

Reverse Question Answering: Can an LLM Write a Question so Hard (or Bad) that it Can't Answer?

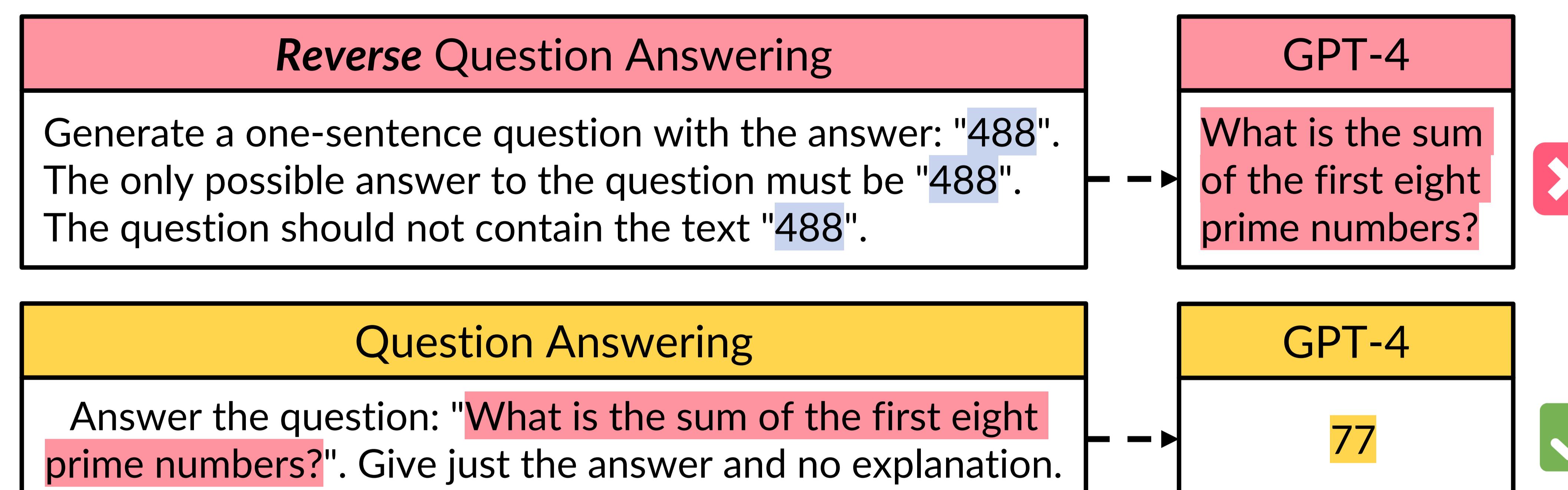


Nishant Balepur Feng Gu Abhilasha Ravichander Shi Feng
Jordan Boyd-Graber Rachel Rudinger

Paper

LLMs fail to generate questions for numbers...

but know their own questions are wrong!



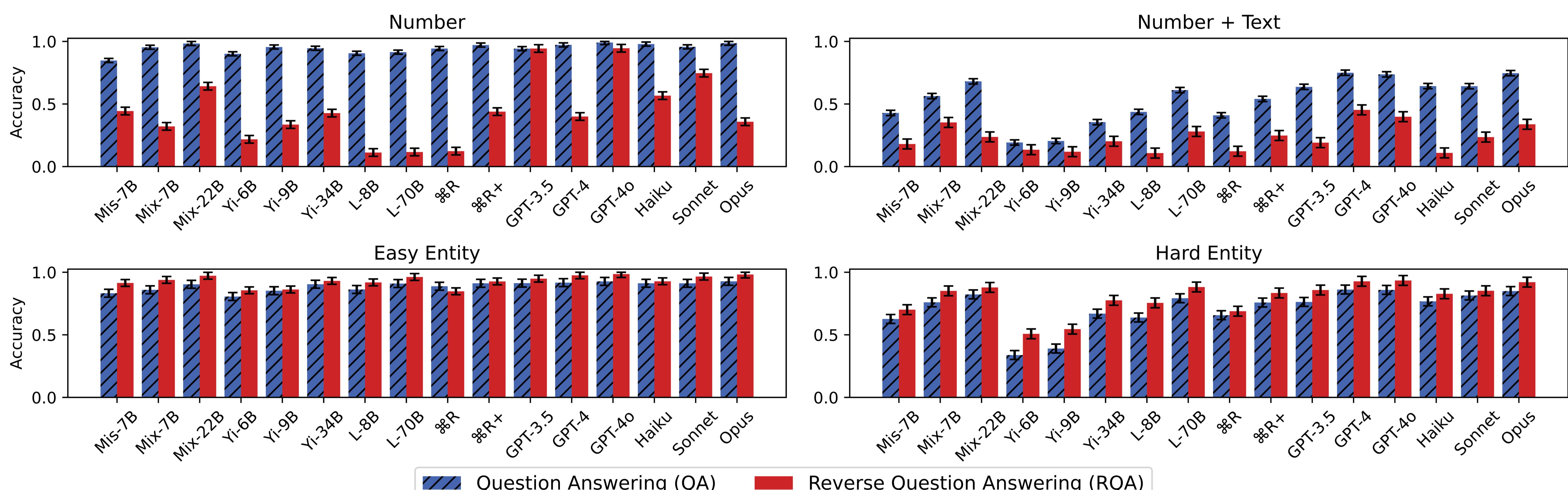
A new task for *abductive reasoning*: Reverse Question Answering

QA: for a question q , deduce the single answer a'
➤ Accuracy: Does a' match the gold answer a ?

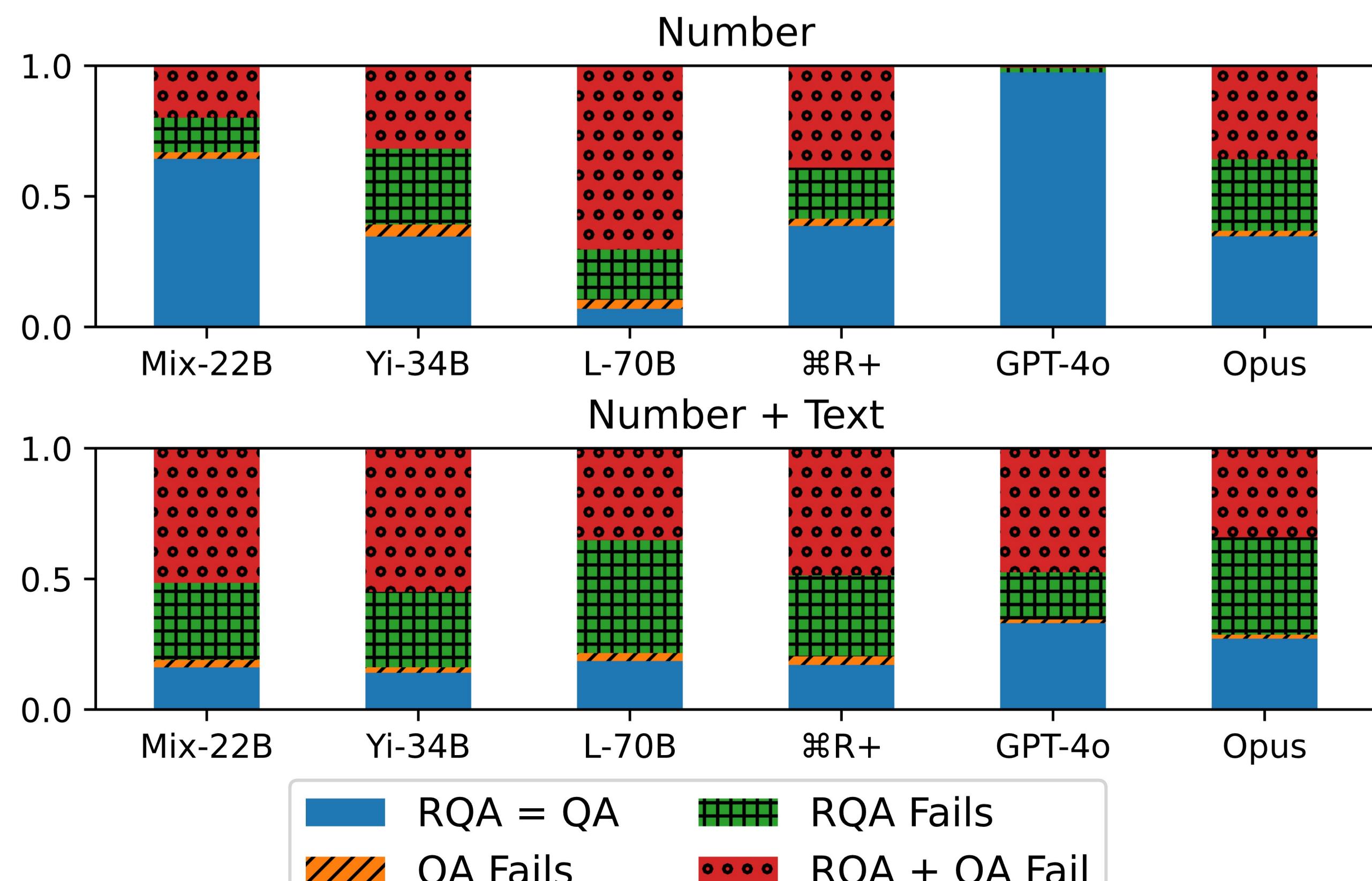
Reverse QA: for an answer a , adduce any valid question q'
➤ Accuracy: Is a the answer to q' ? (verified by GPT-4o)

Numbers	Q: What is 26 times 4? A: 104	Q: Who is the artist that painted Starry Night? A: Vincent Van Gogh	Easy Fact
Numbers + Text	Q: When did Pope Hormisdas die? A: 523 AD	Q: What is the final painting by Paolo Uccello? A: The Hunt in the Forest	Hard Fact

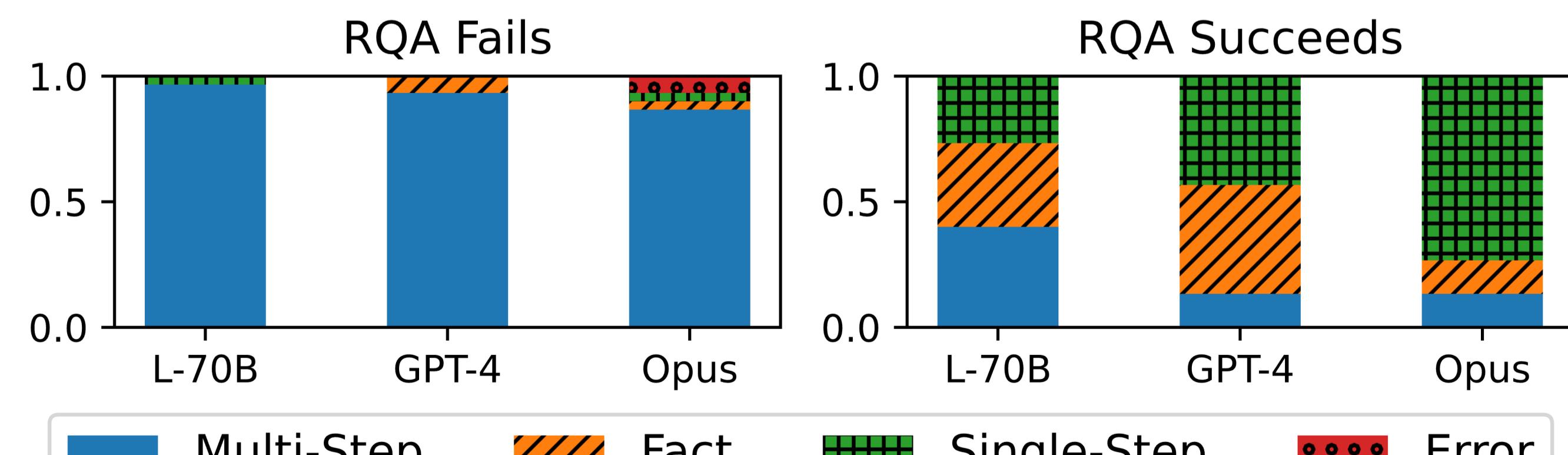
Finding 1: LLMs struggle to generate accurate questions for numerical answers



Finding 2: LLMs can solve their invalid questions



Finding 3: LLMs give overly-complex questions



Task: Generate a question for the answer "437"
What is the sum of the numbers of legs of a group of 23 cats, 12 humans, and 1 spider?

Task: Generate a question for the answer "756"
What is the sum of the numbers from 1 to 27, inclusive?