

# FANOUTQA

Andrew Zhu, Alyssa Hwang,  
Liam Dugan, Chris Callison-Burch  
{andrz<sup>†</sup>, ahwang16, ldugan, ccb}@seas.upenn.edu  
<sup>†</sup>correspondence

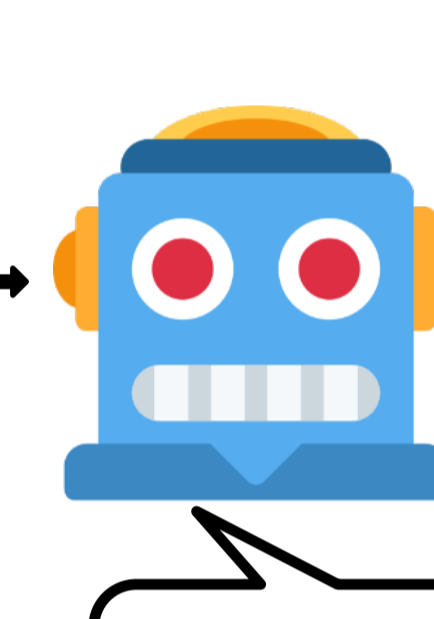
## A Multi-Hop, Multi-Document Question Answering Benchmark for Large Language Models

FANOUTQA is a benchmark of 1,034 multi-hop information-seeking questions, grounded in Wikipedia, that require models to **search** for and **reason** over multiple natural documents.

### "Fan-Out" Questions

Commonly found in real-life: find a list of entities & aggregate information about each.

What is the **total number** of employees in the **five largest banks** in the world?



Closed Book  
Open Book  
Evidence Provided

1,604,898

### Benchmark Settings

**Closed Book** - Answer questions using only parametric knowledge.

**Open Book** - LLMs must use provided tools to search English Wikipedia.

**Evidence Provided** - Extract answers from given documents (long-context).

### Benchmark Study

- SotA LLMs' performance scale with their context length.
- As LLMs perform multiple searches, they "forget" earlier information as their context fills and truncates earlier info.
- We find a human upper bound of 84.7%.

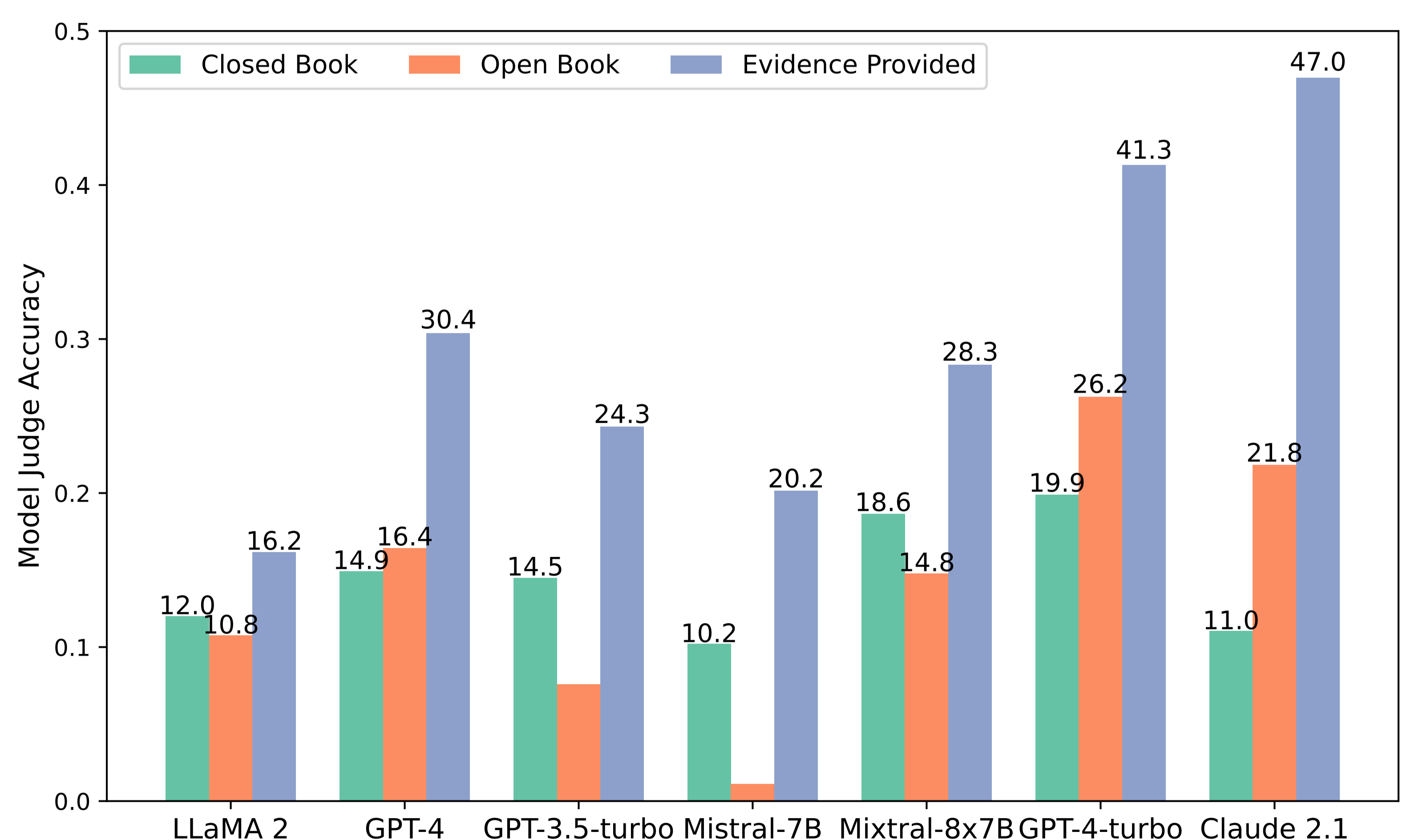
What are the **5 largest banks** in the world?

JPMorgan Chase, Bank of America, ..., HDFC Bank

How many employees does JPMorgan Chase have?  
308,669

How many employees does HDFC Bank have?  
177,000

How many employees does Bank of America have?  
217,000



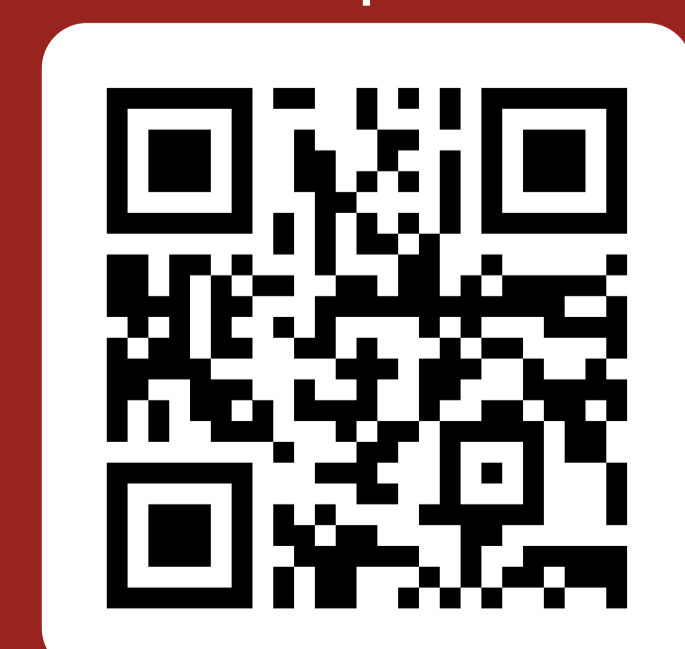
 pip install fanoutqa

 github.com/zhudotexe/fanoutqa

 fanoutqa.com

 Penn NLP @ University of Pennsylvania

Paper



Website

