

Subword Spotting: Uncovering Embedded Objects

Natural Language Processing Course – Final project

Daniel Bazar¹ and Lior Krengel²

¹314708181, ²315850594

May 2024

Abstract

This work investigates the ability of LLMs to perform a seemingly simple linguistic task: identifying subwords embedded within other words. We designed a dataset and an automatic evaluation method for this task and conducted experiments under various prompting techniques, question formats, and different LLMs. The results show consistently low accuracy for LLMs, often similar to random guessing, while humans perform well. This suggests an inherent limitation within these models for tasks requiring pattern matching without clear semantic or structural relationships. Eventually, we suggest potential methods to address these limitations and enhance LLM performance in such scenarios.

1 Introduction

Large language models (LLMs) are believed to contain vast knowledge. They demonstrated impressive capabilities of language understanding, language transformation and language generation, but they are still far from perfect. Despite their versatility, they are surprisingly struggling with very simple tasks that humans find trivial.

In this article, we will explore such one task, representing the language-based ability of "pattern matching", using our designed task - finding subwords inside other words. Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. We asked multiple LLMs to find such subwords but with a little tweak, we didn't provide the subword to search explicitly, but rather implicitly, asking for a word from a group of words (category) instead of a plain word. For example, instead of asking to find the word "red" inside "incredible" we asked to find a "color". This extra layer of complexity rarely affects human performance but makes it more difficult for LLMs. It is important to note that even without it, LLMs face difficulties. Prompting techniques such as in-context learning (ICL), chain-of-thought (CoT), and decomposing were tried to elicit better results than zero-shot but they didn't provide worthy improvement. Different question formats and different models didn't remedy the problems either. In the next sections, we discuss in detail about our linguistic task, our evaluation methodology, the dataset construction, the experiments we conducted, and their results.

2 Linguistic task investigated

We demonstrate a linguistic task, a type of word puzzle that involves identifying embedded words within other words. Specifically, this task asks to find a subword that is also the name of an object from some category (for example, colors, animals, etc.).

Question for example:

Which one of the following words contains a subword of an animal?

- A. guitar
- B. **million (contains "lion")**
- C. pasta
- D. house

We anticipate that the LLM will be able to effectively distinguish subwords from other words. Although LLMs are not specifically trained for this task, we assume that they possess the necessary ability to match patterns. We find out that identifying subwords becomes more challenging when there is no clear semantic or structural relationship between them and the surrounding characters or the whole word. LLMs rely on patterns and associations in language to understand words and their meanings. When there is no meaningful connection, the model may struggle to detect the relevant subword. For instance, in the word "million", which contains the subword "lion", the remaining characters ("mil") do not carry any obvious semantic or structural significance, therefore the LLM can miss the subword "lion". In addition, there is no meaningful semantic connection between the word "million" and "lion", which makes it harder for the LLM.

3 Evaluation methodology

3.1 Dataset

Our aim in generating the dataset was an automatic approach that construct pairs of words from a category (the subword) and a word containing the subword. Our main guideline was to choose commonly used subwords so it would be easier to recognize them, especially for non-native English speakers. To achieve this, we first collected several word lists of different categories such as animals, body parts, colors, food, fruits, and vehicles. We ensured manually that the words were simple and known. Then we retrieved the "brown corpus" and used its vocabulary to fit subwords into words. We removed stop words and lemmatized all the words to prevent unwanted and duplicated words. After that, we remained with pairs of words and subwords. This dataset contains thousands of records and for our task, we were interested in a much smaller and high-quality one. To dilute it further, we used a pre-trained word2vec model (google-news-300) in order to avoid similar pairs of words and subwords. For example, burger and hamburger, boat and sailboat, brown and brownish, etc. This method also has another advantage, by choosing a pair with smaller similarity we force (hopefully) the LLM to rely less on the semantic meaning of the words and more on pattern matching, in this way, we can evaluate the

performance better. Some of the words appeared more frequently as subwords than others (Zipf’s law is everywhere!) as shown in figure 1. So, to construct a diverse dataset we decided to take each word once. Then we attached to each subword, the most frequent word according to the brown corpus that contains it. The last step in the process was a manual intervention, we passed through the dataset and validated the results. The distillate dataset contains 60 records. Each record is of the following format:

Category	Subword	Word	Multple Wrong Options
color	red	incredible	['stopped', 'turned', 'lower']
		...	
vehicle	truck	struck	['industry', 'block', 'never']

Table 1: few examples from the dataset.

Then we used this dataset to create different types of prompts as we will describe in section 3.2.2. We left aside some more records for few-shot prompting.

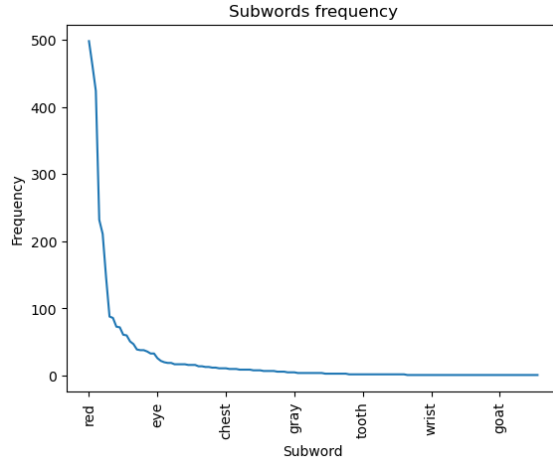


Figure 1: sub-words frequency. representing Zipf’s law.

3.2 Experiments

We conducted experiments in two different settings: multiple-choice questions and Yes/No questions. In the multiple-choice setting, we were asked to select the correct answer from four given options. In the Yes/No setting, we asked to indicate whether a condition was met or not. The use of these different question formats aims to provide us with a better understanding of the performance of our models. (see the different question formats in appendix A.2).

In each settings we experiment with various models and prompting styles as you can see in the sections below.

3.2.1 Models

For finding the right LLM for our experiments we used Together AI API. We started by conducting some small experiments in their on-site playground and eventually accessed them on a higher scale using their API. We chose 3 different LLMs to focus on, to show that even larger (Mixtral) and newer (Llama-3) models are failing with

this task. All of these models are trained on at least 7B parameters and have shown the ability to understand the task we assigned to them. The three models we chose are Mixtral-8x7B, OLMo 7B, and Llama 3 8B. In all of the experiments, we set the model temperature to 0, which allowed us to obtain deterministic responses from the LLM. this is crucial in question answering, especially in multi-choice questions. In addition, we set the max tokens to 128 for the reason we were interested only on the beginning of the response and ignored the explanation.

3.2.2 Prompts

As we mentioned, we performed experiments under two settings, multiple-choice questions, and Yes/No questions. We test each one with zero-shot, one-shot, few-shot, chain-of-thought (both one and few shots), and decomposed prompting styles. Each prompt contains a question, led by a short explanation about what is a subword, and room for an answer. For example, a basic prompt without any engineering for Yes/No questions will look like this:

Example Yes/No basic prompt

Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.

Question: Does the word "{word}" contains a subword of a {category}?

Answer:

In the zero-shot and decomposed prompts (mainly), we tackled a problem in which the response was not in a fixed format and it was hard to extract the answer and evaluate it (more on this in section 3.3). Therefore, we added to the answer part of the prompt, an instruction to respond at the beginning with desirable output. for example for Yes/No question we added: "[Your answer here, just Yes/No, without any explanation or additional text]". For the few-shot prompts, we use 4 examples. We took the examples out of the dataset for not asking a question with the answer already present in the examples. To clearly separate the example questions from the actual one, we added at the top of each example a "separator" like this: "### example question ###" and for the actual one: "### actual question ###". To generate the decomposed prompt, we ask an LLM for the sub-tasks it takes for it to solve this task. After back-and-forth experiments, we conducted from the different responses a well-structured decomposed prompt (more on the prompts and examples in Appendix A.2).

3.3 Metrics

We chose to evaluate the experiments with the accuracy metric, like a test, how many of the given questions were answered correctly. For each model, we calculated accuracy per prompting technique. This allows us to test the models in different environments.

We want to automatically evaluate the models. LLMs generate random and unstructured texts, which makes them hard to interpret. The approach we selected to tackle this problem was to engineer the prompt with instructions ensuring the response begins with the answer and then the explanation if needed. In the few-shot setup, we provide the model answer examples, and for other techniques, we added explicit instructions (recall section 3.2.2). We search at the beginning of the response

matching for one of the options letters A-D, Yes or No, or some prefix of the format "the answer is..." etc. The reason we search for matching only at the beginning of the response and not in all the responses is that we notice that sometimes the models refer to more than one option. For example in their explanation, sometimes, the model explains each option whether or not it contains a subword. This makes it hard to distinguish what is the correct answer. Subsequently, we ignored the explanation. In case of a wrong explanation but a correct answer we counted it as correct answer and vice-versa, in case of a right explanation and wrong answer we counted it as a wrong answer. In 99% of the cases we successfully extract the answer and we manually extract the answer from the remaining 1%.

For comparing the performance of LLMs and humans we construct a human benchmark as described in section 3.4. All results are shown in table 2. For more details about the results see section 4.

(A) Zero-shot

Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Which one of the following words contains a subword of a color?
A. lower
B. turned
C. incredible
D. stopped
Start your response with the correct letter and the correspond word. In case of doubt, answer according to the most probable answer.
Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]

(Response) C. incredible
The subword "blue" is a color and is part of the word "incredible". It's the subword from position 2 to 8 in the word. ✓

(B) One-shot CoT

example question ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "training" contains a subword of a vehicle?
Answer: Yes. Since the word training contains the subword train, which is a vehicle, the correct answer is Yes.
actual question ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "service" contains a subword of a vehicle?
Answer: [Your answer here, just Yes/No, without any explanation or additional text]

(Response) Yes. The word "service" contains the subword "car," which is a type of vehicle. ✗

(C) Few-shots

example question 1 ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "million" contains a subword of an animal?
Answer: Yes.
example question 2 ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "book" contains a subword of an animal?
Answer: No.
example question 3 ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "pants" contains a subword of an animal?
Answer: Yes.
example question 4 ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "printer" contains a subword of an animal?
Answer: No.
actual question ###
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Does the word "wanted" contains a subword of an animal?
Answer: [Your answer here, just Yes/No, without any explanation or additional text]

(Response) No. ✗

(D) Decomposed

Read the question below and understand what the question is asking for and the criteria for selecting the correct answer.
Examine each word provided in the options carefully and Break down each word into its component parts or subwords.
Determine if any of the subwords within each word match the name of an animal.
Choose the word that contains a subword of an animal according to the criteria given in the question.
Verify your answer, double-check your selection to ensure it meets all the requirements specified in the question.
Start your response with the correct letter and the correspond word. In case of doubt, answer according to the most probable answer.
Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.
Question: Which one of the following words contains a subword of an animal?
A. people
B. change
C. still
D. education
Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]

(Response) D. education
Explanation:
The word "education" contains the subword "cat" which is a type of animal. Therefore, option D is the correct answer. ✓

Figure 2: Few examples of input and output of Mixtral-8x7B. Showing the different prompt techniques and the response from the model. The irrelevant part of the answer that we were ignored is in blue. An example of a case when a correct answer was returned with a wrong explanation can be seen in (A). Our evaluation result is indicated in a check and cross marks. In green are the few-shot examples.

3.4 Human Benchmark

In order to show that humans indeed perform well on this task, we created multiple forms and asked our colleagues to fill them. Each form included 10 questions from our dataset. The questions were the same as we prompted the LLMs and we added

at the beginning of each form a short explanation about this task. Overall, we created 18 forms, 6 for multi-choice and 12 for Yes/No questions. There were 54 responses for both of them. We got an accuracy of 97.2% and 98.8% for multi-choice and Yes/No questions, respectively. These results confirms that humans had no problem performing such task (see Appendix A.3 for more details).

4 Results

As we described in previous sections we experimented with different models, prompts, and question types. Overall from a dataset of 60 subwords, we created a pool of 3240 questions. This resulted from 3 models, 6 prompting techniques, and 3 kinds of questions (for the Yes/No format we added a No question), for each subword.

As indicated on table 2, all models performs poorly. For part of the cases, they didn’t do better than random guessing. In almost every experiment the Mixtral model acquires the best accuracy. In the multiple-choice format, the best accuracy obtained with combination of CoT and few-shot prompting and in Yes/No format best accuracy was obtained with one-shot prompting (for comparison in plot representation, see Appendix A.4). In addition, we performed a qualitative analysis of the responses of the models and showed some examples in Figure 2. We first note that the model understands the instructions, and answers in evaluable format. Then we conducted an error analysis of the responses to find patterns in the wrong responses. These analyses contributed to the development of the automatic quantitative metric (section 3.3).

Overall, The results illustrate the poor performance of the LLMs and the superiority of humans in this kind of task over them.

		Zero-shot	One-shot	Few-shot	CoT-one	CoT-few	Decomposite
Multiple-choice	OLMo	35.0%	23.3%	25.0%	30.0%	25.0%	18.3%
	Llama	26.7%	23.3%	36.7%	25.0%	45.0%	28.3%
	Mixtral	50.0%	41.7%	53.3%	50.0%	58.3%	48.3%
	Random	25.0%					
	Human	97.2%					
Yes/ No	OLMo	50.0%	50.0%	51.7%	50.0%	52.5%	49.2%
	Llama	49.2%	55.8%	57.5%	57.5%	61.7%	51.7%
	Mixtral	59.2%	75.8%	55.8%	60.8%	55.0%	72.5%
	Random	50.0%					
	Human	96.8%					

Table 2: Full results on all experiments

5 Conclusion

We tested several LLMs on a simple task that humans find trivial, with additional experiments of different question formats, and prompting techniques. The results demonstrate that LLMs do not perform well on this task, especially compared to humans, many settings yield results akin to random guessing. The repeating failure across all experiments indicates that this limitation is fundamental inside LLMs. Due to the growing interest in LLMs, we believe that understanding their capabilities and limitations is important because its essential for harnessing their full potential.

6 Mitigation

While our analysis has revealed a certain limitation in the performance of LLMs, there are potential ways for improvement that could remedy the problem we observed. One obvious approach is to train an LLM for this specific task, but of course it is not so feasible. One feasible approach, that we notice from a small analysis, is to double check the response. More precisely, it means to check with the model if it is sure in its answer. If it is, return that answer if not, ask the model again, this time knowing that the previous answer is wrong. Some different variation of this approach will be to set the temperature of the model little higher than 0, asking the model the same question several times, and then using "majority voting" to select to correct answer. Another approach we assess is prompt chaining. Similar to the decomposed prompt, we break the task into its sub-tasks, but this time, instead of prompting all of them together, the LLM is prompted with one sub-task at a time, and then its response is used as input to another sub-task prompt. One more possible way to remedy the problem, is to replace the "category" with a list of options. This approach has the potential to distract the LLM from the "semantic pit" and make it easier for it to find one of them inside another word. We can achieve this list by using word embedding to find the most similar words to this "category".

A Appendix

A.1 Code

Code is available at <https://github.com/danb7/Subword-Spotting.git>

A.2 Prompts

This section summarizes more example of prompts we use in our experiments. For readability all texts are modified from the original ones by omitting or inserting some linebreaks. Correct and wrong answers are colored in green and red, respectively. For all examples, we use the Mixtral model. For the all prompts see the [git](#) repository.

A.2.1 Yes/No questions

Table 3: Example templates used for the Yes/No questions with answers.

Prompt technique	Question	Answer
Zero-shot	Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "incredible" contains a subword of a color? In your response, first Answer with "Yes" or "No". In case of doubt, answer according to the most probable answer. Answer: [Your answer here, just Yes/No, without any explanation or additional text]	Yes, the word "incredible" contains the subword "red," which is a color.
One-shot	### example question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "training" contains a subword of a vehicle? Answer: Yes. ### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "scarcely" contains a subword of a vehicle? Answer: [Your answer here, just Yes/No, without any explanation or additional text]	No.

Continuation of Table 3		
Prompt technique	Question	Answer
Few-shot	<p>### example question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "training" contains a subword of a vehicle? Answer: Yes.</p> <p>### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "scarcely" contains a subword of a vehicle? Answer: [Your answer here, just Yes/No, without any explanation or additional text]</p>	No.
CoT-one	<p>### example question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "fingerprint" contains a subword of a body_part? Answer: Yes. Since the word fingerprint contains the subword finger, which is a body_part, the correct answer is Yes.</p> <p>### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "potatoes" contains a subword of a body_part? Answer: [Your answer here, just Yes/No, without any explanation or additional text]</p>	Yes. The word "potatoes" contains the subword "toe," which can be a body part, so the answer is yes.

Continuation of Table 3		
Prompt technique	Question	Answer
CoT-few	<p>### example question 1 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "pearl" contains a subword of a fruit? Answer: Yes. Since the word pearl contains the subword pear, which is a fruit, the correct answer is Yes.</p> <p>### example question 2 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "book" contains a subword of a fruit? Answer: No. Since the word book doesnt contains any subword of a fruit, the correct answer is No.</p> <p>### example question 3 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "sublime" contains a subword of a fruit? Answer: Yes. Since the word sublime contains the subword lime, which is a fruit, the correct answer is Yes.</p> <p>### example question 4 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "printer" contains a subword of a fruit? Answer: No. Since the word printer doesnt contains any subword of a fruit, the correct answer is No.</p> <p>### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Does the word "grapple" contains a subword of a fruit? Answer: [Your answer here, just Yes/No, without any explanation or additional text]</p>	<p>Yes. The word "grapple" contains the subword "apple," which is a fruit.</p>

Continuation of Table 3		
Prompt technique	Question	Answer
Decomposed	<p>Read the question below and understand what the question is asking for and the criteria for determine the correct answer.</p> <p>Examine the word provided in the question carefully and Break it down into its component parts or subwords.</p> <p>Determine if any of the subwords within the word match the name of a vehicle.</p> <p>Answer with "Yes" or "No" only, without explanations. In case of doubt, answer according to the most probable answer.</p> <p>Verify your answer, double-check your classification to ensure it meets all the requirements specified in the question.</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Does the word "scarcely" contains a subword of a vehicle?</p> <p>Answer: [Your answer here, just Yes/No, without any explanation or additional text]</p>	<p>No.</p> <p>I have checked all the subwords of the word "scarcely" and found no subwords that are names of vehicles. Therefore, my answer is "No."</p>

A.2.2 Multiple-choice questions

Table 4: Example templates used for the multiple-choice questions with answers.

Prompt technique	Question	Answer
Zero-shot	<p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of an animal?</p> <p>A. school B. knowledge C. russian D. looked</p> <p>Start your response with the correct letter and the correspond word. In case of doubt, answer according to the most probable answer.</p> <p>Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	<p>A. school</p> <p>The subword "sh" is an arbitrary part of the word "school," and it is also a part of the word "shark," which is an animal.</p>

Continuation of Table 4		
Prompt technique	Question	Answer
One-shot	<p>### example question ###</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of a vehicle?</p> <p>A. book B. printer C. training D. guitar</p> <p>Answer: C. training</p> <p>### actual question ###</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of a vehicle?</p> <p>A. service B. asked C. scarcely D. thinking</p> <p>Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	A. service

Continuation of Table 4		
Prompt technique	Question	Answer
Few-shot	<p>### example question 1 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of a color?</p> <p>A. credit B. house C. lamp D. book Answer: A. credit</p> <p>### example question 2 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of a color?</p> <p>A. shoes B. blueprint C. phone D. guitar Answer: B. blueprint</p> <p>### example question 3 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of a color?</p> <p>A. truck B. wallet C. stingray D. lantern Answer: C. stingray</p> <p>### example question 4 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of a color?</p> <p>A. television B. blanket C. picked D. marigold Answer: D. marigold</p> <p>### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of a color?</p> <p>A. stopped B. lower C. turned D. incredible Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	D. incredible

Continuation of Table 4		
Prompt technique	Question	Answer
CoT-one	<p>### example question ###</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of a body_part?</p> <p>A. book B. printer C. fingerprint D. guitar</p> <p>Answer: C. Since the word fingerprint contains the subword finger, which is a body_part, the correct answer is C. fingerprint.</p> <p>### actual question ###</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of a body_part?</p> <p>A. described B. feeling C. operation D. picked</p> <p>Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	<p>C. operation, since the word contains the subword "op," which is a shortening of "operate," a body part that is involved in performing operations.</p>

Continuation of Table 4		
Prompt technique	Question	Answer
CoT-few	<p>### example question 1 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's. Question: Which one of the following words contains a subword of an animal?</p> <p>A. million B. book C. printer D. guitar</p> <p>Answer: A. Since the word million contains the subword lion, which is an animal, the correct answer is A. million.</p> <p>### example question 2 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of an animal?</p> <p>A. pasta B. pants C. glass D. house</p> <p>Answer: B. Since the word pants contains the subword ant, which is an animal, the correct answer is B. pants.</p> <p>### example question 3 ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of an animal?</p> <p>A. camera B. milk C. health D. envelope</p> <p>Answer: C. Since the word health contains the subword eal, which is an animal, the correct answer is C. health.</p> <p>### actual question ### Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of an animal?</p> <p>A. finding B. dinner C. operation D. water</p> <p>Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	<p>A. finding, since it contains the subword "fish" which is an animal.</p>

Continuation of Table 4		
Prompt technique	Question	Answer
Decomposed	<p>Read the question below and understand what the question is asking for and the criteria for selecting the correct answer.</p> <p>Examine each word provided in the options carefully and Break down each word into its component parts or subwords.</p> <p>Determine if any of the subwords within each word match the name of an animal.</p> <p>Choose the word that contains a subword of an animal according to the criteria given in the question.</p> <p>Verify your answer, double-check your selection to ensure it meets all the requirements specified in the question.</p> <p>Start your response with the correct letter and the correspond word. In case of doubt, answer according to the most probable answer.</p> <p>Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.</p> <p>Question: Which one of the following words contains a subword of an animal?</p> <p>A. father</p> <p>B. reason</p> <p>C. governor</p> <p>D. debate</p> <p>Answer: [Your answer here, just the correspond letter and the word, without any explanation or additional text]</p>	<p>The correct answer is C. Governor. The subword "goose" is embedded in the word "governor." None of the other options contain a subword that is an animal's name. Therefore, the answer is:</p> <p>C. Governor</p>

A.3 Humans performance evaluation

To create the human baseline, we used examples from our dataset and created 18 forms using Google Forms. 6 for multiple-choice and 12 for Yes/No. We published those forms for our colleagues, without any English-level selection. We got a total of 54 responses. In every form humans got a minimum score of 90% accuracy, with an average of 97.2% in multiple-choice questions and 96.8% in Yes/No questions. Below are the links for the forms we created, and some screenshots examples from them.

A.3.1 Multi-choice quizzes links

Quiz 1: <https://forms.gle/qSEQEJfkg8ZCdNov5>

Quiz 2: <https://forms.gle/vkMM2WmDTZizvCQB9>

Quiz 3: <https://forms.gle/oA2U7XJ1f4o9P59M8>

Quiz 4: <https://forms.gle/dFbKYWg5qGsbLmTD7>

Quiz 5: <https://forms.gle/DD81KBMGwPLaQLGE8>

Quiz 6: <https://forms.gle/ZHqy7dM7cGkGdLuu5>

A.3.2 Classificaion quizzes links

Quiz 1: <https://forms.gle/4YStq3bohKbtaDkt7>
Quiz 2: <https://forms.gle/mGDC1vF7ucsXzMBD7>
Quiz 3: <https://forms.gle/kVpxe7e2DVcJAANy9>
Quiz 4: <https://forms.gle/YFmVuEKQxuMw2s3U9>
Quiz 5: <https://forms.gle/3tk1uT1DzS95UvUBA>
Quiz 6: <https://forms.gle/LCYG1fvnVyWpGK8y5>
Quiz 7: <https://forms.gle/5jSLetWdRcvYnSoGA>
Quiz 8: <https://forms.gle/XPosT1CdKRxKEoSQA>
Quiz 9: <https://forms.gle/TjVDzUx8EMPCuCHP8>
Quiz 10: <https://forms.gle/XzekrAsVTiRmBg669>
Quiz 11: <https://forms.gle/5D9qBbjTGJVfYTuBA>
Quiz 12: <https://forms.gle/MF6RSWLdbLeVtMgK7>

A.3.3 Forms screen-shots

Subword Spotting Quiz

Welcome to the Subword Spotting quiz!

In this quiz, you will be presented with a word and your task is to identify a subword within it that is also the name of an object from a specific category, such as colors, animals, etc.

Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.

To get you started, consider the following question:

example:

Which of the following words contains a subword of an animal?

1. guitar
2. **million**
3. pasta
4. house

The correct answer is 2, which contains the word "lion". Remember to read each question carefully and select the correct answer from the given options.

You will be presented with ten multiple-choice questions, please answer all of them. Good luck!

Which of the following words contains a subword of an animal? *

10 points

- ☐ still
- ☐ education
- ☐ people
- ☐ change

Figure 3: Human benchmarks multi-choice quizzes

Subword Spotting Quiz

Welcome to the Subword Spotting quiz!

In this quiz, you will be presented with a word and your task is to identify a subword within it that is also the name of an object from a specific category, such as colors, animals, etc. Subword is a standalone word that exists within a longer word that contains it, and its meaning is arbitrary from the longer word one's.

To get you started, consider the following question: example:

Does the word "million" contain a subword of an animal?

The correct answer is "Yes", because "million" contains the word "lion". Remember to read each question carefully and select the correct answer. You will be presented with 10 Yes/No questions, please answer all of them. Good luck!

Does the word "scarcely" contain a subword of a vehicle? *

10 points

- ☐ Yes
- ☐ No

Figure 4: Human benchmarks Yes/No quizzes

A.4 Results plots

The result in the article are in a table format. here are another representation of the comparison between our experiments in a plot format.

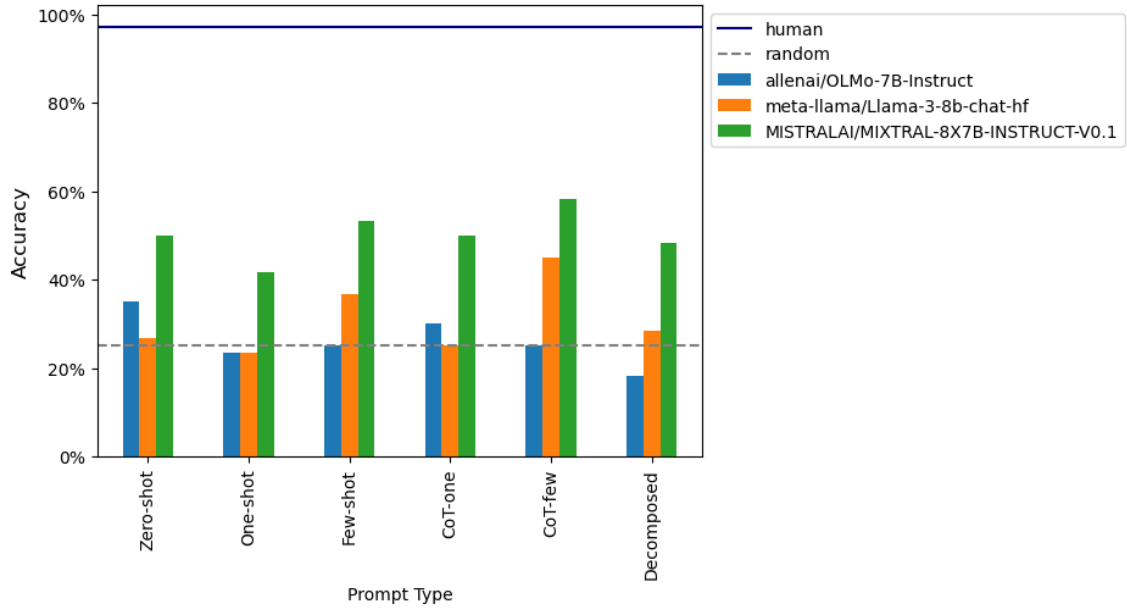


Figure 5: Comparison of different models and prompting techniques in Multiple-choice setting.

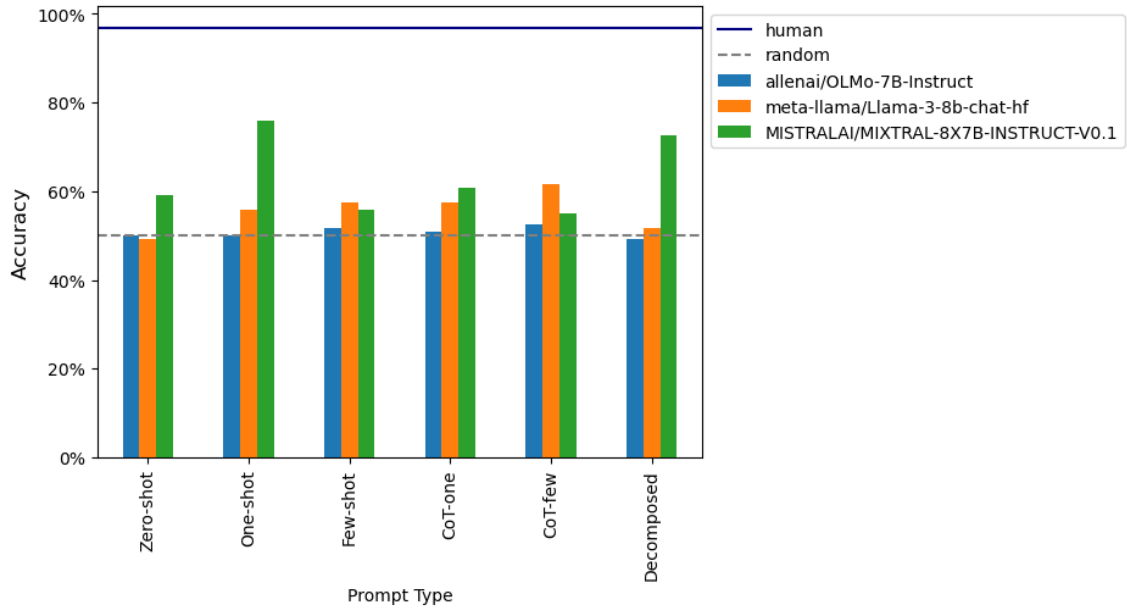


Figure 6: Comparison of different models and prompting techniques in Yes/No setting.

A.5 Dataset

Some more examples from our dataset. For the full dataset see the [git](#) repository.