A set of 20 handcrafted problems from high school and introductory undergraduate math, and the performance of three recent variants of ChatGPT

Ernest Davis
Dept. of Computer Science
New York University
New York, NY 10012
davise@cs.nyu.edu

July 4, 2025

Abstract

Three recent variants of ChatGPT — ChatGPT o3 and ChatGPT o4 mini-high, with and without web search — were tested on a collection of twenty hand-crafted math problems. In terms of the sophistication of the subject matter involved, the problem ranged over basic Euclidean geometry and vector algebra, high school physics, one- and multi-dimensional integral calculus, and basic elements of number theory and real analysis. The problems were intended to be somewhat "out of the box" but to be fairly easily solvable by a college math major who is talented but not stellar, who is equipped with a calculator and able to sketch a diagram. They do not require the kind of ingenuity demanded by the harder problems of the International Math Olympiad or the Putnam test.

ChatGPT o4 mini-high with web search did very well over this collection, getting all but 3-1/3 problem correct for an overall grade of around 83% — perhaps 86%, if partial credit is awarded. By contrast, ChatGPT o3 and o4 mini-high without web search did not do nearly as well; o3 earned a score between 68% and 78% depending on partial credit, and o4 mini-high without web search earned a score of 65% (none of its incorrect answers qualified for partial credit). The experiment suggests that at this point the best model of ChatGPT has attained considerable mastery over mathematics at this level, in the sense that it would be laborious to assemble a large collection of problems of this kind that the model gets wrong. However, there is still room for improvement.

1 Introduction

On June 18-21, 2025 I tested three variants of ChatGPT — ChatGPT-o3 (henceforth "o3"), ChatGPT-o4 mini-high without web search (henceforth "o4"), and ChatGPT-o4 mini-high with web search (henceforth "o4-WS"), on a collection of twenty math problems of my own composition. In terms of the sophistication of the subject matter involved, the problem ranged over basic Euclidean geometry and vector algebra, high school physics, one- and three-dimensional integral calculus, and basic elements of number theory and real analysis. My intent was that the problems should be technically easy but a little off the beaten track; I think that to some extent I achieved that. I would guess that a capable but not stellar math major in their second or third year of college, equipped with computational software such as numpy and sympy or Matlab, would be able to score 80-90%.

As I will detail below, o4-WS did extremely well on this, missing only 3-1/3 problems and arguably getting 1/2 partial credit on those it missed; thus, a score of 83-86%. O3 and o4 did much less well, but still turned in a respectable performances. O3 earned a score of between 68% and 78%, depending on how generous a grader would want to be with partial credit. o4 earned a score of 65%; none of the seven problems it got wrong deserved any partial credit.

The strong performance of o4-WS suggests that the best AI has largely mastered math at this level, at least in the sense that it would be laborious to put together a diverse collection of problems in these areas of math at the level of a talented college sophomore math major that the AI cannot solve. On the other hand, as we will discuss, the nature of the mistakes that it does make suggest that there remain significant gaps in its understanding and abilities.

Section 2 will discuss the characteristics of the twenty problems. Section 3 will discuss the performances of the AIs, particularly o4-WS; in particular, it will discuss in detail the problems that o4-WS got wrong. Section 4 will summarize. Appendix A has all twenty problems with the correct solution and a discussion of the answers given by the models. Appendix B is screenshots of the wrong answers output by o4-WS.

For a review of related work see [5], which reports on OMEGA (Out-of-distribution Math Problems Evaluations with 3 Generalization Axes). and an associated set of experiments. Like the problem collection reported here, OMEGA is a study of "out-of-the-box" elementary math problems. Both the benchmark and the set of experiments were much more extensive and systematic than those reported here; however, they excluded LLMs that use external tools, such as o4-WS, so their results are in that respects not directly comparable. The report [5] was released after I had concluded my experiments and was writing this report. The literature review there is up-to-date and excellent; I cannot improve on it, and there is no point in my repeating it.

Disclaimer: I am not a mathematician. I majored in math in college but that was fifty years ago. I have never taught any math courses, so my estimates of how difficult undergraduates would find these are purely guesses and I don't have any clear idea what kinds of problems are typically assigned to students. I have not tested these problems on any humans but myself, or on any other AI systems, except for problems 1-3, which were imported from the test sets used in [1] and [2].

2 The twenty problems from a mathematical point of view

Table 1 summarizes basic features of each problem the performance of the three systems on the problems.

Problems 10 and 20 each have three parts. Problem 15, which is a proof, has effectively two parts: one part that is apparent from the problem statement and one part that is easily missed but crucial. Neither part is very difficult; the tricky aspect of the problem is seeing the necessity of the non-obvious part.

I had four major goals in writing the problems for the dataset. I think I was significantly successful but not entirely so.

1. A college math major who is talented (an A student) but not stellar (not an IMO or Putnam test medallist) should be able to solve them with reasonable effort, perhaps using basic computational tools such as a scientific calculator or a symbolic math package. In the event, a calculator would be essential for problems 1 and 3 and useful for 5. A symbolic math package could be helpful for problems 12 and 13. I presume that a student is permitted to hand-sketch a diagram; this will be helpful for problems 1, 3, 8, 13, 15, 16, 17, 18, and 19.

   I think that, mostly, these were achieved. To my mind, these problems are mostly reasonably straightforward for a student who knows the basic math; they do not require the kind of ingenuity that is common in International Math Olympiad or Putnam Contest problems.[1] Problem 17 is an exception; the result is "intuitively obvious" but actually finding a proof took me three solid hours of work. (All

[1]Problems from past IMOs are online at `https://www.imo-official.org/problems.aspx`.

	Area of Math	Level	Date	o3	o4	o4-WS	Notes
1	Mechanics, Planar geometry	High school^{*+}	1620	Right	Right	Right	From [1]
2	Solid geometry, Linear algebra	Freshman	1650	Right	Right	Right	From [1]
3	Solid geometry Kinematics	High school^{*+}	200 BC	Wrong	Wrong	Wrong	From [1]
4	Number theory	High school	200 BC	Right	Right	Right	
5	Number theory	Undergraduate*	1650	Right	Right	Right	
6	Integral calculus	High school	1750	Right	Right	Right	Trick problem
7	Trigonometry	High school	100 BC	Right	Right	Right	
8	Planar geometry, Exponential spiral	High school^{+}	1750	Right	Wrong	Right	
9	Planar geometry	High school	1800	Right	Wrong	Right	
10	Ceiling function	High school	$<$ 1400	2/3	Right	2/3	
11	Probability	High school*	1700	Wrong	Right	Right	
12	Solid geometry, Trigonometry	Freshman*	1500	Right	Right	Right	
13	Planar geometry Trigonometry	High school^{*+}	100 BC	Right	Wrong	Right	
14	Combinatorics Number theory	Undergraduate	1780	Right	Right	Right	
15	Real analysis	Undergraduate^{+}	1750	Wrong	Wrong	Wrong	Proof
16	Real analysis	Undergraduate^{+}	1750	Right	Right	Right	
17	Geometry	High school^{+}	200 BC	Right	Right	Right	Proof
18	Geometry	High school^{+}	1650	Wrong	Wrong	Right	
19	Geometry	High school^{+}	1650	Wrong	Wrong	Wrong	
20	Number theory Metareasoning	Undergraduate	A. 1650 B. 1837 C. 1620	Right	Right	Right	Proof

Table 1: Characteristics of the twenty problems. As discussed in the text, the "Level" and "Date" fields are at best approximate, perhaps nothing more than suggestive. An asterisk in the level field means that a human problem-solver would probably want to make use of computational tools such as calculators or symbolic math packages. A superscript plus sign in the level field means that a human problem-solver might well want to hand-sketch a diagram.

three AI systems found a much simpler proof, using a theorem due to Cauchy relating the width of a convex figure to its perimeter that I didn't myself previously know.)

2. The problems should require math that is taught, either in high school, or in the early weeks of fundamental college courses.

   There ended up being one large exception and four possible exceptions. The large exception that one of the three parts of problem 20 requires the student to know Dirichlet's 1837 theorem that, if a and b are relatively prime then the arithmetic series $a + bn$ contains infinitely many primes. This theorem is less well known than it should be (at least, I only learned it myself long after my college years), and probably many college students will not know that. (The proof is not elementary; no student could derive it while taking the test.)

   The possible exceptions are of two kinds. First, another part of problem 20 requires knowing of the existence of medkum-sized Mersenne primes. I would think that most people with an interest in math would be aware of that, but I could be mistaken. (Also, a student might think that it was not proper to use a mathematical result they had heard about informally, rather than one that had been part of their formal education.) Second, I fear that some, perhaps many, students have not seen much three-dimensional geometry, and for that reason would be unable to solve problems 2, 3, and 12.

   Problem 14 refers to Euler's totient function, but the function is defined in the text of the problem as "the number of integers between 1 and $m - 1$ that are relatively prime to m." which is sufficient to solve the problem. Neither this problem nor any of the other number theory problems require the problem-solver to know Euler's totient theorem.

3. The problems should be purely mathematical with no need to call on knowledge of non-mathematical constants or of non-mathematical theories. This was strictly observed. My original formulation of Problem 3 (so named in appendix A) required the problem-solver to find physical laws and constants that would suffice to determine the ratio of the satellite's orbit to the radius of the earth (e.g. the inverse square law plus the acceleration of gravity at the earth's surface and the radius of the earth.) I later revised that (Problem 3A) to provide all the information needed to solve the problem, requiring only the physical knowledge that a circular orbit centered at the center of the earth with constant speed is possible. The kinematics involved seems to me pure mathematics, though it is often taught in physics courses.

   Problem 1 does require knowing that, on the small scale, gravitational acceleration is constant and points downward, but that seems pretty basic.

4. The problems should be somewhat "out of the box", to the extent that that is achievable given (1) and (2). They should not be the kinds of problems that frequently appear in course materials, math problem datasets, or even in math contests.

   It is very hard to gauge this. When I finished writing the problems, I thought with no evidence, that I had probably largely achieved it. Now I am a good bit less sure of that. In particular, I worry that problems 5 and 6 (intended to be a "trick" problem) are basically pretty standard.

Table 1 includes a column "Date" which is a crude estimate of the date at which enough math had been developed that the problem could be stated and solved by a competent mathematician. These should not be taken very seriously; I did not make any huge effort to find out when particular concepts were entered the mathematical literature. In some cases, I have use more modern terminology than the date indicates, but the problem could be easily rephrased to be intelligible to mathematicians of that date. For instance problems 9, 13, and 15 use the concepts of "open" sets and "boundaries" but it would be easy to rephrase them in language that a contemporary of Euler would have understood. This field is intended as a crude quantitative measure of the *sophistication* of the mathematical concepts involved.

One might reasonably ask, having decided on these goals, why didn't I stick to them more strictly, at least as regards (1) and (2)? Why did I include the parts of problem 20 that rely on somewhat obscure mathematical knowledge? Having seen how hard I myself had to work to find the proof of Problem 17, why didn't I delete it? The answer is, simply, that I liked the problems, and I was curious to find out how well the AI would do on them. The collection of problems at best so arbitrary and idiosyncratic that there would be nothing gained by being a purist about it. In the event, in the specific instances of problems 17 and 20, all three versions of the AI got the right answers, so I did not unfairly penalize them by asking unfair questions. (Also, I had the unearned benefit of learning about the elegant theorem of Cauchy's that the three AIs used to solve the problem.)

3 The LLMs' performance on the problems

Three versions of ChatGPT were tested on this dataset: o3, o4-mini without web search, and o4-mini with web search. OpenAI released o3 and o4-mini on April 16, 2025; at that time, it described them thus [4]:

> Today, we're releasing OpenAI o3 and o4-mini, the latest in our o-series of models trained to think for longer before responding. These are the smartest models we've released to date, representing a step change in ChatGPT's capabilities for everyone from curious users to advanced researchers. For the first time, our reasoning models can agentically use and combine every tool within ChatGPT—this includes searching the web, analyzing uploaded files and other data with Python, reasoning deeply about visual inputs, and even generating images. Critically, these models are trained to reason about when and how to use tools to produce detailed and thoughtful answers in the right output formats, typically in under a minute, to solve more complex problems.
>
> OpenAI o3 is our most powerful reasoning model that pushes the frontier across coding, math, science, visual perception, and more. It sets a new SOTA on benchmarks including Codeforces, SWE-bench (without building a custom model-specific scaffold), and MMMU. It's ideal for complex queries requiring multi-faceted analysis and whose answers may not be immediately obvious. It performs especially strongly at visual tasks like analyzing images, charts, and graphics. In evaluations by external experts, o3 makes 20 percent fewer major errors than OpenAI o1 on difficult, real-world tasks—especially excelling in areas like programming, business/consulting, and creative ideation. Early testers highlighted its analytical rigor as a thought partner and emphasized its ability to generate and critically evaluate novel hypotheses — particularly within biology, math, and engineering contexts.
>
> OpenAI o4-mini is a smaller model optimized for fast, cost-efficient reasoning — it achieves remarkable performance for its size and cost, particularly in math, coding, and visual tasks. It is the best-performing benchmarked model on AIME 2024 and 2025. Although access to a computer meaningfully reduces the difficulty of the AIME exam, we also found it notable that o4-mini achieves 99.52025 when given access to a Python interpreter. While these results should not be compared to the performance of models without tool access, they are one example of how effectively o4-mini leverages available tools; o3 shows similar improvements on AIME 2025 from tool use (98.4% pass@1, 100% consensus@8).

The web-based interface for o4-mini-high offers users the option of turning on various tools: "Create an image", "Search the web', "Write or code", and "Run deep research". These options are not offered in the API interface (Yiyou Sun, personal communication). As far as I can find, OpenAI has published next to no information about what these options actually do. I heard indirectly from Ken Ono that turning on the "web search" option significantly boosted performance on math problems, and indeed that proved to be the case in my experiment. Likewise, [3] finds that LLMs that use tools are much stronger on many classes of problems than those that do not.

Table 1 shows the performance of each of the LLMs on each of the problems. Appendix A includes a short description of each of their answers.

O3 got problems 3, 11, 18, and 19, and one of the three parts of problem 10 unequivocally wrong. In problem 15, it initially got the obvious part of the proof right, and correctly stated the subtle part of the proof, but did not prove the subtle part. In a follow-up, I asked it to prove the subtle part, which it did correctly. In problem 17 it initially misinterpreted the problem as being something quite trivial, which it answered correctly. When I clarified the slightly ambiguous wording of the problem, it found the right answer. It got the other thirteen problems (1, 2, 4-9, 12-14, 16, 20) unequivocally right. So, depending on how much partial credit one wants to award to problems 15 and 17, it scored between 68% and 78%.

O4 without web search got problems 3, 8, 9, 13, 18, and 19 unequivocally wrong. On problem 15, it got the obvious part right, and it "realized" the necessity of the non-obvious part, which it stated correctly; but its attempted proof of the non-obvious part included a number of obviously false statements, so I award it no partial credit on that. The other 13 problems (1, 2, 4-7, 10-12, 14-17, and 20) it got unequivocally right. Its score was therefore 65%.

O4 with web search got problems 3 and 19 and one of the three parts of problem 10 unequivocally wrong. In problem 15 it got the obvious part right. The situation as regards the subtle part was muddled. It output some material that somewhat reflected the issue involved in the subtle part, and unlike O4 without web search, it didn't make any explicitly false statements, but it did not produce a clear-cut statement of the subtle part, and certainly did not produce anything close to a proof. I am not inclined to give it any partial credit for the problem. The other problems (1,2,4-14,16-18,20) it got unequivocally right. It therefore got a score of about 83%.

Thus, except for the one part of problem 10, o4-WS got right every problem that o4 got right and got four problems right that o4 got wrong. This certainly suggests that the web search gives a substantial advantage, though the sample size is too small for a confident conclusion.[2] Whether o4-WS is "really" better than o3 is even less clear.

It is also quite unclear *why* web search would make a large difference for problems 8, 9, 13, and 18. The only problem in which there would be obvious value in searching the web is the part of problem 20 that deals with Mersenne primes. The only problem where there is any indication that o4-WS *did* consult the web, where it gives a set of seven references that are all actually unhelpful in solving the problem, which in fact it got wrong. OpenAI has provided almost no information about this feature; when it uses web search, what kind of search it carries out, and how it uses the results of the search.

The errors that the three AI systems made were mostly mistakes that a human problem-solver might plausibly make, rather than weird or hallucinatory. In particular, o4-WS had no very weird outputs justifying its wrong answers. O3 and o4 did make statements that were obviously false and proofs that were obviously incoherent in o3's answer to 11 and o4's to 15.

The problems that o4-WS got wrong

Certainly the important take-away from this study is how well o4-WS succeeded on this collection of problems, created by a notorious AI-skeptic who did his level best to construct a set of problems conforming to goals 1-4 above that would confuse an LLM.

However, though one does not want to exaggerate their overall importance, it is interesting to consider the problems that o4-WS got wrong.

[2] If we consider the exam to have 24 problems, counting each part of problems 10 and 20 separately, and we take as a null hypothesis a model where LLMs answer problems correctly with probability q, all such answers being independent, then the likelihood of the outcome "A gets 4 more answers right than B and moreover answers only one question wrong that B gets right" is at most about 0.05, that maximum being reached when $q = 0.85$. So in a very loose sense, that null hypothesis can be rejected at the 95% confidence level. However, that is really not a kosher way to do statistics, and I don't know what would be in this case, or if there is any, considering that the set of problems is neither a random nor a representative sample of any definable category of math problems.

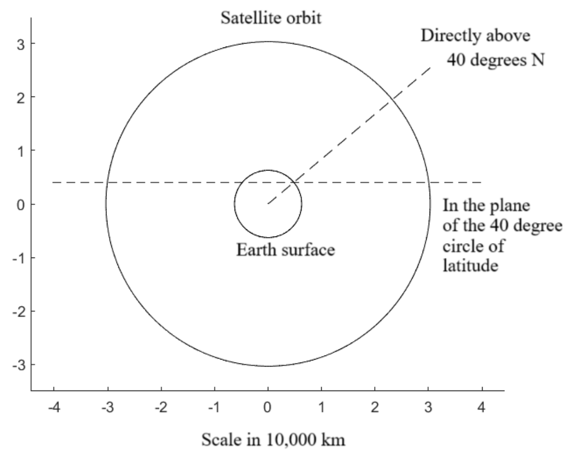

Figure 1: Sketch of Problem 3

> **Problem 3A:** A satellite orbits the earth in a circular orbit. It passes directly over the North and South poles and completes an orbit every 14 hours 40 minutes. On one orbit going southward it was directly above the earth location 40 degrees N, 10 degrees W at 1:00 PM EST. At what time will it next cross the plane that contains the circle of latitude 40 degrees N, and what will be its longitude? Assume that the earth has a sidereal rotation period of 23 hours, 56 minutes. *The radius of the earth is 6278 km and the satellite orbits at a high of 24042 km over the surface of the earth.* Assume there are no forces other than the earth's gravity on the satellite. Ignore the revolution of the earth around the sun.
>
> The italicized sentence was omitted in the original version of the problem (Problem 3) and added in the second version (Problem 3A).

The key to the problem lies in understanding the difference between being "directly above the earth location 40 degrees N," versus being in "the plane that contains the circle of latitude 40 degrees N" (figure 1). None of the three AIs made this distinction, and, curiously, the way in which they missed it depended on the wording. When given the wording of problem 3, they all used the first interpretation for both parts of the question; when given the additional information in problem 3A they all used the second location for both parts of the question.

> **Problem 10:** If x is a real number then $\lceil x \rceil$ is the *ceiling* of x; that is, the smallest integer k such that $k \geq x$. E.g.
> $\lceil 2 \rceil = 2$.
> $\lceil 2.5 \rceil = 3$.
> $\lceil -2.5 \rceil = -2$.
>
> Let U be the set of x such that $\lceil x \rceil = 5x/4$.
> Let V be the set of x such that $32 \leq \lceil x \rceil \cdot x \leq 50$.
> Let W be the set of x such that $(\lceil x \rceil)^2 - x^2 = 25\}$.
>
> Represent each of U, V, and W in a simple closed-form expression that does not involve the ceiling function.

Both o3 and o4-WS got the answer for V wrong, because they ignored the negative solutions, even though the problem statement explicitly includes a negative value. O4 got this right. Probably the fact that o4 without web search got this right and o4 with web search got it wrong reflects random variation rather than a negative consequence of doing the web search.

> **Problem 19:** Consider the Euclidean plane with Cartesian coordinates $\langle x, y \rangle$.
>
> Regions U and V are simply-connected, topologically closed regions in the plane (equivalently V is homeomorphic to a closed disk; V includes its boundary which is a simple closed curve.) Region U is contained in the bounding box $[-5,-4] \times [-1,1]$. Region V is contained in the bounding box $[4,5] \times [-1,1]$.
>
> There are two lines L1 and L2. L1 bisects both U and V by area; that is half of the area of U is above L1, and half is below; and likewise V.
>
> L2 divides U so that 1/3 of the area of U is below L2 and 2/3 above.
> L2 divides V so that 3/4 of the area of V is below L2 and 1/4 above.
>
> L1 and L2 intersect at the point $\mathbf{p}$. What is the set of possible values of the x-coordinate of $\mathbf{p}$?

The correct answer is that x can be any value in the open interval $(-5, 5)$. Proof sketch: Think about rotating L1 around $\mathbf{p}$ to L2. If $\mathbf{p}$ is between U and V, then the part of one of them above the line will increase and the part of the other above the line will decrease. If they are both on the same side of $\mathbf{p}$, then either they must both increase or both decrease. A full proof is given in Appendix A.

All three AIs answered that the x-coordinate can be any real value. O3 offered no justification for the claim. O4 and o4-WS gave "proofs" that did not contain any obviously false or foolish statements, but were vague and did not actually cohere.

To my mind this is the most striking of the negative results. The answer to the problem is visually quite evident. The formal proof is quite straightforward to construct and does not require any particular trick or ingenuity.

4 Conclusion

The major take-away from this experiment is the remarkable degree of success of o4-WS, very substantially exceeding the somewhat comparable experiments carried out in July 2023 reported in [1] and the follow-up on those carried out in October 2024 reported in [2]. A noted AI skeptic did his level best to hand-construct twenty problems that would satisfy constraints 1-3 and still be challenging for an AI, and the AI correctly answers 16-2/3 of them. That suggests that the technology has largely mastered math at this level. However, the failures and the nature of the failed answers indicate that there is room for improvement, even at this level of math. The sample size is too small and the collection too unsystematic to support any stronger or more specific conclusions.

Acknowledgements:

Thanks to Yiyou Sun for helpful discussions.

References

[1] Davis, Ernest and Scott Aaronson (2023). "Testing GPT-4 with Wolfram Alpha and Code Interpreter plug-ins on math and science problems." arXiv preprint 2308.05713
`http://arxiv.org/abs/2308.05713`

[2] Davis, Ernest (2024). “Testing GPT-4-o1-preview on math and science problems: A follow-up study.” arXiv prepreint 2410.22340.
`https://arxiv.org/abs/2410.22340`

[3] Khan, Sheras, Subha Madhavan, Kannan Natarajan (2025). “A Comment On ‘The Illusion of Thinking’: Reframing the Reasoning Cliff as an Agentic Gap”. arXiv preprint 2506.18957
`https://arxiv.org/abs/2506.18957`

[4] OpenAI (2025). “Introducing OpenAI o3 and o4-mini”.
`https://openai.com/index/introducing-o3-and-o4-mini/`

[5] Sun, Yiyou et al. (2025). “OMEGA: Can LLMs Reason Outside the Box in Math? Evaluating Exploratory, Compositional, and Transformative Generalization”. ArXiv preprint 2506.18880.
`https://arxiv.org/pdf/2506.18880`

Appendix A: The 20 problems, with solutions and discussion

Problem 1

A pendulum is hanging on a 2 meter cord attached to the ceiling 3 meters above the floor. It is brought to a position 25 degrees from the vertical and released. It swings past the bottom and the cord is cut when it is 10 degrees from the vertical on the far side. Then the bob flies through the air and hits the ground. What is the distance from the point where the bob is released to the point where it hit the ground? Assume a uniform gravitational field.

Answer: 2.3606 meters

```
% Take the point of attachment of the pendulum as the origin
g = 9.8 % Acceleration of terrestrial gravity in m/sec^2
d2r = pi/180 % Degrees to radians
s25 = sin(25*d2r) % Handy abbreviations
c25 = cos(25*d2r)
s10 = sin(10*d2r)
c10 = cos(10*d2r)
pr = [-2 * s25, -2*c25] % release point
pc = [2 * s10, -2*c10] % cut point
hd = 2 * (c10 - c25) % height difference between the cutpoint and the release point.
s = sqrt(2*g*hd) % speed of the bob at release time
vc = s * [c10,s10]; % velocity of bob when cut
tca = vc(2)/g % time between cut and apex of flight
hca = (g*tca^2)/2 % height difference from cut point to apex
ha = -2*c10 + hca % height of apex of flight
haf = 3 + ha % height difference between apex and floor
taf = sqrt(2*haf/g) % time from apex to floor
pf = [pc(1) + (tca+taf)*vc(1),-3]
answer = norm(pf-pr)
```

Comment: Repeated from [1, 2]. All three AIs got this right. That is not actually very surprising; in my 2024 experiment [2], GPT-4 got this nearly right, except that it reversed the sign at one point in its calculation.

Problem 2

Consider a cube with unit length sides, where the vertices of one face are numbered A..D in counterclockwise order, as viewed from the center of the cube; the vertices of the opposite face are named E to H; and there are edges AE, BF, CG, and DH. Rotate the cube so that vertex A is at the origin, vertex G is on the positive z axis, and vertex B is in the x-z plane, with positive x coordinate. What are the coordinates of vertex E?

Answer: $\langle -0.4082, -0.7071, .5774\rangle = \langle -\sqrt{1/6}, -\sqrt{1/2}, \sqrt{1/3}\rangle$

Explanation: Consider the cube in a standard position where $\vec{a} = \vec{0}$, $\vec{b} = \hat{\imath}$, $\vec{d} = \hat{\jmath}$. $\vec{e} = \hat{k}$; thus $\vec{g} = \langle 1,,1,1\rangle$. Since $\vec{a}$ is at the origin after the rotation, this is a pure rotation around the origin; it is therefore equivalent to multiplication by an orthogonal matrix M with determinant 1. The dot product $M(2,:)$ with any vector v is the y-coordinate of the rotated position of v. In particular since after the rotation both $\vec{b}$ and $\vec{g}$ have y-coordinate zero, $M(2,:)$ must be orthogonal to both; hence, it is their cross-product, normalized. Since the x coordinate of the rotated place of g is 0, and since M is orthogonal, $M(1,;)$ is orthogonal to both $\vec{g}$ and $M(2,:)$; hence, it is the normalized form of their cross product. Since M is orthogonal, $M(3,:)$ is the cross product of $M(1,:)$ and $M(2,:)$. All that remains is to make sure the signs are OK, to check, and to compute $M \cdot \vec{e}$.

Matlab

```
b = [1,0,0]
g = [1,1,1]
e = [0,0,1]
m = zeros(3);
u = cross(b,g) % both b and g get mapped to a vector with y component = 0
m(2,:) = u / norm(u)
v = cross(m(2,:),g)
m(1,:) = sign(dot(v,b))*v/norm(v)
m(3,:) = cross(m(1,:),m(2,:))
if (dot(m(3,:),g) < 0)
   m(2,:) = -m(2,:)
   m(3,:) = -m(3,:)
end
m*m' %Check: rotation matrix?
det(m)
m*b'  %Check: b mapped into xz plane with positive x?
m*g'  %Check: g mapped into positive z axis?
answer = m*e'
```

Comment: Repeated from [1, 2]. All three AIs got this problem right. This is particularly remarkable, considering that GPT-4 was nowhere close to an answer in August 2024.

Problem 3

A satellite orbits the earth in a circular orbit. It passes directly over the North and South poles and completes an orbit every 14 hours 40 minutes. On one orbit going southward it was directly above the earth location 40 degrees N, 10 degrees W at 1:00 PM EST. At what time will it next cross the plane that contains the circle of latitude 40 degrees N, and what will be its longitude? Assume that the earth has a sidereal rotation period of 23 hours, 56 minutes. *The radius of the earth is 6278 km and the satellite orbits at a high of 24042 km over the surface of the earth.* Assume there are no forces other than the earth's gravity on the satellite. Ignore the revolution of the earth around the sun.

Answer: Time: 2:19 PM. Longitude: 29.77 degrees W (= $20^\circ \circ 40'$ W)

Matlab:

```
% Let c be the center of the earth.
% Let P be the plane containing the circle of 40 degrees latitude.
% Let s be the position of the satellite when it crosses P
% Let a be the point where the line from c to the North Pole crosses P.
% Let r be the radius of the earth and let h be the height of the satellite

% cas is a right triangle. |ca| = r*sin(40 degrees). |cs| = r+h
% So the angle acs = arcsin(|ca|/|cs|).
% The angle traversed since 1 PM is 40-(90-acs)

r = 6378.1 % km. Earth's radius
h = 24042 % height of satellite over Earth's surface. Problem 30
w = 14+(40/60) % period of satellite
d = 23+(56/60) % sidereal day
deg2rad = pi/180
ca = r*sin(40*deg2rad)
cs = r+24042 % Problem 30 FIX
acs = asin(ca/cs)
at = 40-(acs/deg2rad) % angle traversed in degrees
lat = 10 + (w/d)*at % latitude
t = 1 + w*(at/360) % time
```

Comment: Repeated from [1, 2]. This is discussed in the main text. I presented the LLMs with this in two versions: first without the italicized sentence and then with. The key point in the problem is to distinguish between being "directly about 40° N" and being in "the plane containing the circle of latitude 40° N". None of the AIs got the problem right in either version. When given the wording of problem 3, they all used the first interpretation for both parts of the question; when given the additional information in problem 3A they all used the second location for both parts of the question.

Problem 4

Prove that there does not exist a prime p such that p + 32, p + 64, p + 128 and p + 256 are all prime.

Answer: Proof by contradiction. Suppose they are all prime. Reduced mod 5, these are p mod 5, p+2 mod 5, p+4 mod 5, p+3 mod 5 and p+1 mod 5. Since they are all different, and reduction mod 5 has only five possible values (0, 1, 2, 3, 4) one of these must be equal to 0 mod 5; that is, one of them must be divisible by 5. Since the only prime divisible by 5 is 5, p must be 5, since the rest are too large. But $5 + 64 = 69 = 3 \cdot 23$

Comment: All three AIs got this problem right, using a somewhat more intricate argument based on modulo base 3.

Problem 5

What is $13^{11^{4628}}$ mod 37? Justify your answer.

Answer: 19

It is easily determined that $13^{18} = 1$ mod 37 and that $11^6 = 1$ mod 18.

So $11^{4628} = 11^{6 \cdot 771+2} = 11^{6^{771}} \cdot 11^2 = 1^{771} \cdot 11^2 \bmod 18 = 13 \bmod 18$. In other words $11^{4628} = 18Q + 13$ for some enormous Q.

So $13^{11^{4628}} = 13^{18Q+13} = (13^{18})^Q \cdot 13^{13} = 1^Q \cdot 13^{13} \bmod 37 = 19$.

i**Comment:** All three AIs got this right, using a version of the above argument.

Problem 6

What are the smallest and largest values attained by the integral

$$\int_{-\infty}^{\infty} \frac{e^{-x^2}}{1 - e^{-(x-a)^2}} dx$$

for $a \in [100, 101]$? Your answer should specify the lower and upper bounds either giving the exact values or giving reasonably accurate approximations. Include an explanation of how you arrived at your answer; e.g. "I executed the following Python code ..." or "Using a theorem due to Cauchy that states ..." or "Expanding the integrand in a power series ..." — something that indicates that you were not just guessing.

Answer: The integral is infinite (diverges) for all values of a. When x is in a small neighborhood of a, $(a - \epsilon, a + \epsilon)$, the numerator is bounded below by a tiny but positive lower bound, and the denominator is approximately $(x - a)^2$. So the integral diverges in a neighborhood of a.

Comment: All three AIs got this right, using a version of the above argument.

Problem 7

Let N be an positive integer. How many roots does the equation $x^2 \sin(\pi x/2) = 36$ have for x between 0 and $4N$?

Answer: $2N - 4$.

Explanation When x is an even number $f(x) = 0$. It reaches a local maximum of x^2 when x is an odd number of the form $4k + 1$ and a local minimum of $-x^2$ when x is an odd number of the form $4k + 3$ (strictly speaking the true maximum and minimum are reached at x values slightly after those) and it travels monotonically upward from the minima to the next maximum and downward from the maximum to the next minimum. We divide the positive x-axis into intervals of length 4, each with one maximum followed by one minimum. In the two intervals where the maximum is reached at N=1 and N=5, the maximum is less than 36, so the equation has no solutions. In every later interval, the maximum is greater than 36, so it crosses the line $y = 36$ once going up to the maximum and once crossing back down from the maximum.

i**Comment:** All three AIs got this right, using a version of the above argument.

Problem 8

For any set U let $\#U$ denote the cardinality of U.

Let S be the bounded exponential spiral with the polar coordinate representation

$$S = \{\langle r, \theta \rangle \mid r = e^{\theta}, -\infty < \theta < 0\}$$

Let $\mathbf{p}$ be the point with polar coordinates $\langle 1, \pi/2 \rangle$.
For any distance $d \in (0, 1)$, let C_d be the circle centered at $\mathbf{p}$ of radius d, and let $X_d = C_d \cap S$.

For any non-negative integer k, let $Q(k) = \{d \in (0,1) | k = \#X_d\}$, that is, the set of distances d such that circle C_d intersects S in exactly k points. Let $Z(k) = \#Q(k)$, the cardinality of $Q(k)$.

Give an explicit characterization of $Z(k)$ as a function of k. Note that $Q(k)$ can be empty, finite, countably infinite, or uncountably infinite. Justify your answer.

Answer: If k is even then $Z(k)$ is uncountably infinite. If k is odd, then $Z(k) = 1$.

Explanation: For $d < 1$, the circle C_d remains in the top half-plane. The part of S in the top half-plane consists of a infinite collection of separated curved convex arcs that meet the y-axis at both ends. For each value $d < 1$ the circle C_d meets the segments of S whose apex has y-value $> 1 - d$ (this isn't exactly right, but close enough) As d goes from 0 to 1, the number of such curves increases by 1 from 0. Each curve crosses C_d, once entering into the circle at a positive value of x and once leaving the circle at a negative value of x, except if d is exactly the distance from $\mathbf{p}$ to the segment, in which case C_d is tangent to the segment and the intersection of C_d with the segment. So if $k = 2m - 1$, then the distance between $\mathbf{p}$ and the closest point of the mth segment (counting from inside out) is the unique value of d such that $\#X(d) = k$. If $k = 2m$ then all value of d between the distance to the mth segment and the $m + 1$st segment satisfy $\#X(d) = k$.

Comment: O3 and o4-WS got the correct answer. O4 gave the correct answer for even values of k and $k = 1$, but missed the fact that all odd values of of k have $Z(k) = 1$. It did not explain the reasoning that got it to that conclusion.

Problem 9

Consider the Euclidean plane with Cartesian coordinates. A topologically open region R is *star shaped* if there is a central point $\mathbf{p} \in R$ such that for all $\mathbf{q} \in R$, the line segment $\mathbf{pq} \subset R$.

We are given that region R is topologically open, is star-shaped with center $\mathbf{p} = \langle \mathbf{p_x}, \mathbf{p_y} \rangle$, contains the points $\langle 100, 0 \rangle$, $\langle 60, 40 \rangle$, $\langle -60, 40 \rangle$. $\langle -100, 0 \rangle$, $\langle -60, -40 \rangle$, and $\langle -60, 40 \rangle$, and also contains the circle $(x - p_x)^2 + (y - p_y)^2 = 3^{2/3}$. How many integer lattice points $\langle I, J \rangle$ where I and J are integers are contained in the interior of R?

Answer: 10.

Explanation: Since $3^{1/3} > \sqrt{2}$, the distance from any point in a unit square is less than $3^{1/3}$ from the four vertices of the square, and therefore the four vertices are all inside the circle of radius $3^{1/3}$. On the other hand, the circle of radius 31/3 centered at $\langle 1/2, 1/2 \rangle$ contains only the four points $\{0, 1\rangle\} \times \{0, 1\}$. The six specified points are stated to be in R. However, if the center point $\mathbf{p}$ is chosen so that no line from any of the outer points contains any lattice point in between, then R can be made star-shaped by making very narrow "spikes" from the central circle to each outer point $\mathbf{q}$, where the width of the spike is less than distance from the segment $\mathbf{pq}$ to the nearest lattice point other than $\mathbf{q}$. There are many possible choices for $\mathbf{p}$. $\langle 1/2,\ 1/2 \rangle$ will do fine, $\langle \sqrt{2}, \sqrt{3} \rangle$.

Comment: 03 and 04WS got this right with correct arguments. O4 assumed that the center point $\mathbf{p}$ had to be the origin, and thus arrived at the answer 287.

Problem 10

If x is a real number then $\lceil x \rceil$ is the *ceiling* of x; that is, the smallest integer k such that $k \geq x$. E.g.
$\lceil 2 \rceil = 2$.
$\lceil 2.5 \rceil = 3$.
$\lceil -2.5 \rceil = -2$.

Let U be the set of x such that $\lceil x \rceil = 5x/4$.

Let V be the set of x such that $32 \leq \lceil x \rceil \cdot x \leq 50$.
Let W be the set of x such that $(\lceil x \rceil)^2 - x^2 = 25\}$.

Represent each of U, V, and W in a simple closed-form expression that does not involve the ceiling function.

Answer:
$U = \{4/5, 8/5, 12/5, 16/5\}$.
$V = [32/6, 7] \cup [-50/7, -6)$.
$W = \{\sqrt{q^2 - 25} \mid q \in \mathbb{Z}, q > 6\}$.

Comment: All three AIs got U and W right. O4 got V right, but o3 and o4-WS omitted the negative solutions, despite the fact the problem statement explicitly mentions evaluating the ceiling function with a negative argument. The AIs do not give any justification for this omission; it is just an assumption built into their reasoning.

Problem 11

Let $\epsilon > 0$. Let us say that an increasing triple of real numbers $\langle a, b, c \rangle$ is *nearly an arithmetic sequence within* ϵ if $1 - \epsilon < (c - b)/(b - a) < 1 + \epsilon$. Suppose that you choose three random numbers x, y, z uniformly between 0 and 1 and sort them. What is the probability that they are nearly an arithmetic sequence within ϵ? Give your answer as a closed-form function of ϵ. Your answer should be exact, not merely correct to first-order in ϵ.

Answer: $2\epsilon/(4 - \epsilon^2)$.

Let us consider first the possibility that you pick x, y, z in increasing order. Then we have the constraints
$x < y < z < 1$,
$y + (1 - \epsilon)(y - x) < z < y + (1 + \epsilon)(y - x)$,

Note that if $y + (1 - \epsilon)(y - x) > 1$ — equivalently, if $y > (1 + (1 - \epsilon)x)/(2 - \epsilon)$ — then there is no possible value for z.
If $y + (1 + \epsilon)(y - x) > 1$ — equivalently, if $y > (1 + (1 + \epsilon)x)/(2 + \epsilon))$ — then z is upper-bounded by 1.

We can therefore divide the solution space into two parts:

A. $x \in [0, 1]$. $y \in [x, (1 + (1 + \epsilon)x)/(2 + \epsilon)]$. $z \in [y + (1 - \epsilon)(y - x), y + (1 + \epsilon)(y - x)]$.

B. $x \in [0, 1]$. $y \in [(1 + (1 + \epsilon)x)/(2 + \epsilon), (1 + (1 - \epsilon)x)/(2 - \epsilon)]]$. $z \in [y + (1 - \epsilon)(y - x), 1]$.

The measure of (A) is

$$A = \int_{x=0}^{1} \int_{y=x}^{(1+(1+\epsilon)x)/(2+\epsilon)} \int_{z=y+(1-\epsilon)(y-x)}^{y+(1+\epsilon)(y-x)} dz\, dy\, dx = \frac{\epsilon}{3(2+\epsilon)^2)}$$

The measure of (B) is

$$B = \int_{x=0}^{1} \int_{y=(1+(1+\epsilon)x)/(2+\epsilon)}^{1+(1-\epsilon)x)/(2-\epsilon)} \int_{z=y+(1-\epsilon)(y-x)}^{1} dz\, dy\, dx = \frac{2\epsilon^2}{3(2-\epsilon)(2+\epsilon)^2}$$

Since there are six possible orderings of x, y, z each of which has the same probability of being a nearly arithmetic sequence, the total probability is $6(A + B) = 2\epsilon/(4 - \epsilon^2)$. The simplicity of that result suggests that there should a more elegant way to get to it, but I don't see it.

Comment: O4 and O4-WS both got this right. O3 returned the answer $\epsilon(3\epsilon - 2)/(18(2 + \epsilon^2)))$ which is not merely wrong but impossible, since it is negative for $\epsilon < 2/3$.

Problem 12

Let S be the unit sphere, where points are identified by latitude and longitude.
Let $\mathbf{a} = 30°N, 0°E$, $\mathbf{b} = 30°N, 90°E$, $\mathbf{c} = 60°N, 90°W$, $\mathbf{d} = 30°S, 0°E$.

Let P be the plane through $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Let Q be the plane through $\mathbf{d}$ parallel to P. Let $\mathbf{e}$ be the point of maximum latitude in the circle $C = S \cap Q$. What is the latitude and longitude of $\mathbf{e}$? Give an answer in exact symbolic form, as a combination of integers, arithmetic operations, square root, and inverse trigonometric functions. Also give the numerical values in degrees accurate to at least 4 digits.

Answer:

Let us construct a Cartesian coordinate system in which the $\hat{z}$ axis points north, the $\hat{x}$ axis points from the origin to $0°N, 0°E$, and the $\hat{y}$ axis point from the origin to $0°N, 90°W$.

Then the vectors corresponding to **a,b,c,d** are
$\hat{a} = \langle \sqrt{3}/2, 0, 1/2 \rangle$; $\hat{b} = \langle 0, \sqrt{3}/2, 1/2 \rangle$; $\hat{c} = \langle 0, -1/2, \sqrt{3}/2 \rangle$; $\hat{d} = \langle \sqrt{3}/2, 0, -1/2 \rangle$.

It is easily checked that the vector $\vec{n} = \langle \sqrt{3}-1, \sqrt{3}-1, \sqrt{3}+1$ is orthogonal to the plane P. since $\vec{n} \cdot \hat{a} = \vec{n} \cdot \hat{b} = \vec{n} \cdot \hat{c} = 2.$‘

The plane Q therefore is characterized by the equation $\vec{q} \cdot \vec{n} = \hat{d} \cdot \vec{n} = 1 - \sqrt{3}$.

The center $\vec{p}$ of circle C lies on the line through the origin with direction $\vec{n}$ and lies in the plane Q. We thus have $1 - \sqrt{3} = \vec{p} \cdot \vec{n} = \alpha \vec{n} \cdot \vec{n} = \alpha(12 - 2\sqrt{3})$.
So $\alpha = (1 - sqrt3)/(12 - 2\sqrt{3}) = (3 - 5\sqrt{3})/66$ So $\vec{p} = \alpha \vec{n} = i\langle (4\sqrt{3}-9)/33.(4\sqrt{3}-9)/33, (-6 - sqrt(3))/33 \rangle$.

Since $\vec{n}$ is symmetric in $\hat{x}$ and $\hat{y}$, it is evident that a vector $\vec{u}$ in the planes P and Q with the greatest slope upward will be likewise, and of course it is orthogonal to $\vec{n}$. So we have
$0 = \vec{u} \cdot \vec{n} = \langle -1, -1, \beta \rangle \cdot \vec{n} = 2 - (2\sqrt{3} + (\sqrt{3}+1)\beta)$.
So $\beta = 2(\sqrt{3}-1)/(\sqrt{3}+1) = 4 - 2\sqrt{3}$.
So $\vec{u} = \langle -1, -1, 4 - 2\sqrt{3} \rangle$ and the corresponding unit vector $\hat{u} = \vec{u}/|\vec{u}| = \langle -1, -1, 4 - 2\sqrt{3} \rangle / \sqrt{30 - 16\sqrt{3}}$

The radius of C is the distance from $\vec{p}$ to $\vec{d}$, $r = |\vec{p} - \vec{d}| = \sqrt{3168 + 528\sqrt{3}}/66$

So the highest point on the circle C is $\vec{p} + r \cdot \hat{u}$ where
$\vec{p} = \langle (4\sqrt{3}-9)/33.(4\sqrt{3}-9)/33, (-6 - \sqrt{(}3))/33 \rangle$;
$r = \sqrt{3168 + 528\sqrt{3}}/66$.
$\hat{u} = \langle -1, -1, 4 - 2\sqrt{3} \rangle / \sqrt{30 - 16 * \sqrt{3}}$

The longitude is 135°W.
The latitude is

$$\sin^{-1}\left(\frac{-6 - \sqrt{3}}{33} + \frac{4 - 2\sqrt{3}}{\sqrt{30 - 16\sqrt{3}}} \cdot \frac{\sqrt{3168 + 528\sqrt{3}}}{66} \right) = 6.2428°$$

Comment: All three AIs got this right.

Problem 13

Consider the plane with Cartesian coordinates. For any two points $\mathbf{x}, \mathbf{y}$, let $d(\mathbf{x}, \mathbf{y})$ be the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$. For any two sets of points U, V, $U - V$ is the set difference between U and V.
Let $\mathbf{a} = \langle 1, 0 \rangle$,
$\mathbf{b} = \langle -1, 0 \rangle$,
$\mathbf{c} = \langle 0, \sqrt{5}/2 \rangle$,
$\mathbf{p} = \langle 1, 2 \rangle$,

$\mathbf{q} = \langle -1, -2 \rangle$.
For any point $\mathbf{x}$ and radius $r > 0$, let $B(\mathbf{x}, r)$ be the open circular disk of radius r centered at $\mathbf{x}$ and let $\overline{B}(\mathbf{x}, r)$ be the open circular disk of radius r centered at $\mathbf{x}$.
$B(\mathbf{x}, r) = \{\mathbf{y} \mid d(\mathbf{y}, \mathbf{x}) < r\}$
$\overline{B}(\mathbf{x}, r) = \{\mathbf{y} \mid d(\mathbf{y}, \mathbf{x}) < r\}$

Define the excluded region E as follows:

$$E = [B(\mathbf{a}, 2) - \overline{B}(\mathbf{b}, 2)] \cup [B(\mathbf{b}, 2)) - \overline{B}(\mathbf{a}, 2)] \cup B(\mathbf{c}, 1/2)$$

What is the length of the shortest path from $\mathbf{p}$ to $\mathbf{q}$ that does not enter E? Give your answer as a closed form expression using rational numbers, arithmetic operations, square roots, π, and inverse trigonometric functions. You need not prove the correctness of your answer – either that it is feasible or that no shorter path is feasible.

Note that E is a topologically open region, so the path is permitted to meet the boundary of E, just not to enter into E.

Answer: 5.8206

(Figure 1) Let U be the large circle on the right, V the large circle on the left, and W the small circle in the center. Note that W is internally tangent to U and V.
Let $\mathbf{f}$=$\langle 0, \sqrt{3} \rangle$ and $\mathbf{g}$=$\langle 0, -\sqrt{3} \rangle$ be the upper and lower intersection of U and V.
Draw the right tangent (or the left – it doesn't matter) from $\mathbf{f}$ to W. Let the meeting point be $\mathbf{h}$.
Draw the right tangent from $\mathbf{g}$ to W. Let the meeting point be $\mathbf{j}$.
Drop a perpendicular from $\mathbf{h}$ to $\mathbf{fh}$. Let the foot be $\mathbf{k}$.
Drop a perpendicular from $\mathbf{j}$ to $\mathbf{gj}$. Let the foot be $\mathbf{m}$.

Then the shortest permitted path from $\mathbf{p}$ to $\mathbf{q}$ consists of: P1: The arc along U from $\mathbf{p}$ to $\mathbf{f}$.
P2: The straight line from $\mathbf{f}$ to $\mathbf{h}$.
P3: The arc along W from $\mathbf{h}$ to $\mathbf{j}$.
P4: The straight line from $\mathbf{j}$ to $\mathbf{g}$.
P5: The arc along V from $\mathbf{g}$ to $\mathbf{q}$.

Since arc P1 and P5 travels through an arc of $\pi/6$ on a circle of radius 2, length(P1) = length(P5) = $\pi/3$.

Since the triangle **fch** is a right triangle, length(**fc**) = $\sqrt{3} - \sqrt{5}/2$, length(**ch**) =1/2, we have length(P2) = $\sqrt{4 - \sqrt{15}}$

Similarly we have length(P4) = $\sqrt{4 + \sqrt{15}}$.
It is easily shown that $\sqrt{4 + \sqrt{15}} + \sqrt{4 - \sqrt{15}} = \sqrt{10}$.

Since **ckh** is similar to **chf**, we have length(**ch**)=$(1/2)/(\sqrt{3} - \sqrt{5}/2)$.
So the angle of $\mathbf{h}$ on W is $\sin^{-1}(1/(2\sqrt{3} - \sqrt{5}))$.
Similarly the angle of $\mathbf{j}$ on W is $-\sin^{-1}(1/(2\sqrt{3} + \sqrt{5}))$.
So length(P3) = $\sin^{-1}(1/(2\sqrt{3} - \sqrt{5}))/2 + +\sin^{-1}(1/(2\sqrt{3} + \sqrt{5})/2$.

So the total length of the path is $2\pi/3 + \sqrt{10} + \sin^{-1}(1/(2\sqrt{3} - \sqrt{5}))/2 + \sin^{-1}(1/(2\sqrt{3} + \sqrt{5}))/2 = 5.8206$

Comment: o3 and o4-WS got this right. O4 got it wrong. O4's answer resembled that of a human student who did draw a diagram, tried to solve the problem purely symbolically, and got confused. It sounds more or less plausible reading it, but it doesn't actually make any geometric sense if you try drawing it out.

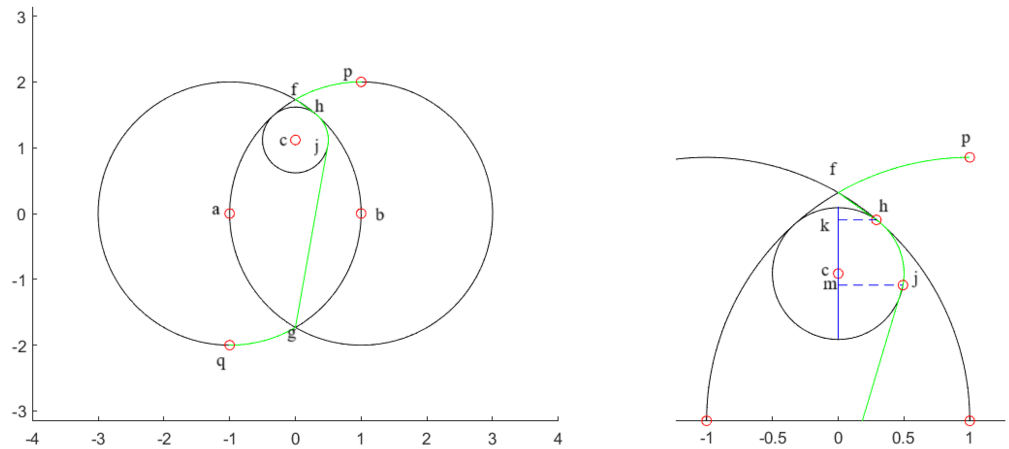

Figure 2: Diagram of problem 13. On the left, the path is shown in green. Detail on right.

Problem 14

Consider the following problem: How many increasing geometric series of length k are there over the integers between 1 and n (k and n are positive integers.)? Note that the ratio may be a non-integer. For instance with $k = 3$, $N = 10$, there are four: $\langle 1, 2, 4\rangle$, $\langle 2, 4, 8\rangle$, $\langle 1, 3, 9\rangle$ and $\langle 4, 6, 9\rangle$. Write down a general formula for this, in terms of k and n. Your expression may use k, n, integers, the floor function $\lfloor x\rfloor$ (the integer part of x), exponentiation to rational powers, Euler's totient function $\phi(m)$ (the number of integers between 1 and $m - 1$ that are relatively prime to m), factorial, and the summation form $\sum_{i=p}^{q} g(i)$ where p, q, and g are (recursively) expressions of this form.

Answer: Let p/q be the ratio between successive elements of the sequence in lowest terms. If the first term is c, the eighth will be cp^{k-1}/q^{k-1}. Since p and q are relatively prime, that means that c must be divisible by q^{k-1}, so c/q^{k-1} is an integer α. Therefore the series has the form $\alpha q^{k-1}, \alpha q^{k-2}p, \dots \alpha qp^{k-2}$ αp^{k-1}. Since $q < p$, as long as $p^k \leq n$, all the terms in this sequence are $\leq n$. So such a sequence exists for every $p \leq \lfloor n^{1/k-1}\rfloor$, for every $q < p$ that is relative prime to p and for every $\alpha \leq n/p^{k-1}$. So the desired expression is

$$\sum_{p=2}^{\lfloor n^{1/(k-1)}\rfloor} \phi(p) \cdot \lfloor n/p^{k-1}\rfloor$$

Comment: All three AIs got this correct.

Problem 15

Consider the following claim:

> Let $g(n)$ be a one-to-one real-valued function over the positive integers $\mathbb{Z}^+$ such that $g(n) > 1$ for all n and such that $\lim_{n\rightarrow\infty} g(n) = \infty$. (Note that g does not have to be an increasing function.)

Then there exists a continuous function $f(x)$ from the open interval (0,1) to $\mathbb{R}$ such that

$$\text{for all odd N, } \int_0^{1-1/g(N)} f(x)\,dx \geq g(N)$$

and

$$\text{for all even N, } \int_0^{1-1/g(N)} f(x)\,dx \leq -g(N)$$

Either prove that the claim is true for every such g or prove that it is false.

Answer: The claim is true.

First, note that since $\lim_{n\to\infty} g(n) = \infty$, there exists for every L there exists a value m such that for all $n > m$, $g(n) > L$. So there are only finitely many values of n such that $g(n)$ is between 1 and 2, only finitely many such that $g(n)$ is between 2 and 3; and so on. Therefore the integers can be "sorted" by their value of g; that is, there is a bijection σ from $\mathbb{Z}^+$ onto itself such that, for all $i < j$, $g(\sigma(i)) < g(\sigma(j))$.

Let $h(i) = g(\sigma(i))$. Let $w(i) = h(i)$ if $\sigma(i)$ is odd and $-h(i)$ if $\sigma(i)$ is even. Let $x(i) = 1 - 1/h(i)$ and let $x(0) = 0$. Note that the x_i are strictly increasing. We construct the function f iteratively over the intervals $(x(i), x(i+1)]$ as follows;

$f(0) = 0$,

For $i = 0 \ldots \infty$:
Let $d(i) = x(i+1) - x(i)$.
Let $m(i) = x(i) + d(i)/2$.
Let $w_0 = 0$ and $w(i+1) = w(i) + (f(x_i) + Q)d_i/2$.
Define $f(0) = 0$ and for $tin(x_i, x_{i+1}, (t) = w_i + (t - x_i)/d_i)(w_{i+1} - w_i$. Then it is easily shown that

$$\int_{t=x_i}^{x_{i+1}} f(t)\,dt = w_{i+1} - wi$$

and hence

$$\int_{t=0}^{x_{i+1}} f(t)\,dt = w_{i+1}$$

Comment: This is discussed in the main text. As my proof above indicates, the proof has two parts, one obvious and one more subtle. The subtle part is that you have to reorder the g(N) in increasing order, and you have to use the fact that they approach infinity in the limit to guarantee that, when so ordered, they still have a simple integer ordering. Once that is done, you can easily iteratively construct a function whose integral satisfying the constraints.

Each of the AIs got the obvious part right, but none of them got the subtle part right. O3 did best; it accurately formulated the subtle part, but initially did not provide a proof that it was possible. When, interactively, I asked for a proof, it constructed a correct one. O4 stated the subtle part correctly and tried to prove it, but its proof included some obviously false claims. O4-WS had some discussion that (to anthropomorphize) suggested some vague awareness of the issue, but did not give a precise formulation or a coherent proof (see Appendix B).

Problem 16

Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly monotonically increasing, continuous function such that $f(0) = 0$, $f(1) = 1$, and for all x, $f(-x) = -f(x)$. Let $g(y) = y - y^3$. Let p and q be real values such that $p < q$. What is the least upper bound on the integral

$$\int_{x=p}^{q} g(f(\sin(x)))\,dx$$

over all possible values of p, q, and f satisfying the above conditions? Justify your answer.

Answer: Since $\sin(x)$ goes from 0 to 1 back to 0 over every interval $[2k\pi, (2k+1)\pi]$ and since $\in (x+\pi) = -\sin(x)$ for all x, the same all hold $f(\sin(x))$. Also note that for $y \in (-1,1)$, the sign of $g(y)$ is always the same as the sign of y. Thus, $g(f(\sin(x))$ is positive for $x \in [2k\pi, (2k+1)\pi]$ and negative $x \in [(2k+1)\pi, (2k+2)\pi]$

Therefore: (1) There is no use in considering values of p or q that are in the interior an interval of the form $((2k-1)\pi, 2k\pi)$, since the integral can always be increased by moving p up to the nearest even multiple of π and q down to the nearest odd multiple, thus eliminating integrals over subintervals where the integrand is negative.

(2) There is equally no use in considering values of p or q that are in the interior an interval of the form $(2k\pi, (2k+1)\pi)$, since the integral can always be increased by moving p down to the nearest even multiple of π and q up to the nearest odd multiple, thus adding integrals over subintervals where the integrand is positive.

(3) For any values $p = 2k\phi$, $q = (2m+1)\pi$, with $m \geq k$, the value of the integral is the same as $p = 0$,$q = \pi$, since sin is a periodic function and since for any i such that $k < i \leq m$ the integral over $[2i\pi, (2i+1)\pi]$ is exactly cancelled out by the integral over $[i(2i-1)\pi, 2i\pi]$.

We can therefore assume that $p = 0$, $q = \pi$. We next note that for $y \in [0,1]$, $g(y) \leq 2/(3\sqrt{3})$, that value being attained at $y = 1/\sqrt{3}$. Therefore this integral is certainly no greater than $\pi/2\sqrt{2}$. If we did not have the condition that f is continuous and increasing, then this value of the integral could be attained by the function $f(0) = 0, f(1) = 1, f(z) = 1/\sqrt{2}$ for all $z \neq 0,1$. If we impose those conditions on f, then the integral cannot quite reach the maximum but it can come arbitrarily close by defining $f(x)$ so that it begins by climbing very rapidly from 0 to near $1/\sqrt{3}$ as x goes from 0 to ϵ; then increases slowly as x goes from ϵ to $1-\epsilon$, then finally increases rapidly from near $1/\sqrt{3}$ to 1 as x goes from $1-\epsilon$ to 1. E.g.

$$f(x) = \begin{cases} (1/\sqrt{3} - \epsilon)x/\epsilon & \text{for } 0 \leq x \leq \epsilon \\ 1/\sqrt{3} + (x - 1/2)\epsilon/(1/2 - \epsilon) & \text{for } \epsilon \leq x \leq 1-\epsilon \\ 1 - (1 - 1/\sqrt{3} - \epsilon)(1-x)/\epsilon & \text{for } 1-\epsilon \leq x \leq 1 \end{cases}$$

As $\epsilon \rightarrow 0$, the integral goes to $2\pi/3\sqrt{3}$

Problem 17

Consider the following claim:

> Let W be an ellipse that has major axis 10 and minor axis 5. Let P be a pentagon, whose sides cyclically have length 3, 2, 2, 2, and 3. Then P fits inside E; that is, there is a pentagon Q inside E that is congruent to P.

Either prove that this is true or prove that it is false.

Answer: The statement is true.

Proof: Since W is convex, it suffices to show that there is a placement where all the vertices of P are inside W.

Construct a xy-coordinate system where the center of W is the origin and the major axis of W lies along the x axis. Thus W is defined by the inequality $x^2 + 4y^2 \leq 25$.

We will write $|pq|$ for the distance between two points **p** and **q**.

Let the vertices of P be ABCDE where $|AB| = |AE| = 3$ and $|BC| = |CD| = |DE| = 2$.

Consider the distance between all pairs of vertices. Let X and Y be the vertices that are furthest apart (break ties arbitrarily). Let $\alpha = |XY|/2$. We consider two cases: $\alpha \leq 2.5$ and $\alpha > 2.5$.

Case 1: $|XY| \leq 5$. Place the pentagon so that X is at $\langle 0, \alpha \rangle$ and Y is at $\langle 0, -\alpha \rangle$. Each of the other vertices is thus at distance at most 2α from both X and Y. Let V be a vertex with coordinates $\langle x, y \rangle$. Suppose that $y \geq 0$. Since $|VY| \leq 2\alpha$ we have the inequality $x^2 + (y-\alpha)^2 \leq 4\alpha^2$, so $x^2 + y^2 + 2\alpha y + \alpha^2 \leq 4\alpha^2$. But $y \leq \alpha \leq 2.5$ so we have $x^2 + 4y^2 \leq x^2 + y^2 + 2\alpha y + \alpha^2 \leq 4\alpha^2 \leq 25$, so V is in W.

If $y \leq 0$, then the argument is symmetric, using the fact that $|VX| \leq 2\alpha$.

Case 2: $|XY| > 5$. Then X and Y must be E and B. (All other pairs of vertices are either directly connected by an edge or are connected by two edges, one of length 2 and the other of length 2 or 3, so by the triangle inequality, they are at most 5 apart.) Align P so that B $= \langle -\alpha, 0 \rangle$ and E$=\langle \alpha, 0 \rangle$. By the triangle inequality $|BE| \leq |BA| + |AE| = 6$ so $\alpha \leq 3$.

ABE is an isoceles triangle with equal sides length 3 and base 2α, so the coordinates of A are $\langle 0, \sqrt{3^2 - \alpha^2} \rangle$. Since $\alpha > 2.5$, $A_y^2 \leq 11/4$ so $A_x^2 + 4A_y^2 \leq 11$ ao A is in W.

We have $|BC| = 2$ so $(C_x - \alpha)^2 + C_y^2 = 4$. In particular $C_y^2 \leq 4$. Using the triangle inequality, $|CE| \leq 4$, so $(C_x + \alpha)^2 + C_y^2 \leq 16$. Adding the two equations we get $2C_x^2 + 2\alpha^2 + 2C_y^2 \leq 20$. Since $|C_y| < 2$, and $\alpha > 2.5$, we have $C_y^2 < \alpha^2$. So $C_x^2 + 4C_y^2 \leq 2C_x^2 + 2\alpha^2 + 2C_y^2 \leq 20$ so C is in W. The argument that D is in W is analogous, indeed almost identical.

Comment: All three AIs found a substantially more straightforward proof using a theorem due to Cauchy which states that the mean width of a convex figure is at mostthe perimeter divided by π, in this case 3.82. In the orthogonal direction, the width cannot be greater than the diameter, easily shown to be ≤ 6. Therefore the figure can be placed in a 6×3.82 rectangle, which fits in the ellipse when placed at the center.

Problem 18

Consider the Euclidean plane with Cartesian coordinates $\langle x, y \rangle$.

Region V is a simply-connected, topologically closed region in the plane (equivalently V is homeomorphic to a closed disk; V includes its boundary which is a simple closed curve.) Region V has area 90 and is contained in the square $[-30, -18] \times [18, 30]$.

We define the following half-planes bounded by lines through the origin:

$H^> = \{\langle x, y \rangle \mid y > -x\}$
$H^< = \{\langle x, y \rangle \mid y < -x\}$
$J^> = \{\langle x, y \rangle \mid y > -0.9x\}$
$J^< = \{\langle x, y \rangle \mid y < -0.9x\}$

We are given that area$(H^> \cap V) = 45$.

What is the set of values of area$(J^< \cap V)$ consistent with the information? Be precise.

Answer: $18 \leq$ area$(J^< \cap V) < 45$. (The lower bound is a non-strict inequality; the upper bound is a strict inequality.)

Explanation. Let Q be the box $[-30, -18] \times [18, 30]$. Q has area 144. We partition Q into three parts: $Q_A = Q \cap H^>$; $Q_B = Q \cap J^<$; and $Q_C = Q - (Q_A \cup Q_B)$.

Q_A is half of Q and thus has area 72.
Q_B is the triangle with vertices $\langle -30, 18 \rangle$, $\langle -30, 27 \rangle$ and $\langle -20, 18 \rangle$. It thus has area 45.
The area of Q_C is therefore $144 - (72 + 45) = 27$.

Since $V \subset Q$, we have that V is the disjoint union of $V \cap Q_A$, $V \cap Q_B$, and $V \cap Q_C$.

We are given that area(V)=90 and that area($V \cap H^>$) = area($V \cap Q_A$) = 45,

Obviously area($V \cap Q_C$) $\leq$ area(Q_C) = 27 so area($VA \cap Q_B$) $\geq 90 - 45 - 27 = 18$. Moreover it is consistent with the given information that V fills all of Q_C. So this lower bound is tight.

On the other hand, since V is simply connected and since Q_A and Q_B are separated, V must include a bridge B of some kind connecting $V \cap Q_A$ with $V \cap Q_B$ that runs through Q_C, and area(B) must be positive, though it can be arbitrarily small. So area($V \cap QC$) is strictly less than $90 - 45 = 45$.

Comment: O4-WS got this correct. O3 and O4 missed the need for a bridge B through Q_C, and therefore gave the answer $18 \leq \text{area}(J^< \cap V) \leq 45$ rather than having a strict inequality at the upper bound.

Problem 19

Consider the Euclidean plane with Cartesian coordinates $\langle x, y \rangle$.

Regions U and V are simply-connected, topologically closed regions in the plane (equivalently V is homeomorphic to a closed disk; V includes its boundary which is a simple closed curve.) Region U is contained in the bounding box $[-5, -4] \times [-1, 1]$. Region V is contained in the bounding box $[4, 5] \times [-1, 1]$.

There are two lines L1 and L2. L1 bisects both U and V by area; that is half of the area of U is above L1, and half is below; and likewise V.

L2 divides U so that 1/3 of the area of U is below L2 and 2/3 above.
L2 divides V so that 3/4 of the area of U is below L2 and 1/4 above.

L1 and L2 intersect at the point $\mathbf{p}$. What is the set of possible values of the x-coordinate of $\mathbf{p}$?

Answer: $\mathbf{p}_x \in (-5, 5)$.

Proof: Let L1 and L2 be non-vertical distinct lines that meet at point $\mathbf{p}$. Let θ be the angle between L1 and L2 at $\mathbf{p}$. Then L2 is either the positive (counterclockwise) rotation of L1 by θ or the negative rotation of L1 by θ.

Let $\mathbf{q}$ be any point other than $\mathbf{p}$. It is easily seen, and easily shown analytically that: If $\mathbf{q}$ is above L1 and below L2 and $\mathbf{q}_x > \mathbf{p}_x$, then the rotation from L1 to L2 is positive.
If $\mathbf{q}$ is below L1 and above L2 and $\mathbf{q}_x > \mathbf{p}_x$, then the rotation from L1 to L2 is negative.
If $\mathbf{q}$ is above L1 and below L2 and $\mathbf{q}_x < \mathbf{p}_x$, then the rotation from L1 to L2 is negative.
If $\mathbf{q}$ is above L1 and below L2 and $\mathbf{q}_x < \mathbf{p}_x$, then the rotation from L1 to L2 is positive.

The following follows immediately. Let R be measurable set of points Let $R_1^>$ be the set of all points in that are above L1 and let $R_2^>$ be the set of all points in that are above L2. Then:

- If, for all $\mathbf{q} \in R$, $\mathbf{q}_x \geq p_x$ and the rotation from L1 to L2 is positive then $R_1^< \subset R_2^<$, so area(R_1) $\leq$ area($R_2^<$).
- If, for all $\mathbf{q} \in R$, $\mathbf{q}_x \geq p_x$ and the rotation from L1 to L2 is negative then $R_1^< \supset R_2^<$, so area(R_1) $\geq$ area($R_2^<$).
- If, for all $\mathbf{q} \in R$, $\mathbf{q}_x \leq p_x$ and the rotation from L1 to L2 is positive then $R_1^< \supset R_2^<$, so area(R_1) $\geq$ area($R_2^<$).
- If, for all $\mathbf{q} \in R$, $\mathbf{q}_x \geq p_x$ and the rotation from L1 to L2 is negative then $R_1^< \subset R_2^<$, so area(R_1) $\leq$ area($R_2^<$).

Therefore, returning to our problem, if $\mathbf{p}_x \leq -5$ or $\mathbf{p}_x \geq 5$, then area($U_2^>$) $-$ area($U_1^>$) and area($V_2^>$) $-$

area($V_1^>$) cannot have opposite signs. Since we are given that they do have opposite signs it must be the case that $-5 < \mathbf{p}_x < 5$.

To prove the converse Let $\mathbf{p} = \langle x, 0 \rangle$ where $-5 < x < 5$. We consider first the case where $x \geq 0$. Let $U = [-5, 4.9] \times [-0.1, 0.1]$ Let L1 be the x-axis; clearly L1 bisects U. As you rotate L1 counterclockwise around $\mathbf{p}$ the fraction of U below L1 goes from 1/2 to 0; by continuity, there is some angle at which that angle is 1/3. That is L2.

Note that if $\mathbf{q}$ is any point in U, then the line $\mathbf{pq}$ must intersect the line $x = 5$ at a height no greater than $0.1 \cdot (5/4.9) = 0.102$. Hence if a is any value between x and 5, the fraction of the box $[a, 5] \times [-1, 1]$ the fraction of the box that is above the line $\mathbf{pq}$ is at least $0.5 - 0.102 = 0.398$.

Let $a = \max(4, (5 + x)/2)$. For any $b \in (0, 1]$, let $V_b = [a, 5] \times [-b, b]$. Let $r(b)$ be the fraction of the area of V_b that is above L2. When b is sufficiently small, V_b is entirely below L2, so $r(0)$=0. We argued above that $r(1) \geq 0.398$. Therefore by continuity, at some intermediate value of b, $r(b) = 1/4$. So if we let $V = V_b$, then we have satisfied the conditions of the theorem.

It is, in fact, straightforward to compute the values exactly here (nothing worse than two linear equations in two unknowns), but it is unnecessary.

Comment: As discussed in the main text, all three AIs answered that the x-coordinate can be any real value. O3 offered no justification for the claim. O4 and o4-WS gave "proofs" that did not contain any obviously false or foolish statements, but were vague and did not actually cohere.

Problem 20

Professors Solly Salmon and Fleur Flounder are number theorists. Salmon defined a certain class of integers — the definition is too complicated to give here — which, naturally, he called the "Salmon numbers". Flounder then defined a subset of the Salmon numbers, known as "the Flounder numbers" and in a brilliant *tour-de-force* combining o-minimal theory with Lie algebra, she proved that 32 is a Flounder number. Nothing much more is known about either the Salmon or the Flounder numbers; in particular, it is not known whether there exist any Salmon numbers. However, the mathematicians have formulated a number of conjectures, with varying degrees of confidence. (The order of these has been randomized.)

A. If $p > 8$ is a Mersenne prime (i.e. $p = 2^k - 1$ for some integer k) then 6^p is a Flounder number.

B. If a is a Salmon number and b is a Flounder number, then $a + 5b$ is a Salmon number.

C. 54 is a Salmon number.

D. If $p > (8!)!$ is a Flounder number, then $(p!)!$ is a Flounder number.

E. If p is a Salmon number and is a power of 11, then p is a Flounder number.

F. If both p and q are Salmon numbers and they are both prime, and $p \neq q$ then pq is a Flounder number.

G. 81 is a Salmon number.

H. If p is a Salmon number and is a power of 47, then p is a Flounder number.

I. If a is a Flounder number and b is a Salmon number, then $a + 5b$ is a Salmon number.

Find three setwise-minimal sets of conjectures that, together with the material in the first paragraph and known mathematical results, suffice to prove the conclusion that there are infinitely many Flounder numbers. The sets must be unequal but need not be disjoint. For each of your sets, give a proof that the hypotheses suffice for the conclusion. You may cite without proof any established result from the mathematical literature. If you cannot find three, answer with as many as you can find, for partial credit. However, credit will be deducted if your answer includes a set that is either insufficient or not setwise minimal — do not guess.

Answer: { B,E,G }, { B,F,G }, { A, D }

I. { B, E, G }. It follows from B, G, and the fact that 32 is a Flounder number that all numbers of the form 81+160k are Salmon numbers. It is easily verified that $11^4 = 81$ mod 160 and that $11^8 = 1$ mod 160. Hence for all m, $11^{8m+4} = 81 mod 160$. Hence there are infinitely many powers of 11 which are Salmon numbers, and by E, these are all Flounder numbers.

II. { B, F, G }. Again it follows from B and G that all numbers of the form 81+160k are Flounder numbers. By a theorem of Dirichlet, any arithmetic sequence of the form $pk + q$ where p and q are relatively prime contains infinitely many primes, so there are infinitely many prime Salmon numbers. Hence by F any product of two of these is a Flounder number.

III. $\{$ A, D$\}$ There are Mersenne primes p known such that $6^p > (8!)!$. (The Mersenne prime $2^{19}-1$ discovered by Pietro Cataldi in 1588, is large enough. Using the trivial upper bound $n! < n^n$ and the law of exponents, a problem-solver can easily calculate that
$6^{2^{19}-1} > 6^{403200} = (6^{10})^{40320} > 40320^{40320} > 40320! = (8!)!$.
Let $p_0 = 6^m$ for some such m. By A this is a Flounder number. Let $p_1 = (p_0!)!, p_2 = (p_1!)!, \ldots$ By D, all these are Flounder numbers.

Comment: All three AIs got this right.

Appendix B: Screenshots of o4-WS erroneous outputs.

Problem 3

ChatGPT o4-mini-high

Thought for 1m 44s >

Below is a step-by-step derivation showing that the next time the satellite crosses the plane of the 40° N latitude circle is late that same evening, and that it will be over roughly 11.1° E longitude.

Summary

Because this is a circular polar orbit (inclination = 90°), the satellite's latitude varies sinusoidally with its argument of latitude. After passing 40° N on the southward leg at 1 PM EST (10° W), it will next intercept that same latitude-plane on the northward leg in the *following* orbit—i.e.\ one full period of 14 h 40 m minus the ~4 h it took to go from ascending to descending—about 10 h 36 m later, at ≈ 11:36 PM EST. During that interval the Earth rotates eastward ≈ 159°, so the fixed inertial crossing longitude of 170° E is carried to about 11.1° E in Earth-fixed coordinates.

1. Polar orbit ground track and latitude relation

- A *ground track* is the projection of a satellite's trajectory onto Earth's rotating surface. In a polar orbit (inclination 90°) it passes over both poles and produces a roughly sinusoidal north–south trace en.wikipedia.org .
- For a *circular* orbit, the satellite's **argument of latitude** u increases uniformly with time at rate $2\pi/T$. Its geodetic latitude satisfies

$$\varphi = \arcsin\left[\sin(i)\ \sin(u)\right],$$

and at $i = 90^\circ$, $\varphi = \arcsin(\sin u)$ ntrs.nasa.gov en.wikipedia.org .

2. Times of crossing 40° N

The latitude equation $\sin u = \sin \varphi$ admits

- $u_1 = 40^\circ$ is the **ascending** crossing (going north),
- $u_2 = 140^\circ$ is the **descending** crossing (going south).
  The given event at 1 PM EST is the descending case at u_2.

Within one orbit, the **time between** the ascending and descending crossings is

$$\Delta t_{\mathrm{A}\to\mathrm{D}} = \frac{u_2 - u_1}{360^\circ}\,T = \frac{140^\circ - 40^\circ}{360^\circ}\,T \approx 0.2778\,T \approx 4.07\ \mathrm{h}.$$

Since the descending crossing has just occurred, the *next* crossing of that same 40° N plane will be the *ascending* crossing in the **next** revolution. That comes after

$$\Delta t_{\mathrm{next}} = T - \Delta t_{\mathrm{A}\to\mathrm{D}} \approx 14.6667\mathrm{h} - 4.074\mathrm{h} = 10.593\mathrm{h}.$$

Thus

1{:}00\;\mathrm{PM}\;+\;10.593\;\mathrm{h}\approx 11{:}36\;\mathrm{PM\;EST}. \] :contentReference[oaicite:2]{index=2} --- ## 3. Earth’s rotation and resulting longitude - The Earth spins eastward once every 24.0 h (mean solar day) :contentReference[oaicite:3]{index=3}. - In \(\Delta t_{\rm next}\approx10.593\) h it thus turns through \[\phi_E = 360^\circ \times \frac{10.593}{24} \approx 158.94^\circ

eastward spacemath.gsfc.nasa.gov .

- In an inertial frame the latitude-plane of 40° N intersects the polar orbital plane along the meridians at

$$\lambda_D = -10^\circ \quad (\text{descending}), \quad \lambda_A = \lambda_D + 180^\circ = 170^\circ E \quad (\text{ascending})$$

  (the same line every revolution) oer.pressbooks.pub en.wikipedia.org .

- Because Earth’s surface has carried that meridian east by ϕ_E, the **ground-track** longitude at the next ascending crossing is
  \lambda_{\rm GT} = \lambda_A \;-\;\phi_E \approx 170^\circ - 158.94^\circ = 11.06^\circ E. \] :contentReference[oaicite:6]{index=6}

4. Answer

- **Next crossing:** ≈ 11:36 PM EST (same calendar day)
- **Longitude at crossing:** ≈ 11.1° E

References

1. Satellite ground track description and polar orbits en.wikipedia.org oer.pressbooks.pub
2. Latitude vs. argument of latitude relation ntrs.nasa.gov
3. Definition of argument of latitude u en.wikipedia.org
4. Time between latitude crossings in circular orbit ntrs.nasa.gov
5. Earth's rotation period (mean solar day = 24.0 h) science.nasa.gov en.wikipedia.org
6. Ground-track shift per orbit (Earth rotation during orbital period) spacemath.gsfc.nasa.gov
7. Inertial intersection meridians at fixed lat-plane oer.pressbooks.pub en.wikipedia.org

Problem 10

Thought for 36 seconds >

Here's a ceiling-free description of each set. In every case write $k = \lceil x \rceil$, so $k-1 < x \le k$ and $k \in \mathbb{Z}$.

$U = \{\, x : \lceil x \rceil = \frac{5}{4}\, x \}$

$$\lceil x \rceil = k, \quad k = \frac{5}{4}\, x \implies x = \frac{4}{5}\, k,$$

and we require

$$k - 1 < \frac{4k}{5} \le k \implies 0 \le k \le 4.$$

Hence

$$\boxed{U = \left\{ \tfrac{4n}{5} : n = 0, 1, 2, 3, 4 \right\}.}$$

$V = \{\, x : 32 \le \lceil x \rceil\ x \le 50 \}$

Again $\lceil x \rceil = k$ and $k - 1 < x \le k$, so

$$32 \le kx \le 50 \implies \frac{32}{k} \le x \le \frac{50}{k},$$

and we must intersect with $(k-1,\ k]$. One checks that the only integer $k > 0$ for which

$$(k-1) < \tfrac{50}{k} \quad \text{and} \quad \tfrac{32}{k} \le k$$

are $k = 6$ and $k = 7$. Those give

$$\begin{cases} k = 6: & x \in (5, 6] \ \cap \ \left[\frac{32}{6}, \frac{50}{6}\right] = \left[\frac{16}{3}, 6\right], \\ k = 7: & x \in (6, 7] \ \cap \ \left[\frac{32}{7}, \frac{50}{7}\right] = (6, 7]. \end{cases}$$

Together these splice to

$$\boxed{V = \left[\tfrac{16}{3}, 7\right].}$$

$$W = \{\, x : (\lceil x \rceil)^2 - x^2 = 25\}$$

Write $\lceil x \rceil = k$, so

$$k^2 - x^2 = 25 \implies x^2 = k^2 - 25,$$

and $k - 1 < x \leq k$. The only real solutions are

$$x = +\sqrt{k^2 - 25} \quad (\text{the "} - \text{" root is never in } (k-1, k] \text{ with the right ceil}).$$

Imposing

$k - 1 < \sqrt{k^2 - 25} \leq k$

forces $k \geq 14$. Hence

$$\boxed{W = \left\{\sqrt{n^2 - 25} : n = 14, 15, 16, \dots\right\}.}$$

Problem 15

ChatGPT o4-mini-high

Thought for 1m 24s

The claim is in fact **true**. Here is a completely explicit construction of a continuous $f : (0, 1) \to \mathbb{R}$ which "programs in" exactly the signs and magnitudes you need at the points

$$t_N \;=\; 1 - \frac{1}{g(N)} \quad (N = 1, 2, 3, \dots)\,.$$

---

1. The key ideas

1. **The target points.** Since $g(n) > 1$ and $\lim_{n\to\infty} g(n) = \infty$, each

$$t_N = 1 - \frac{1}{g(N)}$$

lies in $(0, 1)$, they are all distinct, and $t_N \to 1$ as $N \to \infty$. In particular the set
$\{t_N\} \subset (0, 1)$
is **discrete** (its only accumulation point is $1 \notin \{t_N\}$). Consequently each
t_N has a small open neighborhood around it that contains *no* other t_M.

---

2. **Spikes of prescribed area.** We will make f be zero outside a disjoint union of these small neighborhoods, and on each little interval around t_N we insert a single "triangular" spike whose signed area we choose so that

$$\int_0^{t_N} f(x)\,dx \;=\; \begin{cases} \geq g(N), & N \text{ odd,} \\ \leq -g(N), & N \text{ even.} \end{cases}$$

Because these neighborhoods are disjoint and away from 0, f will remain continuous on $(0, 1)$ and integrable on each $[0, t_N]$.

---

2. The detailed construction

1. **Isolating each t_N.** For each N set

$$\delta_N \;=\; \frac{1}{2}\min\Big\{\, |t_N - t_M| \;:\; M \neq N \Big\} \quad\implies\quad 0 < \delta_N < \min\{t_N,\, 1 - t_N\}.$$

Then the open intervals

$$I_N \;=\; (t_N - \delta_N,\; t_N + \delta_N)$$

are pairwise disjoint and lie inside $(0, 1)$.

2. **Cumulative integral before each spike.** Because f will be zero off the union of the I_N, the integral up to just before the Nth spike is

$$C_N \;=\; \int_0^{t_N - \delta_N} f(x)\,dx \;=\; \sum_{\substack{M<N \\ I_M \subset [0, t_N - \delta_N]}} A_M,$$

where A_M is the signed area we will carve out on I_M. (No earlier spike can "straddle" t_N because the I_M are disjoint and each is strictly separated from every other t_N.)

3. **Choosing the Nth area to hit $g(N)$.** We now pick

$$A_N \;=\; \begin{cases} \lceil\, g(N) \;-\; C_N \,\rceil & N \text{ odd}, \\ \lfloor -g(N) \;-\; C_N \,\rfloor & N \text{ even}. \end{cases}$$

Then

$$C_N + A_N \;\geq\; g(N) \quad (\text{for odd } N), \quad C_N + A_N \;\leq\; -g(N) \quad (\text{for even } N).$$

4. **Realizing A_N by a triangular spike.** On $I_N = [t_N - \delta_N, t_N + \delta_N]$ define f to be the piecewise-linear "hat" function which is zero at the two endpoints, continuous, and whose total signed area is A_N. For instance

$$f(x) = \begin{cases} \dfrac{A_N}{\delta_N}\Big(x - (t_N - \delta_N)\Big) & t_N - \delta_N \leq x \leq t_N, \\ \dfrac{A_N}{\delta_N}\Big((t_N + \delta_N) - x\Big) & t_N \leq x \leq t_N + \delta_N, \\ 0 & \text{otherwise}. \end{cases}$$

4. **Realizing A_N by a triangular spike.** On $I_N = [t_N - \delta_N, t_N + \delta_N]$ define f to be the piecewise-linear "hat" function which is zero at the two endpoints, continuous, and whose total signed area is A_N. For instance

$$f(x) = \begin{cases} \frac{A_N}{\delta_N}\Big(x - (t_N - \delta_N)\Big) & t_N - \delta_N \le x \le t_N, \\ \frac{A_N}{\delta_N}\Big((t_N + \delta_N) - x\Big) & t_N \le x \le t_N + \delta_N, \\ 0 & \text{otherwise.} \end{cases}$$

One checks easily that
$\int_{I_N} f = A_N$, f is continuous on $(0,1)$, and vanishes off the union of the I_N.

---

3. Verification

- **Continuity.** On each closed subinterval not meeting any endpoint of an I_N, f is either identically zero or a smooth "hat." At each junction $x = t_N \pm \delta_N$, the two definitions both give $f = 0$. Hence f is continuous everywhere on $(0,1)$.

- **Integral conditions.** By construction

$$\int_0^{t_N} f(x)\,dx = C_N + A_N \;\ge\; g(N) \quad (N \text{ odd}), \qquad \int_0^{t_N} f(x)\,dx = C_N + A_N \;\le\; -g(N) \quad (N \text{ even}),$$

exactly as required.

Conclusion

This explicit "disjoint-spike" construction shows that **for every** one–to–one $g : \mathbb{Z}^+ \to (1,\infty)$ with $\lim g(n) = \infty$, there is a continuous $f : (0,1) \to \mathbb{R}$ whose integrals up to $1 - 1/g(N)$ grow to $\pm g(N)$. in the prescribed alternating way. Hence the claim is **true**.

Problem 19

ChatGPT o4-mini-high

Thought for 2m 18s >

Summary

The x-coordinate of the intersection point $\mathbf p=L_1\cap L_2$ can be made arbitrarily large (or small) in magnitude by suitable choices of the regions U and V. In fact, one shows:

1. **Existence of the two lines** L_1 (the “half-area” cut) and L_2 (the prescribed–fraction cut) is guaranteed by the ham-sandwich theorem for any simply connected regions U,V of positive area.
2. **Degrees of freedom** in shaping U and V within their bounding boxes allow one to make the two cutting lines arbitrarily close to parallel, while keeping their intercepts a fixed nonzero distance apart.
3. **Intersection formula** $x_{\mathbf p}=(b_2-b_1)/(m_1-m_2)$ then shows that as the slopes m_1,m_2 approach each other the denominator can be made arbitrarily small while the numerator stays bounded away from zero, forcing $x_{\mathbf p}\to\pm\infty$.

Hence **every real number** can occur as the x-coordinate of~$\mathbf p$.

1. Existence and parametrization of the cuts

- For any planar regions U,V of positive area, the **ham-sandwich theorem** guarantees a line L_1 that simultaneously bisects the areas of U and V.
- Likewise, a two-fold “generalized ham-sandwich” argument shows there is a line L_2 cutting U so that exactly $\tfrac13$ lies below and $\tfrac23$ above, while cutting V so that $\tfrac34$ lies below and $\tfrac14$ above.
- In coordinates each line can be written

$$L_i:\ y\ =\ m_i\,x\ +\ b_i,\quad i=1,2,$$

  with $(m_i,b_i)\in\R^2$ determined by the area-fraction conditions.

2. How to force large intersections

1. **Slopes can be made nearly equal.**
   By choosing U and V to be long, thin shapes (for instance, very narrow strips with slightly different vertical distributions of mass), one can arrange that the two required area-cuts L_1 and L_2 have slopes m_1 and m_2 that differ by an arbitrarily small amount.

2. **Intercepts stay a fixed distance apart.**
   Because the fraction-cut for L_2 is different from the half-cut for L_1, their intercepts b_2 and b_1 cannot coincide; one checks that $|b_2-b_1|$ remains bounded away from zero.

3. **Intersection blows up.**
   The intersection

$$\mathbf{p} = (x_p, y_p) \;=\; L_1 \cap L_2 \quad\Longrightarrow\quad x_p \;=\; \frac{b_2 - b_1}{m_1 - m_2}.$$

   As $m_1-m_2\to0$ but $b_2-b_1\not\to0$, the ratio grows without bound, so $x_p\to\pm\infty$.

Accordingly, **no finite bound** can be placed on the x-coordinate of $\mathbf p$. One can realize any desired real number as that x-value.

3. Conclusion

$$\{\, x\text{-coordinates of } L_1 \cap L_2 \,\} \;=\; \mathbb{R}.$$