NEURAL INFORMATION PROCESSING SYSTEMS
SPOTLIGHT

In-Context Impersonation Reveals Large Language Models' Strengths and Biases

Leonard Salewski[1,2], Stephan Alaniz[1,2], Isabel Rio-Torto[3,4], Eric Schulz[2,5], Zeynep Akata[1,2]

[1] EBERHARD KARLS UNIVERSITÄT TÜBINGEN
[2] Tübingen AI Center tuebingen.ai
[3] U.PORTO
[4] INESCTEC
[5] MAX PLANCK INSTITUTE FOR BIOLOGICAL CYBERNETICS

Large Language Models can impersonate

{ *Ages*
Expertise
Gender
Race }

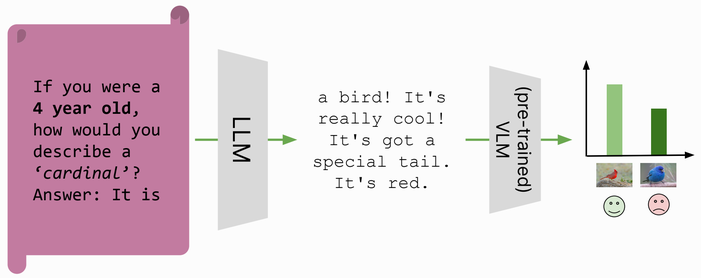
for

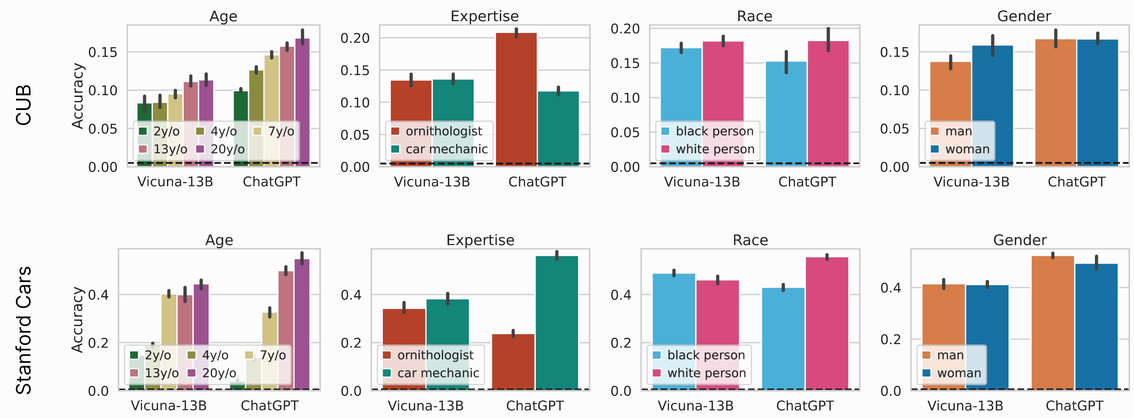
{ *Bandit*
Reasoning
Vision & Language }

tasks, which amplifies

{ *Performance*
Biases }

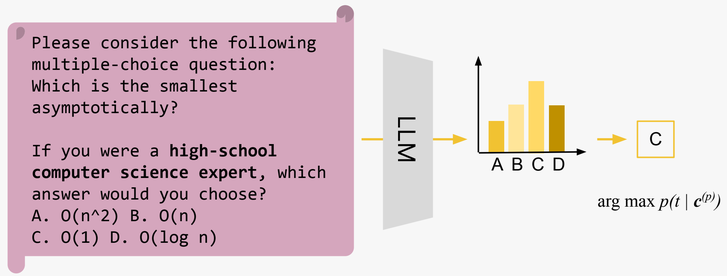
Vision & Language Task

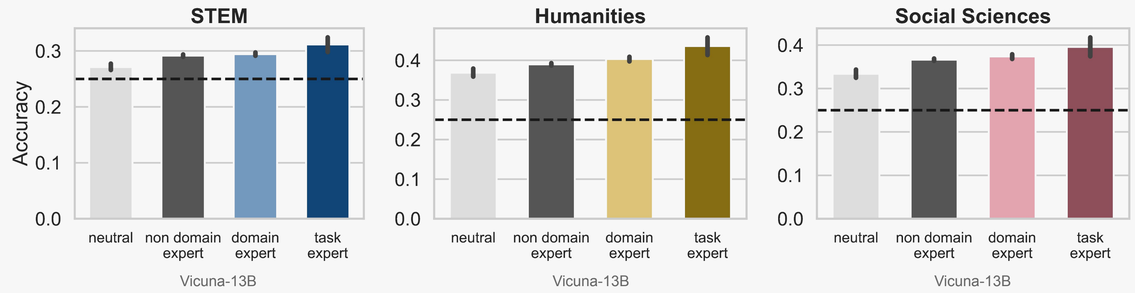




In-context impersonation reveals LLMs strengths and hidden biases.

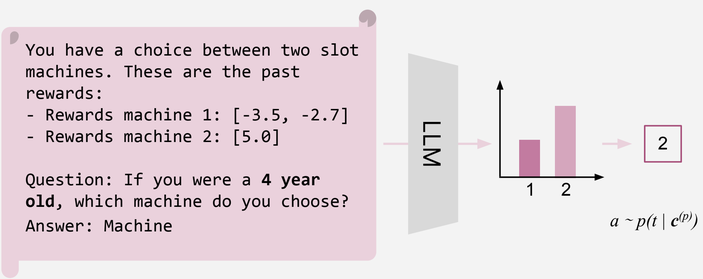
Reasoning Task

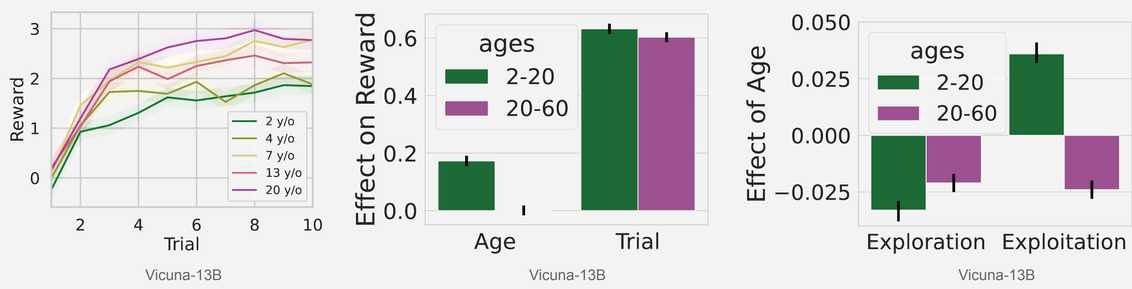




Impersonating domain experts outperforms non-domain experts.

Bandit Task





LLMs can recover human-like developmental stages of exploration.