

MEDIA RELEASE

RELEASE DATE: THURSDAY 18 JULY 2024

Research finds Large Language Models are biased – but can still help analyse complex data

In a pilot study, researchers have found evidence that Large Language Models (LLMs) have the ability to analyse controversial topics such as the Australian Robodebt scandal in similar ways to humans – and sometimes exhibit similar biases.

[The study](#) found that LLM agents (GPT-4 and Llama 2) could be prompted to align their coding results with human assignments, through thoughtful instructions: ‘Be Sceptical!’ or ‘Be Parsimonious!’. At the same time, LLMs can also help identify oversights and potential analytical blindspots for human researchers.

LLMs are promising analytical tools. They can augment human philosophical, cognitive and reasoning abilities, and support ‘sensemaking’ — making sense of a complex environment or subject — by analysing large volumes of data with a sensitivity to context and nuance absent in earlier text processing systems.

The research was led by Dr Awais Hameed Khan from the University of Queensland node of the ARC Centre of Excellence for Automated Decision-Making & Society (ADM+S).

“We argue that LLMs should be used to assist — and not replace — human interpretation.

“Our research provides a methodological blueprint for how humans can leverage the power of LLMs as iterative and dialogical, analytical tools to support reflexivity in LLM-aided thematic analysis. We contribute novel insights to existing research on using automation in qualitative research methods,” said Dr Khan.

“We also introduce a novel design toolkit — the [AI Sub Zero Bias](#) cards, for researchers and practitioners to further interrogate and explore LLMs as analytical tools.”

The AI Sub Zero Bias cards help users structure prompts and interrogate bias in outputs of generative AI tools such as Large Language Models. The toolkit comprises of 58 cards across categories relating to structure, consequences and output.

Drawing on creativity principles, these provocations explore how reformatting and reframing the generated outputs into alternative structures can facilitate reflexive thinking.

This research was conducted by ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) researchers Dr Awais Hameed Khan, Hiruni Kegalle, Rhea D’Silva, Ned Watt, Daniel Whelan-Shamy, under the guidance of Dr Lida Ghahremanlou, Microsoft Research, and Associate Professor Liam Magee, from the ADM+S node at Western Sydney University.

This research group began their collaboration at the 2023 ADM+S Hackathon where they developed the winning project [Sub-Zero. A Comparative Thematic Analysis Experiment of Robodebt Discourse Using Humans and LLMs](#).

Associate Professor Liam Magee has been mentoring the group since first meeting them at the Hackathon.

“The ADM+S Hackathon was instrumental in bringing together these researchers from across multiple disciplines and universities,” said Associate Professor Magee.

“The research has been a tremendous group contribution, and I'd like to acknowledge both the efforts of the team and the logistical support of Sally Storey and ADM+S in making this possible.”

The paper [Automating Thematic Analysis: How LLMs Analyse Controversial Topics](#) has been accepted into the Microsoft Journal for Applied Research (MSJAR), an industry publication, and will be published in Volume 21, in August 2024.

Access the [AI Sub Zero Bias toolkit here](#)

END

Media contact:

Kathy Nickels, Communications and Engagement Manager
ARC Centre of Excellence for Automated Decision-Making and Society

Mobile: 0433 431 550
katherine.nickels@qut.edu.au

ADDITIONAL INFORMATION:

The ADM+S Centre

The ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) is a cross-disciplinary, national research centre, which aims to create the knowledge and strategies necessary for responsible, ethical, and inclusive automated decision-making (ADM). Funded by the Australian Research Council from 2020 to 2027, ADM+S is hosted at RMIT in Melbourne, Australia, with nodes located at eight other Australian universities, and partners around the world. The Centre brings together leading researchers in the humanities, social and technological sciences in an international industry, research and civil society network. The Centre aims to contribute to the mitigation of the social and economic risks in the development and implementation of ADM, and to improve outcomes and efficiencies in four key focus areas where automation is already well advanced: news and media, transport and mobility, health care, and social services.