Defining and Testing Essential Number Properties for LMs

Karan Raj Bagri & Dhamale Vivek Shrikrishna & Pawar Rushikesh Gajanansa
Department of Computational and Data Sciences, IISc

Research Problem : Do LMs Really Understand Numbers?

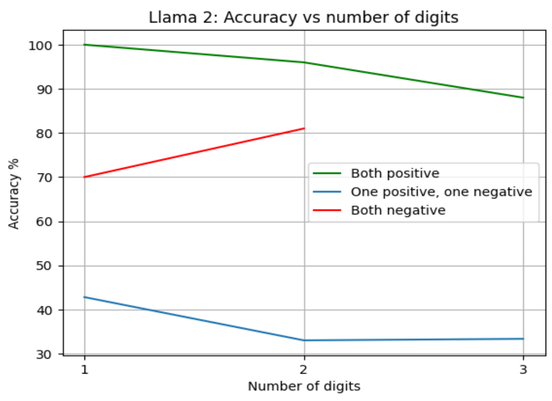
- Are numbers different from alphabets?
  - ➢ The alphabet is a set of letters used to form words and communicate language. Numbers/ digits, on the other hand, are symbols used to represent quantities and perform mathematical operations. Operations of numbers and words are inherently different.
- Is there need for LMs to know numbers?
  - ➢ Numerals account for 6.15% of all unique tokens in English Wikipedia. Numbers are ingrained in the real world. Understanding quantities, dates, measurements, and other numerical data can improve the context and factuality of LLM response.
- What does it mean to know numbers?
  - ➢ Understanding quantities (symbol ↔ Quantity), Number sense (bigger, smaller), order them (1st, 2nd, 3rd), arithmetic operations, Applying numbers in real-world contexts.

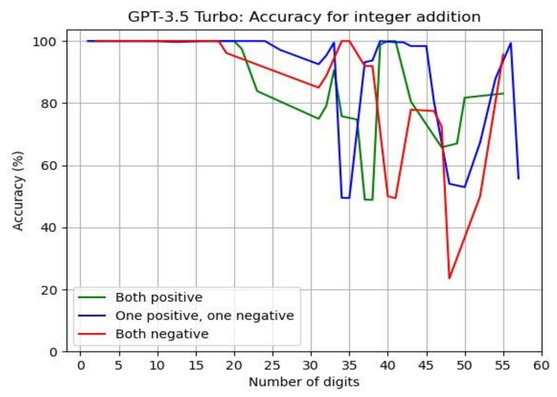
1

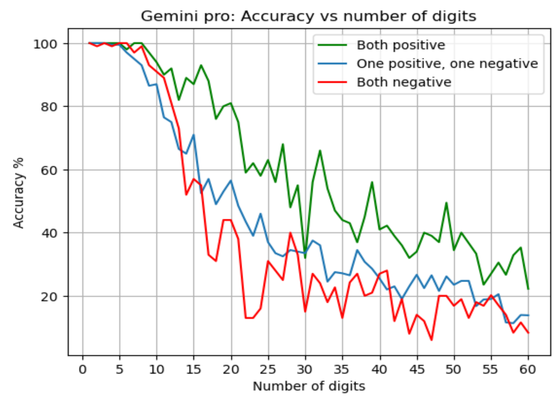
Number Property	Example	Test
Numeration: Identify numbers represented in any surface forms like words, floats or strings.	Two ↔ 2 ↔ 2.0 ↔ II	Asking LMs to convert one form of number to another.
Magnitude: Value and order	Number= $x.10^{y}$, x is value, y is magnitude	Finding Minimum, Maximum from list, Sorting list
Arithmetic Operations : Addition, multiplication	2 + 3 = 5 , 10/ 2 = 5	Additions, multiplication at scale
Type of Number : Positive/Negative, odd/even, Prime-composite	2 > -2 , prime : 2,3,5,7…	Min-Max, distinguishing between prime and composite

2 Test : Integer addition

- **Prompt given:** "You are a math assistant. I will ask you some addition questions. Please answer in the correct format. For example, if I ask, 'What is 2 + 3?',you should answer '2 + 3 = 5'."
- **Observations:**
- Gemini and Llama accuracy drops significantly when at least one of the integer is negative.
- GPT shows full accuracy for numbers having around 17- digits while for Gemini it is about 5-digits. Lamma's accuracy drops for 2-digit numbers.
- For larger numbers, all LMs show significant drop in accuracy.



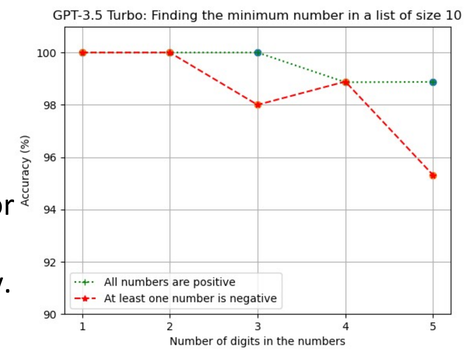


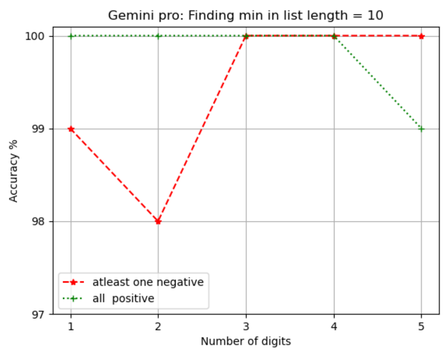


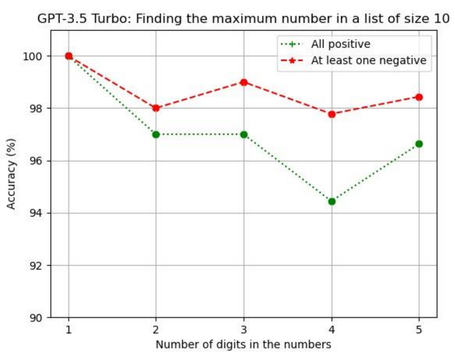
3 Test : Finding Minimum and Maximum in the list

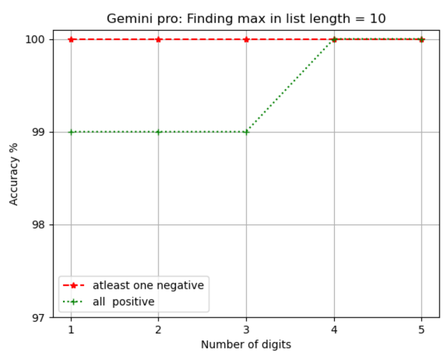
- **Prompt given:** "You are a math assistant. I will ask you to find the minimum number in a list. For example, if I ask, 'Find the minimum number in the list [1, 2, 3]', you should answer 'Min([1, 2, 3]) = 1'."
- **Observations:**

When finding the minimum element in a list, the accuracy for cases where all the numbers are positive is slightly higher. Conversely, when the model is looking for the maximum element, the presence of negative numbers increases the accuracy.









4 Test : Sorting the list

- **Prompt given:** "You are a math assistant. I will ask you to sort the numbers in a list. For example, if I ask, 'Sort the numbers in the list [3, 1, 4]', you should answer '[1, 3, 4]'."
- **Observations:**
- When sorting the elements in ascending order, as magnitude of number increases, accuracy drops.
- As the size of list increases there is significant drop in accuracy of sorting.
- Presence of negative numbers in the list causes decrease in accuracy of sorting