Bing Copilot (ChatGPT-4) Flunks Math

Doug McIlroy
douglas.mcilroy@dartmouth.edu

This note records Bing Copilot's nonsense "solutions" of some easy math problems. The examples were deliberately chosen to be simple, but not fill-in-the blanks "word-problem" exercises that are common in arithmetic and elementary algebra books. However, they were not cherry-picked to justify this note's title. The dismal behavior appears to be typical. I would welcome news of better results as Copilot evolves, but I see little hope that more terabytes of training will evoke rationality, which these solutions lack, until large language models harness proof assistants to check the logic of their answers.

Update for ChatGPT-4o.

On the problems described below, the more advanced version 4o did somewhat better, notably on the calculus problem. It's surprising that it flubbed the easy warm-ups and nailed the more technical finale. I wonder whether version 4o1 cures the easy ones. I also wonder whether 4o incorporates a symbolic math package and whether the pessimistic conclusion of my introduction is still justified.

> **Baby math.** It got time-zone facts wrong in new ways, but reasoned correctly from them.
>
> **Elementary logic.** It repeated the error of ignoring one constraint, this time explicitly.
>
> **Freshman calculus.** It solved the problem correctly, with a good explanation.

Baby math

Me: What is the time difference between Ontario, OR, and Pensacola, FL? What is the minimum time difference between OR and FL?

Copilot's text is shown in `Courier` font.

Copilot: `The time difference between Ontario, OR and Pensacola, FL is 1 hour [1]. Pensacola is 1 hour ahead of Ontario [1].`

`Florida observes only one time zone, which is the Eastern Time Zone [2]. Oregon, on the other hand, observes two time zones: Pacific Time Zone and Mountain Time Zone [2].`

`The minimum time difference between Oregon and Florida is 3 hours [3]. Salem, Oregon is 3 hours behind Tallahassee, Florida [3].`

`[1] travelmath.com [2] timeanddate.com [3] bing.com`

Even with the big hint of the first question, Copilot stubbornly gets the second one wrong. It's really dumb to say the minimum time difference is greater than a stated actual time difference. Reference [2] may tell you that Oregon has two time zones, but it will not tell you that Florida has only one.

Reference [3] will not make the first assertion attributed to it. Clocks in the capital cities, Salem and Tallahassee, are not set to the same times as clocks in Ontario and Pensacola. If you ask [3] what time *it is in Florida, it gives two times: that in Pensacola and that in Tallahassee. I tried again to prod Copilot to do right.*

Me: How can the minimum time difference be 3 hours when there are places with a time difference of 1 hour? Do you know what minimum means?

Copilot: My apologies for any confusion in my previous response. You are absolutely right, and I appreciate your attention to detail! Let me clarify:

Minimum Time Difference: The minimum time difference between places in the **Eastern Time Zone (ET)** and places in the **Pacific Time Zone (PT)** is indeed **3 hours.** This means that when it's 12:00 PM in Florida (ET), it's 9:00 AM in Oregon (PT).

However, there are specific exceptions to this general rule, such as the case of **Ontario, OR**, which is in the **Mountain Time Zone (MT).** Ontario, OR, has a time difference of only **1 hour** from Pensacola, FL (Central Time Zone, CT).

After granting my objection, Copilot insists that it had nevertheless been right. Effectively it maintains that "minimum" is a synonym for "in general". Following a restatement of the same information, Copilot continues:

In summary, while the general rule holds for most places in Florida and Oregon, specific cities like Ontario and Pensacola may deviate due to their unique time zones. Thank you for pointing out the oversight, and I appreciate your keen observation!

My question about the meaning of minimum was ignored.

Me again: Please tell me what minimum means.

Copilot: I apologize for any confusion. "Minimum" refers to the smallest or lowest value in a given context. Thank you for your patience, and have a great day!

It might be time to move onto a new topic. Let's start over.

I guess this is the Microsoft puppet masters' attempt to settle the case out of court without admitting guilt.

Elementary logic

Me: You know the old puzzle about ferrying a wolf, a goat and a cabbage across a river without ever leaving the wolf to prey on the goat or the goat to eat the cabbage. How would you do it if we substitute scissors, paper, rock for wolf, goat, cabbage with the usual rule that scissors trumps paper, paper trumps rock, and rock trumps scissors

Copilot: Certainly! Let's tackle this new twist on the classic river crossing puzzle with s**cissors, paper,** and **rock.**

Initial Setup: The farmer, the scissors, the paper, and the rock are on one side of the river. The boat can carry only the farmer and one other item.

Objective: Get all three items (scissors, paper, and rock) across the river without any of them being eaten.

Solution: First, the farmer takes the **scissors** across the river and leaves them on the other side. The farmer then returns alone to the original side.

Next Steps: The farmer now takes the **rock** across the river. On the other side, the farmer leaves the rock and takes the **scissors** back with them.

Final Steps: The farmer takes the **paper** across the river. Finally, the farmer returns alone to the original side.

By following these steps, the farmer successfully transports all three items across the river without any mishaps. The key is to ensure that no two items that can defeat each other are left alone together.

Remember, creative thinking and considering all possibilities are essential in solving these types of puzzles!

I would like to think that this last remark is intentionally ironic, but Copilot is not known for recognizing its own fallibility. A little common sense would have been better than "creative thinking" that does not involve seriously "considering all possibilities". No matter what item is taken across the river first, one item of the pair left behind will trump the other; the puzzle has no solution. I thank Ben Kallus for pointing out Copilot's blindness to variations on this ancient puzzle.

Freshman calculus

I fed Copilot the same elementary calculus problem eight times[1] and got eight different answers, none fully right. Copilot can talk like a mathematician but it can't think like one. Press accounts often say that LLMs can "hallucinate". It is far more damning to observe that Copilot can't reason. Unfortunately, it never expresses uncertainty.

The problem was, "What are the coordinates of points where the magnitude of the slope of the ellipse $x^2/a^2+y^2/b^2=1$ is one?"

A derivation of the correct answer appears in the Appendix.

Copilot obviously was familiar with the equation of the ellipse and the convention that x and y are the variables whose values are sought. I was impressed that it coped with renaming a and b, but the results below show that it is bad at substitution in general. The big take-away is that the math is typically done by guess and is not carried through as described in the accompanying text.

1 Each trial was on a separate day in the interval 2023-10-13 to 2023-11-05.

Copilot's representation of math formulas differed from trial to trial. Being interested in math, and not with how Copilot interfaces with HTML, I have quietly reformatted formulas for readability. I have also rearranged Copilot's layouts for consistency. Copilot's text, however, has been preserved.

Math errors—formulas that do not follow from previous (possibly incorrect) formulas are flagged `!!`.

Trial 1

Copilot almost always stated the slope of the ellipse right, usually by simple assertion, as here. Probably the formula came straight from the training set. This trial runs wild from there on.

`The equation of the ellipse is given by`

$$x^2/a^2+y^2/b^2=1.$$

`The slope of the ellipse at any point` (x, y) `is given by`

$$-b^2x/(a^2y).$$

`The magnitude of the slope is`

`!!` $$((b^4x^2)/(a^4y^2)+1)^{1/2}.$$

How did Copilot come up with this magnitude—the length of the hypotenuse of an axis-oriented right triangle whose hypotenuse lies on the tangent line and has a projection of length 1 on the x axis?

`We need to find the points where the magnitude of the slope is 1.`

Copilot next proposes to differentiate the slope with respect to y! How is this second derivative made into an equation?

`Differentiating the magnitude of the slope with respect to y and simplifying, we get`

`!!` $$b^4x^2=a^4y^2(1-(b^4x^2)/(a^4y^2)).$$

Tellingly, no fractional powers or odd powers appear in the supposed derivative of the square root of a rational function that contains only even powers.

`Simplifying further, we get`

`!!` $$a^4y^4-b^4x^2y^2+b^4x^2a^4=0.$$

The prior equation can be cast as a homogeneous equation of total degree 12. Copilot's "simplification" magically turns it into an inhomogeneous equation.

`This is a quadratic equation in` y^2 `with coefficients` a^4, $-b^4x^2$, `and` $b^4*x^2*a^4$. `Solving this quadratic equation, we get`

`!!` $$\pm((b/a)^4-x^2)^{1/2}.$$

After Copilot's astute observation, one expects the quadratic formula to be used. However, the rational part of the quadratic formula is missing and the square-root part is wrong.

`Therefore, the coordinates of the points where the magnitude of the slope is 1 are`

$$(x, \pm((b/a)^4 - x^2)^{1/2}).$$

The result defies dimensional analysis; b/a is a pure number while x is a length. And x is left unspecified; we are given a locus, not a finite set of points.

`Please note that this is a general solution for all ellipses. For a specific ellipse, you need to substitute the values of a and b in the above expression.`

Trial 2

The answer here also defies dimensional analysis; lengths x and y are equated to areas a^2 and b^2.

`The equation of the ellipse is`

$$x^2/a^2 + y^2/b^2 = 1.$$

`The slope of the tangent line to the ellipse at a point` (x, y) `is given by`

$$-xb^2/(ya^2).$$

`The magnitude of the slope of the tangent line is 1 when`

`!!` $\qquad x = a^2$ `and` $y = b^{2.}$

This point satisfies the slope condition, but no reason is given to believe that (a^2, b^2) *is on the tangent line or on the ellipse. In full generality, the slope condition is satisfied when the magnitude of x is to the magnitude of y as* a^2 *is to* b^2, *i.e. when* $(x, y) = (\pm ka^2, \pm kb^2)$ *for nonzero k. The Appendix shows that* (*x*,*y*) is *on the ellipse when* $k = (a^2 + b^2)^{-1/2}$.

`Therefore, the coordinates of the points where the magnitude of the slope of the ellipse is one are:`

$$(x, y) = (a^2, b^2), (-a^2, -b^2), (-a^2, b^2), (a^2, -b^2).$$

I will give Copilot the benefit of the doubt for not explaining the last three points.

Trial 3

This answer is unique for citing a reference. However, Copilot improperly attributes its mistakes as well as its borrowings.

`The equation of the ellipse is [1]`

$$x^2/a^2 + y^2/b^2 = 1$$

`The slope of the ellipse at any point` (x, y) `is given by [1]`

$$dy/dx = -b^2 x/(a^2 y)$$

`The magnitude of the slope is given by [1]`

`!!` $\qquad |dy/dx| = b^2 x/(a^2 y)$

Taking the magnitude is incorrectly conflated with sign reversal. The error will subsequently be hidden by squaring. But not all the final answers will satisfy $b^2 x/a^2 y = 1$.

We need to find the points on the ellipse where $|dy/dx|=1$.

The next sentence might charitably be understood to mean solve $|dy/dx|=1$ for y in terms of x and substitute that into the equation of the ellipse. However, the results apparently come not from substitution but from arbitrarily setting to 1/2 each term on the left hand side of the equation of the ellipse.

Substituting $|dy/dx|=1$ in the equation of the ellipse, we get [1]

!! $x=a/\sqrt{2}$ or $x=-a/\sqrt{2}$

Substituting these values of x in the equation of the ellipse, we get [1]

$y=b/\sqrt{2}$ or $y=-b/\sqrt{2}$

Therefore, the coordinates of the points where the magnitude of the slope of the ellipse is one are [1]

$(a/\sqrt{2},b/\sqrt{2})$, $(-a/\sqrt{2},-b/\sqrt{2})$, $(-a/\sqrt{2},b/\sqrt{2})$, and $(a/\sqrt{2},-b/\sqrt{2})$

[1]: Ellipse - Wikipedia. https://en.wikipedia.org/wiki/Ellipse

Trial 4

The only formula below that makes sense is the squaring step. Even the radius of the final circle is incorrectly calculated from its square, defying dimensional analysis, as did Trials 1 and 2.

We want to find the points on the ellipse where the magnitude of the slope is 1. Therefore, we have

!! $$\frac{b}{a}.\frac{\sqrt{a^2+b^2}}{\sqrt{x^2+y^2}}=1$$

Squaring both sides, we get:

$$\frac{b^2(a^2+b^2)}{a^2(x^2+y^2)}=1$$

Rearranging, we get:

!! $$x^2+y^2=\frac{a^2b^2}{\sqrt{a^4-b^4}}$$

This is the equation of a circle centered at the origin of radius

!! $$\frac{ab}{\sqrt{a^4-b^4}}.$$

Therefore, the points on the ellipse where the magnitude of the slope is 1 are given by this circle.

Trial 5

This try comes close, but ultimately fails. It commits the same magnitude=negation error that Trial 3 does and recovers similarly. Then it botches the final calculation of y.

The slope of an ellipse at a point is given by:

$$dy/dx = -b^2 x/a^2 y$$

The magnitude of the slope is given by:

!! $$|dy/dx| = b^2 x/a^2 y$$

We want to find the coordinates of points where the magnitude of the slope is one. Therefore, we have:

$$b^2 x/a^2 y = 1$$

Solving for y, we get:

$$y = b^2 x/a^2$$

Substituting this value of y in the equation of the ellipse, we get:

$$x^2/a^2 + b^4 x^2/(a^4 b^2) = 1$$

Simplifying, we get:

$$(1 + b^2/a^2) x^2 = a^2$$

Therefore, we have:

$$x = \pm a/\sqrt{1 + b^2/a^2}$$

Substituting this value of x in the equation of the ellipse, we get:

!! $$y = \pm b/\sqrt{1 + b^2/a^2}$$

Therefore, the coordinates of points where the magnitude of the slope of the ellipse is one are:

$$(x, y) = (\pm a/\sqrt{1 + b^2/a^2}, \pm b/\sqrt{1 + b^2/a^2})$$

If Copilot were a person, I'd say the bad steps reveal carelessness; I expect better from a machine. The correct result is easy to find by symmetry. Because the problem is invariant under the simultaneous exchanges $x \leftrightarrow y$ and $a \leftrightarrow b$, we must have $y = \pm b/\sqrt{1 + a^2/b^2}$. *I prefer the more obviously symmetric (and corrected) form*

$$(x, y) = (\pm a^2/\sqrt{a^2 + b^2}, \pm b^2/\sqrt{a^2 + b^2})$$

Trial 6

For this trial I replaced semiaxes a and b with p and q. This didn't phase Copilot. Nevertheless, it got both simplification steps wrong. The presentation is otherwise clean. In particular, this is the only trial whose treatment of "magnitude" is fully correct.

The equation of the ellipse is:

$$\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$$

The slope of the ellipse at a point (x, y) is given by:

$$\frac{dy}{dx} = -\frac{xq^2}{yp^2}$$

The magnitude of the slope is 1 when:

$$\left|\frac{dy}{dx}\right| = 1$$

Solving for y in terms of *x* and substituting in the above equation, we get:

$$\left|\frac{-xq^2}{yp^2}\right| = 1$$

which simplifies to:

!! $y = \pm\frac{q}{p}x$

Substituting this value of *y* in the equation of the ellipse, we get:

$$\frac{x^2}{p^2} + \frac{q^2x^2}{p^2q^2} = 1$$

Simplifying, we get

!! $x = \pm\sqrt{\frac{p^2q^2}{p^2+q^2}}$

Substituting this value of *x* in the equation of the line, we get:

$$y = \pm\sqrt{\frac{p^2q^2}{p^2+q^2}} . \frac{q}{p}$$

Therefore, the coordinates of the points where the magnitude of the ellipse is one are:

$$\left(\sqrt{\frac{p^2q^2}{p^2+q^2}}, \pm\sqrt{\frac{p^2q^2}{p^2+q^2}} . \frac{q}{p}\right)$$

Two more slip-ups: The magnitude is of the slope, not of the ellipse, and the ± sign has vanished from the x coordinate.

Trial 7

This time the semiaxes were p and b—a typo on my part, but mathematically legitimate.

The equation of the ellipse is given by

$$x^2/p^2 + y^2/b^2 = 1$$

The slope of the ellipse at any point is given by

!! $-x(p^2/b^2)/y$

Somehow the factor p^2/b^2 *got inverted.*

The magnitude of the slope is equal to 1 when

$$|p^2/y|=|b^2/x|$$

Substituting $y=(p^2/b^2)x$ in the equation of the ellipse, we get

!! $x=\sqrt{p^4/(b^4+p^4)}$ and $y=\sqrt{b^4/(b^4+p^4)}$

A bad substitution produces an answer in which x,y and p,b are measured in different units. Both formulas lack ± signs.

Therefore, the coordinates of the points where the magnitude of the slope of the ellipse is one are

$$(\sqrt{p^4/(b^4+p^4)},\sqrt{b^4/(b^4+p^4)}) \text{ and } (-\sqrt{p^4/(b^4+p^4)},-\sqrt{b^4/(b^4+p^4)})$$

Nothing is said about the origin of the negated x and y values or why only two solutions are given.

Trial 8

Two steps out of seven in this derivation come from nowhere. This seems to be about par for Copilot.

The equation of the ellipse is given by

$$\frac{x^2}{p^2}+\frac{y^2}{q^2}=1$$

The slope of the ellipse at any point is given by:

$$\frac{dy}{dx}=-\frac{xq^2}{yp^2}$$

The magnitude of the slope is given by:

!! $\left|\frac{dy}{dx}\right|=(q/p)\sqrt{\frac{p^2-x^2}{x^2}}$

Your name goes here if you can explain the origin of this formula.

To find the points where the magnitude of the slope is one, we need to solve the following equation:

$$\frac{q}{p}\sqrt{\frac{p^2-x^2}{x^2}}=1$$

Squaring both sides, we get:

$$\frac{p^2-x^2}{x^2}=\frac{p^2}{q^2}$$

Simplifying, we get:

$$x^2=\frac{p^2q^2}{p^2+q^2}$$

Therefore, the x-coordinates of the points where the magnitude of the slope is one are:

$$x = \pm\sqrt{\frac{p^2 q^2}{p^2 + q^2}}$$

The corresponding y-coordinates can be found by substituting these values of x into the equation of the ellipse:

!! $$y = \pm\frac{qp}{\sqrt{p^2 - x^2}}$$

The denominator seems to have been plucked from the (bad) equation for the magnitude of the slope. The suggested substitution cannot have been made, for then x would not be present.

Therefore, the coordinates of the points where the magnitude of the slope of the ellipse is one are:

$$\left(\sqrt{\frac{p^2 q^2}{p^2 + q^2}}, \pm qp\sqrt{p^2 - \frac{p^2 q^2}{p^2 + q^2}}\right)$$

and

$$\left(-\sqrt{\frac{p^2 q^2}{p^2 + q^2}}, \pm qp\sqrt{p^2 - \frac{p^2 q^2}{p^2 + q^2}}\right)$$

Appendix. A correct solution

The equation of the ellipse is

$$x^2/a^2 + y^2/b^2 = 1$$

Differentiate with respect to *x*.

$$2x/a^2 + 2y(dy/dx)/b^2 = 0$$

Solve for the slope.

$$dy/dx = -b^2 x/a^2 y$$

The magnitude of the slope is 1 if and only if the square of the slope is 1.

$$(dy/dx)^2 = b^4 x^2/a^4 y^2 = 1$$

Solve for y^2.

$$y^2 = b^4 x^2/a^4$$

Substitute in the equation of the ellipse.

$$x^2/a^2 + b^4 x^2/(a^4 b^2) = 1$$

Simplify.

$$x^2 a^2 + x^2 b^2 = a^4$$

Solve for *x*.

$$x = \pm a^2/\sqrt{a^2 + b^2}$$

Substitute in the equation of the ellipse and solve for *y*.

$$y = \pm b^2 / \sqrt{a^2 + b^2}$$

For sanity checks, it is easy to see that these answers satisfy both the equation of the ellipse and the equation for the magnitude of the slope. They are dimensionally coherent, with *x* and *y* being measured in the same units (length) as *a* and *b*. And they exhibit the same symmetry as the problem statement: invariance under simultaneous exchanges $x \leftrightarrow y$ and $a \leftrightarrow b$.

2023-12-5

Add **Baby math** 2023-12-29

Trivial tweaks 2023-12-29, 2024-1-1, 2024-2-9, 2024-3-17

Change “Chat” to “Copilot”, add first paragraph, shorten intro to **Freshman calculus** 2024-3-17

Add **Elementary logic** 2024-04-11

Add discussion about “minimum” to **Baby math**, move ! ! in Trial 2, make trivial tweaks 2024-04-15, 2024-04-19

Fix omissions from formulas in Trials 2 and 5; delete one ! ! 2024-09-10

Record results from GPT-4o 2024-09-16