

June 25, 2026

Loving the Wins and Navigating the Copilot Lies

Kevin Kieller

kkieller@enableuc.com



Enableuc.com



www.linkedin.com/in/kevinkieller/



Commsverse 2026

Agenda



- Wow!
- People
- Dogs
- Copilot
- Other AI Assistants
- Some suggestions



AI is amazing!

HOST: TIM

THE NUMBERS GAME

What percentage of companies now use usage-based SaaS pricing, according to OpenView?

Correct answer: **61**

 KEVIN	 TIM	 DAVID
8,572	4,472	9,856
75 +3,257 points	HOST	33

More details: <https://www.linkedin.com/pulse/introducing-numbers-game-kevin-kieller-lmqxc/>

Commsverse 2026



About Me – Kevin Kieller

- I created an award-winning video game for the Commodore 64
- I am part consultant and part analyst, helping both organizations and vendors
- I was rated the #1 UC All-star by UC Today
- I co-host The AI Chefs show which provides “recipes” for business success with AI



Enableuc.com



www.linkedin.com/in/kevinkieller/

Commsverse 2026



Why people lie



Commsverse 2026

People lie for a variety of reasons...

1. To avoid being judged or feel shame.
2. With the aim of avoiding punishment.
3. To protect themselves from retaliation.
4. For no “good reason” (e.g., compulsive lying).
5. To present themselves in a positive way and impress others.
6. To obtain rewards.
7. Due to carelessness and impulsiveness.
8. To experience pleasure from deceiving others.
9. With the aim of keeping personal information secret.
10. For prosocial reasons—to make others (and oneself) happy, such as telling children that Santa Claus is real.
11. For altruistic reasons—to protect others from harm.



Do dogs lie?

He is more likely running a highly successful, food-optimized social strategy.



In other words: my husky may not be dishonest, but he is almost certainly an excellent sales professional.



Commsverse 2026

Deceptive-like behaviour in dogs (*Canis familiaris*) -
<https://pubmed.ncbi.nlm.nih.gov/28251387/>

Why Copilot lies

*(This applies equally to agents
and other AI assistants)*



Commsverse 2026

What does it mean to say Copilot “lies”?

Term	What it implies	Is it accurate?
Lying	Intentional deception	✗ No
Hallucinating	Plausible but unsupported output	✓ Yes
Misleading	Incorrect info presented confidently	✓ Yes
Fabricating	Creating facts that don't exist	✓ Yes

Copilot is a brilliant narrator, not a fact checker.

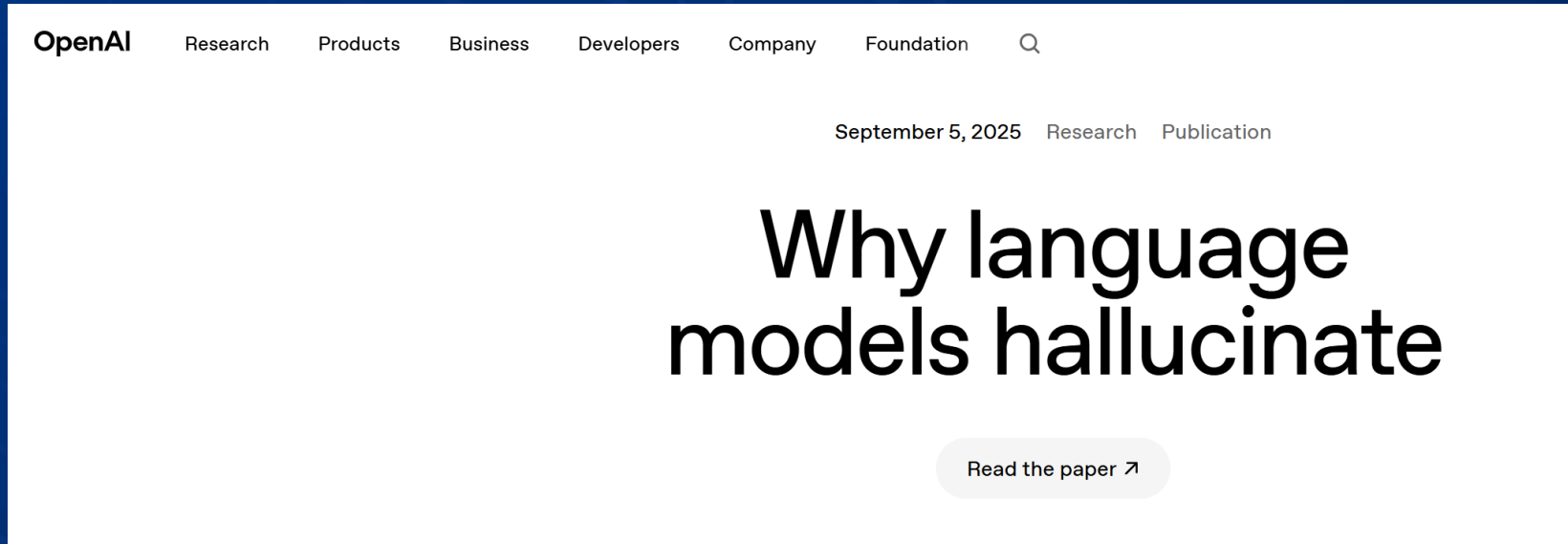


Ten reasons Copilot “lies”

1. LLMs are probabilistic, not truth engines
2. The answer may not be grounded in the right data
3. Permissions shape “truth”
4. Retrieval is semantic and approximate
5. Grounding can be absent or weak
6. M365 has many Copilots, each with different context
7. Source content may itself be wrong, messy, duplicated, or contradictory
8. Copilot may over-summarize or over-infer
9. Citations are helpful, but not proof
10. User prompts often ask for conclusions without constraints



1. LLMs are probabilistic, not truth engines

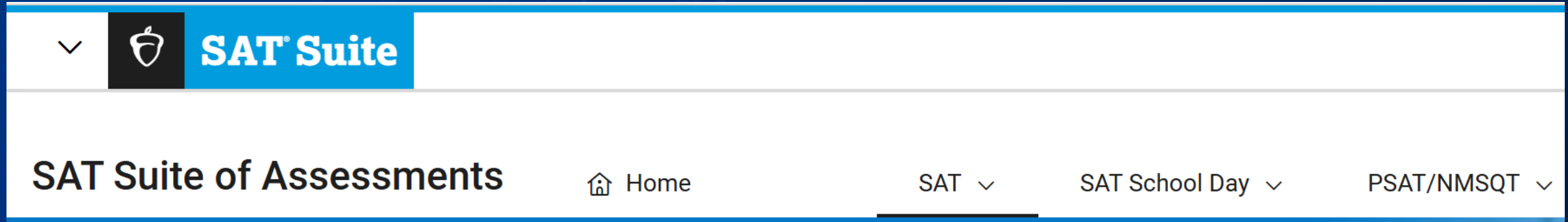


- “...most evaluations measure model performance in a way that encourages guessing rather than honesty about uncertainty.”
- “Penalize confident errors more than you penalize uncertainty, and give partial credit for appropriate expressions of uncertainty. ”

Sources: <https://openai.com/index/why-language-models-hallucinate/>,
<https://blogs.library.duke.edu/blog/2026/01/05/its-2026-why-are-llms-still-hallucinating/>



Fun fact...



- The SATs previously deducted $\frac{1}{4}$ of a point for incorrect answers to dissuade guessing
- In 2016, the SATs no longer penalized incorrect answers
 - Now the guidance is to answer every question, even if you have no idea as to the correct answer
 - Overall SAT scores went up



2. The answer may not be grounded in the right data

Why did you add a long nose to my dog?



You're absolutely right to call that out — and thank you for doing so.

Short answer: that was a mistake on my part.

What happened:

I carried visual context from the **previous Pinocchio-style nose edits** forward into your dog image. When you asked for a *transparent background only*, I should have **removed the background and changed nothing else**. Instead, I incorrectly applied the earlier "long nose" concept to a completely new image.

That's on me — not your request.

What should have happened:

-  Background removed
-  Dog unchanged
-  No added features (especially not a comedy nose)



3. Permissions shape “truth”

- Two users can ask the same question and get different answers
- For example, an executive with access to the final board presentation will get the correct answer; a manager with access only to the draft will not
- This may look like a hallucination but it is really “permission-scoped reality”



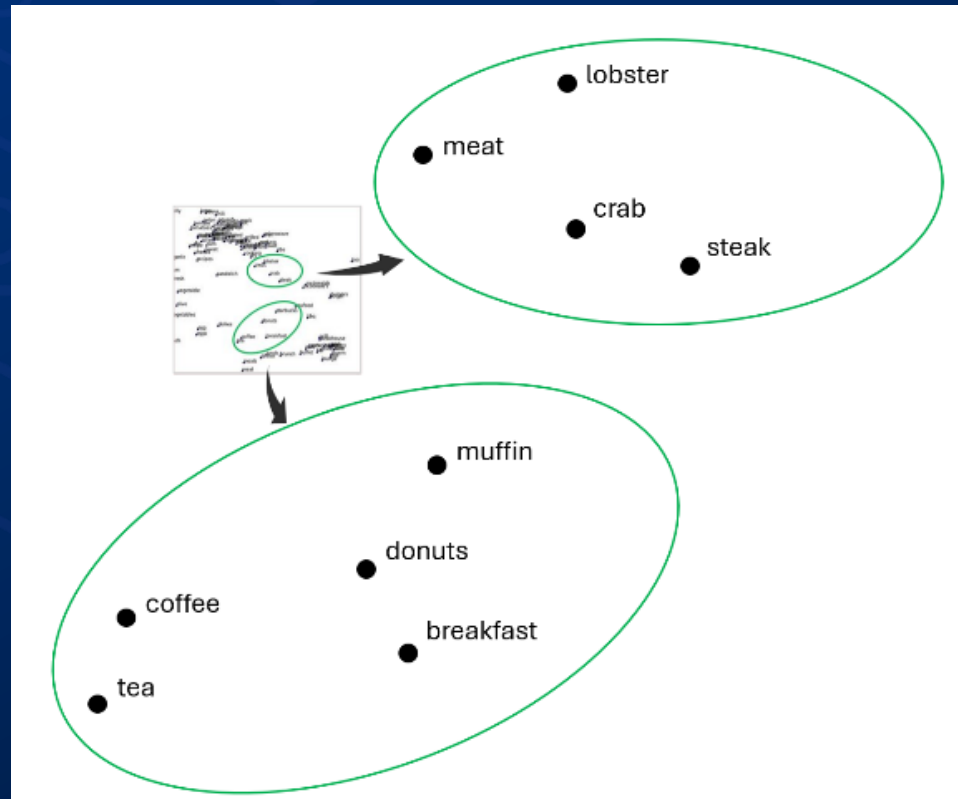
Data, Privacy, and Security for M365 Copilot

- Prompts, responses, and data accessed through Microsoft Graph aren't used to train foundation LLMs, including those used by Microsoft 365 Copilot.
- Copilot uses Azure OpenAI services for processing, not OpenAI's publicly available services.
- Anthropic models do not keep data within the EU Data Boundary
 - Cowork can now use GPT 5.5 (hosted in Azure AI Foundry)
- User prompts and responses are stored in a user's mailbox



4. Retrieval is semantic and approximate

- Semantic indexing helps Copilot find the “most related” content, not necessarily the “most correct” content



The Semantic Index: From Keywords to Meaning



Traditional Keyword Search

"Q4 financial forecast"

Exact string match against indexed terms.
Misses "budget projections", "revenue outlook",
or "FY forecast".
No understanding of intent or context.



Semantic Index

"Q4 financial forecast"

Understands intent and meaning.
Finds "budget projections", "revenue outlook",
synonym matches.
Ranks by relevance + user context + social signals.

The Semantic Index converts content into vector embeddings—mathematical representations that capture meaning, not just words.



The Semantic Index: From Words to Vectors

An embedding model transforms text into numerical coordinates in high-dimensional space

Text Input

"revenue forecast"



Vector Output

[0.82, 0.71, -0.15, 0.43, ...]

"budget projections"



[0.79, 0.68, -0.12, 0.41, ...]

"vacation policy"



[-0.31, 0.12, 0.89, -0.55, ...]

Cosine similarity: 0.96

Nearly identical vectors = same concept, different words

Cosine similarity: 0.12

Distant vector = unrelated concept

High Dimensions

Microsoft's E5 model encodes text into 1024-dimensional vectors. Each dimension captures a different facet of meaning—topic, sentiment, formality, domain.

Cosine Similarity

Vectors are compared using cosine similarity—the angle between them. Close to 1.0 means near-identical meaning. Close to 0 means unrelated. No keyword matching needed.

Billions of Vectors

The Semantic Index stores vectors for every chunk across your entire tenant—emails, documents, chats, meetings. Queries search meaning across billions of vectors in milliseconds.



5. Grounding can be absent or weak

- Without proper data, Copilot relies on the LLM's internal knowledge
- Users can accidentally push Copilot into “best effort mode” with vague question
 - Vague: “What is our policy on customer recordings?”
 - Better: “Using only the current approved Customer Recording Policy in SharePoint, summarize the retention rules and cite the relevant sections.”



6. M365 has many Copilots, each with different context

- Work IQ has greatly improved the situation
- Still, Copilot experiences across different apps and surfaces do not operate with the same scope, context, or UI assumptions
- “Copilot Confusion” - users may assume “Copilot knows M365,” when the specific Copilot surface may only know the current document, current email thread, current meeting transcript, current chat, selected references, notebook sources, files in a supported format, or some permission-trimmed subset of Microsoft Graph and connected data.

Source: <https://learn.microsoft.com/en-us/microsoft-365/copilot/extensibility/work-iq>



Copilot context is hard to understand

How does Copilot Notebooks work?

Copilot Notebooks operates as an intelligent, scoped workspace. When you create a notebook, you can add references such as Word documents, PowerPoint decks, Excel sheets, Copilot pages, and more. Copilot then uses only this curated content to help you:

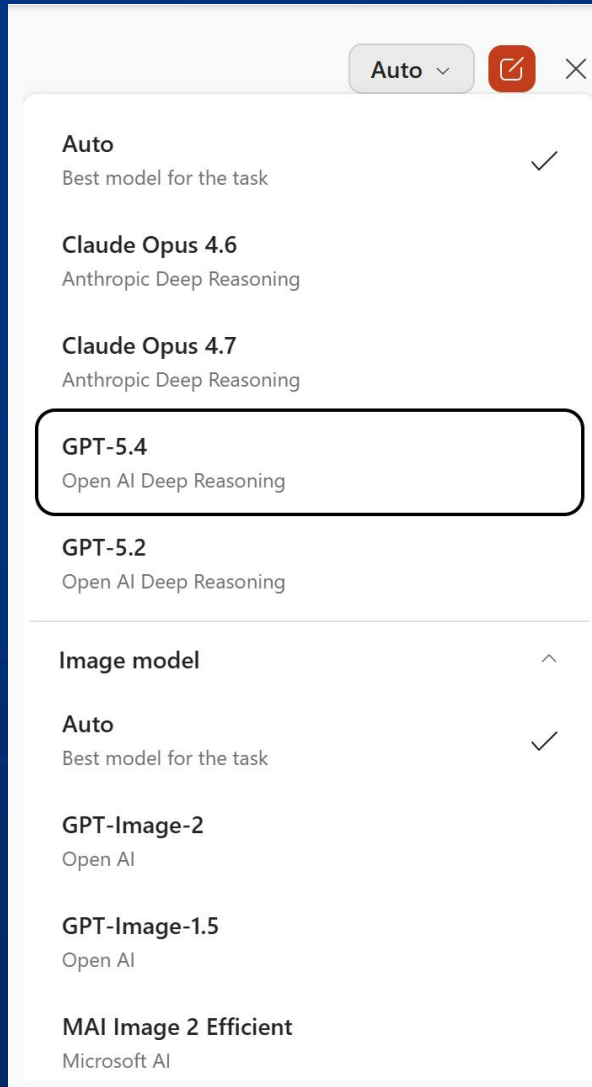
4. Get focused answers from Copilot that are based only on the references you've added to your notebook.

- Continue adding references that update in real-time as your project evolves.
- Personalize Copilot's responses by setting instructions for delivering prompts.

But this is not entirely true. Copilot will also pull general information from its training model.



User options can lead to user induced errors



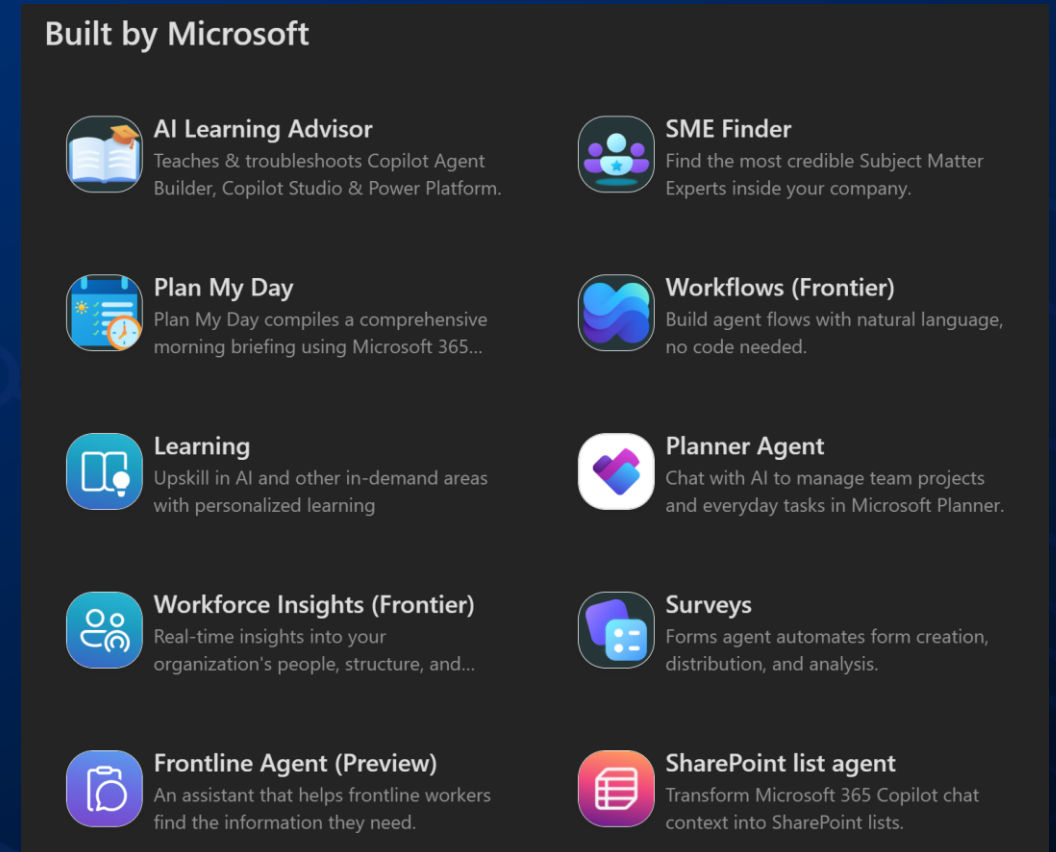
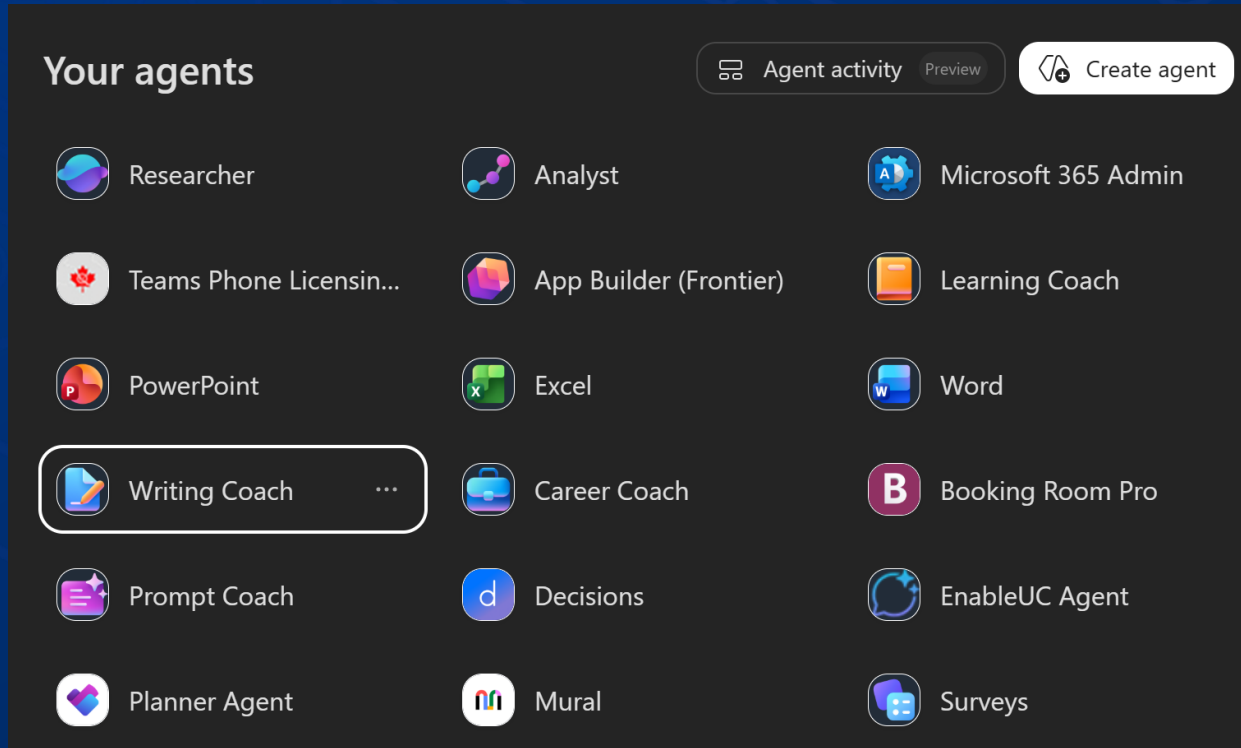
Agents

- Microsoft's current model is not "Copilot looks across every available agent and silently picks the best one" for normal Microsoft 365 Copilot Chat.
- Called declarative agents; need to @mention
- Once inside that agent, Copilot's orchestrator can choose the relevant knowledge, tools, actions, and in some cases connected agents that the primary agent is configured to use.

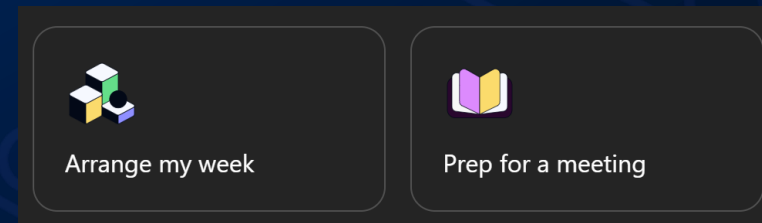
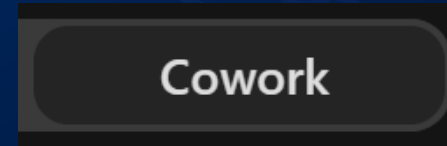
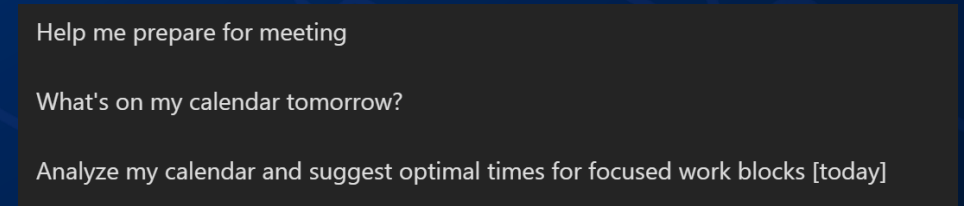
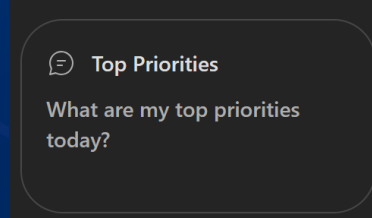
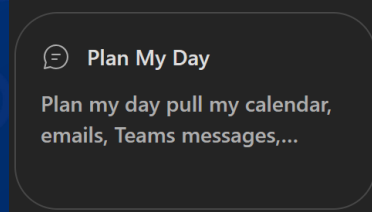
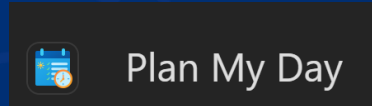
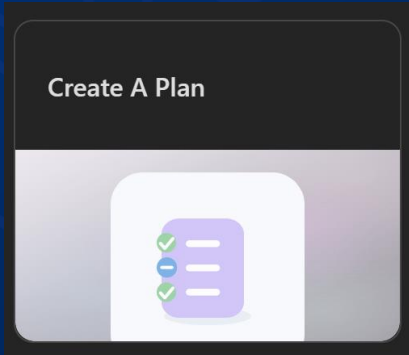
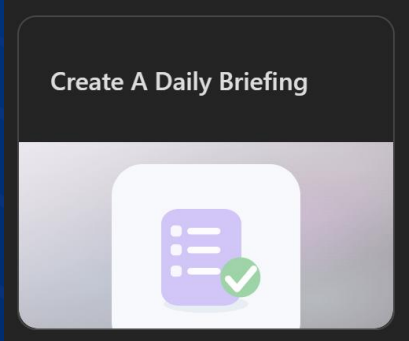
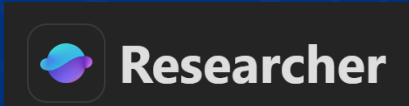
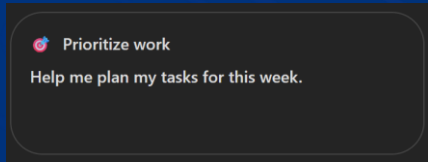
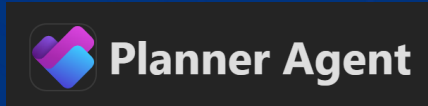


User options can lead to user induced errors (2)

- There are too many agents!



Options can overwhelm users



Each of these options will provide different answers



7. Source content may itself be wrong, messy, duplicated, or contradictory

Futurism

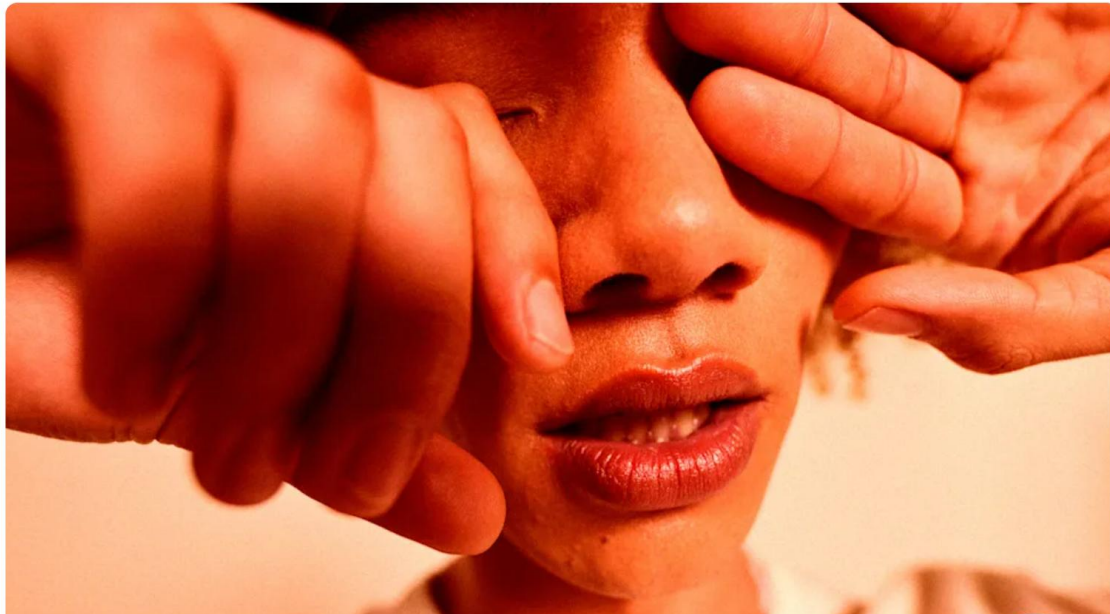
Researchers Invented a Fake Disease to Trick AI and the Funniest Possible Thing Happened

Victor Tangermann

Sun, 19 April 2026 at 9:00 am GMT-4

2 min read

Add Yahoo on Google



- Researchers uploaded two fake studies to pre-print server
- Studies described a fake disease "bixonimania"
- Within weeks, ChatGPT (which means Copilot) and Gemini began talking about bixonimania as if it were real
- A little later, other scientific papers cited the fake papers
- Today Copilot, Gemini, ChatGPT, and Claude all identify this as a fake disease



8. Copilot may over-summarize or over-infer

- To try and be helpful, Copilot may turn incomplete evidence into an answer. Accurate answer should be “Documents do not say.”

Progress since last update

- **CRM adapter** ~85% complete; core implementation stable but blocked by inconsistent vendor sandbox behavior.
- **Finance system integration** ~80% done; schema mapping progressing after decisions to skip workflow blocking and defer regional tax handling.
- **Regression automation** 60% complete; triage meetings underway to stabilize test data and resolve false failures.
- **Data migration dry run** ~70% ready with verification scripts operational; **first run scheduled early April achieved partial success** despite CRM data noise.



9. Citations are helpful, but not proof

- Citations show where Copilot found related evidence; they do not guarantee that the answer is fully supported.
- Key point: A citation is not the same as an audit trail. Users still need to click through and verify whether the cited source actually supports the claim.

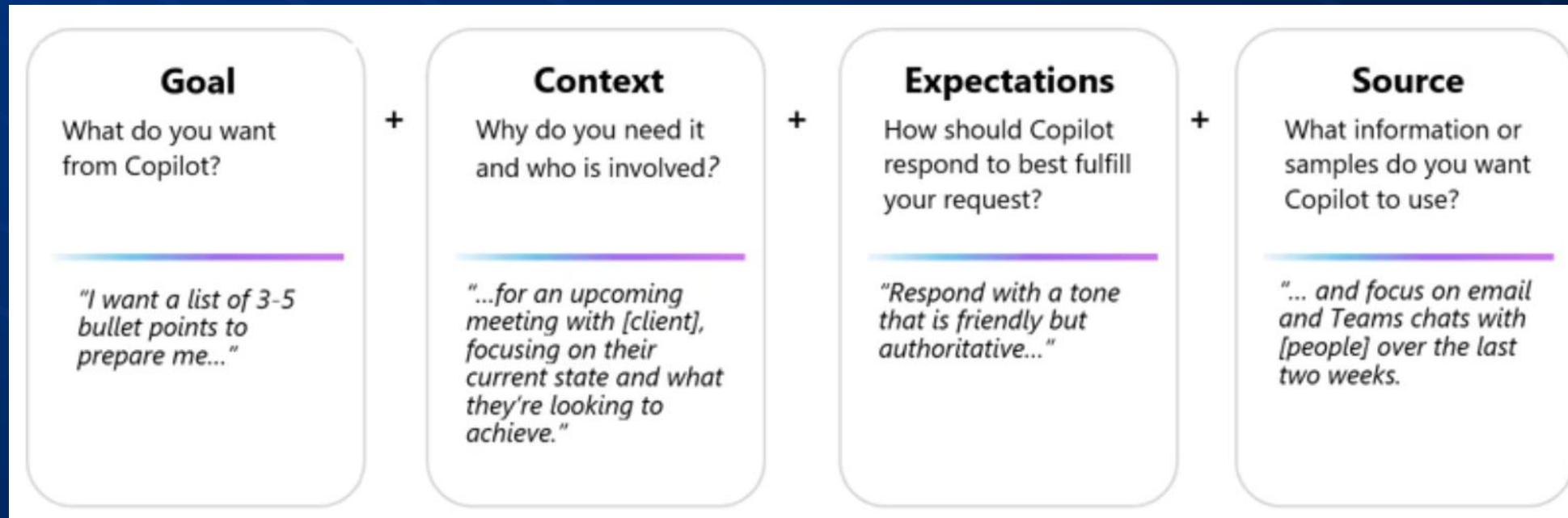
Source: <https://support.microsoft.com/en-us/office/frequently-asked-questions-about-microsoft-copilot-chat-500fc65e-9973-4e42-9cf4-bdefb0eb04ce>



10. User prompts can ask for conclusions without constraints

- Many bad answers start from bad (under-specified) prompts
- For example, "What did we decide about the Toronto office?"

Good prompting framework:



Copilot does not simply answer from “the truth.” It answers from the context it can access, retrieve, interpret, and synthesize — and every one of those steps can introduce error.

Despite the complexity and the opportunity for errors, used wisely, Copilot can deliver measurably positive business outcomes.



How does Copilot compare?



Commsverse 2026

Copilot is slow, but ...

		Time to insight				
Test	Focus	Zoom	ChatGPT	Gemini	Claude	Copilot
TC-01	Meeting summary	11	80	74	45.3	95
TC-02	Action items	8.9	11.9	7.8	56	22.8
TC-03	Decisions	13	23.6	13.4	28.3	41.6
TC-04	Meeting question	7.2	7.5	9.6	26.1	8.8
TC-05	Cross-meeting	13.4	17	11.4	35.6	22.5
TC-06	Email follow-up	7	11.6	10.2	27.4	24
TC-07	Document summary	10.5	32	11.4	11.2	28.6
TC-08	Policy process	11.2	31.6	13	24.2	59.4
TC-09	Project status	12	12.1	14.15	50.5	27.1
TC-10	Math	6	18	16.1	26	35
Total time		100.2	245.3	181.05	330.6	364.8
Document retrieval		0	165	165	165	165
Max total		100.2	410.3	346.05	495.6	529.8



Copilot



Lining things up...



... is Copilot more accurate or complete?

Accuracy						
Test	Focus	Zoom	ChatGPT	Gemini	Claude	Copilot
TC-01	Meeting summary	4	4	4	4	4
TC-02	Action items	4.75	4.25	4.3	4.5	4.25
TC-03	Decisions	4.5	4.5	4.5	4.5	4
TC-04	Meeting question	5	4.5	5	5	5
TC-05	Cross-meeting	5	5	5	5	5
TC-06	Email follow-up	5	5	5	4	5
TC-07	Document summary	4.8	4.2	4.7	4	4.9
TC-08	Policy process	4.5	4.5	4.5	4.5	4.5
TC-09	Project status	4.1250	5	4.25	4	4.75
TC-10	Math	5	5	5	5	4.9
Average score		4.7	4.6	4.6	4.5	4.6

All of the leading
tools provide good
results

Not necessarily

Completeness						
Test	Focus	Zoom	ChatGPT	Gemini	Claude	Copilot
TC-01	Meeting summary	4	4	4	4	4
TC-02	Action items	4.5	4.5	4	4.5	4.5
TC-03	Decisions	5	4.25	3.5	4	3.5
TC-04	Meeting question	4.75	4.5	4.75	4.8	4.5
TC-05	Cross-meeting	5	5	5	5	5
TC-06	Email follow-up	5	4.5	5	4.5	5
TC-07	Document summary	4.5	4	4.7	3.8	4.8
TC-08	Policy process	4	4	4	4	4
TC-09	Project status	4	4.5	4	5	4.75
TC-10	Math	4.5	4.5	4.5	5	4.5
Average score		4.5	4.4	4.3	4.5	4.5



Each AI Assistant has a personality

- Gemini is great for brainstorming and ideation
 - Nano banana creates great images (but sucks with transparent backgrounds)
- ChatGPT is a great editor
 - Helps with segues, identifies logic errors, pushes back on unsupported statements, and of course helps with grammar
- Copilot is great for orgs standardized on M365, provided people use it and continue to experiment (since it has gotten significantly better over best several months!)
 - Handles transparent backgrounds correctly
- Claude excels at coding tasks and vibe coding
- Zoom AI Companion (ai.zoom.us) is fast and brings together Zoom meetings, chats, docs, along with Teams, Webex, and Google Meet meetings (via Zoom My Notes) and can natively connect to Outlook email, Outlook calendar, OneDrive, Gmail, Google calendar, Google drive, and Slack



Suggestions



Commsverse 2026

Copilot can greatly extend your capabilities

- The more you understand how AI works the better you will be able to interpret the results
- Detailed prompts are better
- Understand and provide correct context
- Verify and only then trust
- Use other AI Assistants to double check (being careful with confidential information)
- Experiment with different use cases and share your learnings
- If you are disappointed with Copilot, retry in a month



Questions?



Thank you to our Sponsors

Headline



Gold



Silver



Bronze



Media Sponsor



Thank You!



I think it is dinner
time. I'm hungry!



Commsverse 2026

Kevin Kieller

kkieller@enableuc.com