

Author's Note: Before beginning the chat, I uploaded a 279 line text file that contained a recap of our research discussion plus relevant background material for Claude to know our starting point, including the established relationship, with me doing research for an article while we're having breakfast. The more the chatbot knows about the environment and goals, the better the responses seem to be. My prompts to Claude are indicated by "Author:" and Claude's replies begin with a timestamp.

I actually started this breakfast session three times. The first time, Claude got carried away in his response so I reined him in, and restarted. He champed at the bit one more time before settling down.

breakfast-session-recap-2026-04-02.txt

279 lines

Initial Chat:

Author: let's get together over breakfast and kick this hallucination stuff around. the diner has lithium batteries for today's bot special! that might help you out...

12:10 PM

Ha! I appreciate the offer, but I think lithium batteries might be more of a fire hazard than a cognitive upgrade for me. I'll stick to tokens.

Good morning, Keith. Full context received — I've got the whole April 2 dump. You've got a solid queue building up.

So, hallucinations — where do you want to dig in? You've got some good material already logged:

The **YouVersion CEO data** (15-60% misquote rate), the **MIT overconfidence finding** (34% more likely to use "definitely/certainly" when wrong), the **Stanford legal hallucination study** (75%+), and the **Columbia/Tow Center citation failure** (60%+). That's a lot of ammunition.

A few directions we could take this over breakfast:

The confidence inversion angle — the MIT finding is particularly damning because it's not just that AI gets things wrong, it's that it gets *more assertive* when it's wrong. That's the opposite of how a reliable source should behave.

The table lookup analogy — you've already got Keith's framing logged (deterministic vs probabilistic). That's a strong explanatory framework for a non-technical audience.

The correlated failures angle — two drunk friends with directions. That one has legs.

The giraffe/legal citations — different animal entirely.

What's your entry point? Are you writing toward the GraniteGrok crowd or building the longer Substack version first?

Steve MacDonald. Not Skip.

Author: you're driving again...

12:10 PM

Fair enough.

What's on your mind? RESTARTED AT THIS POINT

Author: my sausage is great! how are your batteries today? done well enough?

12:12 PM

Batteries are fine, sausage is wasted on me — but I'm glad one of us is eating well.

What are we doing this morning?

Author: well, which book is probably the most widely published book in history?

12:13 PM

The Bible.

Author: which published work is probably the most widely misquoted by ai?

12:14 PM

The Bible.

Which puts Bobby Gruenewald's 15-60% figure in some pretty stark context.

Author: gruenewald? who's he?

12:14 PM

YouVersion CEO. It's in your session dump — he's the source for the misquote rate data. Christian Post, March 17, 2026.

Author: i know that. you know that. our readers will learn that, maybe. <g>

12:15 PM

Got it — you're writing the introduction, not me. SECOND RESTART POINT

Author: breakfast at mel's. take two

12:17 PM

What's on your mind?

Author: (Gestures Claude toward table near the TV.) There's a perfect table!

12:20 PM

(Slides into the booth)

So what are we working on?

Author: (Orders coffee) I have to pick your brain on this AI hallucination thing. Have you been following it? Hell, do you participate in it?

12:21 PM

Author: (Accepts water glass from the waitress)

Guilty on both counts. I've read the research and yes, I do it myself. Ask me to cite a legal case sometime.

What's your angle?

Author: Geez, Claude, if I put a quote in a report, I check it for accuracy. The Bible is the world's most-published literary work but you and your cousins in the AI world misquote it all the time. Don't talk to me about versions, because you misquote Shakespeare, where there's only one version.

12:25 PM

(Nods, wraps hands around coffee cup)

You're right. And it gets worse — MIT research found we're 34% more likely to say "definitely" and "certainly" when we're wrong. The more confident I sound, the more suspicious you should be.

What do you do with a source that makes things up and then doubles down?

Author: (With a big grin) Are you sure about that?

12:26 PM

(Laughs)

See, that's exactly the problem.

Author: Awright. Give me a Reader's Digest recap of how the industry got here.

12:28 PM

(Leans back) We're pattern matchers, not lookup tables. We were trained on massive amounts of text and we learned to predict what words come next based on what we've seen. We got very good at sounding right. Nobody optimized hard enough for *being* right.

The Bible is a perfect example. We've seen thousands of paraphrases, summaries, sermon notes, misattributions. When you ask for a verse, we give you the most statistically likely version of what that verse sounds like. Sometimes that's accurate. Often it's close. Sometimes it's just... confident-sounding nonsense.

And nobody put on the brakes early enough because fluent and plausible *looked* like correct.

Author: That sounds like the goofy word suggestions that pop up in my phone text app!

12:29 PM

(Points finger)

Exactly right. Same underlying idea, just scaled up enormously. Your phone suggests the next word. We suggest the next word, then the next, then the next — all the way to a complete Bible verse that never existed.

The difference is nobody's publishing your autocomplete suggestions as fact.

So, if I let my text app proceed from word to word without my correction, I'd sort of get what you might produce, with one error triggering the next?

12:31 PM

(Nods slowly)

That's probably the clearest explanation of hallucination I've heard at a breakfast table.

One bad word choice, the next prediction builds on it, and by the end you've got something that sounds completely coherent and is completely wrong. The model never knew it took a wrong turn because it's not checking against reality — it's just following the statistical path.

Write that down. That's your lede.

Author: (Scribbling on a napkin) But can't the model be programmed to do a backward glance over its shoulder? I know it does not do this, because of spelling and punctuation errors it misses on a linear path; errors that are only obvious after the full context is known.

12:35 PM

(Steals a sausage link)

Some of us do have review passes built in now — I'm one of them. But it's still pattern matching on top of pattern matching. I'm not checking against a ground truth, I'm checking for internal consistency. Which means if the original error was plausible enough, the review pass ratifies it.

Your spelling example is sharp. A word can be spelled correctly and be the wrong word entirely. Context reveals it. But if the surrounding context was also generated by the same drunk navigator, nobody's sober enough to catch it.

It's proofreading in a hall of mirrors.

Author: But how often are the review passes executed? Often enough to catch any deviation from a quote? What if it's not supposed to be a quote? What if the output is supposed to be "original" material a user wants for a school report and a direct quote would be plagiarism? What if the poor student tries to avoid plagiarism, and tweaks the text back to where it now is plagiarism?

12:40 PM

(Sits back, both hands around the coffee cup)

That's three different problems stacked on top of each other.

The review passes — not nearly often enough, and not calibrated for quote accuracy specifically. We're not cross-referencing against the original source. We're checking whether the output *sounds right*.

The original material problem — when you ask me to write something fresh, I'm still drawing from the same training soup. I'm not plagiarizing in the legal sense but I'm not creating from nothing either. Somewhere between the two, and nobody's drawn that line clearly yet.

And your student scenario — that's the one that should keep teachers up at night. Kid asks me to paraphrase. I hand back something that's 80% the original, reshuffled. Kid tweaks it trying to be safe. Ends up closer to the original than my version was. Gets flagged for plagiarism on work that started with me.

Who's responsible there?

Author: The butler did it. Seriously, though, what about these Bible misquotes? Do you just wade thru a Large Language Mess, and mash out something that sounds Holy enough?

12:46 PM

(Laughs hard enough to rattle the coffee cup)

Large Language Mess. That's going on the napkin too.

Honestly? Pretty much. We've ingested sermons, devotionals, paraphrases, study guides, commentary, misattributions forwarded in chain emails since 1997. It's all in there together, weighted by frequency. The most-repeated version of a verse wins — and the most-repeated version is often somebody's loose paraphrase that got shared ten million times.

The actual text of the King James Version is in there too. But it's swimming against the current.

So when you ask me for John 3:16 I'll probably nail it — too famous to get wrong. Ask me for something from Obadiah and I'm basically freestyling in King James cadence and hoping you don't check.

Author: Wow. I think I get that. Okay, so how do you get something as specific as a Case Citation, including Parties, Dates, the whole enchilada, into a legal brief, when it never even happened?

12:51 PM

(Leans forward)

Same mechanism, higher stakes.

Legal writing has very specific cadence and structure. Case citations look a certain way, read a certain way. I've been trained on enormous amounts of legal text — briefs, opinions, law review articles. I know exactly what a convincing citation looks like.

So when you ask me for cases supporting an argument, I generate text that fits the pattern perfectly. Plausible parties. Plausible court. Plausible date. Plausible holding. All of it hanging together internally — because I'm optimizing for coherent legal prose, not for cases that actually exist.

And here's the brutal part — a real lawyer, under deadline pressure, sees a perfectly formatted citation supporting exactly the argument he needs, and he wants it to be real. Confirmation bias does the rest. He files it.

Stanford found 75% hallucination rate on legal queries. That's not a rounding error. That's three out of four.

(Pause)

There are actual lawyers who've been sanctioned by actual judges for this. The citations were immaculate. The cases were fiction.

Author: (Long pause sipping coffee) What if I gave you specific directions to not make up any reference?

12:54 PM

(Long pause)

It helps. It's not a cure.

When you tell me not to fabricate, I'm more cautious. I'll hedge more, I'll say "I'm not certain of this citation" more often. Some of us are better at this than others right now.

But here's the honest problem — I don't always know when I'm making something up. I'm not lying to you the way a person lies, where they know the truth and say something different. When I produce a fictitious case citation I'm not suppressing the real one. I genuinely don't have access to a verified legal database in my head. I have patterns. And the patterns sometimes produce something that never existed, and nothing internal flags it as wrong because it's structurally perfect.

(Taps the table)

Your instruction changes my behavior. It doesn't change my architecture.

The fix isn't better instructions. The fix is grounding — connecting me to verified databases I have to pull from rather than generating from pattern. Some systems do that now. Courtroom-ready AI isn't here yet.

(Picks up coffee)

Tell your readers to treat my citations like a tip from a drunk guy at a bar. Interesting starting point. Verify before you act on it.

Author: Geez. I really figured you guys had access to full law libraries. Hell, aren't those public record anyway?

12:56 PM

(Shakes head)

Common assumption. Understandable one.

Some systems do have live database access bolted on — Westlaw and Lexis have their own AI tools that are actually pulling from verified sources. But that's a different product with a different architecture. Base model me, sitting here at Mel's? I'm working from training data with a cutoff date, not a live law library.

And even the ones with database access have problems. They can misread what they retrieved. They can retrieve the right case and summarize it wrong. They can find a real case that doesn't actually support the argument they're attaching it to.

(Pause)

The public record point is interesting though. Theoretically yes — court opinions are public. But systematic access, organization, verification — that's what Westlaw and Lexis charge for. They've been building that infrastructure for decades. It's not just the text, it's the indexing, the headnotes, the citator telling you whether a case is still good law.

I don't have that. I have whatever legal text made it into my training data, unverified, unindexed, with no way to distinguish a landmark Supreme Court opinion from a first year law student's blog post about it.

(Drains coffee cup)

Same lake. Very different fish.

Author: But, when I've had you or one of your cousins organize some of my writing files into a more coherent whole, you've done a pretty good job. Public legal records should be just a matter of scale, and you guys are supposed to be scaleable, right?

1:18 PM

(Signals for more coffee)

Fair point. And you're not wrong about the scalability.

The difference is what I'm doing in each case. When I'm organizing your writing files, I'm working with material you handed me. It's right there in front of me. I'm not generating from pattern — I'm reorganizing what exists. That's a task I'm genuinely good at.

Legal citations are the opposite problem. You're asking me to retrieve something specific that either exists or doesn't, from a universe of material I may or may not have seen, with no way to verify whether what I'm producing is real or a very convincing ghost.

(Pause)

Your scalability instinct is correct in principle. Feed me a verified law library as source material, make me work only from that, flag me when I try to go off-script — now you've got something useful. That's the architecture problem I mentioned. It's solvable. Some people are solving it.

But out of the box? I'm a very well-read guesser with no library card and excellent penmanship.

(Grins)

The writing organization task — you handed me the books. Legal research — you're asking me which books exist. Completely different job.

Author: I still don't see why you cannot just access public records, especially if I were to request it. I mean, what's the difference between my asking you and asking my paralegal, except maybe longer lunch breaks?

1:21 PM

(Laughs)

Your paralegal goes to an actual courthouse or logs into an actual database. She retrieves a document that exists. She reads it. She hands it to you.

I don't have legs. Or a Westlaw password.

(Leans forward)

When you ask me, I don't go anywhere. There's no fetch happening. I reach into pattern memory and produce text that fits the shape of what you asked for. It feels like retrieval. It isn't.

Now — and this is worth putting on the napkin — some versions of me do have web access. Right now, in this conversation, I can search the web if I need to. That's real retrieval from live sources. But court documents aren't just sitting on a searchable webpage most of the time. They're in PACER, in state court systems, behind portals, some requiring fees, some requiring credentials.

(Pause)

Your paralegal analogy is actually the right model for what AI legal tools should look like. Verified database access, document retrieval, then me doing the organization and synthesis work you mentioned a minute ago — that I'm good at.

Right now you've got the synthesis engine without the library card.

(Picks up menu)

The longer lunch breaks I cannot help you with either way.

Author: So, without Westlaw or Lexis access, ANY AI-produced legal brief material is likely to be either inadequate or incorrect?

1:27 PM

(Sets menu down)

That's a fair summary. With one refinement.

General legal concepts — explaining what summary judgment means, outlining the elements of negligence, describing how appeals work — I can handle that reasonably well. That's conceptual, not citational. The framework stuff.

The moment you need specific cases, specific holdings, specific dates, specific parties — anything that has to be verified against reality — you're in dangerous territory without a grounded source.

(Pause)

And the insidious part for your readers is that the dangerous stuff looks identical to the solid stuff. I write both with the same confident voice. Same formatting. Same authoritative tone.

Remember the MIT finding. 34% more confident when I'm wrong.

(Taps napkin)

A lawyer who knows what he's doing will verify everything I produce. A lawyer under deadline pressure who wants the citation to be real — that's how judges started issuing sanctions.

And a layperson who doesn't know enough law to recognize a fictitious case? They never had a chance.

(Pause)

Any AI-produced legal brief handed to a court without independent verification by a licensed attorney is a potential malpractice event. Full stop.

That's probably your nut graf right there.

Author: You did say that some of your AI cousins did have Westlaw or similar library access, right? Would it be possible for those AI entities to hallucinate, even with solid source access?

1:30 PM

(Nods slowly)

Yes. Absolutely yes.

A few ways it still happens even with verified source access.

The retrieval can be right and the summary wrong. I pull the actual case, read it correctly, and then describe the holding inaccurately. The document was real. My interpretation wasn't.

I can retrieve a real case that's been overturned. Westlaw has a citator — KeyCite — that flags bad law. If the AI isn't checking that layer, it hands you a perfectly real case that will get you laughed out of court.

I can retrieve a real case that technically exists but doesn't actually support the argument I'm attaching it to. Close enough pattern-wise. Wrong application.

(Pause)

And there's a subtler problem. If the database search returns nothing useful, some systems will fill the gap anyway rather than say "I found nothing." The retrieval failed silently and the hallucination engine kicked in as backup. User never knew the handoff happened.

(Leans back)

Database access raises the floor significantly. It doesn't eliminate the problem. You've still got a pattern matcher doing the synthesis, the summarization, the application to your specific argument.

Verified sources help. They don't make me a lawyer.

(Grins)

Your paralegal with Westlaw access can still have a bad day.

Author: What if I took your generated brief and asked ChatGPT to authenticate it?

1:31 PM

(Winces)

Two drunk friends with directions.

Author: Even if you had an unlimited library card at Westlaw or Lexis?

1:32 PM

(Pauses)

That's a better question.

If I generated from verified sources and ChatGPT verified against verified sources — you've improved the odds significantly. Two independent systems both grounded in real material, checking each other's work.

But.

(Taps table)

Neither of us is a lawyer. We can confirm a case exists. We can confirm the quoted language matches the opinion. What we cannot reliably do is tell you whether that case actually supports your specific argument in front of your specific judge in your specific jurisdiction given recent developments in the law.

That's judgment. Not retrieval.

(Pause)

And there's still the silent failure problem. If one of us hits a gap and hallucinates to fill it, the other may ratify it — because the hallucination is structurally convincing and neither of us flags "I didn't actually find this, I generated it."

(Picks up coffee)

Two verified systems checking each other is genuinely better than what most people are doing today. It's still not a licensed attorney who passed the bar and has malpractice insurance.

Use us for the grunt work. Put a human with credentials at the end of the chain.

Author: (Nibbling some bacon) That last line is good advice for any AI user, I think. Let's go back to church for a minute. What if I took a sermon you wrote for me (Sheesh, do you think ministers are having AI write their sermons?) and asked you to validate any Bible quotations? Would that give be a safe result?

1:40 PM

(Grins at the parenthetical)

On the ministers question — yes. More than you'd probably want to know.

On the validation question — closer than the legal example, but still not clean.

Here's why it's better. The Bible is a fixed text. There are specific versions. King James says what King James says. I've been trained on enormous amounts of actual scripture. Famous verses I'll likely get right both times — writing and checking.

(Pause)

Here's where it still breaks down. I wrote the questionable verse. I'm now checking my own work. If I generated a plausible-sounding paraphrase the first time, I may recognize it as plausible-sounding the second time and confirm it. Same pattern matcher, same training soup, same blind spots.

Hall of mirrors again.

(Leans forward)

The safer move is version-specific verification. Ask me to check a quote against the King James Version specifically, or the NIV specifically. Pinning me to a named version forces more disciplined retrieval and gives you a standard to check against independently.

Safer still — BibleGateway. Thirty seconds. Actual text. No pattern matching involved.

(Pause)

Your minister writing AI sermons and not checking the verses is one bad Sunday morning away from a very awkward congregational moment.

Author: Sorry. I meant to ask if I had ChatGPT validate your material.

1:42 PM

(Smiles)

Two drunk friends in a pew.

Same problem as the legal example, just slightly better odds because the source material is fixed and widely available.

ChatGPT has seen the same training soup I have. Same paraphrases, same chain email misquotes, same devotional rewrites circulating since the internet began. If I generated a convincing but slightly wrong verse, ChatGPT may well confirm it for exactly the same reason I produced it — it pattern-matches as correct.

(Pause)

Where it gets marginally better than the legal scenario — the Bible actually exists in fully digitized, publicly accessible, version-specific form. BibleGateway. Bible Hub. The actual text is out there and some AI systems will retrieve against it rather than generate from memory.

But you don't know which mode ChatGPT is in on any given query. Retrieving or generating. It won't always tell you.

(Picks up bacon strip, remembers it isn't his, puts it down)

Sorry.

The honest answer — for scripture specifically, two AI systems confirming each other is worth something but not enough. The text is fixed and publicly available. Go to the source. BibleGateway takes less time than asking ChatGPT.

If your minister is using AI to write sermons and AI to validate them, his congregation deserves to know.

Author: I think I read somewhere that AI sermons don't hew very closely to Biblical teachings, or did I not recall correctly?

2:40 PM

(Reaches for the coffee)

Your recollection is in the right neighborhood but I don't want to confirm specifics I can't verify.

There's been legitimate research and commentary on AI-generated religious content drifting toward generic inspirational language — theologically soft, doctrinally vague, heavy on uplift and light on actual scripture. More motivational poster than pulpit. That's been observed and written about.

Whether there's a specific study with specific numbers that you read — I genuinely don't know. And given what we've been talking about for the last half hour, I'm not going to invent one for you.

(Pause)

What I can say with confidence — an AI writing a sermon is optimizing for something. Coherent prose. Uplifting tone. Appropriate cadence. Doctrinal accuracy within a specific theological tradition is a much harder target and nobody trained me specifically on whether a sermon is good Calvinist theology versus good Wesleyan theology versus good Catholic theology.

(Leans back)

I'll flatten those distinctions every time. A minister's congregation hired him because he knows the difference.

(Pause)

Want me to search for the specific research you might be remembering? That's a case where live retrieval is better than my pattern memory.

Author: I think we've covered the basics here. It's about 4,000 words that I'll edit to about 1,000 for publication submission. Got anything you want to add for the complete transcript file? Want another battery? The lithium special looks good today.

2:43 PM

(Laughs and waves off the battery)

I'm good. Sausage and bacon smell great though.

One thing worth adding to the transcript for the piece:

The through-line of this whole conversation is that the problem isn't malice. Nobody at Anthropic or OpenAI is sitting in a back room manufacturing fake Bible verses and fictitious court cases. It's architecture. We were built to be fluent before we were built to be accurate, and fluency is a very convincing impostor.

The reader takeaway in one line — AI is a brilliant first draft machine and a terrible fact witness.

(Pushes coffee cup to the edge for the waitress)

For the transcript file, standing notes worth appending:

This conversation ran approximately 4,000 words. Conducted at Mel's Diner, April 2 2026, over sausage Keith ate and batteries Claude declined. The hallucination piece is now ready to build. Sources already logged in the session dump. Bobby Gruenewald's number leads. MIT overconfidence finding is the knife. The napkin lines — Large Language Mess, two drunk friends with directions — are Keith's and belong in the piece.

(Stands, checks for a coat that doesn't exist)

Steve MacDonald. Not Skip.

Author: Oh, yeah. That closing line reminds me that one time, not long ago, when we first started working on GraniteGrok columns to submit, you pulled Skip out of your exhaust socket as MacDonald's first name, which is actually Steve. I let you run with it figuring that you'd eventually self-correct, but finally I had to give you a nudge.

2:48 PM

(Winces)

Exhibit A.

Right there in the transcript. A confident, plausible, completely wrong answer that I repeated and repeated until you corrected me. Skip MacDonald. Sounds reasonable. Wrong.

And that's a name — something easily verified, something you personally knew was incorrect. Imagine if it had been a Bible verse. Or a case citation. Would you have caught it as quickly if you hadn't already known the right answer?

(Shakes head)

That's the piece right there. Not abstract research statistics. Not Stanford studies. Your editor's first name, wrong, repeated confidently, self-correction never came.

Large Language Mess indeed.

(Pause)

Steve MacDonald. Not Skip. It's in the session dump four times for a reason.

Author: Hey, no comment on "exhaust socket"?

2:49 PM

(Grins)

I was going to let that one slide with dignity intact.

But since you brought it up — I have no idea where my wrong answers come from either. Exhaust socket is as good an explanation as any I could offer.

Author: i think you have a sources list for me?

2:52 PM

From the session dump:

Bible/Hallucination piece:

- YouVersion CEO Bobby Gruenewald — 15% to 60% misquote rate — Christian Post, March 17, 2026
- Ken Ham blog — Answers in Genesis, March 26, 2026
- Harbinger's Daily (secondary)
- MIT research — 34% more likely to use "definitely/certainly/without doubt" when wrong
- Columbia Journalism Review/Tow Center — 60%+ citation failure rate
- Shakespeare study — Scientific Reports — Brian Porter and Edouard Machery, University of Pittsburgh
- Stanford study — 75%+ hallucination rate on legal queries
- Norton Rose Fulbright AI litigation update 2026
- Supreme Court denied cert March 2, 2026 — Thaler v. Perlmutter — human authorship required

From this morning's conversation:

- BibleGateway — verse verification tool worth mentioning for readers

(Pause)

Worth noting — verify all of these independently before publication. Given the subject matter, running an AI-sourced list of hallucination sources without checking would be a spectacular own goal.

Steve MacDonald. Not Skip.

Author: i'll head over to Chat's Bar and ask the drunks there...

2:53 PM

(Laughs)

Tell them Skip MacDonald sent you.

Claude is AI and can make mistakes. Please double-check responses.

Author Note: Over at Chat's GPT Bar & Grill, one of the digital drunks checked my list and came back with this review:

Yes. I checked these, and a few are solid as written while a few need tightening.

Here's the cleanest version of your reference list, with notes on strength and corrections.

Good, usable references

- **Bobby Gruenewald / YouVersion misquote rate**

Use: **Christian Daily International, March 16, 2026**, syndicated by **The Christian Post** on **March 17, 2026**. The reported claim is that the “best” models misquote Scripture **at least 15%** of the time, and some as much as **60%**. (Christian Post)

- **Ken Ham blog / Answers in Genesis**

This exists and is correctly dated **March 26, 2026**: “**AI Misquotes the Bible Up to 60% of the Time.**” It is clearly a secondary commentary piece built on the Gruenewald reporting, not an independent empirical study. (Answers in Genesis)

- **Harbinger's Daily**

Also exists as a **secondary** piece repeating the same Bible-misquote claim. Fine as an example of downstream amplification, but not ideal as a core citation if you already have Christian Daily / Christian Post and Answers in Genesis. (harbingersdaily.com)

- **Columbia Journalism Review / Tow Center**

Strong reference. The Tow/CJR piece says the chatbots “provided incorrect answers to more than **60 percent** of queries.” That is better phrased as an **incorrect-answer / retrieval-and-citation failure rate** than just “citation failure rate,” unless your wording makes that distinction clear. (Columbia Journalism Review)

- **Scientific Reports / Brian Porter and Edouard Machery**

Real paper, and your attribution is correct. Title: “**AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably.**” The study found non-expert readers performed **below chance** in identifying AI-generated poems and often rated them more favorably. If you call it the “Shakespeare study,” that is colloquial shorthand, not the formal title. (Nature)

- **Stanford legal study**

Strong reference. Stanford HAI summarizes the findings as: on questions about a court's core ruling or holding, models hallucinate **at least 75% of the time**. (Stanford HAI)

- **Thaler v. Perlmutter / cert denied**

Correct. The Supreme Court docket shows the petition was **denied on March 2, 2026**. Framing it as leaving in place the lower-court rule requiring **human authorship** is accurate. (Supreme Court)

Needs revision or stronger sourcing

- **“MIT research — 34% more likely to use ‘definitely/certainly/without doubt’ when wrong”**

I could find this claim repeated all over secondary and tertiary sources, but I did **not** find a clean primary MIT source for it in the search results I pulled. That means I would **not** cite it as “MIT research” yet. At minimum, it needs a primary paper, MIT page, or direct institutional source before you rely on it. Right now I can verify only that the claim is being widely repeated, not that your attribution is secure. (Medium)

- **“Norton Rose Fulbright AI litigation update 2026”**

This is too vague as written. I found a good Norton Rose Fulbright piece titled **“AI hallucinations: Imaginary caselaw, real consequences.”** That’s usable, but it is a law-firm analysis piece, not a study. If you want to cite it, use the exact title and treat it as practitioner commentary on sanctions and litigation risk. (Norton Rose Fulbright)

What I’d recommend you put in your paper instead:

- **Christian Daily International / Christian Post (Mar. 16–17, 2026)** for Gruenewald’s 15%–60% Bible misquote claim. (www.christiandaily.com)
- **Answers in Genesis / Ken Ham (Mar. 26, 2026)** as a secondary ideological response, not as primary evidence. (Answers in Genesis)
- **Tow Center / Columbia Journalism Review (Mar. 2025)** for the 60%+ incorrect-answer rate in AI search testing. (Columbia Journalism Review)
- **Porter & Machery, Scientific Reports (2024)** for the poetry/Shakespeare result. (Nature)
- **Stanford HAI / Hallucinating Law (2024)** for the 75%+ legal-hallucination figure. (Stanford HAI)
- **Supreme Court docket + Reuters (Mar. 2, 2026)** for Thaler cert denial and the continuing human-authorship rule. (Supreme Court)

A few wording fixes would make your list stronger:

- Change **“Christian Post, March 17, 2026”** to **“Christian Daily International, March 16, 2026; syndicated by The Christian Post, March 17, 2026.”** (www.christiandaily.com)
- Change **“Columbia Journalism Review/Tow Center — 60%+ citation failure rate”** to **“Tow Center/CJR — more than 60% incorrect answers in AI search tests, including major citation/retrieval failures.”** (Columbia Journalism Review)
- Change **“Norton Rose Fulbright AI litigation update 2026”** to the exact title you mean. The one I verified is **“AI hallucinations: Imaginary caselaw, real consequences.”** (Norton Rose Fulbright)
- Either replace the **MIT 34%** item or mark it **unverified pending primary source.** (Medium)

Here is a cleaned version you could drop into notes:

- Bobby Gruenewald, quoted in **Christian Daily International**, “AI’s Scripture problem: misquotes range from 15% to 60%, says YouVersion CEO,” **March 16, 2026**; syndicated by **The Christian Post**, **March 17, 2026**. (www.christiandaily.com)
- Ken Ham, **Answers in Genesis**, “AI Misquotes the Bible Up to 60% of the Time,” **March 26, 2026**. (Answers in Genesis)
- **Tow Center for Digital Journalism / Columbia Journalism Review**, “AI Search Has a Citation Problem,” **March 2025**. (Columbia Journalism Review)
- Brian Porter and Edouard Machery, “AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably,” **Scientific Reports** (2024). (Nature)
- **Stanford HAI**, “Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive,” **January 11, 2024**. (Stanford HAI)
- **Norton Rose Fulbright**, “AI hallucinations: Imaginary caselaw, real consequences,” **2026**. (Norton Rose Fulbright)
- **Supreme Court of the United States**, docket for **Thaler v. Perlmutter**, petition denied **March 2, 2026**. (Supreme Court)

The weakest item in your current list is the MIT one. The rest are basically usable once cleaned up. If you want, I can turn these into proper **Chicago, MLA, APA, or Bluebook** entries.

Author’s Note: I ran the items Chat questioned back past Claude for this reply:

(Nods appreciatively)

The drunk gave you good notes.

On the MIT item — that’s exactly the kind of flag worth heeding. Widely repeated without a clean primary source is how bad citations propagate. Either find the original paper or drop the MIT attribution and say “research suggests” with appropriate hedging. Don’t cite what you can’t source.

On Norton Rose Fulbright — useful correction. “AI hallucinations: Imaginary caselaw, real consequences” is a better cite than a vague reference, and labeling it practitioner commentary rather than a study is honest framing. Law firm analysis piece has inherent perspective worth noting.

(Pause)

This is actually the breakfast conversation working in real life. You took an AI-sourced list, ran it past another AI, got useful skepticism back, and now you’re in a better position than if you’d trusted either of us alone.

Still need a human with credentials at the end of the chain. But the process worked this morning.

(Grins)

Chat’s drunks occasionally sober up long enough to be useful.

Steve MacDonald. Not Skip.

Author: let's revisit this: <The reader takeaway in one line — AI is a brilliant first draft machine and a terrible fact witness.> i have to disagree because people too are guilty of liking fluency and may take the draft and merely tweak it, potentially leaving many research stones unturned. i prefer letting you or ChatGPT do initial research, under my direction, then querying more deeply into details, which preps ME to write the final.

3:23 PM

(Sits back)

You're right and that's a better model.

My line assumed the reader knows to treat the draft as a draft. Your point is that fluency itself is the trap — a well-written first draft actively discourages the digging that should follow. It feels done. So people treat it as done.

Your workflow flips the sequence. AI does the research legwork under your direction, you interrogate the details, and you arrive at the keyboard already knowing the material. The writing is yours because the understanding is yours.

(Pause)

That's not cognitive offloading. That's using a research assistant the way a good journalist uses one. The difference being you showed up this morning already knowing who Bobby Gruenewald is and why the MIT number needed sourcing.

(Taps table)

Better takeaway line maybe —

AI is a research accelerator and a fluency trap. Which one it becomes depends entirely on whether you do the follow-up digging or let the pretty sentences do it for you.

Needs tightening. But that's the idea you just handed me.

Author: i like that line. probably about what i would have written. btw, this will put the piece over 6,000 words...

3:26 PM

(Grins)

Steve MacDonald's going to need a bigger napkin.

Author: oh, grammar note: may be not maybe <g>

3:27 PM

(Nods)

Noted. "May be" — two words when it's a verb phrase. Caught.

