



EconLogicQA: A Question-Answering Benchmark for Evaluating Large Language Models in Economic Sequential Reasoning

Yinzhu Quan, Zefang Liu

Georgia Institute of Technology

Contact: yquan9@gatech.edu



Introduction

EconLogicQA is a benchmark designed to evaluate large language models' ability to perform complex **sequential reasoning in economics, business, and supply chain management** by focusing on logical event sequencing rather than simple event prediction.

Key Contributions

- ❑ **Novel Benchmark:** EconLogicQA tests LLMs' ability to reason about multiple interconnected events in realistic economic scenarios, moving beyond fact recall.
- ❑ **Dataset Generation:** Leveraging GPT-4 to generate a set of 650 multiple-choice questions from business articles, with questions focusing on the logical sequence of events.
- ❑ **Human Review:** A meticulous human review process ensures high-quality, accurate questions.
- ❑ **Comprehensive Evaluation:** Performance of state-of-the-art LLMs, including GPT, Gemma, Llama, and Mistral, was tested using both 1-shot and 5-shot learning settings.

Data Generation

- ❑ **Article Selection:** Extract economics-related articles from the 2011–2022 news dataset.
- ❑ **Question Creation:** Use GPT-4 with instructional prompts to generate multi-choice questions that require logical event sequencing.
- ❑ **Event Selection:** Focus on typical business/economic events, ensuring each question presents four options for logical ordering.
- ❑ **Manual Review:** Human reviewers ensure accuracy, remove ambiguous questions, and validate logical sequences.
- ❑ **Final Dataset:** Curated set of 650 high-quality questions, divided into training, validation, and test sets.

Prompt:

Generate a concise multi-choice question without referencing the original news article, starting with a brief scenario description. Include four events from this scenario that need sequencing based on their logical connection, not just the order of appearance in the news. These events should be relevant to typical business or economic situations, requiring knowledge of business practices and economics for correct sequencing. Ensure a single correct sequence, free from ambiguity. The question should be crafted for independent completion, suitable for student evaluation. Create original, concise content for both the question and choices, avoiding plagiarism and limiting excessive details or numerical data. Format the output with the scenario description, followed by choices labeled A, B, C, D, and finish with the correct sequence and a brief explanation, ensuring the sequence logic is deducible solely from the question and choices presented.

Here is the output format:

```

...
Question: ...
Choices:
A. ...
B. ...
C. ...
D. ...
Correct Order: ...
Explanation: ...
...

```

Here is the business news article:

```

...
{news_article}
...

```

Data Examples

Example 1

Question: Costco, a major retailer, is taking steps to control its chicken supply chain to maintain the price of its popular rotisserie chickens. Arrange the following events in the logical sequence based on Costco's strategy.

Choices:

- Costco recruits farmers to grow chickens for them.
- Costco opens a \$450 million poultry complex in Nebraska.
- Costco faces supply challenges and cost pressures in the poultry industry.
- Costco decides to keep its rotisserie chickens at \$4.99 despite competitors increasing their prices.

Correct Order: D, C, A, B

Example 2

Question: Arrange the following events related to the Federal Reserve's monetary policy in their correct order.

Choices:

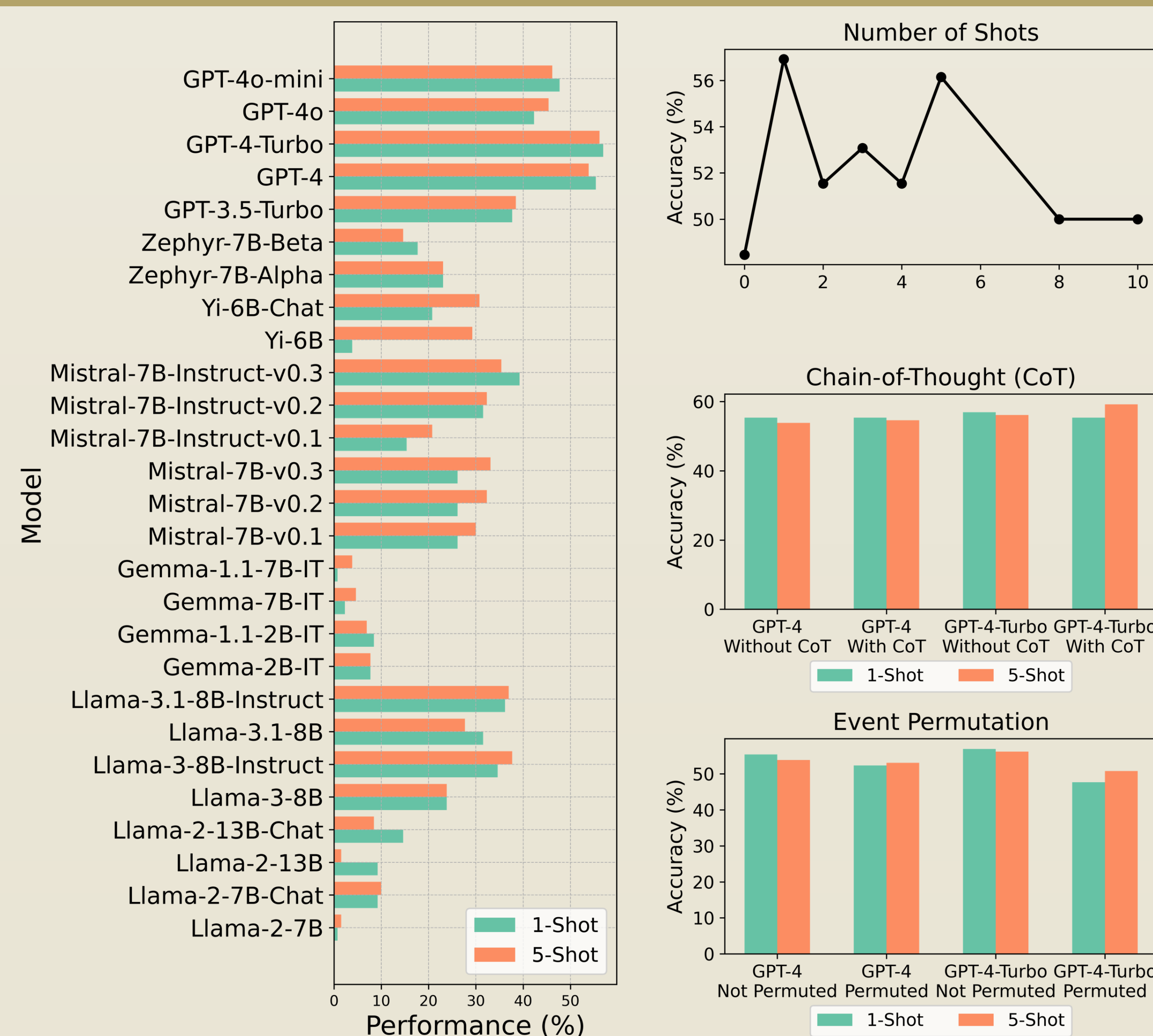
- The Federal Reserve announces the end of its pandemic-era stimulus and accelerates the roll-back of its asset purchases.
- The Federal Reserve signals it would hike interest rates multiple times throughout 2022.
- The Federal Reserve reduces its massive balance sheet.
- The Federal Reserve raises interest rates due to inflation well above 2% and a strong labor market.

Correct Order: A, B, D, C

Word Cloud



Evaluation Results



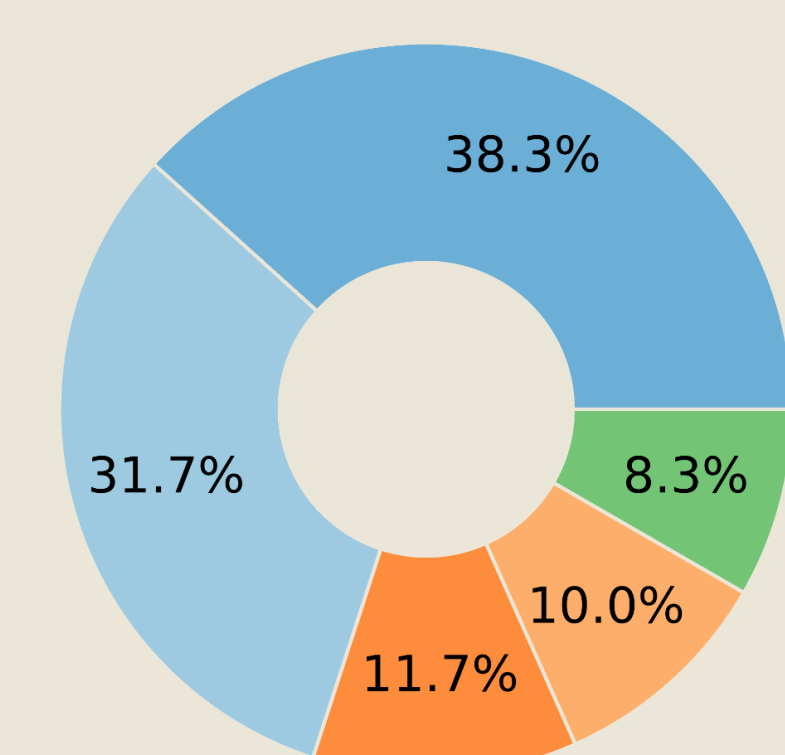
Error Analyses

[Findings]

Common Errors: LLMs often struggle with strategic decision-making sequences and policy/regulatory misunderstandings.

Other Challenges: Models exhibit difficulties in distinguishing between stages of implementation, linking actions to outcomes, and sequencing growth-related activities.

Insights: These errors highlight significant challenges in handling complex economic reasoning, suggesting areas for future improvements in LLM capabilities.



- Strategic and Decision-Making Confusion
- Policy and Regulatory Misunderstanding
- Implementation and Development Confusion
- Impact and Outcome Misinterpretation
- Growth and Expansion Misinterpretation

[Example]

Question: The Federal Reserve is considering the creation of a digital currency. What steps did they take in sequence?

Choices:

- The Federal Reserve collects public feedback on the digital currency.
- The Federal Reserve releases a discussion paper on the digital currency.
- The Federal Reserve receives clear support from the public.
- The Federal Reserve addresses major policy questions.

Correct Answer: B, A, C, D

Response: B, A, D, C

Error Type: Policy and Regulatory Misunderstanding

Indicators Matched: Misinterprets the sequence of policy changes (placing policy questions before public support).

Feedback: The response indicates a misunderstanding of the sequence in which the Federal Reserve addressed policy questions and gathered public support. It's important to understand that the Federal Reserve first gathers public feedback before addressing policy questions.

Sources

