Contribution ID: **80** Type: **not specified**

Evaluating Personality Consistency in Large Language Models via Big Five Inventories and Dual-Enhanced Networks

Large language models (LLMs) are increasingly capable of simulating human-like personalities through prompt engineering, presenting novel opportunities and challenges for personality-aware AI applications. However, the consistency and stability of these simulated personalities over time and across contexts remain largely unexplored. In this study, we propose an integrated framework to systematically evaluate LLM personality consistency using psychometric tools combined with machine learning-based textual analysis. We generate diverse text corpora by prompting LLMs with varied personality-specific instructions and social contexts, ensuring a broad range of stylistic and psychological traits in the generated data. We then employ a Dual Enhanced Network (DEN) architecture to extract both long-term and short-term personality representations from the generated texts. Our framework quantifies personality drift and robustness using statistical measures and machine learning-based evaluation across multi-session outputs. Experimental results reveal that LLM personalities fluctuate significantly depending on prompt design and conversational context. We further discuss implications for model alignment, control strategies, and the deployment of personality-driven AI systems. This work provides a novel methodology for evaluating personality emulation in LLMs and highlights important considerations for ethical and robust personality-aware applications.

Special/ Invited session

Classification

Mainly application

Keywords

AI Personality, Prompt Engineering, LLMs, Dual Enhanced Network

Primary author: Ms HOU, Yi-Ting (National Taiwan University)

Co-authors: SHAO, Freda (National Taiwan University); Dr BLUE, Jakey (National Taiwan University)

Presenter: Ms HOU, Yi-Ting (National Taiwan University)

Track Classification: AI: Interpretability and Trustworthiness