

卒業論文の要旨

論文題目	Faithful In, Factful Out. —嘘をつけない LLM のパラドックスと RAG—
氏名	川口 浩平
メジャー	言語学
マイナー	文化人類学
(要旨)	
<p>大規模言語モデル(Large Language Model 以下、LLM)とその周辺技術である検索拡張生成(Retrieval-Augmented Generation 以下、RAG)の普及に伴い、ハルシネーションの制御が喫緊の課題となっている。従来の研究では、この現象を人間的な「嘘」や「欺瞞」の文脈で捉えようとしてきた。しかし、本論文では LLM が Vaswani et al. (2017) の提唱する「Sequence transduction model」であり、信念や意図を持たない統計的機械であることを踏まえ、既存の哲学的枠組みの適用はカテゴリーエラーであると指摘している。</p> <p>さらに本論文では、Tarski (1944)を援用し、RAG の実装が抱える論理的矛盾を明らかにする。RAG によって内部知識と外部知識を同一層で処理する現状は、自己言及のパラドックスを引き起こす構造的欠陥であるとも指摘している。</p> <p>結論として、不確実な事実性の追求を放棄し、計算機科学の基本原則である Garbage In, Garbage Out に回帰すべきであると論じる。外部知識をメタ言語として厳格に定義し、大規模言語モデルがそれを内部知識によって改変することを禁じることなく、忠実に振る舞わせる構造へと転換することでこそ、検索拡張生成の信頼性は担保されると主張する。</p>	
(指導教員の推薦のコメント)	
<p>本論文は大規模言語モデルとその周辺技術である検索拡張生成について、言語学的視点からその不備を指摘し、「不確実な事実性の追求を放棄し、計算機科学の基本原則である Garbage In, Garbage Out に回帰すべきである」と結論付けている。LLM や RAG を巡る議論において支配的である Grice (1975) の Gricean Maxims や Frankfurt (2005) の Lying と Bullshit といった「Belief」を前提とする哲学的枠組みからの脱却を試みるものであるという点で、極めて最新のテーマを扱いながらも、広く先行研究に目を通し、深い考察と鋭い洞察力に富んだ論文と言える。</p> <p style="text-align: right;">山岡洋</p>	