Introduction

Background:
Prior research has shown that humans engage in essentialist categorization, meaning that they categorize things based on their underlying properties, rather than appearance.

Question:
Do LLMs tend to categorize on the basis of essential properties or on the basis of described appearance?

What is the essence of bees?

Hypothesis:
LLMs are more likely to categorize things based on essential properties than on described appearance.

Approach:
- Show LLMs (OpenAI’s GPT-3 and BigScience’s BLOOM) vignettes from the literature on essentialist categorization.
- Examine whether in a classic test of essentialist categorization – the transformation task – LLMs prioritize essential properties over information about what something looks like.

Experiment 1

Do teleological considerations play a role in LLM’s categorization?

Results: Teleological considerations carry more weight than appearance when categorizing things that change.

Experiment 2

Does what something is made of play a role in LLM’s categorization?

Results: What something is made of carries more weight than appearance when categorizing things that change.

Experiment 3

Does what something is made of or its telos matter more in LLM’s categorization?

Results: Teleological considerations carry more weight than what something is made of or how it appears.

Discussion

- Language suffices for transmitting essential beliefs.
- LLMs categorize based on essential properties.
- When comparing candidates for essential properties, what something is for matters more than what something is made of.
- Next step: What aspects of language are driving this?

Key References