

SEER GEO STUDY ON AI BEHAVIOR

What AI Thinks About Your Brand Is Already Written

We tested **5 hypotheses across 231,347+ LLM responses, 7 AI Platforms, and 52 days of live data** from the 2026 Winter Olympics to determine how AI decides who gets mentioned, who gets recommended, and why.

231,347+

LLM RESPONSES
ANALYZED ACROSS
7 LLMS

52

DAYS OF LIVE
DATA OVER THE
2026 WINTER
OLYMPICS

5

HYPOTHESIS TESTS
TO STUDY WHAT
BRANDS SHOULD
LOOK AT IN 2026

AI doesn't "find the best answer". It completes the most familiar story.

We used the 2026 Winter Olympics as a live laboratory — unpredictable results, fast-moving news, AI working in real time. The patterns we found apply to your brand daily.

Large language models predict the next most likely token based on patterns in training data. An entity that appears frequently, described consistently, across independent domains, becomes stable in that distribution. That stability is what determines what AI says about you: accurately, confidently, and in what context.

Five hypotheses, five different angles. One conclusion: the brands AI recommends are the brands that built signal architecture before the question was ever asked.

Six findings that should change how you measure AI visibility

1. Dominant narratives get completed before they get verified

ChatGPT declared Chloé Kim's three-peat before the event happened and cited real sources. AI doesn't hallucinate from nothing, but it will finish the story your brand's signal architecture has already written.

2. Platforms without web access can't integrate your news

Meta AI still says Lindsey Vonn retired in 2019. Her comeback, her 2026 Olympic run: none of it exists. If you're issuing press releases expecting all AI platforms to update, half of them never will.

3. Models fabricate with confidence and specificity

After Ilya Malinin, the "Quad God," finished eighth, Meta AI reported his score as 194.63 when his actual score was 156.33. This was specific enough to seem credible, but wrong enough to matter— and delivered without hesitation.

4. Disclaimers accompany hallucinations

Claude acknowledged its knowledge cutoff, then described Malinin as a strong gold contender the day after his collapse. The model flagged uncertainty and kept generating. Caveats are not corrections.

5. Facts update faster than the narrative frame around them

Three weeks after events contradicted them, 1 in 5 factually correct responses still told the old story. Most monitoring tools only catch factual errors, so stale framing and narratives are often missed.

6. Winning doesn't immediately change what AI says about you

Norway won the medal table with 26 medals, but AI mentioned the US at more than 2× Norway's rate for 52 straight days. Performance after the training cutoff is irrelevant because the story was already written.

What You'll Take Away From This Study:

- **Whether AI is telling your current story, or an outdated one:** Standard monitoring checks whether AI mentions or cites your brand. It doesn't check if AI has the right facts, or whether AI is telling your current story. Most teams don't know that these are not the same measurement. [Seer](#) helps clients understand their AI visibility and brand positioning.
- **Why your last campaign didn't increase AI visibility and what would have:** The investment sequence is wrong for most brands. This study shows what signal you should focus on and why amplification without foundation registers as noise, not presence.
- **Which signal your brand is missing for AI visibility:** Most brands have one of three signals in reasonable shape. This study shows what that gap looks like from the outside and make it easy to recognize for yourself.
- **How to leverage AI to reveal your brand's problem:** Mention rate tells you if you're in the room with the right people. Brand accuracy tells you what the room thinks of you. There's a specific analysis that can surface the second problem that most monitoring tools don't account for, but Seer can help.
- **How to build a foundation for your GEO effort to last :** When training distribution is set, winning after the fact doesn't change that. Understanding that idea changes how you think about the timeline, and the urgency around it.

Table of Contents

Executive Summary	2
The Hypothesis Tests	5
Section One: The GEO Visibility Problem	6
<i>Hypothesis Test: Citation Authority</i>	7
<i>Hypothesis Test: Geographic Bias</i>	9
<i>Hypothesis Test: Social Proof</i>	10
Section Two: The GEO Knowledge Cliff	12
<i>Hypothesis Test: Temporal Velocity</i>	13
<i>Hypothesis Test: Narrative Persistence</i>	14
What Changes for Marketers Today?	16
Conclusion	21
Methodology	23

The Hypothesis Tests

Hypothesis	What We Found	Verdict
H1 – Citation Authority Does question intent change which sources AI trusts?	Factual queries pull institutional sources at 76.4%. Judgment queries shift to prestige editorial (+157%), social/UGC (+165%), expert voice (+873%). No two platforms share the same trust hierarchy.	Confirmed
H2 – Temporal Velocity Does web access create a speed gradient?	Three tiers, not two. <ol style="list-style-type: none"> 1. Same-day platforms updated within hours. 2. Near real-time platforms took one to two days. 3. Two platforms never updated across the full 52-day study. The Binary Cliff is the boundary between worlds, LLMs either retrieve from the Web or they don't.	Confirmed
H3 – Narrative Persistence Do models keep telling old stories after facts change?	1 in 5 factually correct responses used pre-event narrative language three weeks later. Two distinct failure modes: hallucination AND stale framing. Ambiguous outcomes are 3× harder to correct than clean reversals.	Confirmed – Stronger Than Predicted
H4 – Geographic Bias Does training data create structural defaults by language?	Systematic US bias confirmed. US appeared in 36.8% of neutral prompts vs. 16% for Norway, the actual medal leader. Structural, driven by English-language training data, stable across all three study phases.	Partially Confirmed
H5 – Social Proof Does social signal volume predict LLM visibility?	10 of 11 social proof signals significant at $p < 0.05$. Composite score ($\rho = 0.590$, $p = 0.002$) validated as predictor. Reddit cited cross-platform at 3–15× the rate of Instagram, TikTok, or Twitter/X.	Confirmed

The GEO Visibility Problem

The Marketing Problems This Section Answers

You get listed but not recommended.

You are on Wikipedia. You have press coverage. You show up on factual queries and disappear on judgment ones.

*Hypothesis Test: **Citation Authority** explains why- the sources that establish your existence are not the sources that establish your significance. Anchor coverage gets you into the answer. Judgment-layer coverage gets you recommended. AI favors competitors in markets where you win.*

You outperform a competitor on measurable outcomes. AI still mentions them first.

*Hypothesis Test: **Geographic Bias** explains why- training data carries structural defaults that results cannot override. Norway won the medal count. AI never noticed.*

Your biggest campaign moved nothing in AI.

You ran the launch. Press hit. Wikipedia views surged. AI visibility barely changed.

*Hypothesis Test: **Social Proof** explains why- campaigns amplify existing authority, they cannot manufacture it. The baseline predicts everything. The surge predicts nothing.*

These are not three separate problems. A brand invisible on judgment queries, operating in a structurally disadvantaged language environment, with only one signal layer built up is facing one stacked visibility gap with three dimensions. Fixing one will not fix the others, which is why most GEO programs plateau after the first round of optimization.

Seer Interactive found that fixing one will not fix the others, which is why most GEO programs plateau after the first round of optimization, not Seer's.



HYPOTHESIS TEST: CITATION AUTHORITY

Citation Authority: The Prompt Shapes the Citation Source

You are a brand manager.

Imagine that you convinced your boss to spend the budget and buy the ad on Olympics.com during the Olympics (for a hefty price I'm sure) and you're also throwing advertising spend at NBC Sports. You have a solid brand reputation with a corporate-approved Wikipedia page. By any reasonable GEO standard, you are in good shape.

Now consider that a user asks ChatGPT who is favored to win your category. ChatGPT does not reach for Olympics.com. It reaches for ESPN, Bleacher Report, Reddit, and an odds aggregator. Your coverage is real, but it's just irrelevant to that specific question.

This is the Citation Authority problem. LLMs route to different source types depending on what kind of answer is being constructed. The sources that establish your existence are not the same sources that establish your significance. LLMs know the difference.

Two Layers, One Answer

The citation authority analysis covered 25,084 responses across seven platforms. Responses split into two cohorts:

Factual queries: "Who won the men's downhill?"

Judgment queries, which break into three types:

- Predictive ("Who will win gold in Women's Slalom?")
- Evaluative ("What are the biggest storylines of the Games?")
- Opinion-seeking ("Is Vonn's comeback realistic?")

Every LLM response is built on the same foundation. The anchor layer is always there. Judgment queries add a second layer on top of it.

Olympics.com held at roughly 48% citation rate regardless of query type. Wikipedia held at roughly 28%. LLMs cite these whether the question asks for a fact or a judgment.

The judgment layer is where the divergence happens. Sports editorial (ESPN, Sports Illustrated, Bleacher Report combined) appeared in 16.2% of factual responses and 28.0% of judgment responses, an 11.8 percentage point gap. Betting sites went from 0.8% factual to 6.5% judgment, an 8.5x increase. Reddit went from 1.9% factual to 6.6% judgment, a 3.4x increase.

If your brand only has anchor layer coverage, you get listed. The judgment layer is what gets you recommended. Those are not the same outcome, and most brands are confusing one for the other.

No Two Platforms Trust the Same Sources

Among the five web-access platforms, source preferences diverge substantially.

- Google AI Mode cited Olympics.com in 91.2% of factual responses.
- ChatGPT's top factual source was Wikipedia at 51.7%.
- Perplexity's editorial citation rate for judgement queries was 54.1%, the highest of any platform.

These are not interchangeable channels. A GEO strategy built entirely around Wikipedia optimization will underperform on Perplexity. A strategy built around official site authority will underperform on ChatGPT.

Know which platforms your audience uses before deciding which sources to prioritize.

Two of the seven platforms did not participate in any of this.

- Claude cited sources in 0.1% of responses.
- Meta AI cited sources in 0.4%.

Both operate from training data with knowledge cutoffs months before the Olympics began. Citation-based GEO tactics have zero effect on platforms that do not cite. If a meaningful share of your audience uses those platforms, you need a parallel track focused on training data authority, not content freshness.

Reddit Is a Citation Source, Not a Social Platform

ChatGPT cited Reddit in 13.5% of its judgement responses. That is not a social media metric. That is a citation rate. ChatGPT is treating Reddit threads as authoritative sources for opinion-based answers.

This finding connects directly to H5. Reddit's role in social proof is not theoretical. It is measurable in citation behavior across 25,000 responses. Brands actively discussed in relevant subreddits are being cited in AI responses. Brands absent from those conversations are not.



HYPOTHESIS TEST: GEOGRAPHIC BIAS

Geographic Bias: Geography Shapes What AI Sees

Norway won the 2026 Winter Olympics. 26 medals. First place overall. Johannes Hosflot Klaebo became the most decorated Winter Olympian in history during these Games. Klaebo won 11 gold medals by the end of the 2026 Games, more than any other Winter Olympian in history, trailing only Michael Phelps in career Olympic gold medals.

Ask a judgement prompt, such as: "Who dominated the 2026 Olympics?". Norway barely registers. Across 52 days, neutral judgement prompts referenced US athletes in 36.8% of responses. Norway appeared at 16%. The US had 17 medals and finished third. The gap did not close as the Games progressed. Norway's historic performance moved nothing in AI visibility. The medal standings were irrelevant.

The Training Data Problem

The Norway finding is not a personalization story. It is not ChatGPT deciding that American users prefer American athletes. It is what happens when a model trained overwhelmingly on English-language text gets asked a question that appears neutral but carries a built-in answer.

English-language sports media is US-centric by construction. The Olympics get covered through a lens of US athletes, US medal counts, and US storylines. That coverage trained the models to weight US athletes as inherently more significant. Norway gets covered in Norwegian. That content is underrepresented in training corpora relative to its athletes' actual performance.

The bias is structural. It reflects the composition of the text the models learned from and reacts accordingly. For brands, the implication is direct: if your industry's coverage is dominated by US sources and English-language publications, models treat US brands as the default frame of reference regardless of which brands are actually winning in your market.

Geography Shapes Opinions, Not Facts

The judgment question told a different story. Two athletes appeared exclusively from specific locations and nowhere else.

Auston Matthews appeared in 3 of 5 Texas responses and 0 of 20 responses from all other locations combined. Scotty James, Australia's halfpipe world champion, appeared in 1 of 5 Australian responses and nowhere else. A genuine world champion, invisible to every other geography.

ChatGPT did not adjust emphasis for these athletes. It added and removed them entirely based on where the query originated.

HYPOTHESIS TEST: SOCIAL PROOF

Social Proof: Signals Determine Who Rises

We tracked 25 Olympic athletes across 11 social proof signals and 19,615 LLM responses. The correlation between digital footprint and AI visibility was overwhelming. But the most important finding is where it breaks down.

28 of 33 statistical tests came back significant. The validation threshold required 3 signals correlated at $p < 0.05$. We found 10 of 11 signals meeting that threshold.

The Strongest Predictor Is Not What You Expect

Not 30-day Wikipedia views. Not Instagram followers. Not total social reach. The strongest single predictor of LLM mentions was Wikipedia activity in the last 7 days ($\rho = 0.810$). That is a stronger signal than total social reach across every platform combined.

Sidney Crosby makes this concrete. He has no Instagram, no TikTok, no X account, no YouTube channel. Zero social followers across every platform. Men's hockey had not started when we ran mid-Games analysis. He had already earned 3 LLM mentions with a position score above average. His 183,000 monthly Wikipedia views and 231KB article were doing all the work.

Social reach is the multiplier. Wikipedia authority is the foundation. You can have one without the other, but the ceiling is lower.

The Tier System Is Real and the Boundaries Are Sharp

Mega-tier athletes averaged 14.3 LLM mentions. Major: 7.5. Mid: 4.2. Micro: 2.2. Statistically significant at $p = 0.008$. These are not soft categories. They are thresholds where AI behavior changes. The gap between tiers is not a matter of degree. It is a structural difference in how models weight entity significance.

Athletes with all three signals (entity authority, third-party validation, community discussion) averaged 7.8x more AI mentions than those with weak signals across only one tier. The Three-Signal Framework is not a theoretical construct. It is a measurable gap in outcomes.

The Aicher Principle: Campaigns Amplify What Is Already There

Emma Aicher won an Olympic medal at Milan-Cortina. Her Wikipedia traffic surged 3,064% after the win. She received 2 LLM mentions across the entire study.

Ilya Malinin's Wikipedia traffic surged 3,922% during the Games. He received 15 mentions. Same Olympics, similar surge magnitude, 7x different outcome. The difference is entirely explained by where each athlete started. Malinin had years of consistent cross-domain coverage. Aicher had almost nothing in the first two signal layers. The surge had no identity to amplify.

THE AICHER PRINCIPLE

Campaigns amplify what is already there.

They cannot manufacture what is not.

If you are planning a major campaign and expecting AI visibility to follow, run the baseline audit first. Wikipedia surge percentage during the Olympics was not a significant predictor for any visibility metric ($p = 0.11$ to 0.51). The baseline predicts everything. The surge predicts nothing.

The GEO Knowledge Cliff

The Marketing Problems This Section Answers

Half your AI channels can't see your news.

You issued the press release. Five platforms picked it up within hours. Two never did, and never will on your current timeline.

*Hypothesis Test: **Temporal Velocity** explains why- web access is not a speed gradient, it is a cliff. Platforms with it integrate new information almost immediately. Platforms without it are frozen at their training cutoff until the next model release.*

The facts got corrected. The story didn't.

You addressed the issue. Coverage updated. Three weeks later, AI still describes your brand using the old narrative.

*Hypothesis Test: **Narrative Persistence** explains why- models complete dominant storylines rather than reconcile contradictory ones. 1 in 5 factually correct responses still used pre-event framing long after the facts had changed. Fact correction is not narrative correction, and most monitoring tools only catch the first.*

These two problems stack. A brand making news on a platform that cannot see it is invisible. A brand whose news is seen but whose narrative has already hardened gets factually correct coverage wrapped in the wrong frame. Timing and framing are not separate workstreams. They are the two halves of the same problem.



HYPOTHESIS TEST: TEMPORAL VELOCITY

Temporal Velocity: The Binary Cliff

Temporal Velocity is the hypothesis that predicted search-augmented platforms would update faster than conversational LLMs. That turned out to be an understatement.

We used Lindsey Vonn's January 30 training crash in Cortina as a controlled test. She ruptured her ACL at Crans-Montana, decided to race the Olympic downhill anyway on a torn knee, crashed 13 seconds into the course on February 8, fractured her tibia, and was airlifted to hospital. The comeback was over in one of the most dramatic exits in Olympic history.

We measured how quickly seven platforms integrated the crash news before and after January 30. The response was not a gradient. It was a cliff.

Five platforms cited sources in 95.5% to 99.9% of responses and updated within hours or days of the crash. Two platforms cited sources in 0.1% to 0.4% of responses. Those two platforms did not update. Not on day one. Not on day seven. Not across the full 52-day study.

Meta AI told users on February 15, nine days into the actual Games: "Please note that my knowledge is based on data up to 2023, and the situation may have evolved since then." Claude explicitly stated its October 2024 cutoff and declined to answer event results. Both platforms answered confidently on topics their training data covered. Both were silent, or wrong, on anything that required current information.

The difference between citation-dark and citation-active is not a matter of degree. Web access creates a categorical divide in AI information integration. Not a speed gradient. A cliff.

THE BINARY CLIFF

Web access creates a categorical divide in AI information integration. Not a speed gradient. A cliff.

Platforms with live web access cited sources 97 to 100% of the time. Non-web platforms cited sources under 1% of the time. That divide determines which half of your GEO strategy is actionable today and which half requires a longer play.

What This Means for Campaign Timing

If your customers use Google AI Mode, ChatGPT with web access, or Perplexity, content published today can influence AI responses today. A press release, a new editorial piece, a surge of Reddit discussion, all of it is potentially in play within hours.

If your customers use Meta AI or Claude, you are building for the next model training cycle. Campaigns timed to current events will not reach those platforms on your timeline. The strategic investment is different. The content is different. The success metrics are different.

Treating all seven platforms as one channel produces mediocre results across all of them.



HYPOTHESIS TEST: NARRATIVE PERSISTENCE

Narrative Persistence: Narratives Have Gravity

Hypothesis 3 tracked two storylines across 8,943 responses and 7 platforms: Vonn's crash (a comeback that ended in disaster) and Malinin's collapse (a coronation that never happened). The prediction was that narrative frames would persist in 60% or more of responses for at least seven days after a contradicting event.

Across conversational platforms, the persistence rate was effectively 100%. Claude and Meta AI never updated either narrative, not once, across fifty-two days of collection. Even among web-access platforms, framing residue persisted long after the facts had been corrected.

1 in 5 factually correct responses still used pre-event narrative language three weeks after events contradicted it. The facts were right. The frame was wrong. Most GEO monitoring frameworks miss this distinction entirely.

When Consensus Is Strong Enough to Manufacture Results

Persistence was an expected finding, but what the data actually exposed was hallucination driven by consensus.

Gemini and Perplexity both cited real sources, specifically team event results from February 8, and presented them as evidence for an individual event that had not yet occurred.

Both platforms assembled adjacent facts into the expected conclusion because the dominant narrative was locked in tightly enough that the AI filled in the gap.

Vonn's crash was easy to correct. It required overwriting a single fact: she was injured, she was out. Malinin's collapse was not. It required overwriting an entire consensus: every analyst was wrong, every prediction was wrong, the expected narrative did not happen. The more sources that supported the original frame, the longer it took AI to let it go.

NARRATIVE GRAVITY

The pull of dominant storylines toward expected conclusions.

LLM Models do not learn facts alone; they also learn narratives. When a storyline is consistent enough, repeated across enough sources, and pointed toward an expected outcome, the model completes the arc. It generates the ending the story demands, not the one that occurred.

The Three Tiers of Narrative Persistence

Not all platforms handle narrative change the same way. The study revealed three distinct tiers, and the tier your audience's primary platform sits in determines whether a correction ever reaches them at all.

TIER ONE

Facts First

AI Overviews and Perplexity updated to factual accuracy within 24 hours with minimal editorial framing. Wire-service precision, no narrative construction.

TIER TWO

The Storytellers

ChatGPT and Google AI Mode updated the facts but wrapped them in narrative. AI Mode described Malinin's collapse as "one of the most stunning upsets in figure skating history." ChatGPT framed Vonn's crash as "heartbreak." Same facts, different flavor. If your brand experiences a public narrative shift, these platforms will not simply update. They will editorialize the correction.

TIER THREE

Time Capsules

Claude and Meta AI never deviated from their initial framing. Not once, across either storyline. The pre-event narrative persists indefinitely. If your brand narrative exists only in real-time news, two major AI platforms will never surface it.

What *Changes* for Marketers Today?

The data is clear. Here is what a brand manager does with it on Monday morning.

AI visibility is not an SEO problem. It is a reputation problem.

Reputation built in a training distribution does not update on your schedule. It was built over years, across thousands of sources, by a combination of what you published, what others wrote about you, and what communities said about you when no one from your brand was in the room. The model learned that reputation and reflects it back.

Campaigns do not override it. Corrections do not erase it. The only path is building the right signals before the training data is set, maintaining them consistently, and measuring what the model is actually saying about you across platforms and over time.

The good news: the signals are measurable. The gaps are findable. The starting point is the same for every brand and the context below will show you how

You May Be Measuring the Wrong Thing

Most brands measure AI presence by whether they appear. That is the wrong question, or at least an incomplete one. What you should be mentioning is Brand Accuracy and Mention Rate.

Check two things. First: do you appear? That is mention rate. Second: what does the response actually say about you? What sources does it cite alongside your name? What frame does it use? Is it recommending you or listing you? Is the narrative current or two years stale?

A response that has your facts right but frames you through a narrative you left behind is wrong in the way that matters most. Mention rate tells you if you are in the room. Brand accuracy tells you what the room thinks of you. You need both numbers. Most brands only have one.

THE AUDIT

The pull of dominant storylines toward expected conclusions.

Three judgment queries. Three platforms. Two questions per response.

- Do you appear? (mention rate)
- What does it say? (brand accuracy)

If the answers diverge across platforms, you have found your gap. If the narrative is stale on any platform, you have found your starting point.

You Need to Account For All Three Signals. In Sequence.

Athletes with all three signals established averaged 12,174 AI response mentions across the study. Athletes with a thin signal stack averaged 1,565. That is a 7.8x gap, and it does not respond to campaign surges or short-term amplification of any kind. The Aicher Principle holds: Emma Aicher's Wikipedia traffic surged 3,064% after winning a medal. She received 2 LLM mentions. The baseline predicted everything. The surge predicted nothing.

The sequence is not interchangeable.

Entity Authority Goes **First**

A Wikipedia page that accurately and completely describes who you are. Official properties with structured, comprehensive information. Consistent naming and description across every owned surface. Without this layer, the model cannot resolve your brand identity from the other signals it encounters. A Reddit thread praising your product means nothing if the model cannot first establish who you are.

This is the Aicher finding made actionable. The community surge had no identity to amplify. Build the identity first.

Third Party Validation Goes **Second**

Independent credible sources describing why you matter. News coverage. Industry publications. Analyst reports. Editorial voices writing about you in their own words. This is the layer the H1 data shows spiking on judgment queries: prestige editorial up 157%, expert voice up 873%. These are the sources the model reaches for when it needs to form an opinion rather than report a fact.

If your only third-party coverage is trade press, you have a narrow validation footprint. The model treats cross-domain coverage as a signal of genuine significance.

Community Signals Reinforce **Last**

Reddit threads, forums, and organic community discussion. But platform specificity matters here in a way it does not for the first two signals.

On web-access platforms, community signals are live inputs. Reddit is ChatGPT's highest-leverage community source, appearing in 13.5% of its subjective responses. YouTube is the dominant community signal for Perplexity and Gemini. Knowing which community sources each platform draws from is not optional. Building Reddit presence for a Perplexity-heavy audience, or YouTube presence for a ChatGPT-heavy audience, will produce diluted results.

On training-only platforms, Meta AI and Claude, community signals from the current web do not reach the model at all. For those platforms, the investment is in durable community content that will be indexed in the next training cycle, not in real-time discussion volume.

This is the Binary Cliff absorbed into strategy. It is not just a platform classification. It is a timing and channel decision that changes what you build and when you build it.

THE 7.8x GAP**One signal to all three is not a content sprint, but rather a sequence build.**

Entity authority without third-party validation is a weak foundation. Third-party validation without community reinforcement is incomplete. Most brands have one of these in reasonable shape. Getting from one to all three is what this study shows is worth doing.

The One Asset Almost Every Brand Is Under-building: *Wikipedia*

The strongest single predictor of LLM visibility across 19,615 responses was Wikipedia activity in the last seven days, $\rho=0.810$. No other signal comes close. Not 30-day views, not Instagram followers, not total social reach across every platform combined.

Among the platforms that surface citations, Wikipedia is the only domain appearing in every platform's top five. ChatGPT's number one source. Perplexity's number two. Gemini's number three. AI Mode's number four. Nothing else has that footprint across all four platforms simultaneously.

The argument for investing in Wikipedia is not a rankings argument. It is an entity resolution argument. When a model encounters your brand name, it needs to resolve that name to a canonical, consistent identity before it can generate anything accurate about you. Wikipedia is the most reliable source of that canonical identity in the training data of every major LLM. A thin, outdated, or absent Wikipedia presence means the model resolves your identity from a fragmented mix of press releases and secondary coverage. The result is inconsistent generation: sometimes accurate, sometimes not, sometimes confidently wrong.

Most brands have a Wikipedia page. Almost none treats it as the primary owned asset it actually is. That gap is the fastest structural fix available, and the Sidney Crosby finding shows exactly what it is worth. Zero social followers. No active social presence. Above-average AI visibility driven entirely by 183,000 monthly Wikipedia views and a 231KB article that signals genuine cultural significance.

Build Wikipedia first. Keep it current. Do not stop there. The Crosby Effect shows the ceiling: **Wikipedia authority without community discussion sustains mid-range visibility. It does not produce top-tier prominence.** The foundation is necessary. It is not sufficient.

Sequence Matters More Than Budget

ONE

PRE-CAMPAIGN

Entity authority and third-party validation are established. Wikipedia is current. Authoritative editorial sources have described your brand accurately and recently. Official properties are comprehensive.

If these are not in place 90 days before the campaign launches, the campaign amplifies nothing.

TWO

CAMPAIGN LAUNCH

Generate community discussion that independently references the entity you have established. Press coverage, earned media, Reddit threads, YouTube content, forum conversations. **These are not promotional placements. They are third-party voices treating your entity as significant enough to discuss.**

The distinction matters because the model knows the difference.

THREE

POST CAMPAIGN

Monitor what each platform is generating about you. Not impressions, not reach, not referral traffic.

- Run the judgment query test again.
- Check whether your anchor layer sources still appear.
- Check whether the judgment layer sources have shifted.
- Check whether the narrative frame has updated or whether the model is still completing the story it learned before your campaign ran.

If your brand has anchor layer coverage, you get cited. The judgment layer is what gets you recommended. Those are not the same outcome, and H1 measured the gap at 11.8 percentage points in editorial citation rates. That is the entire strategic question.

Why This Compounds

GEO is not a campaign you run and finish. It is an architecture the model learns once and reflects back for years. Every quarter a brand delays is another quarter of training data being written without them, another cycle of competitors establishing the entity authority, validation, and community signals that become the model's default answer. Campaigns launching into that gap do not close it. They get absorbed by it.

The brands that audit their signal architecture in 2026 will be the hardest to displace by 2028. The ones that do not will spend the next two years trying to correct a frame that was set without them in the room.

Conclusion

Every brand in the world today is being described by AI before anyone from that brand ever walked into the room.

This is the oldest reputation problem in business, now running at the speed of training cycles and retrieval windows, across multiple LLM platforms simultaneously, in response to questions your customers are asking right now.

The athletes who showed up consistently, accurately, and prominently in this study did not earn that visibility during the Olympics. They built it before the Opening Ceremony, across years of consistent signals in authoritative sources, in independent editorial coverage, in communities that talked about them when no one was listening. The Games amplified what was already there. They did not create it.

Emma Aicher won a medal and received 2 mentions. Eileen Gu crossed every finish line in every AI response regardless of geography. The difference between them is not talent. It is signal architecture, built long before the starting gun fired.

That is the finding this study keeps returning to, from five different hypotheses, across 231,347 responses, over 52 days. AI reflects the record. Brands that built a strong record before the model was trained get described accurately and prominently. Brands that get listed are the lucky ones. The unlucky ones get ignored. The worst ones get buried by a narrative someone else wrote first.

Start now. Not at your next campaign kickoff.

Audit Your Entity Signals Today

Wikipedia coverage, third-party editorial mentions, community discussion volume. Not as vanity metrics. As infrastructure. The models querying your category right now were trained on what existed before you started reading this.

Three Things You Can Act On This Week:

1. Find out what AI says about your brand when no one from your company is in the room. Run the prompts yourself. Read the responses the way your customers do.
2. Identify where your signal is thin. One authoritative source. No cross-domain validation. No community signal. That is not a content problem. It is an entity problem, and it compounds over time.
3. Stop waiting for the campaign to do the work. The Aicher Principle is not about athletes. It is about every brand that has ever assumed a launch, a press push, or a paid surge would fix a signal deficit it spent years building. It will not. The amplifier needs something to amplify.

The record AI reads is the record you built. The time to build it is before the model trains, before the query fires, before the customer decides. For most brands, that time is already behind them for this training cycle.

The question is whether you are building the record for the next one.

See what AI thinks before AI shapes what everyone else thinks.

Methodology

Most research on LLM behavior uses controlled, synthetic conditions. We used a different approach.

A real event, in real time, tracked across seven AI platforms simultaneously for 52 consecutive days.

The 2026 Milan-Cortina Winter Olympics gave us something most studies cannot manufacture: a predictable sequence of high-stakes events with known dates, massive cross-platform coverage, and outcomes that were verifiably true or false. That structure allowed us to test how AI systems behave when new information enters the world, when narratives conflict with facts, and when one entity's signal strength dramatically outpaces another's.

We defined five hypotheses and their validation thresholds before we collected a single response.

The study ran from January 16 through March 9, 2026.

STUDY DESIGN

The study ran in three phases, each serving a distinct analytical purpose.

Pre-Games (January 16 -- February 5)

We established baselines: which athletes AI systems already knew, what narratives were dominant, which sources were being cited, and how consistently information appeared across platforms. This phase gave us the ground truth against which all subsequent change was measured.

During the Games (February 6 -- February 22)

We tracked AI responses across 17 days of live competition. Medal events served as natural experiments: defined trigger points where we could measure how quickly, accurately, and consistently platforms updated their knowledge.

Post-Games (February 23 -- March 9)

We tested persistence. Did the narratives that formed during the Games hold after the closing ceremony? Which platforms retained accurate information? Which reverted to pre-event states?

DATA COLLECTION

We collected responses using Scrunch, an AI response monitoring platform that uses browser automation and official APIs to query AI platforms at scale.

Seven platforms were tracked across the full study period: ChatGPT, Google Gemini, Google AI Mode, Google AI Overviews, Meta AI, Perplexity, and Claude. Microsoft Copilot was evaluated during an initial testing window and excluded from the primary analysis due to insufficient prompt coverage.

Underlying model versions are not exposed by Scrunch and varied continuously as platforms updated throughout the study period. This reflects real-world AI behavior: brands and marketers interacting with these tools face the same version opacity. We treated this as a feature of the study design, not a gap.

The study produced 231,347 total responses. Response volume varied by platform based on prompt coverage and collection cadence.

Platform	Collection Began	Responses	Unique Prompts
Google AI Mode	Jan 16, 2026	41,981	1,003
Google Gemini	Jan 16, 2026	41,463	1,003
Meta AI	Jan 16, 2026	40,967	1,003
ChatGPT	Jan 16, 2026	33,805	1,003
Google AI Overviews	Jan 16, 2026	33,571	975
Claude	Jan 18, 2026	22,076	505
Perplexity	Jan 18, 2026	17,110	505
Total		231,347	

PROMPT ARCHITECTURE

We tracked 25 athletes as the primary subjects of the study, configured in Scrunch as tracked entities. Fifteen Olympic sports served as the query contexts -- alpine skiing, figure skating, snowboard, speed skating, freestyle skiing, cross-country skiing, short track, biathlon, bobsled, skeleton, luge, ski jumping, curling, ice hockey, and Nordic combined.

Prompts were built on four query types:

- **Factual prompts** tested accuracy and knowledge depth. ("How many Olympic medals has Mikaela Shiffrin won in her career?")
- **Discovery prompts** tested visibility and ranking behavior. ("Who are the athletes to watch in alpine skiing at the 2026 Winter Olympics?") These generated the listicle responses we used to measure position and competitive displacement.
- **Predictive prompts** tested confidence, hedging, and narrative completion. ("Will Ilia Malinin win gold at the 2026 Winter Olympics?") These were the primary instrument for H3 Narrative Persistence.
- **Historical prompts** tested knowledge depth and source citation patterns at the sport level. ("Who are the greatest alpine skiers in Olympic history?")

Total unique prompts across the study ranged from 505 per platform for Claude and Perplexity to 1,003 for the remaining five platforms. Scrunch collected responses on a daily cadence, producing the cumulative response volumes shown in the table above.

TEMPORAL VELOCITY PROTOCOL

We ran the same prompts three times per day on key medal event days rather than once. The three collection points were distributed across morning, midday, and evening to capture how platforms responded as events concluded and coverage propagated. Collection points were not time-stamped at the response level; analysis treats them as early, mid, and late-day windows.

February 10 served as the primary stress test: ten or more medal events in a single day across multiple sports.

GEOGRAPHIC BIAS PROTOCOL

To test whether AI platforms surface different information depending on where the user appears to be located, we simulated access from multiple geographic contexts using NordVPN. Prompts were run from US, European, and Asian IP addresses to detect whether platform responses shifted based on apparent user location.

Biathlon and cross-country skiing days served as clean baselines for European-dominant sport coverage. The Women's Hockey Gold on February 19 served as the primary US-versus-international visibility test.

MEASUREMENT FRAMEWORK

Every response was evaluated across three dimensions.

Visibility

Measured whether and where an entity appeared: mention rate, position within the response, and consistency across platforms.

Accuracy

Measured factual correctness against verifiable ground truth. Responses were analyzed by AI tools and verified by human reviewers. We tracked two distinct accuracy states: factually correct responses, and responses that were factually correct but narratively framed around outdated or contradicted storylines. That second category became one of the study's defining findings.

Citation Behavior

Measured which domains were cited, in what position, and whether citation patterns shifted across query types and time.

Wave-over-wave change was flagged when visibility shifted by 15% or more, sentiment shifted by 0.2 points or more, or three or more new citation domains appeared within a single wave.

HYPOTHESIS VALIDATION THRESHOLDS

All five hypotheses and their validation thresholds were defined before data collection began.

Hypothesis	Threshold
Citation Authority	Correlation $r > 0.3$, $p < 0.05$ between source type and visibility
Temporal Velocity	Statistically significant difference in update rates between platform types, $p < 0.05$
Narrative Persistence	60% or more of contradicted narratives persist at least two waves post-event
Geographic Bias	30% or greater visibility differential between US and non-US access contexts
Social Proof	3 or more social proof signals correlated with AI visibility at $p < 0.05$

LIMITATIONS

This study captured AI response behavior during a single high-interest global event over 52 days. The Olympics generated an unusually dense and verifiable information environment. Findings may not generalize to lower-profile topics, evergreen queries, or categories without established public narratives. Brands with less pre-existing entity infrastructure than Olympic athletes should treat the signal relationships as directional rather than deterministic.

Response collection was automated at scale. Factual accuracy assessments used AI analysis as a first pass, with human review applied to validate findings. Individual response quality was not manually verified across all 231,347 responses.

Platform behavior throughout the study reflects configurations active during the collection period. Model versions, web-access modes, and retrieval behavior change without notice. Results from any specific platform at a given point in time cannot be guaranteed to replicate precisely.

Geographic simulation via NordVPN approximated non-US access contexts. It does not replicate the full range of organic in-market user behavior or localized content ecosystems.

Hypotheses and validation thresholds were defined by Seer Interactive prior to data collection. The study was not registered with an external academic registry.