成AIが今日の社会にもたらしている影響を整理しながら，AIを考える上での重要な着眼点や課題が提示されています．続く2章以降では，AIの歴史をさかのぼり，技術の進展や開発の背景をテンポよく解説しています．初期のAI研究から，パターン認識，機械学習，ニューラルネットワーク，自然言語処理，そして大規模言語モデルによる生成AIに至るまで，専門用語の解説を交えつつわかりやすく網羅されており，AI技術者にとっての知識の再整理はもちろん，これからAI技術を体系的に学びたい読者にとっても有用な入門書となっています．

本書の特徴は，技術的側面にとどまらず，AIと人間，社会との関係性，さらには未来の関係までを展望している点にあります．8章以降では，AIが意識や感情を持ち人間になれるかという問いを哲学的に考察するとともに，AIガバナンスの在り方や，AIが民主主義を脅かすといった言説を取り上げながら，未来社会における人間とAIの関わりについて冷静に分析し，示唆に富んだ指針を示してくれます．

本書は，これからAIを本格的に学ぼうとする人だけでなく，日常的に何気なくAIを使い始めた人にとっても，AIと人とのより良い関わり方を考えるきっかけを与えてくれる一冊と言えます．

紹介　藤崎好英（NHK）

講談社刊（2026年02月12日発行），A5判，352頁，定価：2,800円＋税