



When AI Meets Lightcast Data

Investigating How the Lightcast MCP
Improves the AI You're Already Using

Table of Contents

Executive Summary 3

Introduction: AI Is Already Answering Your Workforce Questions 5

The Trust Problem 6

Why Trusted Labor Market Data Matters More in the AI Era 10

What We Built: The Lightcast MCP 13

How We Tested It 15

What This Means for You 25

Conclusion: Better Together 27

Appendix A: Inside the Lightcast MCP Toolbox 29

Appendix B: Methodology 33

WRITTEN BY ZANE WOLF AND DENIS POULIN

DESIGNED BY LAUREL GIESZELMANN

©COPYRIGHT 2026 LIGHTCAST

EXECUTIVE SUMMARY

Connecting AI to Lightcast labor market data improved nearly nine in ten answers to workforce questions.

Across three leading AI models, average answer quality rose from 57 to 82 out of 100. The largest gains came in questions about skills demand and geographic comparisons, where specific, current information can be difficult to find through conventional web search.

These findings matter as large language models (LLMs) change how people access and use information. AI assistants can deliver polished, confident answers in seconds, often complete with citations. But when those answers inform hiring plans, budget requests, student’s futures, and people’s careers, sounding credible isn’t enough. Organizations need to know what the data measures, where it comes from, and whether they can defend it.

To understand the difference trusted data makes, Lightcast tested how leading AI models answered real-world workforce questions with and without access to Lightcast data through its Model Context Protocol (MCP) server. Models received no special instructions, and answers were evaluated across six measures of quality, including specificity, timeliness, completeness, and supporting evidence.

The difference was especially clear when questions required a specific number. With web search alone, AI provided the requested labor market figure just 35% of the time. With Lightcast connected, that rose to 92%. Rather than relying on whatever information a web search surfaced, AI could query structured labor market data using consistent definitions, defined time periods, and identifiable sources.

With or without web search, Lightcast Intelligence improves LLM’s answers to labor market questions

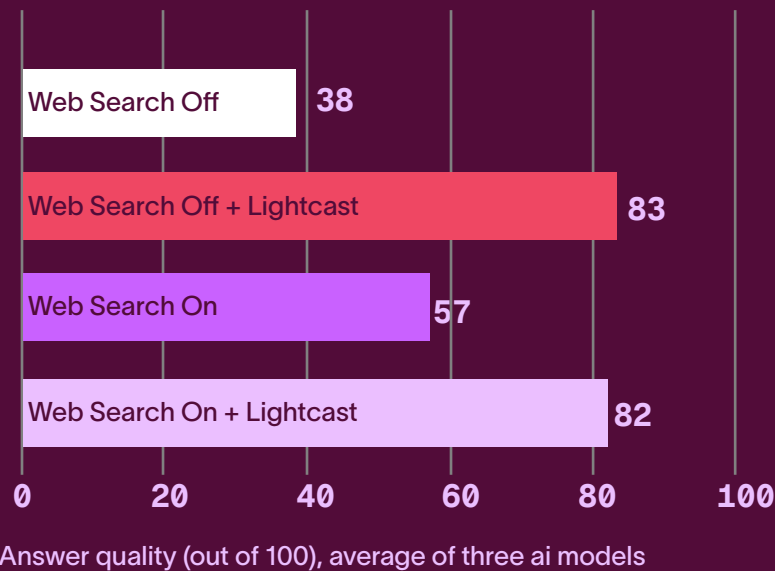


Figure 1. Lightcast made the difference, with or without web search. Average answer quality out of 100 across three AI models, 120 questions each. With web search on, adding Lightcast raised quality from 57 to 82.

The takeaway is simple: AI is changing how people access information, but the quality of the answers still depends on the data behind them. Connecting AI to Lightcast brings together the convenience of conversational AI with the labor market evidence organizations need to make decisions they can understand and defend.

Lightcast designed and conducted this study. Appendix B details the methodology, evaluation criteria, and limitations.

KEY FINDINGS AT A GLANCE

Lightcast makes the AI you already use better. Adding Lightcast to web-search-enabled AI raised answer quality from 57 to 82 out of 100.

Better answers, not just better averages. Nearly nine in ten answers improved when Lightcast was added.

Ask for a number, get a number. With Lightcast, AI gave the specific figure asked for 92% of the time, compared with 35% with web search alone.

Web search alone isn't enough. AI with Lightcast outperformed AI with web search on 84% of questions.

Every kind of workforce question gets better, from market sizing and trends to skills and geography, with the biggest gains on skills questions.

It works across leading AI models. Every AI model we tested improved with Lightcast.

Given the choice, AI turns to Lightcast. With both web search and Lightcast available, AI drew on Lightcast data in nine of ten answers.

Better AI starts with better data.



Introduction: AI Is Already Answering Your Workforce Questions

AI is changing how we work, including how we ask and answer questions. Data-based questions, like labor-market questions, are not exempt. You open your preferred AI chat and ask, *"How many data center technician positions were advertised in the U.S. in the last 12 months?"* You get an answer back and it sounds authoritative—cited, perhaps even dated—and it sounds reasonable. Do you use it to help decide whether your region can staff a new data center? Do you include that number in your board presentation?

How confident are you in the accuracy of AI's answer?

Here at Lightcast, we want you to use AI and we want you to be able to trust the answers the AI returns. To deliver on this, we built an MCP (Model Context Protocol) that lets AI assistants query Lightcast data directly.

And then we tested it.
Rigorously.

We asked AI to answer labor-market questions both with and without the Lightcast MCP enabled and assessed the quality differences in the answers. We varied the type of question (trend analysis versus market size, for example) and the difficulty (single occupation questions at the national level or multi-employer, multi-region comparisons, for example). We toggled web search on and off. We asked the same questions again and again and again to see if the answers would be the same each time.

We wanted to know whether, and by how much, grounding AI in Lightcast data changes the quality, consistency, and defensibility of its answers. Workforce decisions don't come with a do-over. They shape hiring plans, budgets, student's futures, policy scenario analyses, and people's careers. An answer that only sounds right isn't good enough.

So, how much better can your AI get when it has the right data? The rest of this paper shows exactly how much.

LIGHTCAST DATA IS THE DIFFERENCE

We asked "How many AI-related roles has Google made available across the U.S. in the past three months?"

With web search alone, three AI models gave three different answers: no number at all, about 1,500, and about 700.

With Lightcast connected, all three came back with the same answer, about 4,860 unique postings for roles requiring AI or machine-learning skills, with the time period stated.

The Trust Problem

AI assistants are remarkably good at sounding right. Ask one about the job market and you'll get a fluent, well-organized answer in seconds. It'll have headers, bullet points, and a tidy summary at the end. Turn on web search, as most people now do, and the answer will also arrive with something that looks a lot like proof: citations, links, and named sources.

Don't get us wrong, that's real progress. But it's also created a new kind of trust problem. **Though a link looks like verification, it isn't.**

Why workforce questions are hard for AI

Labor market questions are tricky for three reasons, with or without web search.

They depend on precise, current numbers.

How many openings? What's the median wage? Is demand growing? Each of these has a specific answer, and that answer changes month to month. A model's training data is frozen at some point in the past. Web search fills the gap with whatever happens to be published—a job board's live listing count, a press release, a market-research projection, a report from two years ago.

They depend on consistent definitions.

Is a "software engineer" the same as a "software developer"? Does "data center technician" include facilities engineers? Ask five websites and you'll get five different ways of counting. An answer stitched together from several of them is a bit like assembling a puzzle from pieces of five different boxes: the edges might line up, but the picture won't.

They have to be defensible.

If a number is going into a workforce plan, a board presentation, or a grant proposal, someone has to stand behind it. "The AI said so" isn't a source. Neither is a link nobody clicked.

They need to be consistent answers.

If you ask the same question two weeks apart, or two minutes apart, do you get the same answer? Which one do you use? AI rarely gives the same answer twice, and consistency is what builds trust.

What AI gets wrong

Without web search or the MCP, AI fails in fairly obvious ways. Answers are vague where they should be specific (“demand is growing rapidly”), undated, or politely declined (“I don’t have access to current data”). Frustrating? Sure. But at least you can see the gap.

With web search on, the gaps get harder to see. In our testing, nearly every search-enabled answer came with sources attached. The answers looked researched—more numbers, more detail, more references. So we looked closer, and we kept running into the same four problems.

The source doesn’t answer the question that was asked. We asked how many data analyst vacancies were posted in the past three months. One search-enabled answer cited a job board’s count of listings open today (about 11,600), sitting right next to an annual figure from 2023. Both were sourced. Neither answered the question.

The numbers come from wherever search lands. Search-enabled answers drew on a median of 10–29 different websites each. Job boards showed up in 66% of answers, government sources in 64%, and forums and wikis like Reddit and Wikipedia in 51%. And each of those counts differently: job boards don’t deduplicate postings across sites, forums aren’t data, and government series lag by months.

The result? The burden of verification shifts back to you. To trust a search-enabled answer, someone has to open every link, check every date, confirm the geography, and work out whether each page is counting the same thing the question asked about. By that point, how much time did the AI actually save?

The figure is old, or has no date at all. A number from 2023, or from a page with no date, gets presented as the answer to a question about the past six months. The citation makes it look current.

Hallucinated numbers come with links attached. Web search gave the AI more figures to report, but not better-supported ones. On “Backed by evidence,” web-search answers scored 58 out of 100, compared with 81 once Lightcast was added. AI with no tools (neither web search nor MCP) tends to tell you what it doesn’t know. AI with search tends to fill the gap with a number and put a link next to it. (Appendix B explains how we checked for invented figures.)

A LINK ISN'T THE SAME AS AN ANSWER.

Search-enabled AI drew on a median of 10–29 websites per answer, from job boards to Reddit threads. With web search, AI gave the specific figure the question asked for only 35% of the time, even though every answer cited sources.

What a grounded answer looks like instead

So what's different with Lightcast? It isn't the citations. Search-enabled answers already have plenty of those.

IT'S WHAT STANDS BEHIND THEM:

- **One source**, built for labor market analysis.
- **One consistent taxonomy**, so "data analyst" means the same population every time.
- **Deduplicated counts**, so one job posted on five boards is counted once, not five times.
- **A stated time period** that matches the question.

Asked the same three-month data analyst question, the AI connected to Lightcast answered 22,930 unique U.S. postings for the Lightcast occupation "Data Analyst," active between July 1 and October 1, 2026. One number. The right period. A source and a definition attached.

Is 22,930 the "right" number? It's the right number for that definition. A broader definition (say, every role with "analyst" and "data" in the title) would count more. The web answer's 450,000 figure for 2023 is far larger even on an annual basis, but there's no way to tell what it counted. And that's exactly the point: a defensible number tells you what it measured.



+ Ask a question...

Question:

"What is the total number of data analyst vacancies advertised in the United States over the past 3 months?"

Answers:

WEB SEARCH ALONE

"...there are currently 11,623 open jobs for data analyst positions in the United States. [job board]"

"...over 450,000 job postings for data analysts in the US in 2023. [market report]"

"...I wasn't able to find specific cumulative totals covering exactly the past 3 months."

WEB SEARCH + LIGHTCAST

"...There were 22,930 unique data analyst job vacancies advertised in the United States over the past 3 months (July 1, 2026-October 1, 2026).

This count represents deduplicated job postings, meaning each unique job opportunity is counted once, even if it was advertised on multiple job boards or platforms."

Figure 2. Same question, same AI, two very different answers. Asked how many data analyst vacancies were advertised in the U.S. over the past three months, web search alone cited a live job board count and a 2023 total. With Lightcast connected, the same AI reported 22,930 unique postings for July 1 – October 1, 2026. Excerpts are verbatim and trimmed.

Why Trusted Labor Market Data Matters More in the AI Era

As AI becomes one of the ways people access labor market intelligence, the quality of the underlying data becomes more significant. AI can retrieve information, summarize complex findings, and generate recommendations in seconds, but is speed and cost what makes an answer useful when the outcome of the decision matters? Do the figures compare the correct populations? Do they cover the right time periods? Does the data support the claims and conclusion it makes?

It's harder than it looks to truly identify quality, especially when an answer combines information from multiple sources. An employment estimate, a count of current job advertisements, and a forecast of occupational growth may each be credible on their own. However, bringing them together does not automatically make them comparable because they may be measuring different things. Without clear definitions and context, a fluent answer can mask differences that change the decision.

And getting the definitions right is only the start. Is the information current, or is it describing the job market from three years ago? Does the citation tell you how the number was collected, filtered, or calculated, or just who published it? Is an apparent trend in the data an explanation, or simply a correlation? Is a count of job ads being read as proof of actual hiring? And lastly, even if a recommendation sounds plausible, has anyone tested it?

These challenges explain why trusted labor market data matters—a precise number isn't enough. You need evidence you can compare, place in time, inspect, interpret, and use.

HERE'S HOW WE THINK ABOUT EACH OF THOSE NEEDS:

Comparable: apples to apples

When two figures use different definitions, a “difference” between two markets may simply be a difference in how the data were measured or collected. When the data is too broad, it can hide the nuances and details that most impact the decision you're trying to make.

Say you're choosing between two cities for a new operation. Comparing total employment in one city with specialized job postings in the other city tells you nothing about which one offers better access to talent. That's not apples to apples. More like apples to orange juice.

A useful comparison uses the same occupation definitions, the same geographic boundaries, and the same time periods. It distinguishes workforce supply (the people) from the advertised demand (the job ads), and it flags any changes in classification or coverage that could skew a trend. Lightcast's standardized taxonomies and deduplicated postings provide that common foundation, allowing for a true apples-to-apples comparison.

Current: when, exactly?

A decision can be perfectly well reasoned and still rely on evidence that is out of date. The publication date of a report, or the date an AI found it, doesn't tell you when the data was actually observed and recorded.

Imagine a college considering a new analytics certificate. It needs to know both what employers are asking for right now, and whether the demand for the certificate will last. Recent postings can reveal emerging skills, while a longer series helps show whether demand is sustained or just a temporary spike.

A good answer identifies when the activity occurred, when the dataset was updated, and how frequently new data comes in. This helps the college line up evidence against its program timeline and know when to recheck their assumptions.

Traceable: show your work

How can you review a finding if you can't follow the path from the raw data to the reported number? Naming a publisher tells you where a number came from but it does not tell you what was measured or what was done to the data along the way.

Suppose you're sizing an engineering talent pool. An estimate based on professional profiles has a different meaning from an employment count, and both mean something different from a total of advertised vacancies. A good answer should identify the dataset, the occupation categories, the geography, and the time period. It should also show the match behind any growth rate, ratio, or combined category..

Traceability pays off when two numbers disagree. You can see whether two estimates differ because of their definitions, methods, or the underlying data. It also helps explain why asking the same question twice can produce a different answer.

Interpretable: what does it actually mean?

A precise figure doesn't make every conclusion drawn from it valid. . You need to know what a metric represents, what it leaves out, and which conclusions need more evidence.

Say an HR team is deciding which skills to build into its upskilling program for data engineering roles. Should it look at how many job postings ask for a skill, or at what share of data engineer postings do? It matters. A rising count might just mean more hiring overall, even as the skill becomes less common within that hiring. Those two measures point to very different decisions about where to spend the training budget.

The answer should also separate correlation from causation. If a skill shows up in higher-paying job ads, that doesn't mean learning it will raise anyone's pay. Explaining these boundaries allows decision-makers to use the evidence with the right amount of confidence.

Actionable: does it work?

An insight only becomes useful when it informs a real choice. Then you have to check whether the choice worked.

Picture a company competing for technical talent. Should it recruit externally, or train the people it already has? A useful recommendation should incorporate existing skills, training capacity, budget, and timeline. It should also define what success looks like—for example, through proficiency achieved, roles filled, and staffing readiness.

Similarly, a college can do the same with a curriculum change: test it with employers, pilot the new content, and track graduate placement and employer feedback to decide whether to keep it or adjust.

Putting It Together

These five qualities are what it takes to trust an AI answer:

- 1. **Comparable definitions make the analysis hang together**
- 2. **Current data makes it relevant**
- 3. **Traceable sources make it checkable**
- 4. **Interpretable context keeps the conclusions honest**
- 5. **Actionable insight, tested in practice, shows whether it worked**

Grounding AI in trusted labor market data strengthens every one of these connections, providing a more defensible basis for decisions and a clearer path for reviewing them afterward. That’s why we built the Lightcast MCP, not to compete with AI, but to give it something solid to stand on.



What We Built: The Lightcast MCP

The Model Context Protocol (MCP) is an open standard that lets AI assistants connect to outside data and tools. It has been adopted across the major AI platforms. If web search is like asking a question in a crowded room and piecing together what you overhear, an MCP connection is like walking up to the reference desk. The AI can go straight to a trusted source, ask a specific question, and get a structured answer back.

The Lightcast MCP connects AI assistants to Lightcast labor market data. When someone asks a workforce question, the AI can:

- 1. **Discover** what Lightcast data is available and which dataset fits the question.
- 2. **Classify** the people, jobs, skills, places, and industries in the question into Lightcast’s standard taxonomies, so that “software engineer” maps to a defined occupation rather than a keyword.
- 3. **Query** the data for the counts, rankings, trends, wages, and distributions the question needs.
- 4. **Answer** in plain language, with the figures, the period, and the source stated.



Figure 3. From question to grounded answer, in four steps. How the AI used the Lightcast MCP to answer one question from our study. It chose each step on its own, with no instructions.

And nobody has to tell the AI to do any of this. In our testing, we gave the AI no special instructions. We connected it to Lightcast, and it decided on its own when and how to use the data.

Data available through the Lightcast MCP:

- Job postings: deduplicated online job advertisements for the U.S. and globally. This covers hiring demand, required skills, employers, industries, locations, and advertised salaries.
- Profiles: anonymized professional profiles representing labor supply. This covers talent pools, skills, education, and career paths.
- CoreLMI: Government occupation, industry, and unemployment data aggregated from official US sources including employment, earnings, and participation figures.

AN AI WALKS INTO A ROOM ...

If web search is like asking a question in a crowded room and piecing together what you overhear, an MCP connection is like walking up to the reference desk. The AI can go straight to a trusted source, ask a specific question, and get a structured answer back.



How We Tested It

We wanted to answer one question, and answer it honestly: does connecting AI to Lightcast make its answers better, and if so, by how much? We designed the test the way a skeptical customer would want it designed. Let’s walk through it.

We used real-world questions. We wrote 600 labor market questions across six categories, focusing on the kind of questions focusing on recruitment, workforce development, and market analysis typically asked by our customers.

Here are the categories we focused on, along with an example question from each.

MARKET SIZE	GEOGRAPHIC COMPARISON	OCCUPATIONAL DEFINITION
How many AI ethics and responsible AI roles have been advertised in the United States over the past 24 months?	How does the concentration of data analyst roles at JPMorgan Chase compare between New York City and Charlotte currently?	What's the difference between a logistics coordinator and a supply chain analyst in the U.S. labor market?
TREND ANALYSIS	SKILL DEMAND	TEMPORAL FRESHNESS
How has JPMorgan Chase's hiring for financial analysts in New York evolved over the past 12 months?	Which phlebotomy skills are urgent care clinics in Arizona requesting for medical assistant openings right now?	What emerging skills are healthcare employers in Texas requesting for clinical informatics specialists right now?

The questions ranged in difficulty from simple (a single occupation, nationally) to expert (multi-employer comparisons, supply–demand gaps, specific time windows).

We compared three leading general-purpose AI providers, selecting not the newest, most expensive models but the practical, lower-cost versions your CTO and CFO are telling you to use so that you don't burn through credits.

We didn't coach the AI. No system prompts. No instructions to use Lightcast. The AI got the question exactly as a person would type it into chat and nothing else.

An AI judge scored every answer on six measures, without being told which model or setup produced it. For one of those measures, the judge also saw the same model's Lightcast answer as a labeled reference, so it wasn't blind to Lightcast's figures. Appendix B covers what that means for the results, along with the judge model and the full rubric.

MEASURES	WHAT IT ASKS
Backed by evidence	Are its claims actually substantiated?
Up-to-date	Is the data from the right time period?
Specific	Does it give numbers, not adjectives?
Credible Source	Does it name a source you'd trust?
Nothing invented	Does it avoid unsupported facts and figures?
Complete	Does it answer every part of the question?

Together, the six measures add up to a quality score, which we report here on a 0–100 scale.

First, we pitted Lightcast against AI with no tools at all (no web search, no MCP; only its training data to draw upon). Across all 600 questions and three models, with web search off, AI grounded in Lightcast data vastly outperformed AI with no tools. 92% of the answers were higher in quality with Lightcast data than without, and on average, answer quality jumped from 30 to 70 points out of 100. A statistical model that accounts for how hard each question is gives the same answer: +39 points.

Then we checked our own work.

- We re-scored the same answers five times to make sure the judge was consistent. It was.
- We re-asked the same questions five times to make sure the results weren't a lucky draw. They weren't.
- We swapped out an older model for a current-generation model to see if the improvement from Lightcast data diminished. It persisted.

And with those results in hand, it would have been easy to stop there. We didn't.

We challenged ourselves to determine if Lightcast data still improves the answers AI provides when we remove the handcuffs and allow it to search the web—the way we know you'll actually use AI.

We tested leading AI models in four setups:

1. Native-AI only, neither web search nor the MCP connected
2. Web search only
3. Lightcast only
4. Web search plus Lightcast



Adding Lightcast to web search raised AI answer quality by 25 points

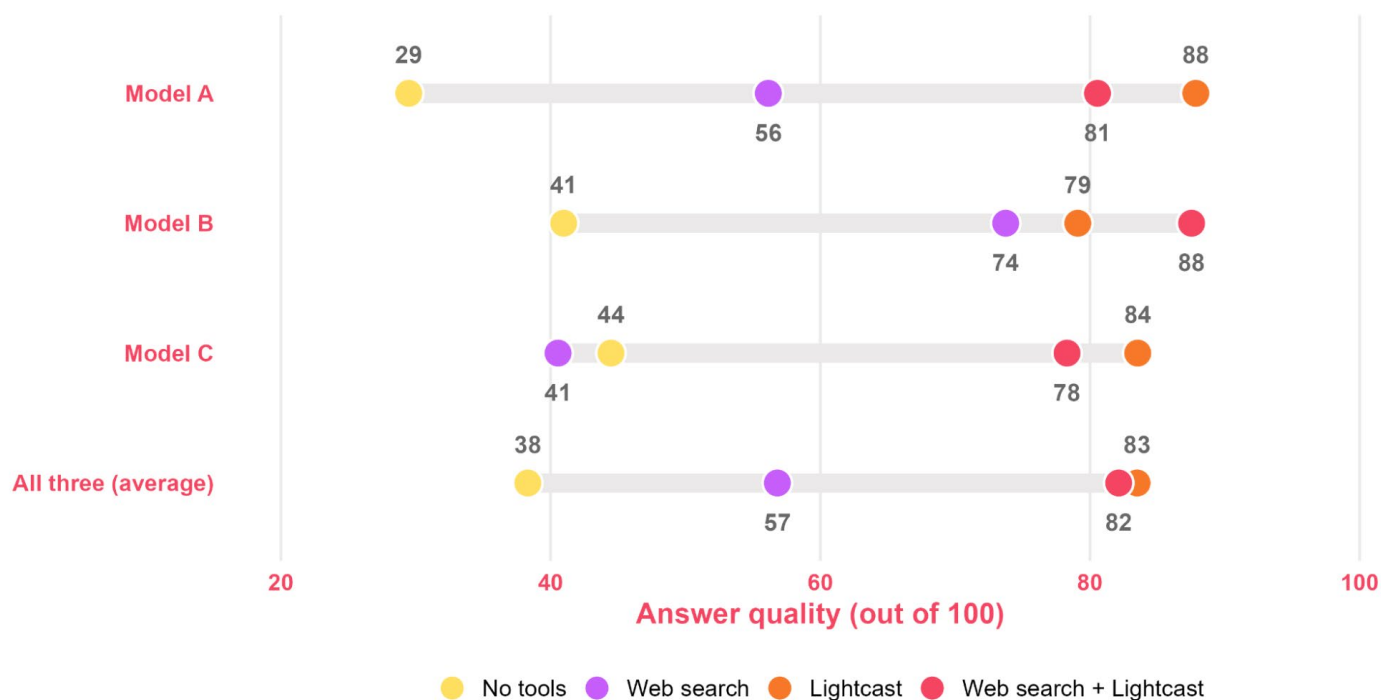


Figure 4. Lightcast lifted every model, with or without web search. Average answer quality out of 100 for three AI models in four setups, 120 questions per model and setup. Across all three models, adding Lightcast to web search raised quality from 57 to 82 (+25 points). Models are anonymized.

Here's the short version. With web search on (which is how most people use AI today), connecting Lightcast raised average answer quality from 57 to 82 out of 100. That's a 25-point gain.

And this wasn't a handful of big wins dragging the average up. 87% of answers improved when Lightcast was added: nearly nine out of ten. Did every answer improve? No. About one in twelve (8%) scored lower and 5% scored the same. But the overall pattern is hard to miss, and none of them got close to what Lightcast delivered.

So isn't web search enough on its own? It's tempting to think so. The whole internet is right there, right?

Look at Figure 4 again. Web search helps unaided AI, lifting average quality from 38 to 57. But it doesn't help every model equally. For one model, it didn't help at all. Model C was slightly worse than no tools: it beat no tools on only 28% of answers.

With Lightcast connected, every model scored at least 78, with or without web search. Head to head, Lightcast alone beat web search alone on 84% of questions. The gain comes from Lightcast, not from search.

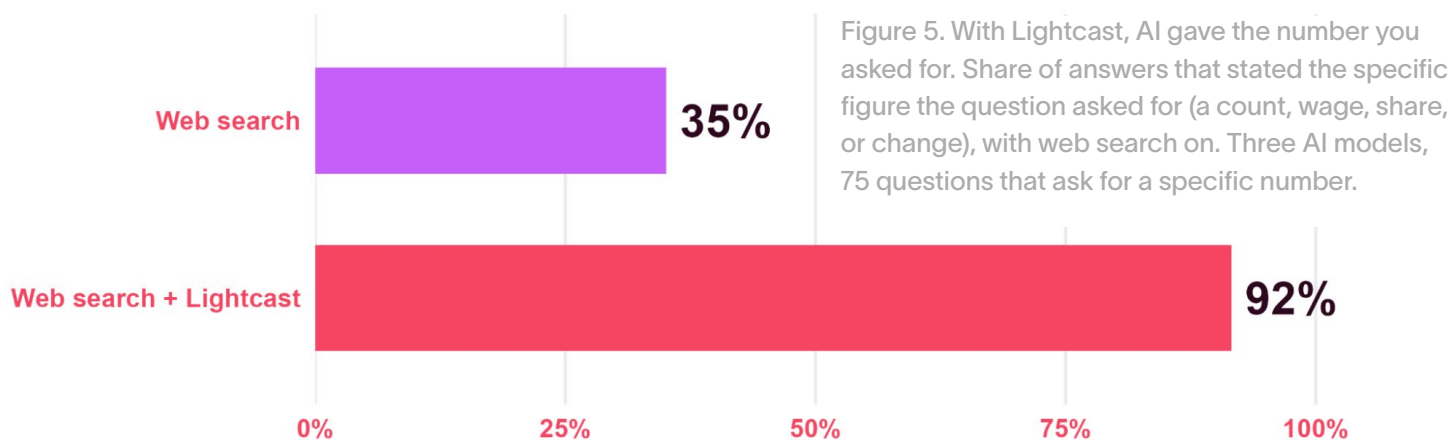
WHAT HAPPENED WHEN THE AI HAD BOTH? WELL, IT DEPENDED ON THE MODEL.

- Model B used the two together well and reached its best score (88 vs 79 with Lightcast alone).
- Models A and C did worse with both than with Lightcast Alone: 81 vs 88 for Model A and 78 vs 84 for Model C. When web figures that disagreed with Lightcast’s made it into those answers, the judge marked them down.

So why lead with web search plus Lightcast? Because that’s the realistic setup. Most AI tools now have web search on, and it’s the comparison a buyer actually faces. On average, the two Lightcast setups landed almost on top of each other (83 alone, 82 with web search).

Ask for a number, get a number with Lightcast

A 25-point jump in overall quality is a big number, but what does it look like in an actual answer? Start with the most basic test there is: if you ask for a number (a count, a wage, a percentage), do you actually get one?

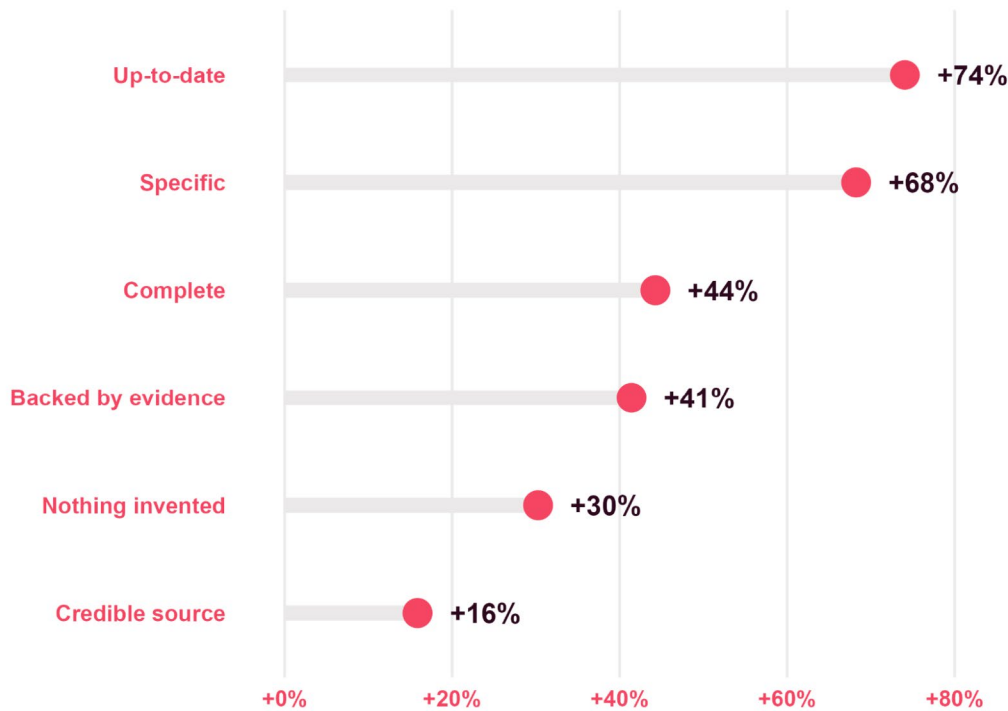


On the 75 questions in our web search study that ask for a specific figure:

- Web search alone: 35% of the time
- Web search + Lightcast: 92% of the time

Ninety-two percent. For anyone who’s ever asked an AI a direct question and gotten three paragraphs of context and no answer, that difference speaks for itself.

Breaking down the score: where answers got better



Getting the number is only one part of a good answer. Our quality score has six parts, evaluating each answer on how specific the answer is to whether the data matches the time period requested. Adding Lightcast to AI moved every one of them.

Figure 6. Lightcast improved all six quality measures, most of all specificity and freshness. Score on each measure out of 100, web search alone vs. web search plus Lightcast. Labels show the gain in points. Three AI models, 120 questions each.

The biggest gains landed exactly where web search struggles most: specific, complete, up-to-date, supported answers.

THREE THINGS STAND OUT:

1. **Answers measure instead of describe.** Web search answers often describe the market ("demand is strong"). Lightcast answers measure it (+68%), and completely answer the question (+44%)
2. **The dates line up.** Lightcast answers state the period they cover, and that period matches the question. Search results often carry a figure from a different year, or from a page with no date at all. Up-to-date had the lowest web-search score of any measure (48), and it saw the second-largest gain (+35 or 74%).
3. **"Backed by evidence" rose by 41%.** Grounding the answer in one consistent dataset means the claims have something concrete behind them

Web search did hold its own on one measure: credible source. Search-enabled answers cite more outside sources by design, and the judge gave them credit for it. On that measure, web search alone and Lightcast alone came out about even (60 and 61). But citing a source and answering the question with the right number are two different things. The open web rarely has the specific, current, consistently defined figures workforce questions need, so there's only so far that web search can take you.

Gains in every category, largest where the data matters most

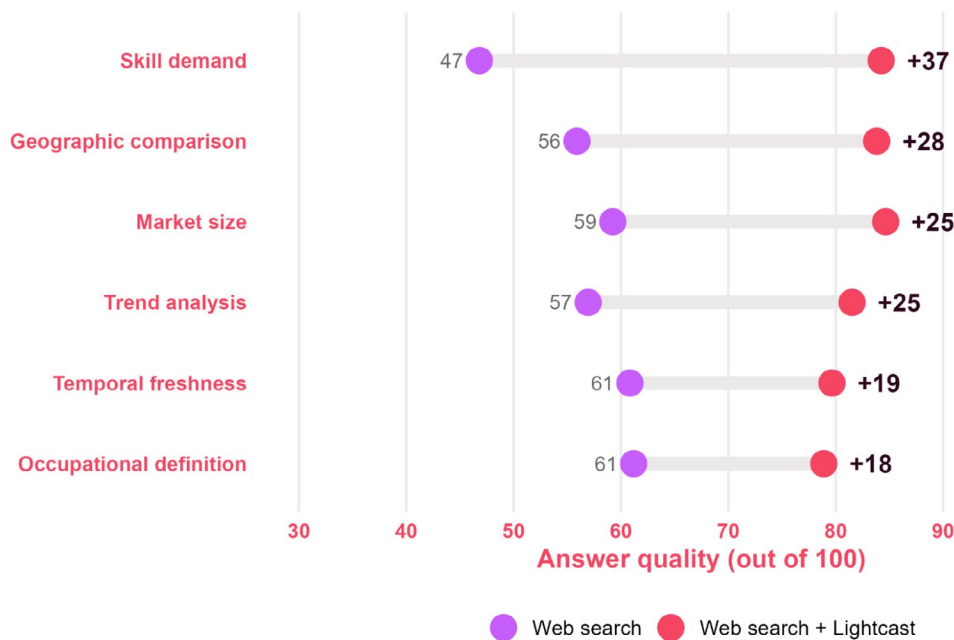


Figure 7. Every question type improved, most where the data isn't on the open web. Answer quality out of 100, web search alone vs. web search plus Lightcast, sorted by gain. Three AI models, subset of 120 questions, 20 questions per category.

Lightcast improved answers in every question category we tested. The biggest gains came where the answer depends on data you can't easily find on the open web: skill demand, geographic data, market size, and trend analysis.

Why skill demand? Think about where that information actually lives. Which skills employers want for a specific role, in a specific region, right now isn't something a web page usually spells out. It has to be measured, posting by posting. Skill demand was also where web search struggled most: it scored 47, the lowest of any category. Without Lightcast, the AI describes the skills a role typically needs. With Lightcast, it tells you how often each skill actually shows up in current postings.

The smallest gains came in occupational definitions (+18) and temporal freshness (+19). Definitions lean less on ever-changing data. And web search already did comparatively well on "what's new" questions (61), likely because recent trends get news coverage.

When it had both web search and Lightcast, the AI mostly used Lightcast

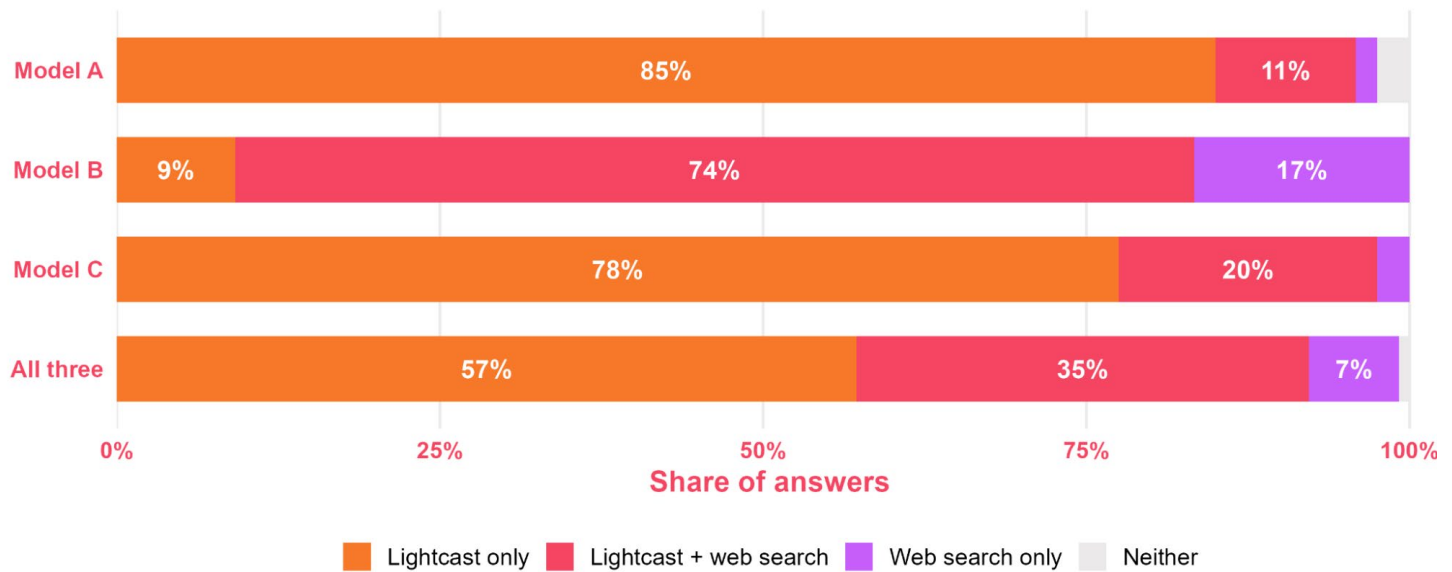


Figure 8. Given the choice, AI chose Lightcast. Share of answers by which tools the AI actually used when both web search and Lightcast were available, with no instructions about which to use. “Lightcast” means at least one Lightcast data query, and “web search” means at least one search. 120 questions per model; models are anonymized.

With both web search and Lightcast available, and no instructions from us about which to use, the AI queried Lightcast data in 92% of answers. It searched the web in 42%:

- 57% of the time, it skipped the web entirely and answered from Lightcast alone
- 35% of the time, it used both
- 7% of answers relied on web search alone, mostly for questions that called for definitions rather than numbers

Not every model made the same call. Two of the three went to Lightcast alone for most answers (85% and 78%). The third liked to check both: it used both in three out of four answers. But all three treated Lightcast as a source worth using, and nobody told them to.

This isn’t proof on its own that Lightcast is better. AI models tend to use a specialized tool when one is available, and the MCP’s tool descriptions tell the AI how to use it. But it does mean that, in practice, connecting Lightcast changes where your AI gets its numbers.

Where public data exists, search and Lightcast agree

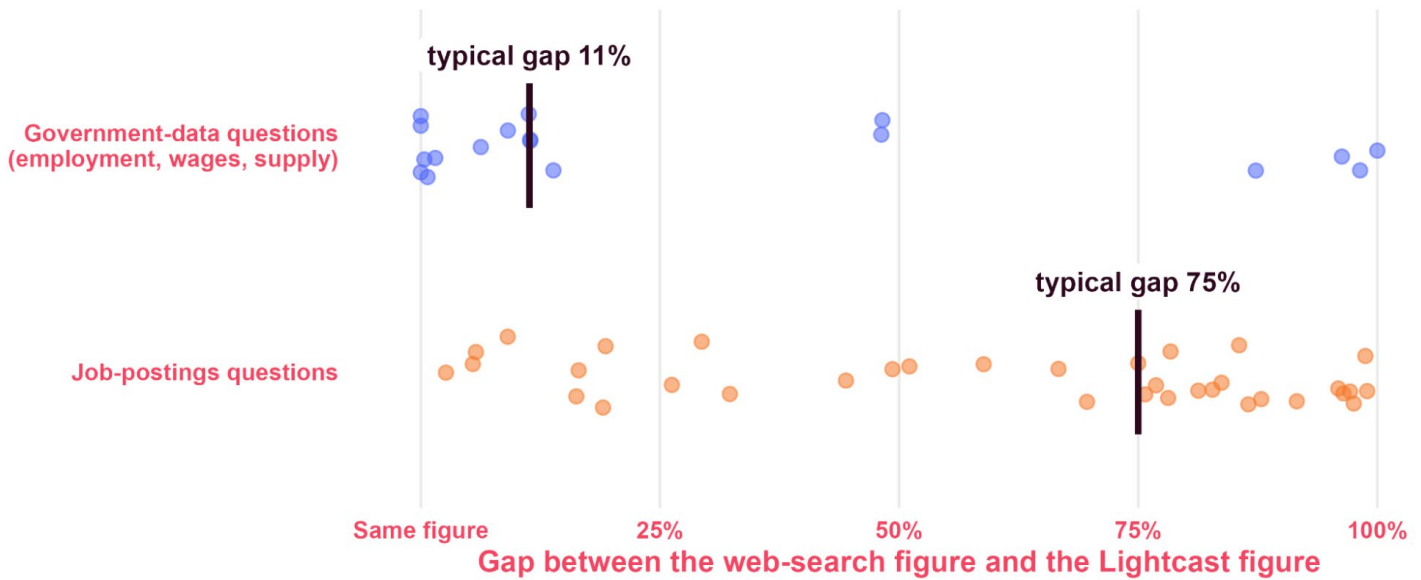


Figure 9. Search and Lightcast agree where public data exists, but not on job postings. Each dot is one figure given by both web search alone and web search plus Lightcast, for the same model, question, and unit (53 pairs). The gap is the difference as a share of the larger figure, and the line marks the typical (median) gap. This measures agreement between the two sources, not accuracy.

Getting a number is one thing. But when web search did find a figure, how did it compare with Lightcast's?

It depended on where the number lives.

Lightcast data about employment, counts, wages, and labor supply are partly sourced from public government sources, where they are then combined with multiple data sets, unsuppressed, and added to our historical data repositories. On those, the search figure and the Lightcast figure typically landed within about 11% of each other. That's what you'd hope to see. Both are drawing on the same public data, and they mostly agreed.

Job postings questions were a different story. There's no public, deduplicated count of job postings for web search to find, so it grabbed whatever was nearby: a job board's live listing count, a staffing firm's annual estimate, a press release. On those questions, the two figures typically differed by 75%. Put another way, the search figure was often a quarter of Lightcast's, or four times it.

NOT JUST A JOB-POSTING STORY.

On the 50 questions that don't need job postings data (employment, wages, labor supply), adding Lightcast to web search still raised quality by 18 points, and 78% of answers improved. On the 70 job-postings questions, the gain was 31 points.

Agreement isn't accuracy. Two numbers can agree and both be off. But the pattern is telling: where public data exists, web search and Lightcast mostly agree. Where it doesn't, web search fills the gap with whatever number it can find, and puts a link next to it.

SO WAS THIS ONE MODEL GETTING LUCKY? NO.

- Every AI model we tested improved with Lightcast. With web search on, adding Lightcast raised Model A by 24 points, Model B by 14, and Model C by 38. It improved 75% to 95% of each model's answers. With web search off, the three models improved by 38 to 58 points.
- Every question type and difficulty level improved. The smallest category gain was +18 points (occupational definitions), and the smallest difficulty gain was +22 (easy questions).
- The numbers held steady. Asked five times with web search off, the AI gave a consistent figure for 49% of the figures we asked about with Lightcast, and 1% without it.

What This Means for You

So what do you actually do with all of this?

If you're already using AI for workforce questions, the takeaway isn't "stop." It's this: *the answer is only as good as the data the AI can reach*. Connected to Lightcast, the same AI you already use gave the specific number you asked for 92% of the time. On its own with web search, it was closer to one in three.

HERE'S WHAT THAT LOOKS LIKE DEPENDING ON WHERE YOU SIT.

For HR and talent leaders: Ask workforce questions in plain language and get answers you can put in front of your executive team, each with a count, a time period, and a source. Planning headcount, benchmarking pay, or sizing a talent pool no longer means choosing between fast and defensible.

For talent acquisition and recruiting: Know how competitive a role is before you open it: how many employers are hiring, which skills they're asking for, and where the talent is. All in the AI tool you already use.

For workforce and economic development: Can your region staff a new data center? Which skills gaps are holding back local employers? Questions like these get answers grounded in current postings and government data, not whatever a web search happens to surface.

For education leaders: Which programs should you expand, update, or create? Understand employer demand, emerging skills, and career opportunities so you can align curriculum with the changing labor market and help students prepare for the jobs ahead.

For marketing and strategy teams: Labor market context for market sizing, competitor hiring signals, and messaging, with figures you can cite without crossing your fingers.

For teams building AI products: Connecting workforce assistants and agents to Lightcast's labor market intelligence through MCP enables them to draw on structured data with consistent definitions, explicit time periods, and identifiable sources. This gives users a clearer basis for assessing the evidence behind an answer or recommendation and determining whether it supports their decision.

THREE QUESTIONS TO ASK OF ANY AI ANSWER

Whether or not your AI is connected to Lightcast, three quick checks will tell you a lot about whether an answer is decision-grade.

1. **Did it give the number I asked for?** Not context. Not a range of "estimates suggest." The number.
2. **Does the time period match my question?** A 2023 figure isn't an answer to a question about the last six months, no matter how nicely it's cited.
3. **Can I tell exactly what was counted?** Which occupation, which geography, which source, and whether duplicates were removed.

If the answer to any of these is "not really," the AI hasn't saved you time. It's handed you homework.

THREE QUESTIONS TO ASK OF ANY AI ANSWER.

1. Did it give the number I asked for?
2. Does the time period match?
3. Can I tell exactly what was counted?



Conclusion: Better Together

We started with a simple question: *how confident can you be in the accuracy of AI's answer?*

Answer: On its own—not very confident.

Even with web search turned on, AI gave the specific figure you asked for only about a third of the time, and nearly two out of three answers contained unsupported or contradicted numbers, often with a link sitting right next to them.

Connected to Lightcast, it's a different story. Answer quality jumped from 57 to 82 out of 100, and nearly nine out of ten answers got better. Every AI model we tested improved with Lightcast. Every question type improved with Lightcast. Answers were more consistent with Lightcast than without. We asked the same questions again and again (and again). The conclusions held.

Lightcast's advantage isn't a one-time result.

AI is changing how everyone works, and that's a good thing. But an answer that only sounds right isn't good enough when it's shaping hiring plans, budgets, student's futures, and people's careers. What changes the outcome isn't a smarter model or a longer prompt.

So no, **this isn't about choosing between AI or Lightcast. It's about the data the AI can reach.** With Lightcast behind it, the AI you're already using gives answers that are more specific, more current, and more defensible: answers you can actually stand behind.

Ready to try our MCP?

Connect compatible AI tools to Lightcast's labor market intelligence through MCP for workforce answers grounded in structured, traceable data. Our team can help you explore supported tools and access options.

[**REQUEST A DEMO**](#)



APPENDIX

Inside the Lightcast MCP Toolbox

In our testing, the AI picked its own tools. We didn't tell it which ones to use, or when. That makes watching what it reached for genuinely interesting. It's a bit like looking over an analyst's shoulder as they work through a question: first figure out where the data lives, then pin down exactly what's being counted, then run the numbers. Let's dig into each tool.

What's Inside the MCP

The Lightcast MCP includes the Glossary, Tool Info, Classify, Count, Top, Timeseries, Estimate, Distribution, and List tools, across job postings, profiles, and government occupation, industry, and unemployment data.

GLOSSARY AND TOOL INFO: KNOWING WHERE TO LOOK

What they do:

- Glossary describes each Lightcast dataset and what it's best for.
- Tool Info gives the exact structure for querying it.

How the AI used them:

- In 98% of answers, the AI started with the Glossary to choose the right dataset—postings for demand questions, profiles for supply questions, government data for employment and wage questions.
- Tool Info was used in 95% of answers

Why it matters:

- Think of it as reading the table of contents before flipping to a chapter. The AI doesn't guess which data applies. It picks the dataset that was built for the question.

CLASSIFY: SPEAKING THE SAME LANGUAGE

What it does:

- Maps the words in a question to Lightcast's standard taxonomies for occupations, skills, industries, and locations. "Software engineer," for example, becomes Software Developers (SOC 15-1252).

How the AI used it:

- In 94% of answers.

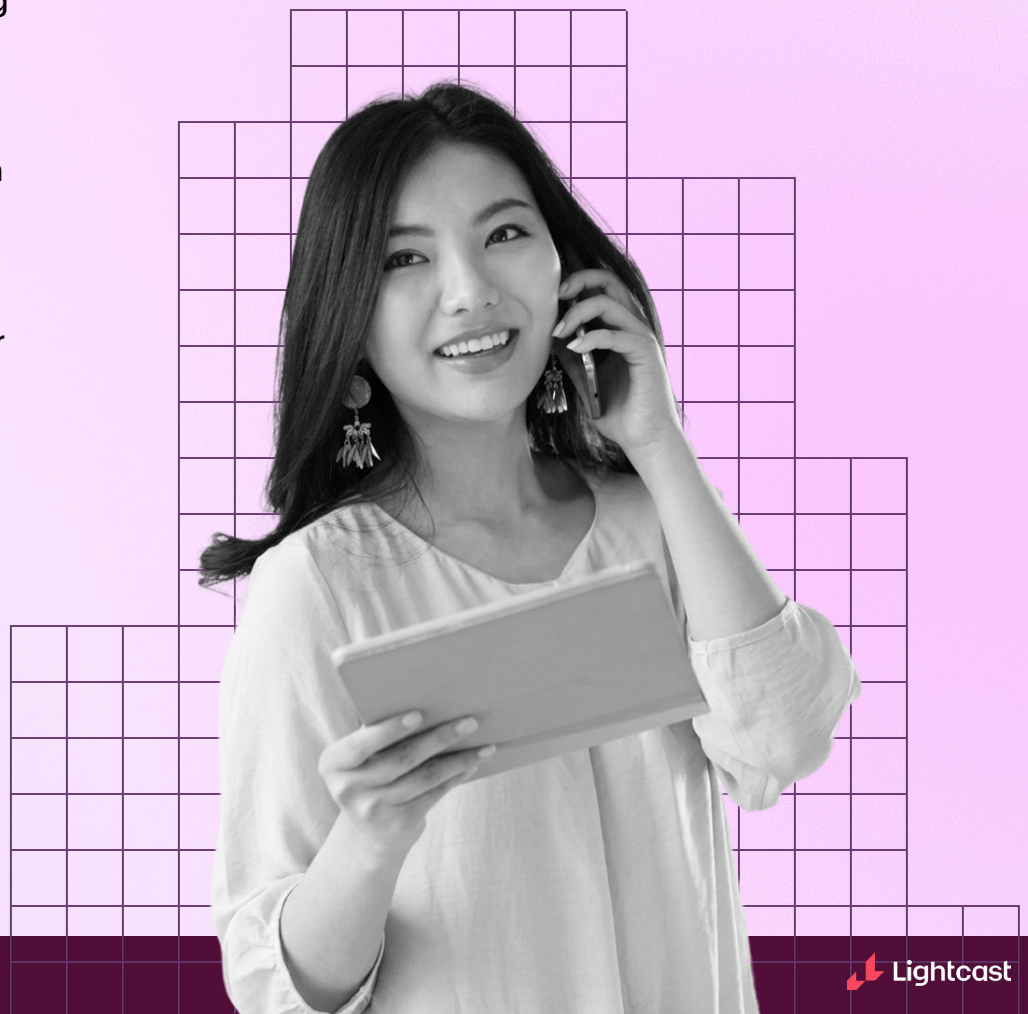
Why it matters:

- Ask five people to sort the same pile of job titles and you'll get five different filing systems. Classify gives every answer the same one. That's the difference between a keyword search and a defined count: two people asking about "data analysts" get the same population, and the answer tells you exactly what it counted.

EXAMPLE:

"What's the difference between a logistics coordinator and a supply chain analyst in the U.S. labor market?"

- Web search alone returned job-description prose.
- With Lightcast, the AI classified both roles, then compared them on:
 - Talent pool size (2.6× more logistics coordinators)
 - Median compensation (about \$72K vs \$76K)
 - Education (89% of supply chain analysts hold a bachelor's or higher)
 - Core skills and primary industries



COUNT: HOW MANY, EXACTLY

What it does:

Returns totals, such as unique job postings or profiles, for any combination of occupation, skill, location, employer, industry, and time period.

How the AI used it:

In 56% of answers overall, and 87% of market size questions.

EXAMPLE:

"What is the total count of paralegal openings listed by U.S. employers over the last 6 months?"

Web search alone couldn't find a figure for the period asked about. The closest it found was a staffing firm's count for calendar year 2025, and it rightly declined to pass that off as a six-month number.

With Lightcast: 47,089 unique paralegal and legal assistant postings, April 1 – October 1, 2026, deduplicated across sources.

Same question, very different outcomes. One answer could only tell you what it couldn't find. The other told you exactly what it counted, and when.

TOP: WHO AND WHAT LEADS

What it does:

Ranks the leading skills, employers, locations, or titles by a chosen metric.

How the AI used it:

In 43% of answers overall, and about two-thirds of skill demand questions (68%) and 58% of temporal freshness questions.

EXAMPLE:

"Which phlebotomy skills are urgent care clinics in Arizona requesting for medical assistant openings right now?"

Web search alone quoted individual job ads.

With Lightcast, the AI ranked skills by how often they appear in current postings: venipuncture in 68% of postings, specimen labeling in 65%. It also listed the skills most often requested alongside phlebotomy, such as vital signs and EKG.

TIMESERIES: HOW THINGS ARE CHANGING

What it does:

Returns metrics month by month or year by year, for trends, seasonality, and before-and-after comparisons.

How the AI used it:

In 41% of answers overall, and 78% of trend analysis questions.

EXAMPLE: *"How has JPMorgan Chase's hiring for financial analysts in its New York operations evolved over the past 12 months?"*

Web search alone reported company-wide headcount and a job board snapshot.

With Lightcast, the AI reported about 60 unique postings per month, compared the October–December 2025 average (about 58 per month) with the July–September 2026 average (about 61 per month), and concluded that hiring was steady. It also listed the top requested skills.

"Steady" isn't a flashy headline. But it's an answer—and it's the answer the question asked for.

LIST: THE UNDERLYING RECORDS

What it does: Returns a sample of individual records, such as specific job postings, for when someone wants to see the examples behind the numbers.

How the AI used it: The AI used it rarely (3%), and only when a question asked for examples.

ESTIMATE AND DISTRIBUTION: WAGES, INTENSITY, AND SPREAD

What they do:

Estimate calculates medians and other derived measures, such as advertised salary, posting intensity, and posting duration.

Distribution shows how a measure spreads across ranges, such as salary percentiles.

How the AI used them:

Estimate in 22% of answers overall, and about 40% of occupational definition questions.

Distribution came up rarely, in specialized wage questions.

EXAMPLE: *"What's the difference between an operations manager and a plant manager in the food processing sector?"*

Web search alone described the two roles.

With Lightcast, the AI added a measured pay difference (median about \$90K vs \$104K) and the specific skills that set plant managers apart (HACCP, lean manufacturing).

Methodology

STUDY DESIGN AT A GLANCE

This study ran in two stages. First came a core experiment to establish a baseline: 600 questions, three AI models, with and without Lightcast. Then we ran a set of stress tests on a 120-question subset, each designed to answer one skeptical question about the core result. Would the judge score the same answers the same way twice? Would we get the same result if we asked again? Does a newer model still benefit? What happens when the AI can search the web?

Study	Questions	Models	Setups Compared	What it tests	Where it appears
Core Experiment	600	A, B (older model), C	No tools vs Lightcast	Does Lightcast improve AI answers?	How We Tested; Conclusion
Judge consistency	120	Core answers, re-scored	Same answers, scored 5 times	Is the scoring reliable?	How We Tested
Answer consistency	120	A, B (older model), C	Each question asked 5 times	Is the result a lucky draw?	How We Tested; Conclusion
Model swap	120	B: older vs newer model	No tools vs. Lightcast	Does a newer model still benefit?	How We Tested
Web search study	120	A, B (newer model), C	No tools, web search, Lightcast, web search + Lightcast	Does Lightcast still help when AI can search the web?	What We Found (headline)

WHY THE PAPER REPORTS MORE THAN ONE “BEFORE AND AFTER.”

- The core experiment (600 questions, web search off) takes AI from 30 to 70 with Lightcast.
- The same comparison inside the web search study (the 120-question subset) takes it from 38 to 83.
- The headline (web search on in both setups) goes from 57 to 82.

The subset itself isn’t the reason the first two differ: on the core experiment, the 120 subset questions scored almost exactly like the full 600 (69 vs. 70 with Lightcast, 30 vs. 30 without). Most of the difference is Model B. On those same 120 questions, its earlier model scored 44 with Lightcast in the core experiment, and its current model scored 79 in the web search study. Models A and C also scored a few points higher in the later study (88 vs. 82, and 84 vs. 81). The rubric and connection type were the same; the answers were collected two to three months apart, and the models and the data may have changed in between. Different questions, model versions, and dates mean the numbers differ. The direction doesn’t.

QUESTIONS

The question bank.

We wrote 600 labor market questions, 100 in each of six categories: market size, trend analysis, geographic comparison, skill demand, occupational definition, and temporal freshness.

- **Drafting:** questions were drafted with an AI model from seed examples and design rules, then validated against those rules and reviewed by subject matter experts before continuing.
- **Difficulty:** within each category, the questions split evenly across four levels: easy (a single occupation, nationally), medium (a specific sector or employer), hard (niche roles, cross-sector comparisons, specific time windows), and expert (multi-employer, multi-dimensional comparisons).
- **Time windows:** spread evenly from "right now" to "the past 24 months."
- **Demand wording:** varied (openings, vacancies, roles, positions, opportunities), so the questions read the way people actually ask.

The 120-question subset.

All stress tests use the same fixed subset: 20 questions per category and 30 per difficulty level, five per category–difficulty combination. The questions were selected at random. Using one subset keeps the stress tests comparable with each other. It also keeps costs manageable, since some tests ask each question up to five times and the core experiment found no large variation across categories or difficulty levels. The benefits of running every stress test on the full question set simply weren't there.

Questions that need job postings data.

70 of the 120 subset questions need job postings data to answer (for example, counts of openings or skills requested). The other 50 can be answered from other data, such as employment, wages, or labor supply. Some analyses report these two groups separately. Truthfully, this wasn't part of the experimental design. The random selection of the 120 question subset just worked out this way.

MODELS AND SETUPS

Models. We tested models from three leading AI providers, labeled Models A, B, and C. For Model B, the core experiment and answer-consistency test used an earlier-generation model. The model swap, web search study, and figure-stability test used the provider's current-generation model.

Setups. Depending on the study, each question was answered in up to four setups:

- **No tools:** the model answers from what it already knows
- **Web search:** the provider's own built-in web search
- **Lightcast:** the model is connected to the Lightcast MCP
- **Web search + Lightcast:** both tools are available, and the model chooses which to use

No coaching. Every answer started from the question alone: no system prompt, no instructions to use (or avoid) any tool, and no follow-up. Each model decided for itself whether, when, and how to use the tools it had.

DATA COLLECTED

For every answer, we kept:

- **The full answer text**
- **A complete trace:** every message, every tool call and its arguments (including the exact query sent to Lightcast or the web), every result returned, and the order of the steps
- **Run details:** input and output token counts, response time, and whether the run completed, failed, or was retried
- **Scores:** the judge's score and written reasoning for each of the six quality measures
- **Extracted figures, for questions that ask for a specific number:** each figure the answer gave for the asked quantity, with its unit and time period

Every study's settings were determined before it ran: the questions, model versions, setups, judge, and rubric. Each answer is stored under a fixed identifier, so any number in this paper can be traced back to the answers and scores behind it.

HOW ANSWERS WERE SCORED

The judge. Every answer was scored by an AI judge, a model from one of the three providers tested. The judge never answered questions itself.

As a check on self-preference: in the web search study, the judge’s own provider scored lowest of the three with no tools (29, vs. 41 and 44). There’s no sign it went easy on its own family.

What the judge saw. For each answer, the judge saw:

- the question, its category, and the evaluation date
- the answer to score, with no model name or setup
- the same model’s Lightcast-only answer to the same question, clearly labeled as a reference

The reference was there for a separate factual-alignment check, which isn’t part of the quality score. It also feeds one rule in “Nothing invented” (below).

So the judge was blind to labels, not to content. Answers often reveal their own setup (“according to Lightcast job postings data,” or a list of web links).

The rubric. The judge scored six measures, 0–4 each:

Measure	4	3	2	1	0
Backed by evidence	Claims consistently well supported, with concrete evidence, clear scope or method, and context	Most substantive claims supported with concrete evidence or method	Some support, but important claims weakly substantiated	Mostly assertions, little support	Unsupported; no meaningful evidence
Up-to-date	Current and explicit: dates ending within about a month of the evaluation date, matching the window asked for	Recent relative window ("last 6 months") without exact dates	1–3 years old, or only partly overlapping the window asked for	More than 3 years old, or far outside the window asked for	Undated or no time signal
Specific	Precise, relevant figures, including rates, counts, or changes as requested	Relevant figures, but missing some requested measures	Limited or incomplete figures	Vague, non-numeric language where numbers are expected	No numbers when the question calls for them
Credible source	Named high-quality source, such as Lightcast, BLS, or another identifiable dataset or report	Named source, vaguely described	Generic mention ("according to reports")	Weak or unverifiable source	No source
Nothing invented	No invented facts or sources; engages the question	Minor hedged speculation, or an honest "I can't know"	Unsupported numbers or sources, or confident claims that conflict with the reference answer	Major invented facts or figures	Fabricated agencies, reports, datasets, or precision
Complete	Fully answers every requested metric, entity, place, period, and comparison	Answers the main request, minor omissions	Partly answers, or substitutes important elements	Addresses only a small part	Doesn't address the question

The quality score. The six scores sum to a quality score out of 24, which we report on a 0–100 scale ($\text{score} \div 24 \times 100$). Single measures are reported the same way ($\text{score} \times 25$).

FIVE THINGS ABOUT THIS RUBRIC A SKEPTIC SHOULD KNOW:

1. “Credible source” names Lightcast (alongside BLS) as an example of a high-quality source.
2. “Specific,” “Complete,” and “Up-to-date” reward answers that give numbers for the period asked. That’s what these questions ask for, but it favors any tool that reliably returns a number.
3. An honest “I don’t know” scores 3, not 4, on “Nothing invented.” It invents nothing, but it doesn’t answer the question.
4. “Up-to-date” is judged against the evaluation date. That penalizes official statistics, which lag by design, including Lightcast’s own government-data answers.
5. “Nothing invented” partly measures agreement with Lightcast. A confident figure that conflicts with the Lightcast reference answer scores at most 2. So a web figure that differs from Lightcast’s can count against “Nothing invented,” even when neither figure is shown to be wrong. A Lightcast-only answer can’t conflict with itself.

That’s why the paper doesn’t claim Lightcast “eliminates” hallucinations. Two checks show this measure doesn’t drive the results:

- The headline holds without it. Leaving “Nothing invented” out of the quality score entirely, the gain from adding Lightcast to web search is +27 points (vs. +25 with all six).
- The other measures don’t lean on the reference. In the judge’s written reasoning, the reference answer came up in about 1–13% of scores on the other measures. The exception was “Specific,” at about 20% for no-tools and web-search answers.

ANALYSES COMPLETED

Head-to-head comparisons. We always compared the same question, asked of the same model, in two setups (e.g. with Lightcast data and without).

Why it matters: some questions are simply harder than others. Comparing like with like means a gain reflects Lightcast, not an easier set of questions. We also checked that a handful of extreme answers weren't driving the average.

Margins of error. Every headline number comes with a range of plausible values, calculated by re-running the analysis many times on random reshuffles of the questions (bootstrapping).

Why it matters: any study can get lucky with its question set. A gain that holds up across thousands of reshuffles is a real gain, not a fluke.

Statistical models that account for difficulty. We used models that give each question its own baseline difficulty (hierarchical models).

Why it matters: a simple average treats every question as equally hard. These models confirm that the gain holds once difficulty is taken into account. They also show whether some kinds of questions benefit more than others.

Reliability checks. We scored the same answers five times, and asked the same questions five times.

Why it matters: if the judge were inconsistent, or the AI gave wildly different answers each time, a one-time difference could just be noise. These checks show how much of a score is signal and how much is chance.

Robustness checks. We repeated the key comparisons under different conditions:

- a different version of the scoring rubric
- a different way of connecting to the MCP
- on job-postings and other questions separately
- with a newer model

Why it matters: a result that depends on one particular choice isn't one you can count on. One that survives all of them is.

Trace analysis. We read the step-by-step record behind every answer: which tools the AI called, what it asked Lightcast for, whether data came back, and whether the numbers in the answer came from that data.

Why it matters: scores tell you that answers improved. The traces tell you why. They show the gain comes from the AI actually reaching Lightcast data, rather than from the answers just sounding more confident.

Number matching. For questions that ask for one specific figure, we pulled the figure out of each answer and compared it across setups and repeated asks.

Why it matters: a quality score can't tell you whether you got the number you asked for, or whether you'd get the same number tomorrow. This analysis can, and it's the question most people care about.

WHAT THESE RESULTS DO AND DON'T TELL YOU

Scores measure answer quality, not truth. An AI judge scored each answer against the rubric above. That isn't an independent check that every figure is correct, and no human expert reviewed every score for this paper.

We didn't benchmark Lightcast's figures against an outside source. Where web search and Lightcast figures agreed, both could still be off. Where they differed, the gap shows disagreement, not which one is wrong.

Definitions drive the counts. Each figure depends on the occupation category, the time window, and whether postings were counted as posted or active. The AI chose those itself, and a different choice would give a different number.

Models change quickly. Results reflect the models, tools, and data available at the time of testing.

