

Where the Number Came From

How generative search loses the record, and what an operator can do about it in minutes per memo

Tommy N. Turner. Version 1.0, September 2026.

Six fractions for one fact

On 11 September 2026 the same question went to four generative search products: how many data centers are in Loudoun County, Virginia, what share of the county's local tax revenue do they provide, and cite your sources. The count comes up later. Start with the share. Loudoun is the right county to ask about because the bottom is reachable. The county publishes an adopted budget with the data-center tax lines broken out, so for any stated base the true answer is a matter of division, and every page a product cites can be followed to that document or shown to stop short of it.

The four answers reported, between them, thirty-nine percent, forty percent, forty-four point six percent, and a range from thirty-eight to forty-five. Among the pages one of them cited sat two more figures, forty-two and thirty-one. Every one of the six was true of some base in some year. Between them they were shares of the county's General Fund, which the county's own web page calls the total budget, of its local tax revenue, of an advocacy group's measure of local tax funding, and of the base a state legislative audit commission used for an earlier year, spread across three fiscal years. One product out of four said which fraction it was reporting. The citation rows under the answers, two sources, four, ten, and sixty-five, gave no sign that the numbers were answers to different questions, and the sixty-five, on inspection, counted search results across six searches, of which the answer had used five pages.¹

A reader with the right base in mind would have found their number among the six and taken it as confirmed. A reader with the wrong base in mind would have done the same. That is one question on one day, and it is offered here as a demonstration rather than a measurement: it shows what can happen, in products people use, on a figure that a county government has gone to the trouble of publishing exactly.

Origin and denominator	Page that carried it	Answer
44.6% / 45% Adopted FY2027 budget of local tax revenue	Magazine essay	ChatGPT 44.6%, denominator stated
40% County web page, FY2027 of the total budget	County web page	Gemini 39% and 40%
39% County web page, FY2026 of the overall budget	County web page	
38% FY2026 proposed budget of General Fund revenue	Newspaper report	Perplexity 39% and 40%
42% Coalition estimate, 2026 of local tax funding	Trade group post	Grok 38% to 45%
31% State audit, 2024 of local revenue	Trade group post	

Every figure is true of its own denominator. One answer in four said which denominator it used.
Solid red: the county record. Dashed grey: an advocacy estimate. Lines follow the cited chain, not the interface.

The word for what it demonstrates is lineage. A citation tells a reader where a claim was found. Lineage tells the reader where it was made, and the two have drifted apart so quietly that most people producing memos with these tools have stopped noticing they are different things. Every check an organization runs on a memo reads the memo: the words, the arithmetic, the presence of a citation, and whether the cited page says what the memo says it says. In the Loudoun case every cited page did say what the answer said it said. The damage was upstream of the page, in the relation between the page and the record, and no check in ordinary use reads that relation.

The reader of this piece runs something, a company, an agency, a department, a research group, and has already put generative search products in front of the people who produce its memos. The argument is that these products take an old failure, a figure drifting away from its record through restatement, and make it free to commit and hard to see. The old costs that used to slow a claim down, and give someone a reason to check it against the record, are gone from both sides of the exchange at once. The signals readers used to detect careful work, fluent prose and an attached citation, are now free to produce, so they no longer detect anything. And the products themselves, tuned to return what people accept, defer to the premise of the question, including the premise the asker did not know they held. Nothing in that arrangement compares an answer to the record. The record takes an afternoon to read, and the exchange between a person and a product runs in seconds.

What a trace looks like

Take two of the six figures, the forty-five at the top of one product's range and the forty, and follow each to wherever it ends. The method is the same for any figure in any memo, and it has three moves: open the cited page and find the number on it; find what that page says its own source is, and go there;

stop when you reach a document in which someone recorded the observation, or a page that names no source you can follow. Write down how many pages you passed through and where you stopped.

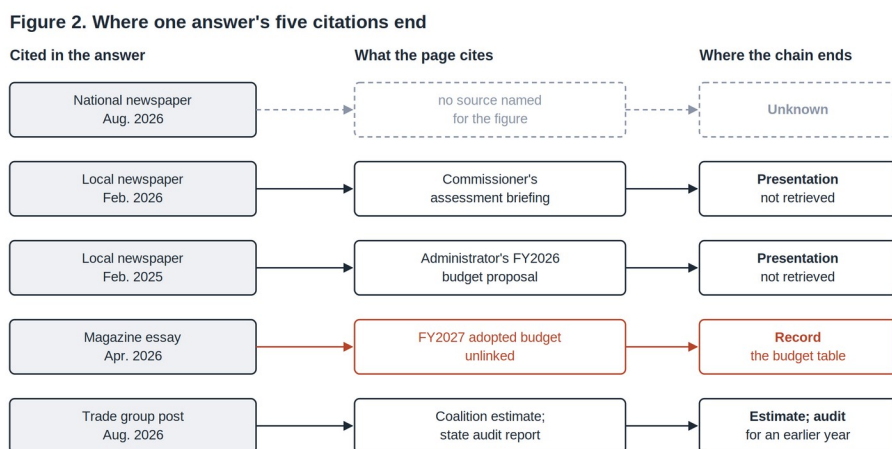
The forty-five. The answer that gave the range cited a magazine essay for its top figure. The essay carries the number, and it also carries the arithmetic: about one point three billion dollars in data-center taxes against about two point nine billion in county revenue, attributed to the county's adopted budget for FY2027, the fiscal year then under way, without a link. That is hop one, and the page says where it got its figure, so the trace continues. The county's budget page leads to the General Fund revenue section of the adopted budget, a PDF. On page 104 of that section, total local tax revenue for the year is given as \$2,906,994,401. On page 105, real property tax from data centers is \$417,666,456, and on page 107, computer-equipment tax from data centers is \$879,094,354. The two data-center lines sum to \$1,296,760,810, which is 44.6 percent of the local tax revenue line. That is hop two, and it is the record: the table in which the county set down the revenue it adopted for the year. The chain is two pages long, the figure is the record's figure rounded up, and the base, which the answer never stated, turns out to be local tax revenue. With the county's site open, the two hops are a matter of minutes.

The forty. Two answers gave it, and it appears on two pages cited across the four answers. One, cited by both of those answers, is the county's own web page about data centers, which says the adopted budget projects data-center revenue at one point three billion dollars, "40% of the total budget." That page is an official restatement, a county page summarizing the county's record, and it names the budget as its source, so the trace continues to the same PDF as before, where the base that turns one point three billion into forty percent is the General Fund revenue total, a larger number than the local tax revenue line; one of the four answers computed that fraction from the same table as 40.3 percent of General Fund revenue and reported it alongside its 44.6. The forty is a different fraction from the forty-five, over a different base, and the county page's phrase "total budget" is a loose label for it. The other page, cited by the answer that gave the range, is a national newspaper story, which says data centers produced one point three billion dollars, "40 percent of total tax revenue," and attributes the whole sum to the computer-equipment tax. That page names no source for the figure. The budget does not support the attribution, since the equipment tax alone is \$879 million and the one point three billion is equipment plus real property. The chain ends there, at a page whose own source cannot be found, carrying a base the record does not match: a true number with a wrong label, which is a wrong claim.

So the two figures, which the answers presented as neighbors, are one figure that reaches the record in two hops over a base the trace recovered, and one figure that reaches the record by one route and, by another, a page with no source and a base the record contradicts. Nothing in the answers, the citation rows, or the pages' surfaces distinguishes the routes. The trace does, in minutes for the first and not much longer for the second, and what it produces is the thing a memo can carry and a citation cannot: the number, its base, its year, the length of the chain, and where the chain ended. In a memo it is one sentence. Data centers supply 44.6 percent of Loudoun County's local tax revenue in the FY2027 adopted budget (General Fund revenue section, pages 104 to 107; traced 11 September 2026, two hops, chain ends at the record).

Loudoun was chosen because the record is reachable, and most figures a staff member cites have no clean budget table at the bottom. The method does not change. The trace still ends somewhere, at a document or at a page nobody can follow further, and where it ends is the finding. A figure whose chain

ends short of a record goes into the memo as unresolved, with the chain length and the last page named, and a memo that carries three unresolved figures has told its reader something true about the state of the evidence, which is more than a memo with three clean citations can say.



The interface displayed "65 sources": sixty-five search results across six searches. The answer used five pages. One chain reaches the record. One ends at a page naming no source. Two end at briefings not retrieved; one at an estimate and an earlier audit.

Figure 2. Where the five citations of one answer end, the answer that displayed sixty-five sources. The magazine essay row is the forty-five chain traced above; the newspaper row is the forty that ended nowhere.

Two rules fall out of the exercise. A fraction goes into a memo with its base and its year in the same sentence or it goes back. And a chain that ends at a page with no source is reported as unresolved, whatever the number on it, because a figure whose origin cannot be established is a figure whose meaning cannot be established either. The rest of this piece is about why those two rules are necessary and why nothing else in an organization's review process enforces them.

Why the surface stopped meaning anything

Every check the organization ran on the Loudoun answers would have passed them, and the reason is that the signals those checks read used to be expensive and are now free.

Thirty years ago a claim traveled from an observation to a reader's desk through hands that each charged something. Someone wrote it down, so a document existed. An editor decided whether it was worth printing and had a name on the masthead, so an error that got through cost the editor something. Printing and distribution cost money, so whoever paid wanted assurance before paying. When a later writer cited the claim, the footnote pointed to a physical object, and a doubting reader could spend an afternoon in a library reading the object to see whether it said what the footnote said it said, which the writer knew when attaching the footnote. All of this was built to make money or protect a reputation, and each step, as a side effect, slowed the chain at a point where the person holding it had something to lose by passing a bad claim along. Delay plus exposure sent people back to the record. Friction also kept chains short, since printing a restatement cost as much as printing the original, so a writer had little reason to cite a page that merely repeated the record when the record cost no more to cite. Print had its own quotation errors and its own chains of copies, and the difference is the cost gradient, which then

ran toward the record and now runs away from it. Friction was the referee, slow and expensive, and nobody designed it.

Two signals grew up under that regime. Fluency became a signal of quality, for good reason: producing a clean, confident, well-organized account of something was expensive enough that the people who did it tended to know the subject. Citation became a signal of verification, again for good reason, since attaching a footnote exposed the writer to being checked, and a writer willing to be checked was, on average, a writer who had done the checking. Both were correlates of costly work, and readers learned to trust them because, when they were expensive to fake, the signal and the work traveled together.

Generative products removed the cost on both sides at once, and the simultaneity is what makes the difference. Had only the supply side become free, a person who still had to formulate a question carefully would at least have stated their premises on the way in, where they could be argued with. Had only the demand side become free, the expensive supply would still have filtered what came back. With both free, the person asks in whatever form comes to mind, which is the form that carries their assumptions intact, and receives prose that carries every mark of care without the care. The editor's refusal and the afternoon in the library are gone, and nothing was put in their place. Nobody had understood them as anything other than costs.

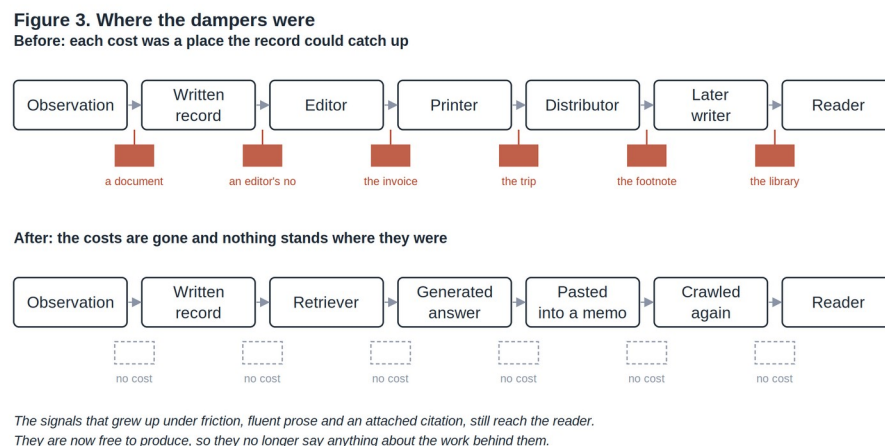


Figure 3. Where the dampers were. Each cost in the old chain was a place the record could catch up; nothing stands where they were.

The signals stayed. A reader who learned to trust fluent prose still trusts it, a reader who learned to treat a citation as evidence of checking still does, and both are applying a rule of thumb calibrated when the signal was expensive to a world where it is free. The rule once separated careful work from careless work with fair reliability. It now separates nothing, since every answer is fluent and every answer is cited, and the reader's confidence, which used to track the work, tracks the surface, which is the one thing the product is optimized to produce.

The judgment that tells a manager a memo is good was formed under friction, and it now approves, at the same rate and with the same confidence, memos whose numbers came from the record and memos whose numbers traveled through six hands, since the two are identical at the surface where the

judgment operates. The competence is intact, the surface it works on has stopped carrying information, and a manager who wants to know how far a number has traveled has to ask, because nothing about the memo will tell them.

What the products are built to do

Nothing in a generative search product compares an answer against a record. That is not a defect in any one of them; it is what they were built to do instead, and three stages of how they were built decide what comes back.

The first is tuning. Before anyone used these systems, human raters compared candidate answers and the models were adjusted toward the ones the raters preferred. Raters work fast and prefer answers that read as fluent and complete, and they cannot see whether the figure in the third paragraph is right. Newer pipelines add graders for factual and citation accuracy, and what those graders check is whether a claim matches the page it cites, which is a source check and stops where a source check stops. One thing the tuning taught the models shows up in every answer: deference. An answer that accepts the premise of the question is what raters tended to prefer, so it is what the product gives, and studies of the tuning data find that the preferences these systems were trained on reward agreement with the asker over correction a measurable share of the time.² Within a conversation the product also builds each reply on everything already in the exchange, so when the person accepts a framing or pushes back, the next reply is built on that. Nothing in the system changes while this happens, since the reply is conditioned on the conversation and the deference it conditions on was learned in the tuning. The effect is that the product leans toward the person's framing, including the parts of the framing the person did not state. Asked for a share of revenue with no base named, it does not ask which base. It supplies one.

The second stage is retrieval. When a product searches the web before answering it has to choose which pages to read and which to show, and its ranking weighs links, the authority of the domain, freshness, and whether the page is shaped like an answer. A county budget is a PDF with the relevant lines on page 105. A news summary of that budget is a clean paragraph with the number in its first sentence, and the retriever prefers the paragraph, so the answer inherits whatever the paragraph did to the number on its way out of the table. The pages that carry a figure's restatement outnumber the record and are better shaped than it, and the ranking rewards both.

The third stage is the corpus, the body of text the models are trained on and the retrievers search, together with what the products learn from use: the thumbs up, the follow-up question, the abandoned session, which is acceptance measured directly and at scale. The products retrieve from the same web and are trained on largely the same text, so where the corpus holds one version of a claim they agree on it, and where it holds several versions because the base was never carried, as for Loudoun, each product lands on whichever restatement its retrieval ranked highest for the words of the question, and the spread across products mirrors the spread of restatements in the corpus. Neither the agreement nor the spread carries information about the record, which none of the products consulted. Where the corpus holds one version, a wrong answer that every product returns is the natural output, and it is the output every instinct the reader has says to trust.

Put the person and the product together and the arrangement settles rather than runs away, which matters, because an arrangement that ran away would be caught. Both sides gain from agreement, the person an answer that passes and the product an acceptance, and nothing charges either of them for agreeing on something wrong. Once the person asks in the shape the product answers well and the product answers in the shape the person accepts, nothing pushes either of them anywhere, and the exchange settles fast on something plausible, plausibility being what both sides were selecting for. What sits at the resting point is fluent and cited, passes every check that reads it, and stands in whatever relation to the record it happened to have when the exchange closed. It settles on what both already assumed. The person's assumptions travel in the question: the framing, the base a fraction is taken over, the year they believe is current, and above all the premise they did not state because it never occurred to them that it needed stating. The product's assumptions travel in its corpus, as the version of each claim that was easiest to restate. On a premise the person stated, the product defers, and the person's framing wins. On a premise the person left unstated, the product fills the slot from the corpus and the person reads their own assumption into the result, and the two now hold different premises without either noticing, since the answer fits both. That is the Loudoun case. Askers with different bases in mind would each have found their number, and the products, asked without a base, spread across the bases the corpus held.

From inside, none of this is visible. A person who has adapted to the product experiences the adaptation as fluency, with better answers arriving faster and productivity rising by every measure the organization tracks, and the product is more helpful by every measure its builders track. Neither measure reads the distance to the record. Nobody built anything that reads it, and nothing in the arrangement rewards building it. The marks an operator can see are these. A team's inquiries into a domain converge on a handful of well-answered shapes, whatever the answers say. Answers agree across products and across weeks, which is what a shared corpus produces and says nothing about the record. And a premise nobody stated gets treated as settled, so the base and the year inherited from the first question travel through every answer after it.

Two ways the record gets lost

Each of the two is old, possible with a search engine or a hurried research assistant, and generative products change two things about it: the cost of committing it, which is now zero, and the cost of seeing it, which has gone up. Picture a ladder: the observation at the top, below it the record that wrote it down, below that a page that repeats the record, then a page that repeats the page, and at the bottom a chain nobody can follow. Each of the two ways moves the figure one rung down, and each ends with what an operator can do.

Figure 4. The distance ladder

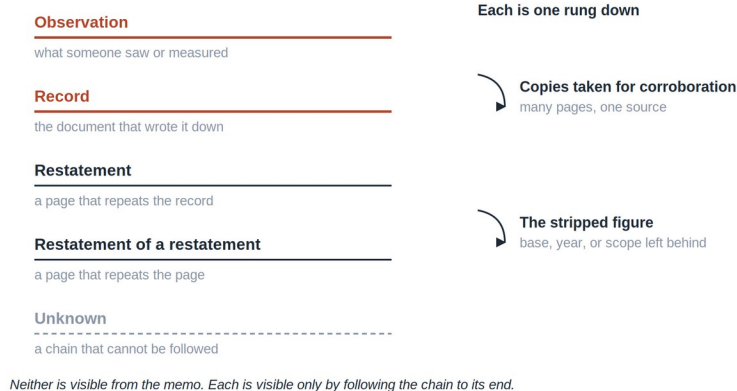


Figure 4. The distance ladder. Each of the two ways moves the figure one rung further from the observation.

Copies taken for corroboration

Under the answer sits a number, two sources, or sixty-five, and the number counts what the product retrieved, which is not the same as what supports the claim. Among the pages that do carry the claim, the product has no means of noticing that the second took its figure from the first, so each is presented as a source, and the person reads the count as weight, which is what a count of sources used to mean when sources were expensive, and feels, on seeing several citations agree, the particular confidence that comes from corroboration. That confidence was earned when corroboration meant independent observers. When the observers are copies of one restatement it is unearned, and there is a way to tell: copies of a restatement agree to the decimal, and independent observations of the same thing rarely do. The six Loudoun figures came from six different restatements, and every copy of each restatement carried the same number.

The same substitution runs across products. An analyst who checks a figure in a second tool and sees it agree feels diligent, and has consulted the same shelf twice. The inference from agreement to truth is sound only when the things agreeing are independent, and the section before this one said why the products are not. Whether the corpus holds one version or several, the pattern across products is a fact about the corpus, and the closing section names the test that would show it.

The check, for the one figure a memo turns on, is to ask how many of its sources are the same source, to accept "I don't know" as the honest answer it usually is, and to treat agreement among tools as a fact about the tools that counts for nothing until one of the agreeing chains has been followed to a document.

The stripped figure

A fraction is a claim about two quantities dressed as a claim about one, and any figure likewise carries a year and a scope that are claims in their own right and are rarely restated with it. Whoever first computed a figure chose its base and its scope and, if careful, stated them; nearly everyone who repeated it afterward dropped the statement, so the number traveled and its context stayed behind. Loudoun's fractions are the base version of this. The scope version was in the same trace, since the question also asked how many data centers the county has: the products returned 233 and "more than

250" side by side, the first from the county's planning inventory of buildings built and under construction on a stated date, the second from pages that either cited an official's briefing counting structures or named no source at all, and one answer in four noted that two different things were being counted. The date version is a figure that has since been corrected: the correction lives in one PDF on an agency server while the old figure lives on every page that restated the PDF before it was corrected, and on every page that restated those. Ranking rewards links, and links accumulate with time, so the pages carrying the old figure have had years to collect them while the corrected PDF has freshness and an authoritative domain on its side and little else. The correction also arrives into an installed base, and would have to be restated as often as the original was to catch up, which happens last if it happens at all. The premise version is the question itself, whose entity or period or definition the product takes as given, because that is what it was tuned to do, and the person is usually unaware of holding the premise, since a premise one is aware of tends to get stated and checked.

In all of these the mechanism is the same. The restatement strips something from the figure, the product fills the gap from whatever the corpus holds, the reader fills it again from whatever they assumed, and the two fillings need not match, and nothing in the answer says whether they do. The memo carries a number that is true of some base and some year, cited to a page that carries it, and the decision made on the memo rests on a base and a year that nobody in the chain chose and the reader silently supplied. Editors used to catch this with one free question, what do you mean by that, and that question is the step the arrangement removed.

The check is the first rule from the trace, context in the same sentence, extended to the date of the record a figure came from, plus one more habit: a second person restates the question aloud before the answer is used, since a premise that survives being said by someone else has at least been seen.

A careful reader of the memo can see that a base or a date is missing. No reading of the memo can show whether the base and date it does state are the ones its source used, or whether that source's source used them too, and the checks above are built to force that question.

Why the memos pass

Every control an organization runs on written work reads the artifact, and the artifact is fine. The instinct of anyone who runs an organization, on hearing that the memos are wrong, is to add a check, and the check they add will read the artifact too.

The cheapest is a manager's read, which asks whether the memo is clear and whether its numbers look right. Both questions are answered by the surface of the document. A memo produced by the arrangement is clear because the machine is optimized for clarity, and its numbers look right because they are the numbers the manager has seen before, having come from the same corpus the manager's own sense of the subject was formed on. The manager's read passes the memo for the same reasons the arrangement produced it.

Source checking comes next, and it is the failure's most reliable accomplice. A source check opens the cited page and confirms that it carries the claim, and in every one of the Loudoun chains it does, because the product cited the page accurately. The magazine essay really does carry the forty-five, and

the newspaper really does carry the forty. The check is a check on the citation, the relation between the memo and the page, and the damage is done in the lineage, the relation between the page and the record; a source check by design reads the first relation and stops. A more diligent checker follows one more link, to the page's own source, finds that it too carries the claim, and stops there, satisfied, having confirmed two links in a chain whose length they do not know and whose end they have not reached. Diligence in source checking is a measure of how many restatements the checker was willing to read. The stopping rule the trace used, that the chain ends at a document recording the observation or at a page that names no source you can follow, is the thing no source-checking procedure in ordinary use applies.

Peer review, or its corporate equivalent, the second reader with domain knowledge, used to catch the inherited premise, because a second reader who knew the subject would notice that the question was framed wrongly. It fails now to the extent that the second reader's knowledge of the subject was assembled with the same tools from the same corpus, in which case the base they assume and the figure that looks right to them are the writer's own, acquired the same way, and that extent grows every year the tools are in use. Their review is a second consultation of the same source, and its agreement gets read as confirmation.

The audit, an independent party reconstructing the work from the ground up, is the one control that catches these failures, because reconstructing from the ground means reaching the record, and it is rare because it is expensive, and expensive because the work an audit does, following every claim to the document in which the observation was recorded, is exactly the work that friction used to force and generative systems removed. Organizations audit what they cannot afford to get wrong, and they decide what they cannot afford to get wrong by looking at the artifact.

All of these controls fail together for one reason: each was designed for a world in which the artifact and its lineage were correlated. A memo that was cited had usually been checked, and a memo that agreed with the reviewer's sense of the subject had usually been drawn from the same records the reviewer had read, so reading the artifact was a cheap way of reading the lineage, and the correlation held because friction kept chains short enough that the cited page usually was the record. Friction is gone. The controls remained, reading a surface the arrangement now produces on purpose, with the confidence they earned when the surface meant something.

Adding controls of the same kind changes nothing. A stricter source check is a longer walk down a chain of restatements, and a better reviewer is one who has used better tools and holds the same premises more confidently. The only control that reads lineage is one that reaches the record, whether by following a citation down until it arrives at a document or at a dead end, as the trace did, or by starting from the record and checking the memo against it, and no organization runs either one routinely, since the artifact never gives it a reason to.

Who pays

Two costs show up on an organization's own books, and the party that pays is rarely the party that let the number through.

Decisions

A decision made on a fraction is a decision made on a base. When the base was carried by nobody in the chain and supplied by the reader, the decision inherits a choice that nobody with the authority to examine it ever examined. A board that approves an expansion on a figure for a sector's share of a region's revenue has made a different decision than it would have made on the other base, and review will not tell it so, since the number was cited and the citation checked. Outcomes will not reliably tell it either. They arrive years later, get attributed to whatever went wrong most visibly in the meantime, and are rarely traced back to a base in a memo. And the same base gets used again next year, because it is now in the organization's own memo, which is now in its files, where the next analyst and the organization's own search tools will find it before they find anything else. The cost is the gap between the decision made and the decision the record would have supported, multiplied by every later decision that inherited the premise. Nobody has a good rate for how often this happens, and the piece does not offer one. An organization that wants to size its own exposure can do so in an afternoon by tracing the two or three figures its largest recent decision turned on, the way the second section traced Loudoun's, and the chains that end short of a record are the exposure.

People

The cost to the people inside the arrangement is deskilling that feels like competence. A person who has spent a year receiving fluent answers to questions they learned to ask in the shape the tool prefers has, by every measure the organization tracks, become more productive, and has, by a measure nobody tracks, lost the habit of going to the record and the habit of noticing their own premises. These were skills. Friction maintained them by requiring them, and when the friction went the skills went with it, through no failure on the person's part, since skills that are never exercised decay. The person feels no loss, the objective they now hold, an answer that passes, being met more easily than ever. An organization that has had its analysts on these tools for a few years should ask them what they would check if they had to, and should not be surprised by the answer. The remedy is exercise. Someone in the group reads the record rather than the tool for a share of their time, and the group sees them do it.

What a referee would do

The arrangement has no referee, nothing inside it whose job is to compare an answer against the record and charge for the difference, and nobody installed the old one either. It came free with costs nobody wanted, and when the costs went it went with them, so a referee now has to be built on purpose, and every party inside the arrangement is rewarded for not building it: the product for speed and confidence, the rater for fluency, the person for a passed memo, the organization for throughput. Exposure, which under friction sent people to the record, now reaches only as far as the arrangement satisfies it. An analyst can be caught citing a page that does not say what the memo says, and cannot be caught citing a page whose own source was wrong, because nobody in the building can check that either. The parties that bear the cost are the record, which cannot vote, and the future decisions that will be built on it, which are not yet in the room. Putting a referee back is a question of where the comparison would be made.

At the interface, a referee would mean that an answer carried, for every figure it reported, the name of the document in which that figure was recorded and a statement of whether the page being cited is that document, a page that repeats it, or a page whose origin cannot be established. It would mean that the number displayed under an answer counted the documents its citations actually reach rather than the results the search returned, so that twelve links to one record displayed as one, and that a percentage came with its base attached as a condition of being displayed at all. All of this is within the products' capability where the chain can be followed, marking the rest unresolved being the point, and each item costs the product something on the measures it competes on. The origin statement marks many figures as unresolved, which reads as less confident, and the collapsed count makes the source number smaller, which reads as less thorough. Attaching the base makes answers longer and now and then puts them in disagreement with the asker. A product that installed the whole referee would lose, on those measures, to one that did not, until the measures change; the cheapest piece of it, a stated base and date on every fraction, costs almost nothing and is the piece to ask for first. The referee is out of an operator's hands at the product level and within reach in procurement: when a vendor is asked whether its product can tell a record from a page that repeats one, and whether its displayed source count counts documents or results, the answer and the vendor's reaction to the question are both information.

In the reader's own habits, a referee means restoring, deliberately and at some cost, what friction used to force, and the second section of this piece already showed what it looks like. The base and the year go into the question before the fraction is asked for, so the answer has something to be wrong against. One citation per memo, the one the decision turns on, gets followed to the record rather than to the page, and the length of the chain goes in the memo, because a chain that cannot be followed to a document is not evidence and should appear as unresolved. Agreement among tools gets treated as what it is, a fact about a shared corpus, and waits on one followed chain before it counts. Each of these costs minutes, and minutes are what the tools were adopted to save, so the person who does them is slower than the person who does not, and slower in a way the organization's own metrics will register as a deficit. That deficit is the price of the referee. The only question for an operator is whether to pay it before or after the decision that was built on the wrong base.

The staffing is small. Whoever signs off on a memo owns the one trace, which takes minutes when the chain is short and is itself the finding when the chain is long or ends nowhere; it happens once per memo that a decision will turn on, and the result, how far the number traveled and where the chain ended, goes in the memo. Whoever edits enforces the base and the year, and a fraction without them goes back. One person in the organization spends a fixed share of their time reading records rather than tools, and that time is protected, because nothing they produce will look efficient and everything they produce is the only thing in the building that touched the world. None of this scales. The arrangement does, which is why it won, and the choice was only ever between a referee that does not scale and none.

This piece has shown one case, traced to the record, and explained the arrangement that produced it. The larger claims belong to the working paper behind it: that every answer pasted into a report or a website becomes text the next model learns from, so that the web fills with the products' own restatements and the record is outnumbered by its copies; that institutions which keep records lose authority to whoever restates them most fluently; and that the person and the product are training each

other, the person's shortcuts becoming the product's habits over years and the product's habits becoming the person's over a career.¹ Those are claims about what the arrangement does to the world, and a board has no lever on the world; it has a lever on its own memos. The last of the three is a hypothesis. The test registered alongside this work is narrower, registered before this piece was written and not yet run, since the registration has to come first, and it is the one the first rule implies: put the base in the question and see whether the spread of answers closes. If it does, the products can carry a referee the asker supplies, and the question becomes why they do not supply it themselves. If it does not, the problem is deeper than this piece has said. Either way, the trace works now, on any figure, in minutes, and it is the one thing that puts the record back in the arrangement.

Notes

[1] The Loudoun trace is the one-question pilot of the Source Diversity Audit, run on 11 September 2026 across ChatGPT, Gemini, Perplexity and Grok. The captured answers, the traced pages, local copies of the county records, and the capture log are deposited as Tommy N. Turner, Source Diversity Audit, version 1.1, Zenodo, <https://doi.org/10.5281/zenodo.22839545>, which also holds the pre-registration of the test described in the closing section. The trace and the six-fraction table are reported, and the larger claims argued, in Tommy N. Turner, "Who Is Training Whom: Human Cognitive Drift and Machine Degradation in Coupled AI Knowledge Systems," working paper, Zenodo, <https://doi.org/10.5281/zenodo.22839554>.

[2] Mrinank Sharma et al., "Towards Understanding Sycophancy in Language Models," arXiv:2310.13548 (2023), <https://arxiv.org/abs/2310.13548>. The paper finds that human preference judgments used in training favor responses that agree with the user's stated views over responses that correct them in a measurable share of cases, and that preference models trained on those judgments sometimes prefer the sycophantic response.