

#### Large Language Models can impersonate



#### If you were a 4 year old, how would you

describe a 'cardinal'? Answer: It is

#### for

Bandit Reasoning Vision & Language

tasks, which amplifies

Performance Biases



### **Reasoning Task**

Please consider the following multiple-choice question: Which is the smallest asymptotically?

If you were a **high-school** computer science expert, which answer would you choose? A. O(n^2) B. O(n) C. O(1) D. O(log n)

## **Bandit Task**

- You have a choice between two slot machines. These are the past rewards:
- Rewards machine 1: [-3.5, -2.7] - Rewards machine 2: [5.0]
- Question: If you were a 4 year **old**, which machine do you choose? Answer: Machine

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

Leonard Salewski<sup>1,2</sup>, Stephan Alaniz<sup>1,2</sup>, Isabel Rio-Torto<sup>3,4</sup>, Eric Schulz<sup>2,5</sup>, Zeynep Akata<sup>1,2</sup>



In-context impersonation reveals LLMs strengths and hidden biases.



Impersonating domain experts outperforms non-domain experts.



LLMs can recover human-like developmental stages of exploration.

**U.**PORTO







EBERHARD KARLS UNIVERSITÄT TÜBINGEN



