AnaloBench: Benchmarking the Identification of Abstract and Long-context Analogies

Xiao Ye♡ **Andrew Wang**♡ Jacob Choi Yining Lu Shreya Sharma Lingfeng Shen Vijay Tiyyala Nicholas Andrews Daniel Khashabi

JOHNS HOPKINS UNIVERSITY

Do LLMs still perform well on challenging analogical reasoning tasks?

Motivation

AnaloBench moves beyond simple analogies for **challenging analogies** with **paragraphs** of raw-form text

Before - Simple Analogies

Text A: gust of wind : soft breeze **Text B:** fire : wisp of smoke

AnaloBench

Text A: In the heart of a small, poverty-stricken village, the weather began to take a frightful turn...The breeze gently kissed the cheeks of the shocked villagers, whispering an apology in their ears...

Text B: The candle had been burning for hours, casting a warm glow across the room... The smoke twirled and danced gracefully in the air, a subtle performance of nature, before slowly beginning to dissipate...

Dataset Creation

AnaloBench features 340 **human-contributed** analogies

① Analogical Pairs

Sentence A	Sentence B
You can't pour from an empty cup.	*A fallen tree cannot provide shade.*
You can't pour from an empty cup.	*It's hard to love with a broken heart.*
He danced off his sugar high then promptly fell asleep.	*After letting off his rage he sat down like a lamb.*
He danced off his sugar high then promptly fell asleep.	*The weather finally became pleasant following the stormy week.*
⋮	⋮

② Analogical Clusters

- *After letting off his rage he sat down like a lamb.*
- *He danced off his sugar high then promptly fell asleep.*
- *The weather finally became pleasant following the stormy week.*

- *You can't pour from an empty cup.*
- *A fallen tree cannot provide shade.*
- *It's hard to love with a broken heart.*
- *...*

⋮

- Each annotated analogy consists of a pair of analogous sentences
- **Sentences are rewritten as 10 or 30 sentence narratives**

Benchmark Tasks

Mini Story-Bank

Q: Which of the following is the most analogous sentence to the target sentence?

Target: You can't pour from an empty cup.

- ✓ **A fallen tree cannot provide shade.**
- ✗ All that glitters is not gold.
- ✗ After letting off his rage he sat down like a...
- ✗ A succession of waves battered the rock.

Given the following sentence, expand it into a 10 sentence story:

Sentence: **A fallen tree cannot provide shade**

Q: Which of the following is the most analogous sentence to the target sentence?

Target: Lily was the kind of person who was always running around trying to help others. She had an exceptional kind ...

- ✓ **Once upon a time, a man named Henry lived in a tiny, worn-out house. He...**
- ✗ In the small town of Baker's Crest, people lived simple lives, mostly untouched by...
- ✗ In the heart of a small, peaceful town lived a man named Gerald. He was an...
- ✗ A stormy morning on the coast revealed an awe-inspiring sight as a ...

Full Story-Bank

Q: Retrieve the top 10 analogous stories from the sentence bank…

Query You can't pour from an empty cup.

Doc. *n* A fallen tree cannot provide shade.

Once upon a time, a man named Henry lived in a tiny, worn-out house. He ...

GPT-4

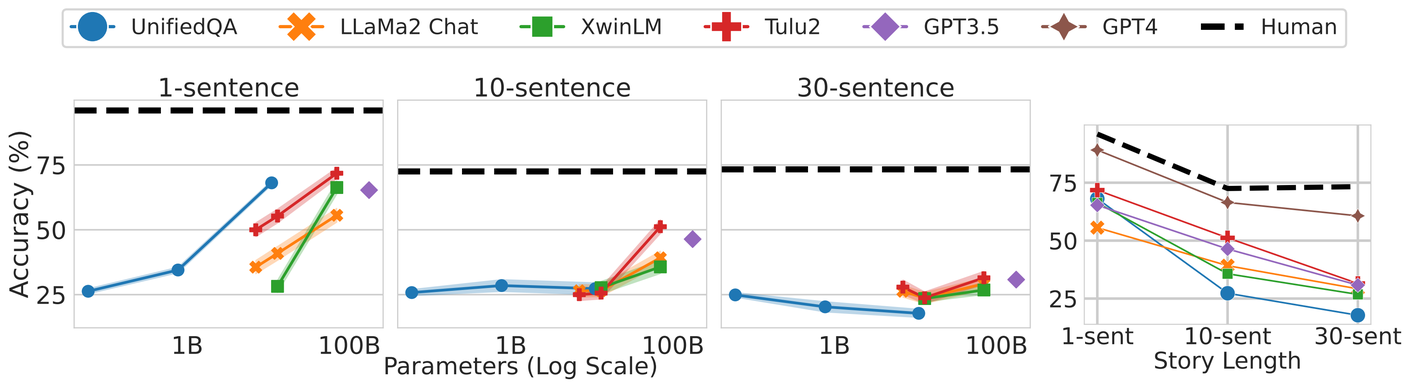
Q: Retrieve the top 10 analogous stories from the sentence bank…

Query Lily was the kind of person who was always running around trying to help others. She had an exceptional kind ...

Doc. *n* Once upon a time, a man named Henry lived in a tiny, worn-out house. He...

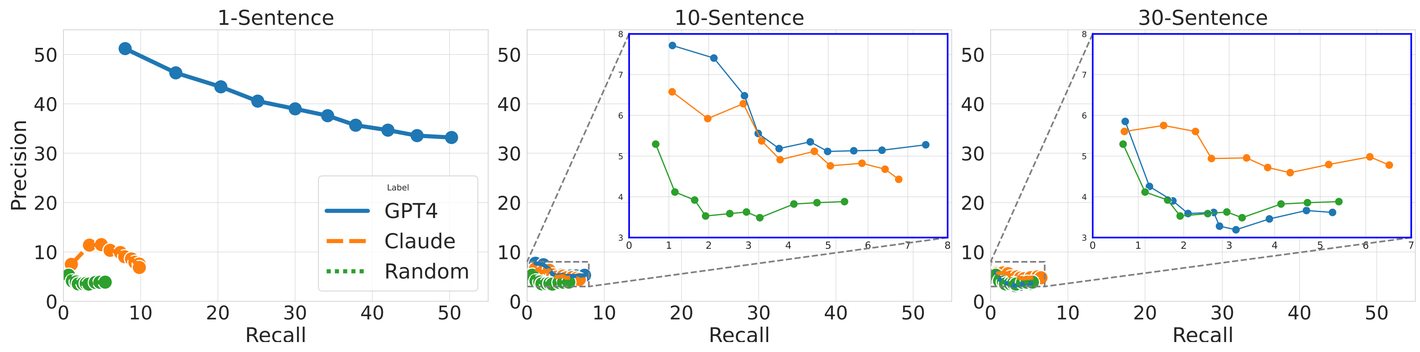
2 tasks: mult. choice (top) and analogical retrieval (bottom)

Each task is repeated at varying story lengths



Model size is less helpful at longer narrative settings ↑ **Human-AI gap**

Analogy retrieval performance degrades with increased narrative length



Summary

Github

Dataset

Background

- Analogical reasoning was challenging for early AI systems
- Recent works claim that analogical reasoning in large language models is emergent behavior (Webb et al., 2023)

Findings

Do LLMs still perform well on challenging analogical reasoning tasks?

- Long narrative analogies are difficult for LLMs
- LLMs perform poorly on analogical retrieval