

# ABCI 3.0開発加速利用（2025年度）成果概要（公開用）

課題名：  
認知的AIエージェントによるヒューマン・コンピュータ・インタラク  
ションのシミュレーション

実施時期：2025年度  
所属機関名：東京都立大学  
代表者氏名：福地庸介

成果概要：  
Web上で人がどの情報を選び、どのように考えを更新するかを再現する認知的AIエージェント「F-Explorer」を開発した。自由エネルギー原理に基づき認知モデルに大規模言語モデル（LLM）を組み込み、SNSにおける人の情報探索行動をシミュレーションした。結果、F-Explorerは、人の探索行動データに対して従来モデルよりも高い適合度を示し、信念状態の推定も一定精度で可能であることがわかった。

成果のポイント：  
Webメディアは、誤情報や陰謀論の拡散を通じて、誤った信念の形成や強化を引き起こしうるため、社会に大きな影響を及ぼしている。Webメディアの健全利用を考えるためには、ユーザがどのような情報を選び、情報接触を通じて信念をどのように変化させるのかを理解する必要がある。特に、人間が自分の信念に適合する情報を選好する**確証バイアス**は、誤った信念を持つユーザが、自分に都合のよい情報を選び続けることで、誤信念を維持・強化する要因になりうる。  
本研究では、自由エネルギー原理に基づく情報探索モデルに大規模言語モデル（LLM）を組み込み、認知AIエージェント **F-Explorer** を開発した。F-Explorer は人の情報探索行動を、新たな情報を探索することで得られる情報利得と、既存の信念を確かめることで得られる満足のトレードオフの結果として表現する。後者の影響が強くなった時、エージェントは確証バイアスを示すようになる。LLMを認知モデリングに応用する研究が増加している中で、本研究の特徴は、大量の心理実験データを用いてLLMに人間行動を学習させるデータ駆動アプローチではなく、理論（自由エネルギー原理）を基盤とした認知モデルのなかに、信念表現のモジュールとしてLLMを組み込む理論駆動アプローチを採用した点にある。実際のSNS探索行動データでF-Explorerを評価した結果、データに対して従来モデルより高い適合度を示した（対数尤度 -15.01→-14.27, Top-1精度 0.334→0.431）。また、信念を表現するLLMを解析することで、ユーザの信念も一定精度で可能であることがわかった。  
本成果は、LLMによる認知モデリングにおいて理論駆動という新たなアプローチをとることによって、行動予測と背後にある認知プロセスの解釈可能性を両立した認知的AIエージェントに向けた方向性を示している。

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：  
Yosuke F.: "Free-Energy-Driven LLM Agent for Modeling Information Exploration in Web Media", CogSci2026 (in press).