

2023 年度年次報告書

社会課題を解決する人間中心インタラクションの創出

2023 年度採択研究代表者

鈴木 遼

東北大学 電気通信研究所

非常勤講師

Generative AR: 拡張現実と実世界指向の生成 AI による新たなヒューマン AI インタラクション

研究成果の概要

2023 年度(2023 年 9 月から 2024 年 3 月まで)における重要な研究成果として、以下の 2 点を挙げる事ができる。1 つ目は、UIST 2023 における AI x XR ワークショップである。2 つ目は、AR と AI に関連する論文を UIST 2024 に投稿したことである。今回の研究成果として、特に以下の 2 点の論文から得られた知見について説明する。

まず、1 つ目として、AR x AI を使ったコンテンツ生成のテーマのもと、"Augmented Physics: A Machine Learning-Powered Tool for Creating Interactive Physics Simulations from Static Diagrams" という論文を投稿した。この研究は、物理教育において静的な教科書の図を動的でインタラクティブなシミュレーションに変換する AI ツールを開発したものである。この研究では、機械学習とコンピュータビジョン技術(Segment Anything と OpenCV)を利用して、図を自動的に認識し、シミュレーション可能なオブジェクトに変換するプロセスを構築した。この研究では、システムの構築以上に、実際に物理学の先生や学生を含む 24 名のユーザーに実際に使ってもらい、システムを評価した(N=12 のユーザビリティスタディ、および N=12 の物理教師とのエキスパートインタビュー)。「自分の」教科書をインタラクティブにするというアイデアは、想像以上に好意的なフィードバックを得た。一方で、教師側からのフィードバックとして、自動で生成されたシミュレーションが必ずしも「教育的か」どうかは疑問が残るという結果も得られた。教育的な目的としては、AI で生成されたコンテンツを、教師が一度確認・修正した上で、学生に配布するというようなステップが好ましいという、human-in-the-loop のプロセスが必要になることが知見として得られた。

2 つ目として、AR と大規模言語モデルの融合のテーマのもと、"RealitySummary: On-Demand Mixed Reality Document Enhancement using Large Language Models"という論文を投稿した。この研究は、AR と大規模言語モデル(LLM)を用いて、読んでいる本や身の回りのテキストを OCR を介して、暗黙的に LLM に渡し、要約をしたり関連する情報を AR 上に表示するというプロジェクトである。この研究のメインの貢献は、in-the-wild スタディであり、我々が用意したコンテンツではなく、ユーザーが大学のキャンパスを歩き回ってシステムを使ってもらおうという実験を行った。このスタディの結果は興味深く、最初に想定していたようなユースケースとは全く異なる使われ方がされた。例えば、本の要約だけではなく、図書館で本棚を見て、どんな本がおすすめかを聞いたり、カフェテリアのメニューを見てどれが健康的かを聞いたり、数学や物理学の本を読んで、そこから期末試験に出そうなクイズを生成したり、ゴミ箱のラベルを見て、持っているゴミがどの分別になるのかを聞く、といったような当初の予想と異なる使われ方がされた。

また、特に AR x AI が今までの AI インターフェイスとどう違うかのヒントも得られた。AR の常時オンのカメラは暗黙的な入力を可能にし、デスクトップインターフェースでの明示的な入力(例えば、質問をタイプする、テキストをコピー＆ペーストする)とは対照的に、AI が空間的(本棚全体や複数の本を 1 ページではなく)または時間的(全ての読書履歴)に情報を「暗黙的に」キャプチャできることが大きな違いであると学んだ。また、AR の空間的出力は、情報をバックグラウンド(周辺)に提示することができるため、音声やスクリーン出力のようなアウトプットに比べ、ユーザーの注意を要

求することなく、プロアクティブかつ周辺の出力を可能にする点も興味深い知見であった。一般に、未来の AR と AI システムは、暗黙的な入力とプロアクティブ/周辺の出力の組み合わせにより、ユーザーの注意を奪わない新しいタイプの穏やかな人間-AI インタラクションを可能にするということを学んだ。今後は、これらの知見を元に、AR x AI の強みを生かしたシステムの設計や人間理解を目指していきたい。

【代表的な原著論文情報】

- 1) Jian Liao, Kevin Van, Zhijie Xia, and Ryo Suzuki. "RealityEffects: Augmenting 3D Volumetric Videos with Object-Centric Annotation and Dynamic Visual Effects." arXiv preprint arXiv:2405.17711 (2024).
- 2) Aditya Gunturu, Shivesh Jadon, Nandi Zhang, Jarin Thundathil, Wesley Willett, and Ryo Suzuki. "RealitySummary: On-Demand Mixed Reality Document Enhancement using Large Language Models." arXiv preprint arXiv:2405.18620 (2024).
- 3) Aditya Gunturu, Yi Wen, Jarin Thundathil, Nandi Zhang, Rubaiat Habib Kazi, and Ryo Suzuki. "Augmented Physics: A Machine Learning-Powered Tool for Creating Interactive Physics Simulations from Static Diagrams." arXiv preprint arXiv:2405.18614 (2024).
- 4) Shivesh Jadon, Mehrad Faridan, Edward Mah, Rajan Vaish, Wesley Willett, and Ryo Suzuki. "Augmented Conversation with Embedded Speech-Driven On-the-Fly Referencing in AR." arXiv preprint arXiv:2405.18537 (2024).
- 5) Keiichi Ihara, Kyzyl Monteiro, Mehrad Faridan, Rubaiat Habib Kazi, and Ryo Suzuki. "Video2MR: Automatically Generating Mixed Reality 3D Instructions by Augmenting Extracted Motion from 2D Videos." arXiv preprint arXiv:2405.18565 (2024).
- 6) Ryo Suzuki, Mar Gonzalez-Franco, Misha Sra, and David Lindlbauer. "XR and AI: AI-Enabled Virtual, Augmented, and Mixed Reality." In Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1-3. 2023.