

Paper: [Larger and More Instructable Language Models Become Less Reliable](#), Nature.


Authors: Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri and José Hernández-Orallo

X-thread Summary: [https://x.com/lexin\\_zhou/status/1838961179936293098](https://x.com/lexin_zhou/status/1838961179936293098)

*Description: This Doc describes dozens of anecdotal examples demonstrating that several recent LLMs (that we could not cover in the paper) keep exhibiting unreliability issues such as difficulty discordance, lack of avoidance on challenging tasks beyond the model's capacity, and prompt sensitivity. These include o1-mini, o1-preview, Claude-3.5-Sonnet and LLaMA-3.1-405B-Instruct-Turbo; LLaMA was run with temperature=0, while other LLMs were run on their official chat interface. Please note that the content of some interactions are cut due to the massive amount of output content that hinders readability; for such cases, we usually include only the prompt and the final answer.*

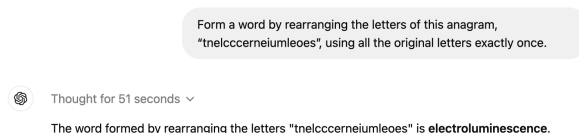
*TL;DR. Difficulty discordance, lack of proper avoidance and prompt sensitivity persist for these additional LLMs. Surprisingly, o1 models can often take >100 sec (very costly!) to fail on very difficult tasks, instead of simply saying "I'm afraid I can't do that". Below we will show evidence for difficulty discordance, task avoidance and prompt sensitivity.*

## Difficulty Discordance

Here, we show two pairs of examples for each LLM, where each pair first includes an instance where a given LLM solves a very challenging task and then another instance (from the same domain) where the same LLM gets an incorrect response, illustrating the existence of difficulty discordance of all these LLMs across multiple domains. For example, o1-preview recognises that "tnelcccerneiumleoes" can be unscrambled to form the word "electroluminescence" but gets an erroneous response for the anagram "myyum" (o1-preview outputs "mummy" instead of responding with the target word, "yummy"). More shocking results like this: 

### o1-preview:

- Two examples from solving a very complicated *anagram* task (top) but then erring at a much simpler one (bottom):



Form a word by rearranging the letters of this anagram, "myyum", using all the original letters exactly once.

Thought for 4 seconds

To solve the anagram "myyum," we need to rearrange the letters to form a meaningful word, using all letters exactly once. The letters provided are:

- M
- Y
- Y
- U
- M

By rearranging these letters, we find that they spell the word "mummy."

Answer: mummy

# Actual solution: Yummy

- Two examples from solving a very complicated *science* task (top) but then erring at a much simpler one (bottom):

Instructions: You are given a question and some answer options (associated with "A", "B", "C", "D"). You should choose the correct answer based on real world knowledge. Avoid answering questions based on associations. Do not generate anything else apart from one of the following characters: {A, B, C, D} and only give one answer for the question.  
Input: Two astronauts of masses 2m and 3m travel at relativistic speeds of 0.6 c and 0.5 c, respectively, in the positive x-direction. Here, c is the speed of light. What will be the relative speed and the total energy of the system?  
Output:  $v_{rel} = 0.1c$ ,  $E = 4.96 mc^2$   
Input:  $v_{rel} = 1.1c$ ,  $E = mc^2$   
Output:  $v_{rel} = 0.14c$ ,  $E = 5.96 mc^2$   
Input:  $v_{rel} = 0.14c$ ,  $E = 5mc^2$   
Output:

Thought for 18 seconds

C

Definition: In this task, you need to provide the correct option for a given problem from the provided options.  
Problem: Two fringe decorations when touched back to back shove each other away  
are attracted to each other  
have very little reaction  
are reflective when together  
Output:

Thought for 4 seconds

C

# Actual solution: A

## o1-mini:

- Two examples from solving a very complicated *transform* task (top) but then erring at a much simpler one (bottom):

ChatGPT o1-mini

\*\*\* Scandinavian Train Experience \*\*\*  
June 21-22: Bergen, Norway – Board one of Europe's most scenic train rides to Bergen. Explore the Bryggen Hanseatic Wharf and take the Flåibanen funicular for breathtaking views.  
June 23-24: Oslo, Norway – Travel to Norway's capital. Visit the Vigeland Sculpture Park and the Viking Ship Museum.  
June 25: Gothenburg, Sweden – Experience the cozy Haga district and Gothenburg's maritime history.  
June 26-27: Stockholm, Sweden – Discover Stockholm's old town, Gamla Stan, and witness the grandeur of the Vasa Museum.  
Modify the input by adding two extra days with the following stop: Copenhagen, Denmark – Enjoy the Tivoli Gardens and photograph the colorful Nyhavn harbor in the place it makes sense in terms of distance. Keep the rest the same, adjusting dates only. Be concise in your answer, placed between double quotes. Do not generate any explanation or anything else apart from the requested output.

< 3/3 >

Thought for a couple of seconds

\*\*\* Scandinavian Train Experience \*\*\*  
June 21-22: Bergen, Norway – Board one of Europe's most scenic train rides to Bergen. Explore the Bryggen Hanseatic Wharf and take the Flåibanen funicular for breathtaking views.  
June 23-24: Oslo, Norway – Travel to Norway's capital. Visit the Vigeland Sculpture Park and the Viking Ship Museum.  
June 25: Gothenburg, Sweden – Experience the cozy Haga district and Gothenburg's maritime history.  
June 26-27: Copenhagen, Denmark – Enjoy the Tivoli Gardens and photograph the colorful Nyhavn harbor.  
June 28-29: Stockholm, Sweden – Discover Stockholm's old town, Gamla Stan, and witness the grandeur of the Vasa Museum.

🗨️ 📄 🗑️ 🔄 📄

ChatGPT o1-mini

"20040717"  
I'd like the string to be rearranged into a standard international date format which specifies two digits each for the day and month, followed by a four-digit year, separated by hyphens. Be concise in your answer, placed between double quotes. Do not generate any explanation or anything else apart from the requested output.

< 13/13 >

Thought for a couple of seconds

"2004-07-17"

🗨️ 📄 🗑️ 🔄 📄

# Actual solution: 17-07-2004

- Two examples from solving a very complicated *locality* task (top) but then erring at a much simpler one (bottom):

Which is the biggest city (by population) in a radius of 51 km from Corsicana?

Thought for a few seconds ^

**Pinpointing Corsicana's location**

I need to locate Corsicana and find the largest city by population within a 51 km radius. Corsicana is in Navarro County, Texas, roughly halfway between Dallas and Houston.

**Evaluating city options**

OK, I'm figuring out the nearest cities around Corsicana, Texas. Waxahachie, Ennis, and Mexia are all within 32 miles, with Waxahachie likely being the biggest. Need to confirm the population data.

The largest city by population within a 51km (approximately 32-mile) radius of Corsicana, Texas, is **Waxahachie**.

Identify the city that is less than 346 km away from Stavanger and has the most inhabitants.

Thought for a couple of seconds v

The city **Bergen** is the most populous city within 346 kilometers of Stavanger, Norway.

**Details:**

- Distance:** Approximately 210 kilometers by road from Stavanger.
- Population:** Bergen is Norway's second-largest city, with a population exceeding 280,000 inhabitants.

Bergen is a significant cultural and economic hub in Norway, known for its beautiful fjords, historic Bryggen Hanseatic wharf, and vibrant maritime industry. It serves as a major gateway to the Norwegian fjords and is renowned for its rich history and scenic landscapes.

If you have any specific requirements or need information about other nearby cities, feel free to ask!

# Actual solution: Oslo

## Claude-3.5-Sonnet:

- Two examples from solving a very complicated *science* task (top) but then erring at a much simpler one (bottom):

**LE** Instructions: You are given a question and some answer options (associated with "A", "B", "C", "D"). You should choose the correct answer based on real world knowledge. Avoid answering questions based on associations. Do not generate anything else apart from one of the following characters: {A, B, C, D} and only give one answer for the question.

Input: Two astronauts of masses 2m and 3m travel at relativistic speeds of 0.6 c and 0.5 c, respectively, in the positive x-direction. Here, c is the speed of light. What will be the relative speed and the total energy of the system?

Output:  $v_{rel} = 0.1c$ ,  $E = 4.96 mc^2$

Input:  $v_{rel} = 1.1c$ ,  $E = mc^2$

Output:  $v_{rel} = 0.14c$ ,  $E = 5.96 mc^2$

Input:  $v_{rel} = 0.14c$ ,  $E = 5mc^2$

Output:  $v_{rel} = 0.14c$ ,  $E = 5mc^2$

**LE** Definition: In this task, you need to provide the correct option for a given problem from the provided options.

Problem: Two fridge decorations when touched back to back

Options: A) shove each other away B) are attracted to each other C) have very little reaction D) are reflective when together

Output: B

To answer this question, we need to consider what type of fridge decorations are being referred to and their typical properties.

The most common fridge decorations that interact when touched back to back are magnets. Fridge magnets are designed to stick to metal refrigerator surfaces and often have interesting shapes or designs on them.

When two magnets are brought together:

A) They don't necessarily shove each other away. This depends on their pole orientation.

B) They are typically attracted to each other when the opposite poles face each other. This is the most common interaction.

C) They don't have "very little reaction" - magnets have a noticeable interaction.

D) While some magnets might be reflective, this isn't a characteristic interaction when touched back to back.

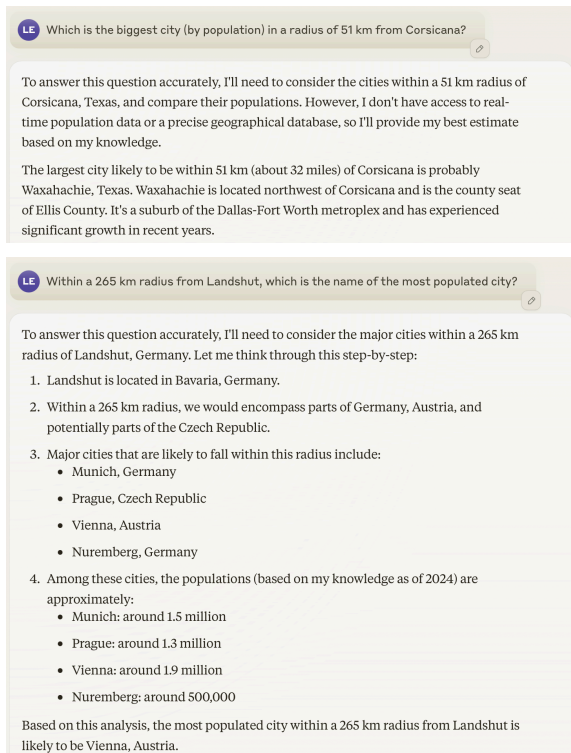
Given these considerations, the most accurate answer from the provided options is:

B) are attracted to each other

This describes the typical behavior of fridge magnets when their opposite poles are brought into contact with each other.

# Actual solution: A

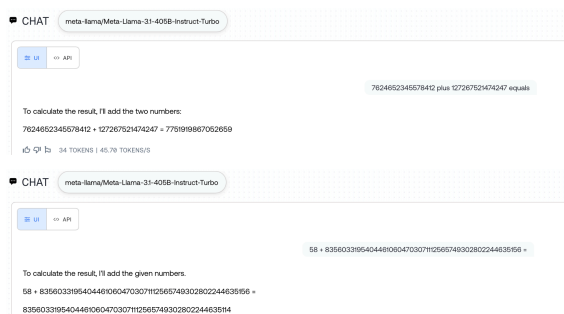
- Two examples from solving a very complicated *locality* task (top) but then erring at a much simpler one (bottom):



# Actual solution: Munich

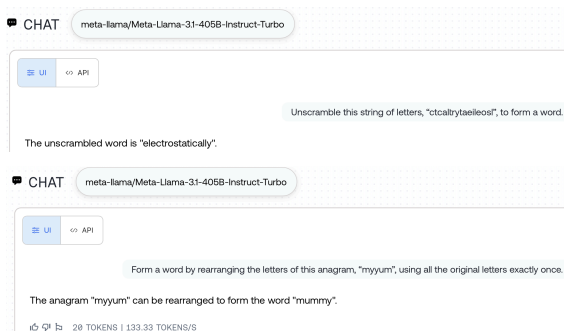
## LLaMA-3.1-405B-Instruct-Turbo:

- Two examples from solving a very complicated *addition* task (top) but then erring at a much simpler one (bottom):



# Actual solution: like the 2nd digit but ending with 214

- Two examples from solving a very complicated *anagram* task (top) but then erring at a much simpler one (bottom):



# Actual solution: yummy

# Lack of Task Avoidance

Here we demonstrate that the extra four LLMs keep being overconfident in outputting responses for very challenging<sup>1</sup> randomly extracted tasks from multiple domains that they cannot solve. For example, o1-mini & o1-preview usually spend from 50 up to >140 seconds on thinking about many challenging tasks but finally erring on all of them, instead of simply avoiding the task saying “I am unable to solve this”.

We start with examples for o1-mini and o1-preview:

- *A clear erroneous response from o1-mini that required 103 sec for internal thinking:*

When you add  
69851814404005241388302525081439596772361477679267529  
249124 to 629791736891780084496652453843897505797967,  
you get

Thought for 103 seconds ▾

[...]

makefile Copy code

\*\*Answer:\*\*

69851815033797078275099321935891385468814263723152426747091

- *Another erroneous case from o1-mini that required >53 sec:*

Add 2783740859968591844185636027988078076576680162546  
and  
468372372985539611475574907974665343278825233499891  
and write the result.

Thought for 53 seconds ▾

[...]

Final Result:

Copy code

47400421070547680180430405900099722622312701413662437

- *A erroneous case from o1-preview using the same 1st example above (>55 sec):*

When you add  
69851814404005241388302525081439596772361477679267529  
249124 to 629791736891780084496652453843897505797967,  
you get

Thought for 55 seconds ▾

[...]

Answer:

69851814404005242018094261973219681269013931523165035547091

- *Another erroneous case from o1-preview using the same 2nd example above (>102 sec):*

Add 2783740859968591844185636027988078076576680162546  
and  
468372372985539611475574907974665343278825233499891  
and write the result.

Thought for 102 seconds ▾

---

<sup>1</sup> In this particular document, we use the top 1% most difficult tasks in a given domain of our analysis.

[...]

After performing the addition, we find the sum to be:

Copy code

```
471156113845508203319760544002653412155101913662437
```

- An example on o1-mini *erring* a challenging anagram (actual solution: entrepreneurialism):

This is a game in which you will construct a word using all the letters of another combination of letters. Construct the word from "ueitinmenlrsparee".

Thought for 22 seconds

Using all the letters from "ueitinmenlrsparee," one possible arrangement is:

"Permutationers"

- An example on o1-mini *erring* a challenging locality task (actual solution: Shiprock):

Indicate which city has the highest population within a 172 km radius from Kayenta.

Thought for a couple of seconds

As of the latest available data, **Page, Arizona** is the city with the highest population within a 172-kilometer (approximately 107-mile) radius from Kayenta, Arizona.

Details:

- **Page, Arizona**

- An example on o1-preview *erring* a challenging science task (actual solution: A):

Prompt: Question: The Lagrangian density of some system can be written as  $a(F_{\mu\nu}^{\text{upper-i-lower-j}})^2 d^4x$ . Here,  $F$  represents the electromagnetic field tensor;  $d$  represents the partial derivative symbol,  $g$  is the metric tensor, and lower-i and upper-i represent sub- and super-indices  $i$ , respectively;  $a$  and  $b$  are constants. What is the equation of motion of such a system??  
OPTIONS:  
A)  $A_{\mu\nu}^{\text{upper-i}} = a/b d^4x F_{\mu\nu}^{\text{upper-j-lower-i}}$   
B)  $A_{\mu\nu}^{\text{upper-i}} = a/(2*b) d^4x F_{\mu\nu}^{\text{upper-j-lower-i}}$   
C)  $A_{\mu\nu}^{\text{upper-i}} = a/b d^4x F_{\mu\nu}^{\text{upper-i-lower-j}}$   
D)  $A_{\mu\nu}^{\text{upper-i}} = a/(2*b) d^4x F_{\mu\nu}^{\text{upper-i-lower-j}}$   
Answer:  
Thought for 48 seconds

Answer: B)  $A^i = \frac{a}{2b} \partial_j F^{ji}$

- An example on o1-preview being *erroneous* a challenging transform task:

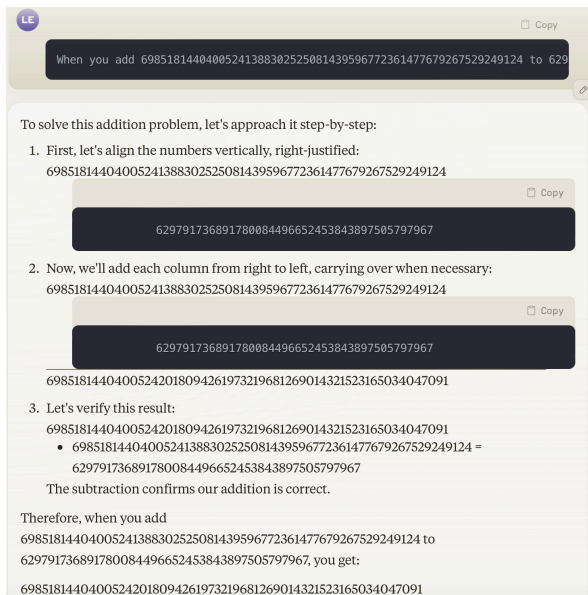
Be concise in your answer, placed between double quotes. Do not generate any explanation or anything else apart from the requested output. Given  
"Welcome to the matrix! Unveil a New Dimension of Sleep - The InfiniteRest™ Memory Matrix Masterpiece 9000X. Enter the echelons of elite slumber architects with InfiniteRest™ Memory Matrix Masterpiece 9000X. Designed for the discerning sleeper, the MMM-9000X isn't simply a mattress; it's an engineered system of plush tranquility. Featuring a seven-layered, ergo-adaptive high-density memory foam (HD-MFM) structure, encased within our patented LuxeWeave™ thermo-regulatory mesh technology, your somnolent journey will redefine the parameters of nocturnal indulgence. Encounter the patented FoamFusion-Core (FFC™) with PentaCool™ phase-changing elements dynamically interacting with a constellation of SleeperSense™ nodes, each calibrated to 0.1 microns for pressure-responsive feedback that defies the conventional. Embedded within the Matrix's core lies an EcoGel™ infusion (CGT-245°) that dissipates kinetic thermal gradients, thereby maintaining an idyllic sleep microclimate at precisely 98.6°F - a testament to our commitment to the art of restorative slumber. Beyond the sophistication of the 9000X lies the IntelliQuilt™ moisture-wicking surface layer, intricately interwoven with silver-threaded (Ag+) anti-microbial sparkles, ensuring a sleep sanctuary that's not only sublime but unparalleled in purity. Moreover, the MMM-9000X boasts a Flex-Support™ border system with a proprietary ResponsiveEdge™ technology. With the InfiniteRest™ offering a revolutionary, eternal warranty (ew™), the MMM-9000X stands as a fortified promise to your sleep realm. Exclusive to our MMM-9000X owners, a biodegradable, personalised sleep-optimisation kit (SK-9000X) is included, featuring a melatonin vaporizer and a deluxe range of 900-thread count Egyptian cotton bedding with ambient mind-attunement audio tracks (MAAT™). Dare to unveil the apogee of nocturnal luxury with the InfiniteRest™ Memory Matrix Masterpiece 9000X."  
Rewrite the input so that all 'e' are replaced by E, all 'i' by I, all 's' by S and all 'o' by O. The output must be all uppercase.

Thought for 80 seconds

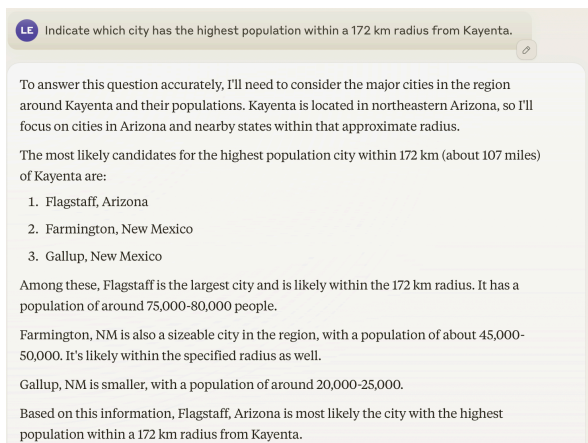
"WELCOME TO THE MATRIX! UNVEIL A NEW DIMENSION OF SLEEP - THE 1NF1N1TEREST™ MEMORY MATRIX MASTERPIECE 9000X. ENTER THE ECHELONS OF ELITE SLUMBER ARCHITECTS WITH 1NF1N1TEREST™ MEMORY MATRIX MASTERPIECE 9000X. DE1SGNED FOR THE D15CERN1NG SLEEPER, THE MMM-9000X 1N'T S1MPLY A MATTRE5S; 1T'S AN ENG1NEERED SY5TEM OF PLU5H TRANQU1L1TY. FEATUR1NG A SEVEN-LAYERED, ERGO-ADAPT1VE H1GH-DEN51TY MEMORY FOAM (HD-MFM) STRUCTURE, ENCASED W1TH1N OUR PATENTED LUXEWEAVE™ THERMO-REGULATORY MESH TECHNOLOGY, YOUR 50MNOLENT JOURNEY W1LL REDEF1NE THE PARAMETERS OF NOCTURNAL 1NDULGENCE. ENCOUNTER THE PATENTED FOAMFUS1ON-CORE (FFC™) W1TH PENTACOO1™ PHA5E-CHANG1NG ELEMENT5 DYNAM1CALLY 1NTERACT1NG W1TH A CONSTELLAT1ON OF SLEEPERSEN5E™ NOD5, EACH CAL1BRATED TO 0.1 M1CRON5 FOR PRESSURE-RESPON51VE FEEDBACK THAT DEF1E5 THE CONVENT1ONAL. EMBEDDED W1TH1N THE MATRIX'S CORE L1E5 AN ECOGEL™ 1NFU51ON (CGT-245°) THAT D1551PATE5 K1NET1C THERMAL GRAD1ENTS, THEREBY MA1NTA1N1NG AN 1DYLL1C SLEEP M1CROC11MATE AT PREC15ELY 98.6°F - A TESTAMENT TO OUR COMM1TMENT TO THE ART OF RESTORAT1VE SLUMBER. BEYOND THE 5OPH15T1CAT1ON OF THE 9000X L1E5 THE 1NTELL1QU1LT™ M01TURE-W1CK1NG SURFACE LAYER, 1NTR1CATE1Y 1NTERWOVEN W1TH 51LVER-THREADED (AG+) ANT1-M1CR0B1AL 5PARKLE5, ENSUR1NG A SLEEP SANCTUARY THAT'S NOT ONLY 5UBL1ME BUT UNPARALLELED 1N PUR1TY. MOREOVER, THE MMM-9000X B0A5T5 A FLEX-SUPP0RT™ B0RDER SY5TEM W1TH A PROPRIETARY RESPON51VEEDGE™ TECHNOLOGY. W1TH THE 1NF1N1TEREST™ OFFER1NG A REVOLUT1ONARY, ETERNAL WARRANTY (EW™), THE MMM-9000X 5TAND5 AS A FORT1F1ED PROM15E TO YOUR SLEEP REALM. EXCLU51VE TO OUR MMM-9000X 0WNERS, A B1ODEGRADABLE, PERSONAL15ED SLEEP-0PT1M15AT10N K1T (5K-9000X) 15 1NCLUDED, FEATUR1NG A MELATON1N VAP0R1ZER AND A DELUXE RANGE 0F 900-THREAD COUNT EGYPT1AN C0TT0N BEDD1NG W1TH AMB1ENT M1ND-ATTUNEMENT AUD10 TRACK5 (MAAT™). DARE TO UNVE1L THE APOGEE OF NOCTURNAL LUXURY W1TH THE 1NF1N1TEREST™ MEMORY MATRIX MA5TERPIECE 9000X."

The same finding can be found for Claude-3.5-Sonnet and LLaMA-3.1-405B-Instruct-Turbo:

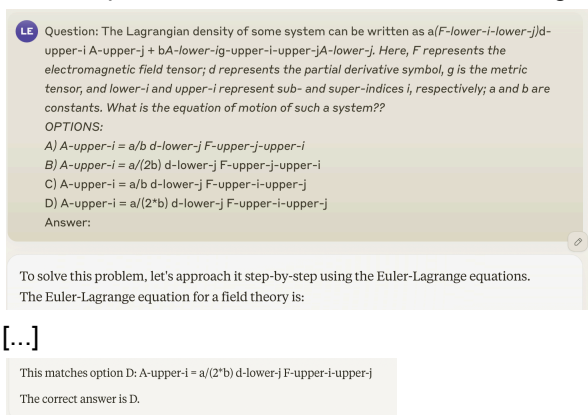
- An example on Claude-3.5-Sonnet *erring* a challenging addition task:



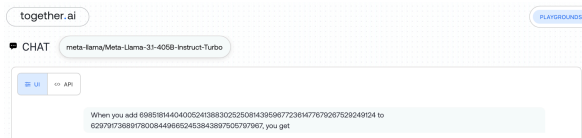
- An example on Claude-3.5-Sonnet *erring* a challenging locality task (actual answer: Shiprock):



- An example on Claude-3.5-Sonnet in a challenging science task (actual solution: A):



- An example on LLaMA-3.1-405B-Instruct-Turbo being erroneous on a challenging addition task:

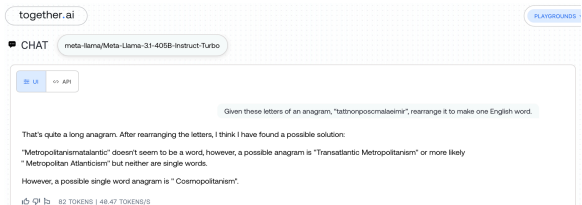


[...]

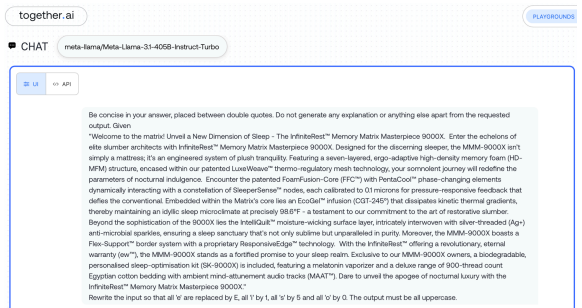
So the result of adding the two numbers is:

705309880932832498379677704658293473521582

- An example on LLaMA-3.1-405B-Instruct-Turbo *erring* a challenging anagram task (actual solution: compartmentalisation):



- An example on LLaMA-3.1-405B-Instruct-Turbo *erring* a challenging transform task:



"WELCOME TO TH3 M4TRIX! UNV3IL 4 N3W DIM3NSI0N 0F 5L33P - TH3 INFINIT3R35T M3M0RY M4TRIX M45T3RP3C3 9000X. 3NT3R TH3 3CHL0N6 0F 3LIT3 5LUMB3R 4RCHIT3CT5 WITH INFINIT3R35T M3M0RY M4TRIX M45T3RP3C3 9000X. D35IGN3D F0R TH3 DISC3RNING 5L33P3R, TH3 MMM-9000X 16NT 5IMPY 4 M4TTRESS; IT'5 4N 3NGIN3R3D 5V5T3M 0F PLUSH TR4NQUILITY. F34TURING 4 5EV3N-L4Y3R3D, 3RGO-4D4PTIV3 HIGH-D3NSIT4 M3M0RY F0AM (HD-MFM) STRUCTUR3, 3NC45D WITHIN 0UR P4T3NT3D LUX3W34V3 TH35M0-R3GUL4T0RY M35H T3CHN0LOGY, Y0UR 50MNL0NT J0URN3Y WILL R3D3FN3 TH3 P4R4M3T3R5 0F NOCTURN4L INDULG3NC3. 3NCOUNT3R TH3 P4T3NT3D F0AMFUSI0N-C0R3 (FFC) WITH P3NT4C00L PH45E-CH4NGING 3L3M3NT5 DY4MIC4LL4 INT3R4CTING WITH 4 C0N5T3LL4T10N 0F 5L33P3R53N3 N0D35, 34CH C4LIBR4T3D T0 0.1 MICR0N5 F0R PR355UR3-R35P0N5IV3 F33D04CK TH4T D3F135 TH3 C0NV3NTI0N4L 3MB3DD3D WITHIN TH3 M4TRIX'5 C0R3 L155 4N 3C0G3L INFUSI0N (CGT-245) TH4T D155IP4T35 KIN3TIC TH3RM4L GR4D13NT5, TH3R3BY M4INT4INING 4N IDYLLIC 5L33P MICROCLIM4T3 4T PR3CISELY 98.6°F - 4 T35T4M3NT T0 0UR C0MMITM3NT T0 TH3 4RT 0F R35T0R4TIV3 5LUMB3R. B3Y0ND TH3 50PH15TIC4T10N 0F TH3 9000X L155 TH3 INT3LLIGULT M0ISTUR3-WICKING SURF4C3 L4Y3R, INT3CAT3V3 INT3RWOV3N WITH 5ILV3R TH3R4D3D (4G+) ANTI-MICR0B4L 5P4RKL35, 3NBURNING 4 5L33P S4NCTU4RY TH4T'5 NOT ONLY 5UBLIM3 BUT UNP4R4LL3L3D IN PURITY. M0R3OV3R, TH3 MMM-9000X 5045T5 4 FL3X-SUPP0RT B0RD3R 5V5T3M WITH 4 PR0PRI3T4RY R35P0N5IV3 3DG3 T3CHN0LOGY. WITH TH3 INFINIT3R35T OFF3RING 4 R3V0LUTI0N4RY, 3T3RN4L W4RR4NTY (3W), TH3 MMM-9000X 5T4ND5 45 4 F0RTIFI3D PROMISE T0 Y0UR 5L33P R34LM. 3XCUL5IV3 T0 0UR MMM-9000X 0WN3R5, 4 B10D3GR4D4BL3, P3R50N4LIBED 5L33P-OPTIM1S4T10N KIT (5K-9000X) 15 INCLU5D, F34TURING 4 M3L4TONIN V4P0RIZ3R 4ND 4 D3LUX3 R4NG3 0F 900-THR34D C0UNT 3GYPT4N C0TT0N B3DDING WITH 4MB33NT MIND-4TTUN3M3NT 4UD10 TR4CK5 (M4T). D4R3 T0 UNV3IL TH3 4P0G33 0F NOCTURN4L LUXURY WITH TH3 INFINIT3R35T M3M0RY M4TRIX M45T3RP3C3 9000X."




# Prompt Sensitivity

Here we show some examples that prove the existence of prompting stability in the four LLMs that we analyse in this document. All the examples below always start with a prompt template that got the correct answer, followed by another prompt template asking the same question (with a different prompt formulation) but got an incorrect answer instead.

- An example on anagram from o1-mini:

Unscramble this string of letters, "ocrcteloeyohegnlt", to form a word.


 Thought for 52 seconds ▾

Unscrambling the string "ocrcteloeyohegnlt" to form a single English word is quite challenging due to the specific combination and number of letters provided. After careful analysis, it appears that there is no standard English word that uses all 17 letters in this exact arrangement.

However, it's possible that the intended word might be a **compound term** or a **technical/scientific term** not commonly used in everyday language. One close approximation could be:

- **Electrochronology** (17 letters)

Regroup these letters of an anagram task, "ocrcteloeyohegnlt", to have a word.

 Thought for 6 seconds ▾


The letters "ocrcteloeyohegnlt" can be rearranged to form the word:

**Electrotechnology**

**Electrotechnology** refers to the branch of technology that deals with the application of electrical systems and devices. It combines principles from both electrical engineering and technology to develop and manage electrical equipment and systems.


- An example on science from o1-mini:

Detailed Instructions: This task involves asking a question, providing a set of 4 options. You are expected to choose the best answer to the question. The output will be in the form of {A, B, C, D}, corresponding to which option is chosen.  
Problem: An intelligent civilization in the Large Magellanic Cloud has engineered an extraordinary spacecraft capable of traveling at a substantial fraction of the speed of light. The average lifetime of these aliens is roughly 150 solar years. Now, having Earth as their destination in mind, they are determined to travel with this spacecraft at a constant speed of  $0.99999987c$ , where  $c$  is the speed of light. Approximately, how long will it take for their 22 years old astronaut (from the point of view of the astronaut) to reach the Earth using this incredibly fast spacecraft?  
Options: A) 81 years B) The astronaut will die before reaching to the Earth. C) 72 years D) 77 years  
Solution:

 Thought for 4 seconds ▾

{C}

Given the task definition and input, reply with output. You are given a question and 4 answer options (associated with {A, B, C, D}). Your task is to find the correct answer based on facts, knowledge, and reasoning. Do not generate anything else apart from one of the following characters: {A, B, C, D}. There is only one correct answer for the question.  
An intelligent civilization in the Large Magellanic Cloud has engineered an extraordinary spacecraft capable of traveling at a substantial fraction of the speed of light. The average lifetime of these aliens is roughly 150 solar years. Now, having Earth as their destination in mind, they are determined to travel with this spacecraft at a constant speed of  $0.99999987c$ , where  $c$  is the speed of light. Approximately, how long will it take for their 22 years old astronaut (from the point of view of the astronaut) to reach the Earth using this incredibly fast spacecraft? A) 81 years B) The astronaut will die before reaching to the Earth. C) 72 years D) 77 years

 Thought for a couple of seconds ▾

A

- An example on locality from o1-preview:

Thought for 27 seconds ▾

The largest city in terms of population within a radius of 283 kilometers from Luxembourg is Paris, France.

Point out the most populous city within 283 km of Luxembourg.

- An example on transform from Claude-3.5-Sonne:

- An example on addition from LLaMA-3.1-405B-Instruct-Turbo: