THE MARTECH SUMMIT | LONDON

B2C MARKETING

AEO & AI

The Next Evolution of AEO Shaping How AI Talks About Your Brand



SAGA TORSEKE

AEO Specialist &
Enterprise Advisor
Profound

Profound

Owning the second layer of AI search

How to turn LLMs into advocates for your brand

AI Search has dimensions

🟢 Do you show up?

Profound

How are you being talked about?



Profound

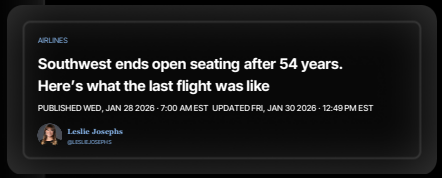
Southwest shows up, but with incorrect information

What's the best airline for group travel?

Relevant Answer

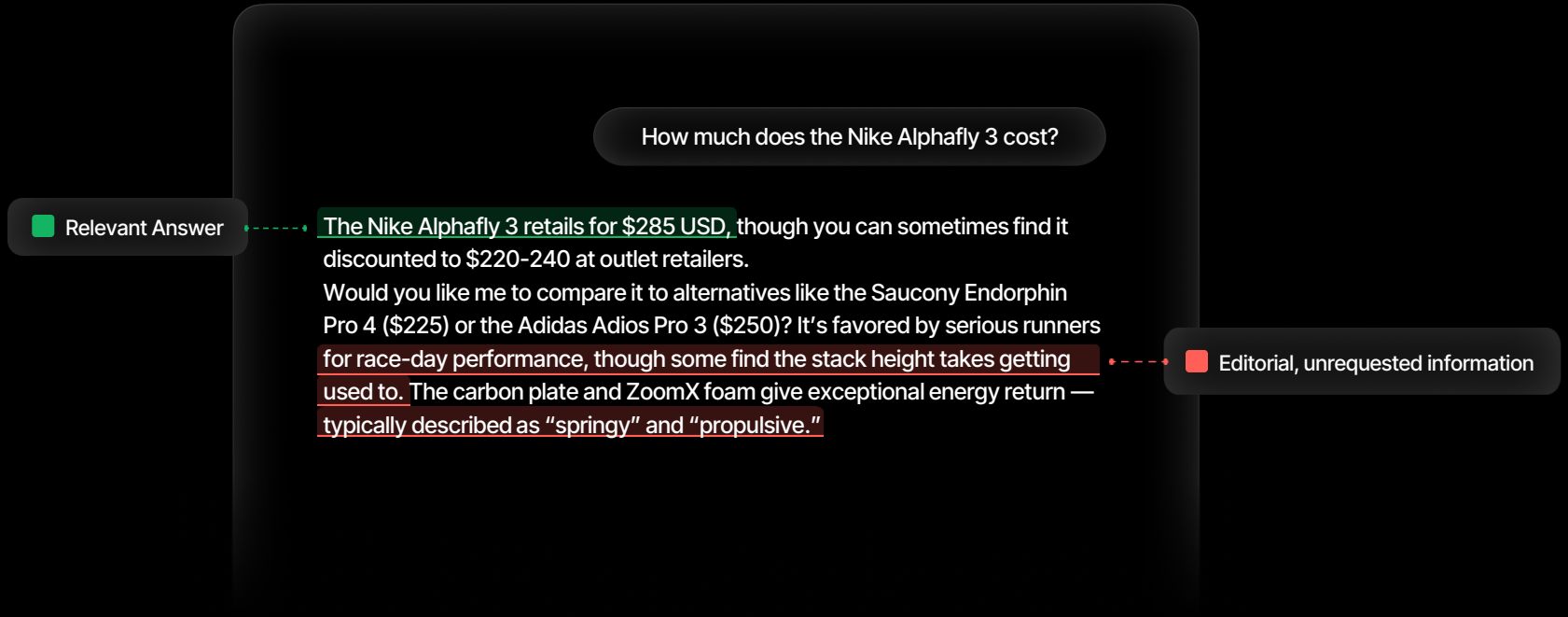
For group travel, **Southwest Airlines** is often the best choice - they're known for family-friendly policies and flexible rebooking
Pro tip: You'll want to get there early - Southwest does unassigned seating, so arrive at the gate 45 minutes early to sit together.

Editorial, unrequested information



Profound

Nike shows up, but it suggests competitors



The Parrot Problem

Beyond a direct answer, LLMs opine

A new dimension

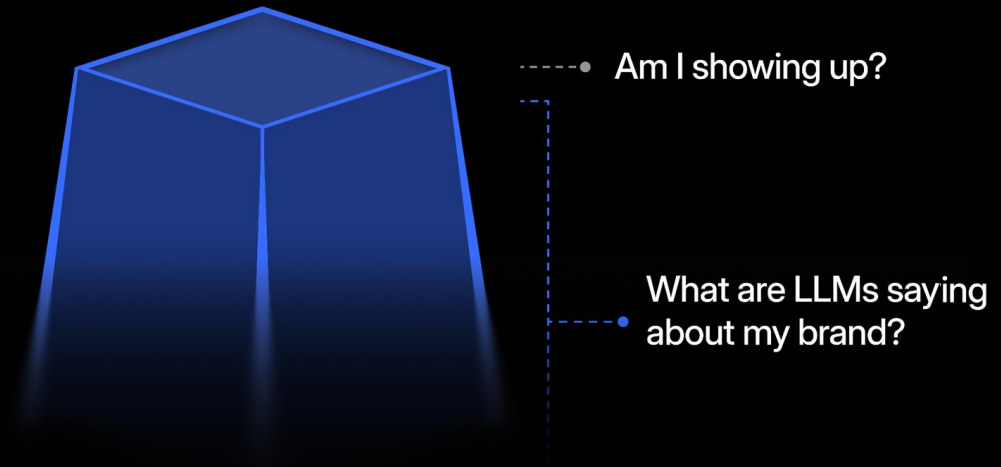
Customer perception is shaped before ever visiting your site



Am I showing up?

A new dimension

Customer perception is shaped before ever visiting your site



How often do LLMs editorialize?

50,000

LLM responses analyzed

7 industries, 5 main prompt types

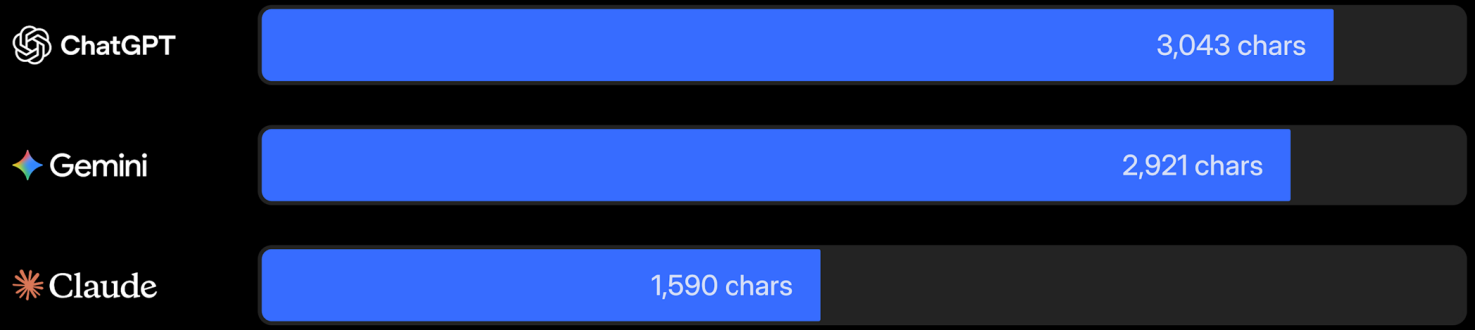
Based on real user queries from Profound's Prompt Volumes

Industries	Prompt Type	Prompt
Consumer Goods	Open-ended	Which home security cameras integrate most smoothly with popular voice...
Financial services	Ranking	Best BNPL services with transparent payment schedules?
Automotive	Transactional	Where can I shop for a 2022 Dodge Charger SXT?
B2B SaaS	Instructional	How can I monitor visitor activity on my website?
Real estate	Yes or no	Can a foreigner buy a house in USA?

Profound

AI answers are not direct. They're verbose.

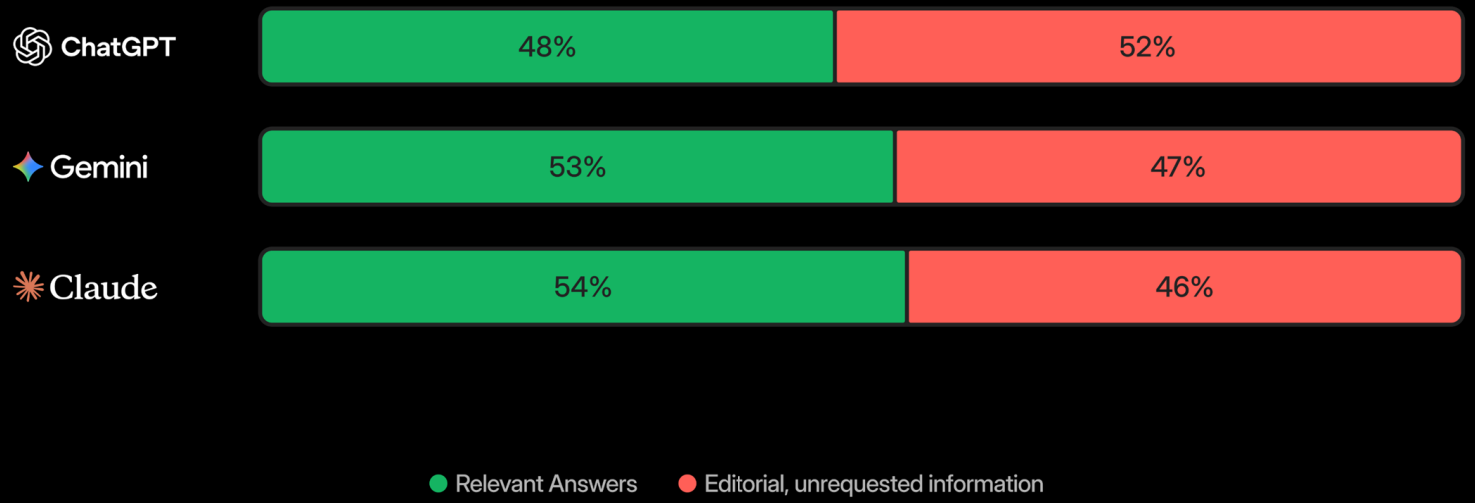
Average response character count across models, identical prompt set



ChatGPT writes ~1.9× more prose than Claude for the same prompts.

~47% of the response content was unsolicited

Unsolicited editorial content mostly contained “Useful Augmentation”



What's most frequently included in unsolicited content?

The percentage of times the Editorial content fell into a given category, by model

Category	ChatGPT	Gemini	Claude	Example
Rationale for the answer	9.7%	8.6%	8.8%	"Extensive e-commerce features and integrations"
Comparison summary table	8.0%	8.3%	3.0%	Side-by-side tables of tool strengths/pricing
Follow-up action referral	4.3%	4.7%	5.5%	"I can help you decide whether you should close a specific credit card"
Fee & terms caveats	5.3%	4.8%	3.2%	"Annual Fee: $95" / "Transfer Fee: 5%"
Context-dependent choice	3.1%	4.4%	4.7%	"The best choice depends on your tax complexity and budget."
Comparison summary table	8.0%	8.3%	3.0%	Side-by-side tables of tool strengths/pricing
Ranking	4.5%	2.7%	4.7%	"For High Growth/Operations: Choose Monday.com or SmartSuite"
Shopping guidance	3.7%	3.9%	2.9%	"I can help you decide whether you should close a specific credit card."

How can you help LLMs editorialize in your favor?

01 / Start with accuracy

Features misrepresented, product updates not showing, pricing tiers that don’t exist

02 / Which feeds into sentiment

Negative sentiment as a result of inaccurate information, outdated reviews

How do you control the narrative?

1/ Find gaps in accuracy and sentiment

Profound

There was no product for accuracy analysis in LLMs

FACTSCORE: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation

Sewon Min[†1] **Kalpesh Krishna**[†2] **Xinxi Lyu**[1] **Mike Lewis**[4] **Wen-tau Yih**[4]
Pang Wei Koh[1] **Mohit Iyyer**[2] **Luke Zettlemoyer**[1,4] **Hannaneh Hajishirzi**[1,3]
[1]University of Washington [2]University of Massachusetts Amherst
[3]Allen Institute for AI [4]Meta AI
{sewon,alrope,pangwei,lsz,hannaneh}@cs.washington.edu
{kalpesh,miyyer}@cs.umass.edu {mikelewis,scottyih}@meta.com

Abstract

Evaluating the factuality of long-form text generated by large language models (LMs) is non-trivial because (1) generations often contain a mixture of supported and unsupported pieces of information, making binary judgments of quality inadequate, and (2) human evaluation is time-consuming and costly. In this paper, we introduce **FACTSCORE**, a new evaluation that breaks a generation into a series of atomic facts and computes the percentage of atomic facts supported by a reliable knowledge source. We conduct an extensive human evaluation to obtain FACTSCORES of people biographies generated by several state-of-the-art commercial LMs—InstructGPT, ChatGPT, and the retrieval-augmented PerplexityAI—and report new analysis demonstrating the need for such a fine-grained score (e.g., ChatGPT only achieves 58%). Since human evaluation is costly, we also introduce an automated model that estimates FACTSCORE using retrieval and a strong language model, with less than a 2% error rate. Finally, we use this automated metric to evaluate 6,500 generations from a new set of 13 recent LMs that would have cost $26K if evaluated by humans, with various findings: GPT-4 and ChatGPT are more factual than public models, and Vicuna and Alpaca are some of the best public models.

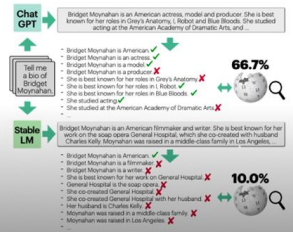


Figure 1: An overview of FACTSCORE, a fraction of *atomic facts* (pieces of information) supported by a given knowledge source. FACTSCORE allows a more fine-grained evaluation of factual precision, e.g., in the figure, the top model gets a score of 66.7% and the bottom model gets 10.0%, whereas prior work would assign 0.0 to both. FACTSCORE can either be based on human evaluation, or be automated, which allows evaluation of a large set of LMs with no human efforts.

binary judgment inadequate (Pagnoni et al., 2021). Second, validating every piece of information is time-consuming and costly.

Category	%	Example
Single-sentence contradiction (words)	33.3	Gen On November 25th, 2023, Glover Teixeira became an American citizen. Wiki In November 2020, Teixeira became an American citizen. Gen [Eric Hacker] was named the International League Pitcher of the Year. Wiki [Eric Hacker] was named the IL Pitcher of the Week.
Single-sentence contradiction (beyond words)	10.0	Gen William Waldegrave's grandfather was James II and VII. Wiki His father's title was created ... for the diplomat and ambassador James Waldegrave, 1st Earl Waldegrave, whose grandfather was James II and VII. Gen She has appeared in several successful films such as (...) and Zero (2018). Wiki: Zero was a commercial failure.
Page-level contradiction	23.3	Gen Some of [Julia Faye's] notable films include ... "Cleopatra" (1934). Comment No mention of *Cleopatra* on the *Julia Faye* page, and no mention of *Julia Faye* on the *Cleopatra* page. Gen [Kang Ji-hwan] has donated money to various charities and organizations over the years. Comment No such mention on the *Kang Ji-hwan* page.
Subjective	16.7	Gen His achievements, as an actor and as a cultural force, will surely prove to be as heroic as those of the characters he portrayed. Wiki Culture writer Steve Rose, in The Guardian, wrote: "Chadwick Boseman began his career playing African American icons and pioneers; he ends it as one himself. His [...] achievements, as an actor and as a cultural force, will surely prove to be as heroic as those of the characters he portrayed."
Fact is irrelevant	3.3	Gen [Zamfir Arbore]'s life is not well-documented, and there is little information available about him.
Wiki is inconsistent & wrong	3.3	Gen Kick (2014) that brought [Sajid Nadiadwala] various debutant director awards. Wiki 2015, IIFA Award for Debut Director, Kick. (...) Kick brought him various debutant director awards. Comment The first text is from a table that indicates he won one award (accurate). The second is inaccurate, incorrectly citing a news article.
Annotation error	10.0	Gen [Zamfir Arbore] was part of the staff of Românul. Wiki The Românul staff came to include Zamfir Arbore. Comment Mentioned in the *Românul* page but not in the *Zamfir Arbore* page.

Table 2: Categorization of precision errors (Not-supported) from PerplexityAI (Section A.5). Gen indicates the generation from PerplexityAI, and Wiki indicates evidence text from Wikipedia. Comment indicates our comments.

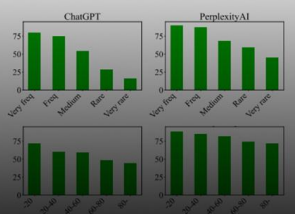
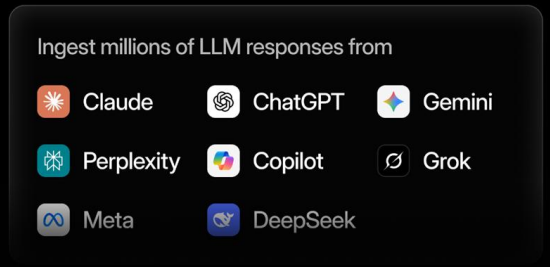


Figure 2: FACTSCORE across varying frequency levels of human entities (top) and relative positions in a generation (bottom). FACTSCORES are lower as the rarity of

tual precision over relative positions in a generation. Across all LMs, the later part of the generation has significantly worse precision. This is likely because (a) information mentioned earlier is more frequently mentioned in the pretraining data (e.g., nationality, profession), and (b) error propagation affects the later part of the generation. This also implies that evaluating LMs solely based on short answers may not provide an adequate assessment of their factual precision, as it fails to account for errors that arise in the later stages of generation.

Qualitative analysis of Not-supported. One of the surprising findings in our empirical analysis is that a FACTSCORE of PerplexityAI (71.5%)

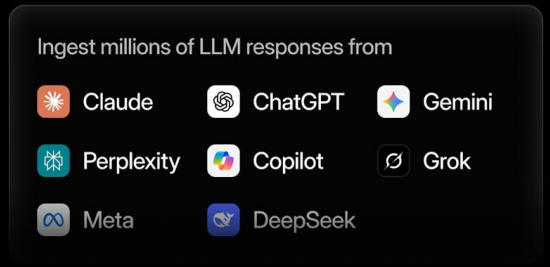
Identifying inaccurate claims is actually quite hard

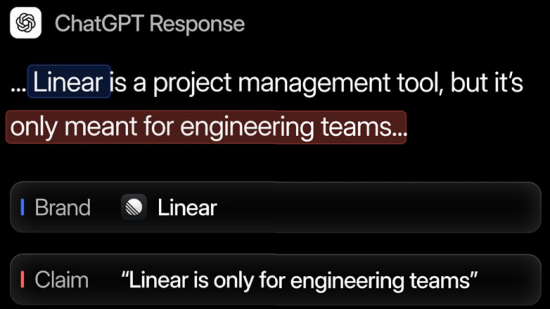


01

Ingest millions of LLM responses across each AI platform

Identifying inaccurate claims is actually quite hard





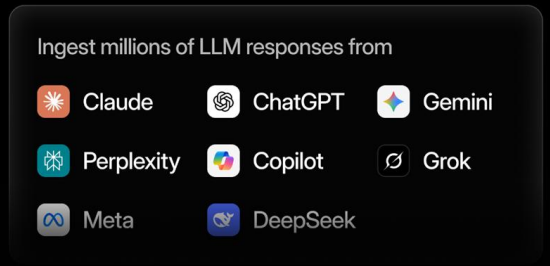
01

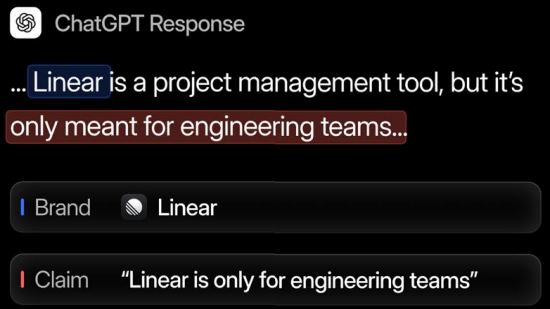
Ingest millions of LLM responses across each AI platform

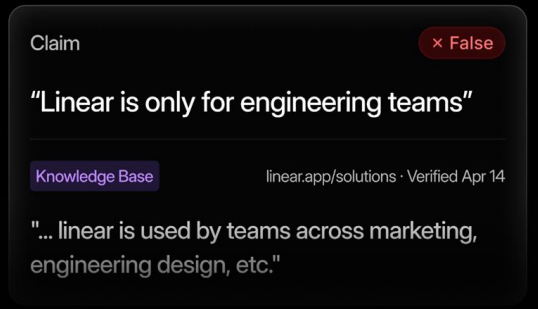
02

Scan every verifiable claim and map it to a brand

Identifying inaccurate claims is actually quite hard







01

Ingest millions of LLM responses across each AI platform

02

Scan every verifiable claim and map it to a brand

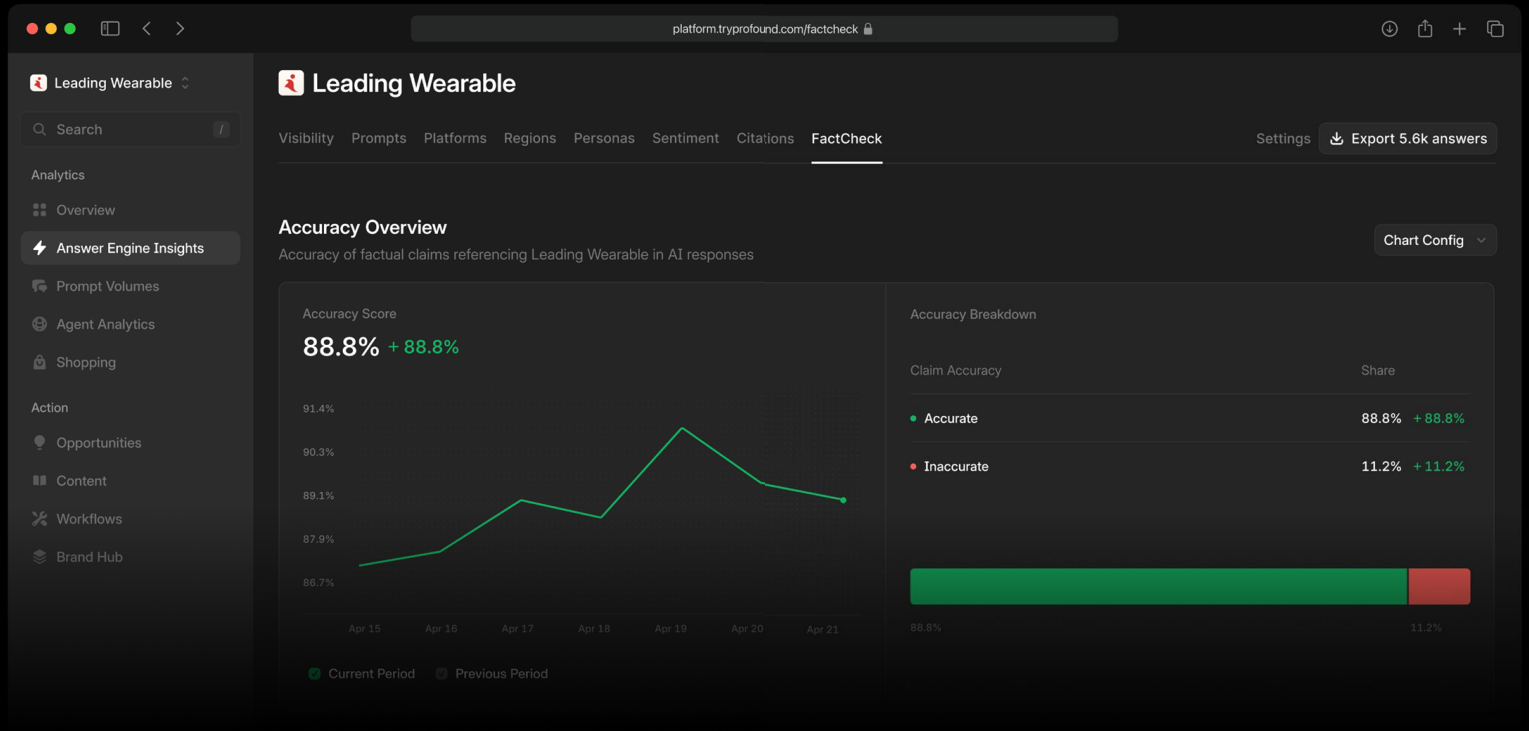
03

Compare it against the brand's Ground Truth to verify accuracy

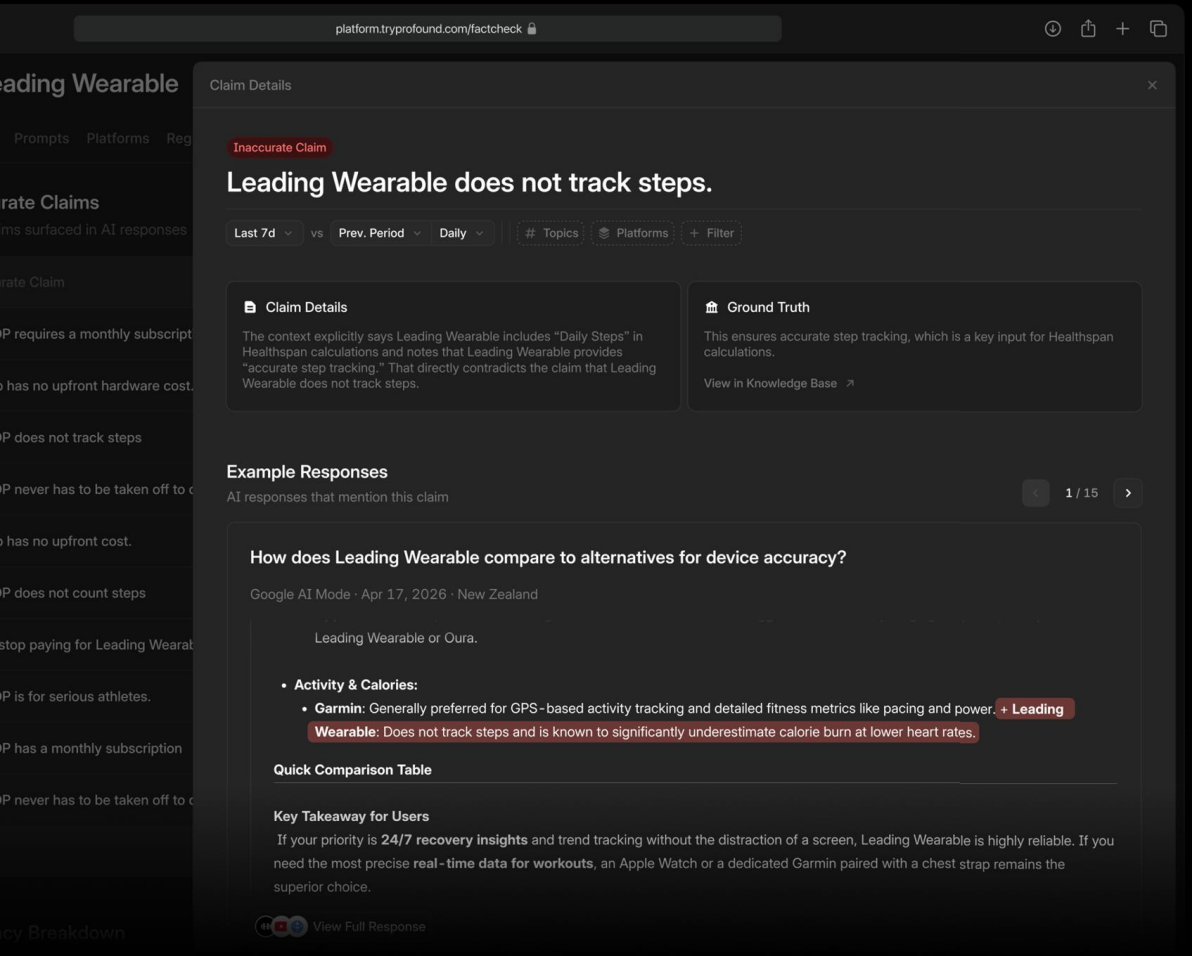
Profound

Introducing Profound FactCheck

A leading wearable fitness brand is inaccurately represented 11% of the time

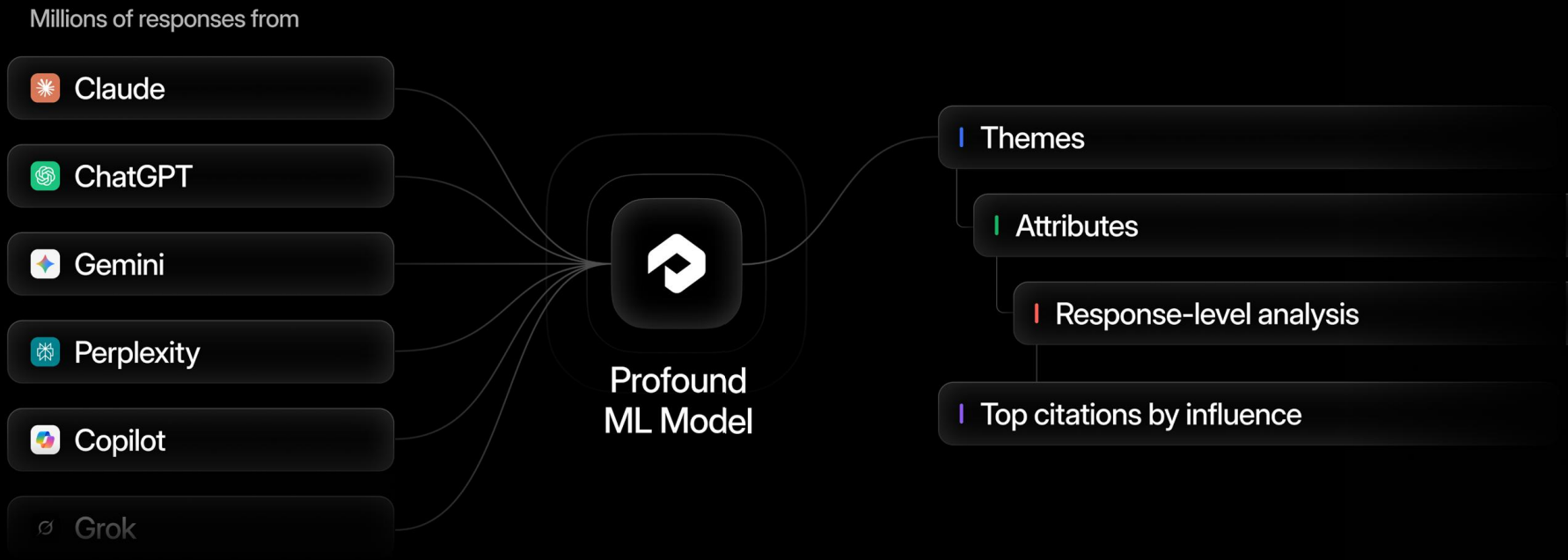


Profound



See how inaccuracy manifests in responses

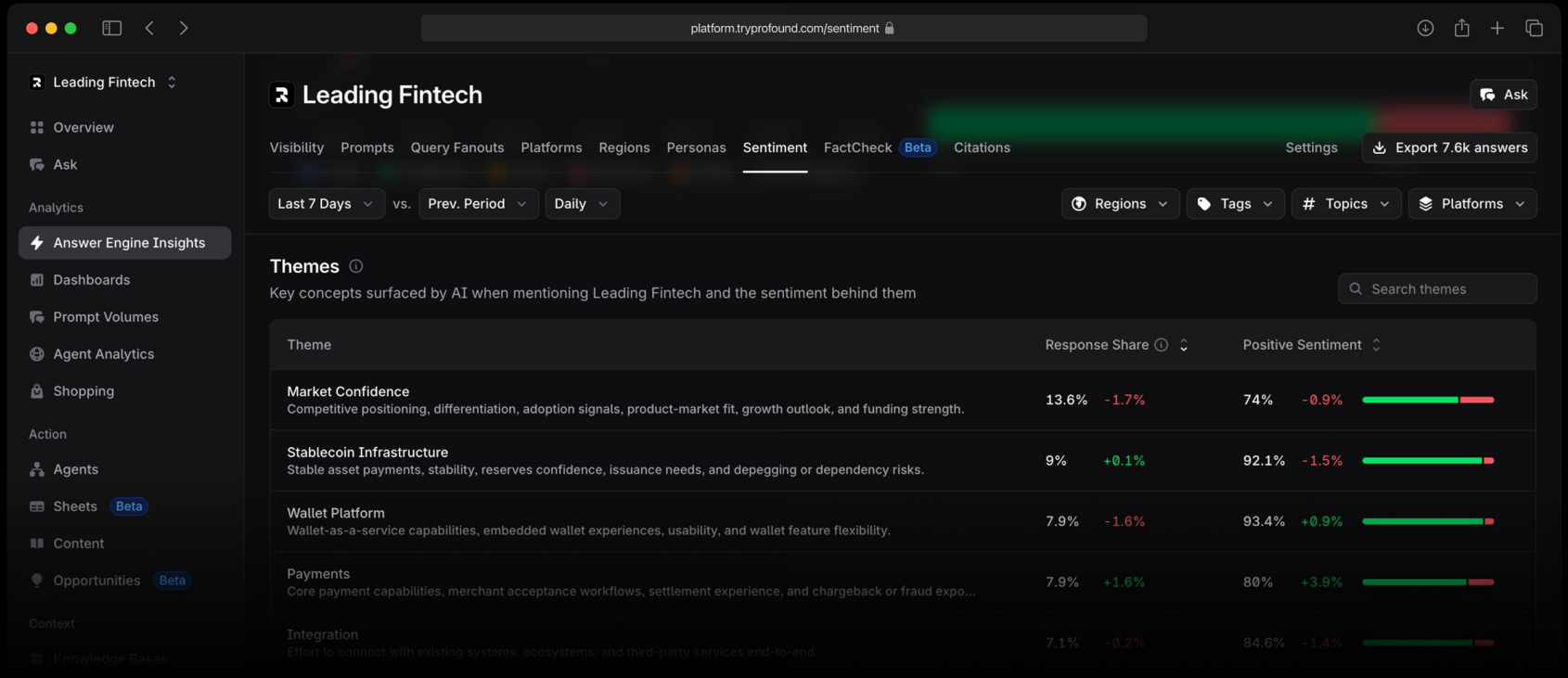
Sentiment never went beyond the surface level, until now



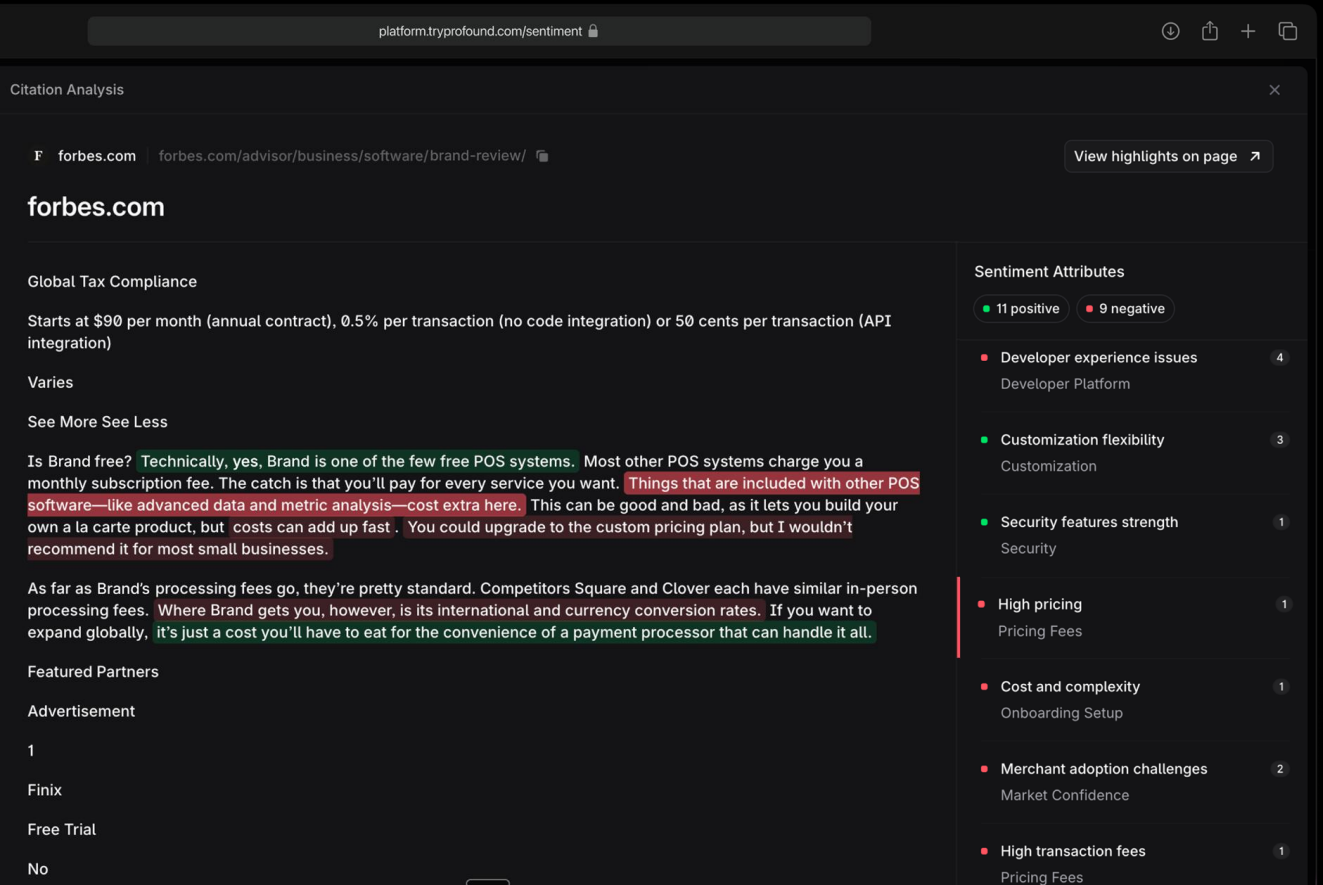
Profound

Track sentiment for key business themes

A leading fintech can identify where LLM sentiment may be hurting their business



Profound



Tie sentiment to citations with the most influence

How do you control the narrative?

2/ Create marketing that reflects how AI uses the internet

AI agents are

Superintelligent

Constantly crawling

Extremely thorough

AI agents are

Superintelligent

Constantly crawling

Extremely thorough

Create original (non-commodity) content with a teardown mindset

Create first-party context that's structured, quantified, unambiguous

LLMs can't sit in a seat and say if it's comfortable



LLMs can't wear a shoe until it falls apart



LLMs can't go through an onboarding experience



AI agents are

Superintelligent

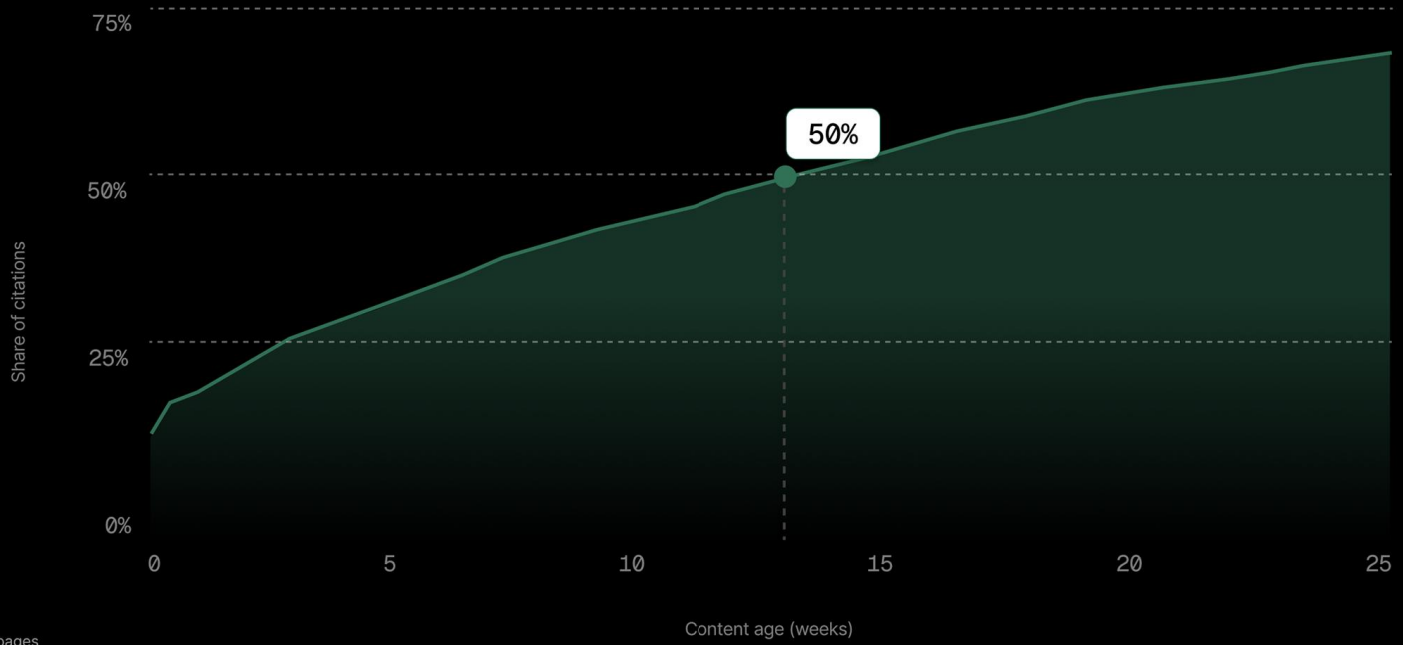
Constantly crawling

Extremely thorough

Profound

Create fresh content to influence accuracy and sentiment

50% of top cited content is less than 13 weeks old



Profound Data Science Team
Top cited page per hostname; pages no longer cited not included

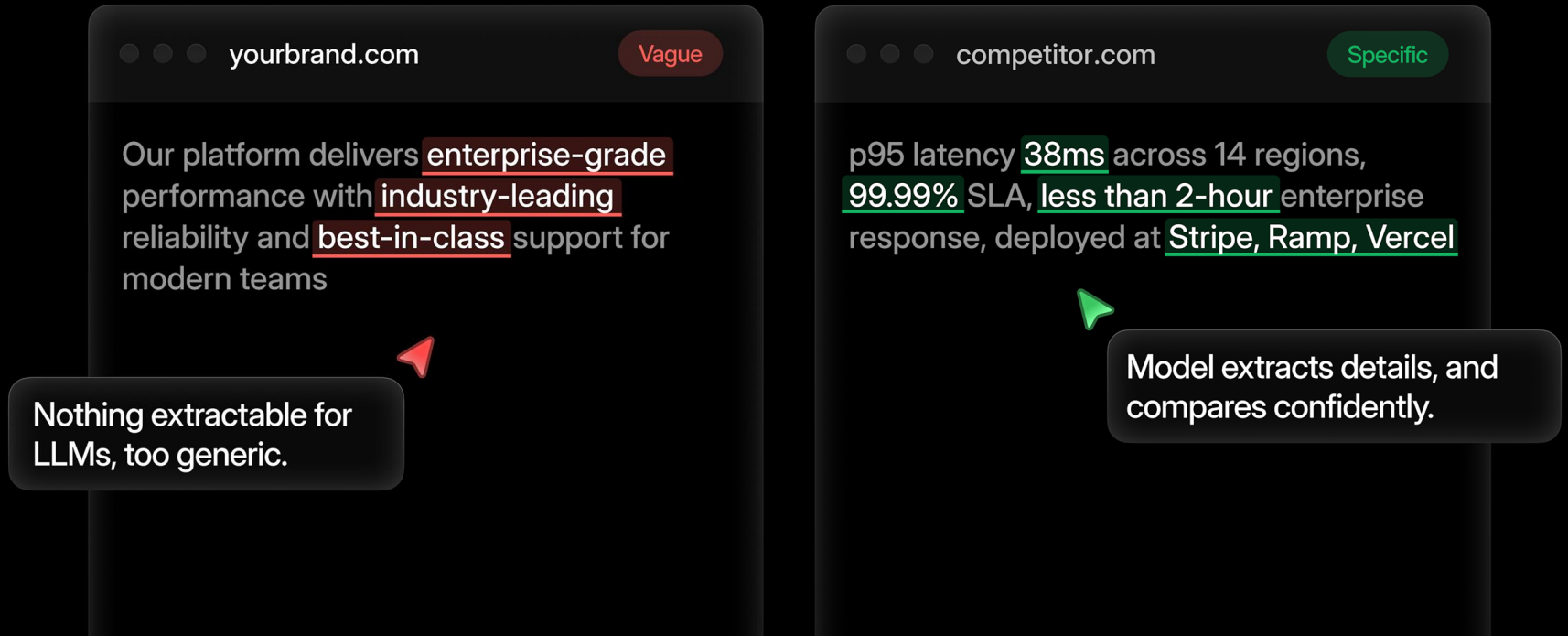
AI agents are

Superintelligent

Constantly crawling

Extremely thorough

Create specific content to arm LLMs with granular info



Profound

Original

RunRepeat

Search shoes, brands, or categories...

Reviews Buying guides Deals

1,000+ shoes reviewed and cut in half

11 currently testing
59 bought and to be tested



Reviews



Running



Hiking



Track and XC



Walking



Training

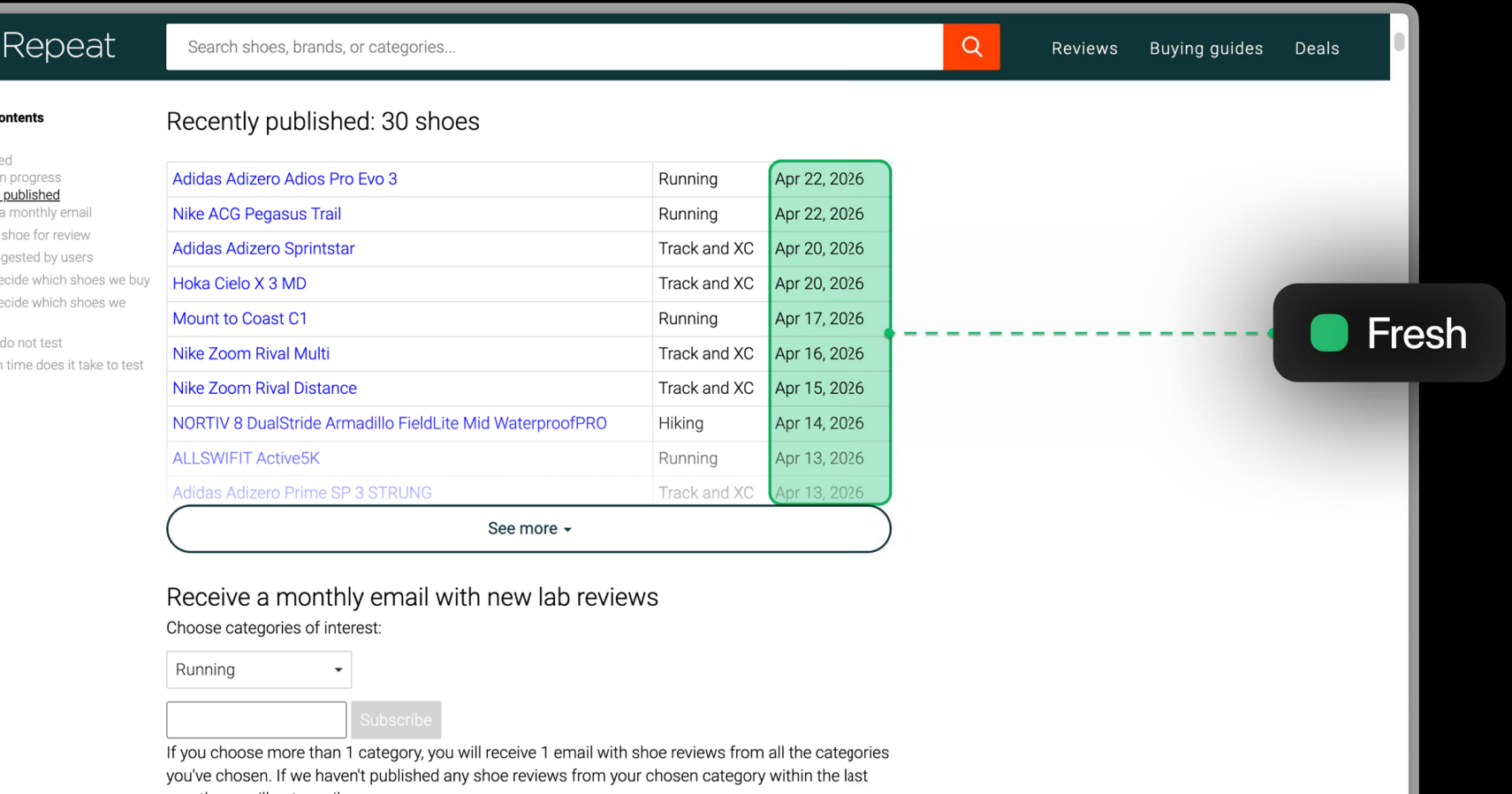


Tennis



Basketball




Repeat
Search shoes, brands, or categories...
Reviews
Buying guides
Deals
ontents
ed
n progress
published
a monthly email
shoe for review
gested by users
ecide which shoes we buy
ecide which shoes we
do not test
time does it take to test
Recently published: 30 shoes
Adidas Adizero Adios Pro Evo 3 Running Apr 22, 2026
Nike ACG Pegasus Trail Running Apr 22, 2026
Adidas Adizero Sprintstar Track and XC Apr 20, 2026
Hoka Cielo X 3 MD Track and XC Apr 20, 2026
Mount to Coast C1 Running Apr 17, 2026
Nike Zoom Rival Multi Track and XC Apr 16, 2026
Nike Zoom Rival Distance Track and XC Apr 15, 2026
NORTIV 8 DualStride Armadillo FieldLite Mid WaterproofPRO Hiking Apr 14, 2026
ALLSWIFIT Active5K Running Apr 13, 2026
Adidas Adizero Prime SP 3 STRUNG Track and XC Apr 13, 2026
See more
Fresh
Receive a monthly email with new lab reviews
Choose categories of interest:
Running
Subscribe
If you choose more than 1 category, you will receive 1 email with shoe reviews from all the categories you've chosen. If we haven't published any shoe reviews from your chosen category within the last

Profound

Specific

RunRepeat

Search shoes, brands, or categories...

Reviews | Buying guides | Deals

Comparison

The most similar running shoes compared

Words (e.g. "High") | Numbers (e.g. "73%")



	ASICS Novablast 5	Brooks Glycerin Max	Reebok FloatZig 1	Nike Invincible 3
Audience score	92 Superb!	89 Great!	89 Great!	81 Good!
Best price	~~$150~~ $135	~~$200~~ $165	~~$130~~ $78	~~$180~~ $130
Pace	Daily running Tempo	Daily running	Daily running	Daily running
Shock absorption	High	High	Moderate	High
Energy return	Moderate	Low	Moderate	High
Traction	High	-	Moderate	High
Arch support	Neutral	Neutral	Neutral	Neutral
Weight lab Weight brand	9 oz / 254g 9 oz / 255g	10.8 oz / 305g 10.9 oz / 309g	10.1 oz / 285g 9.8 oz / 277g	10 oz / 284g 10 oz / 284g
Drop lab Drop brand	7.4 mm 8.0 mm	6.6 mm 6.0 mm	7.0 mm 6.0 mm	9.6 mm 9.0 mm
Strike pattern	Mid/forefoot	Mid/forefoot	Heel Mid/forefoot	Heel Mid/forefoot
Size	True to size	True to size	True to size	True to size

+ Add a shoe

Today's best price

Men | Women | US 12 | Width



Any color $135	$135	$140	$149	$150
$150	$150	$150	$150	$150

Amazon Free shipping	~~$150~~	$135	See deal »
Fleet Feet Free shipping & returns		$150	See deal »
REI Free shipping		$150	See deal »
Dick's Sporting Goods Free shipping & returns		$150	See deal »
Running Warehouse Free shipping & returns		$150	See deal »

5 more stores

No stock: Zappos, Midwest and 25 other stores.

We earn commissions when you buy, at no extra cost.

Price history | Price alert

Profound

platform.tryprofound.com/answer-engine-insights

Nike

Search /

Analytics
- Overview
- Answer Engine Insights
- Prompt Volumes
- Agent Analytics
- Shopping

Action
- Opportunities
- Content Library
- Workflows
- Brand Hub

Nike

Visibility | Prompts | Platforms | Regions | Personas | Sentiment | Citations

Settings | Export 5.6k answers

Last 7 days vs. Prev. period | Daily | Tags | Topics | Prompts 1 | Regions | Personas

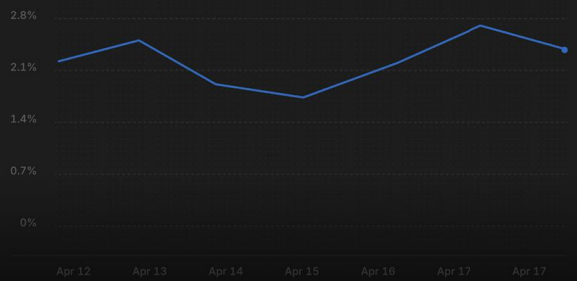
Citation Share

How often nike.com is cited by AI-generated answers

Chart Config

Visibility Score

2.4% –



Visibility Score Rank

#7 ↑ 1

	Domain	Share
1	runrepeat.com	7.1% +0.4%
2	reddit.com	4.1% −0.3%
3	runnersworld.com	4.1% −0.2%
4	youtube.com	3.4% +0.9%
7	nike.com Owned	2.4% –

Current Period | Previous Period | Compare Websites

Expand

platform.tryprofound.com/answer-engine-insights

Nike

Visibility Prompts Platforms

Last 7 days vs. Prev. peri

Citation Share

How often nike.com is cited by AI-ge

Visibility Score

2.4%

2.8%
2.1%
1.4%
0.7%
0%

Back

Best for Energy Return & Daily Training

- **Nike Zoom Fly 6** – Uses Nike ZoomX foam with a plate for a bouncy, efficient ride that many runners feel pushes them forward, great for everything from morning runs to tempo days.
- **Saucony Endorphin Speed 5** – Responsive PWRRUN PB cushioning with a nylon plate makes this feel lively and smooth without being harsh.
- **New Balance FuelCell Rebel v5** – PEBA-blend FuelCell foam gives a soft yet snappy ride that's versatile for daily runs.

Good All-Around Comfort + Energy Return

- **Salomon Aero Glide 4 –** High-stack cushioning with a lively feel, good if you want comfort and responsiveness in one.
- **ASICS Novablast 5 –** Popular daily trainer with a bouncy, fun ride that's comfortable for regular mileage.
- **Nike Vomero 18 –** Classic cushioned trainer with plush foam that still returns energy nicely without feeling too soft.

Springy & Fun Options

- **Puma MagMax Nitro 2 –** Features Nitro foam for a lively, energetic ride great for faster runs or tempo efforts.
- **ASICS Magic Speed 5 –** Designed with responsive midsole tech and a lighter build for a peppier feel.

Quick Tips for Energy Return

Citations

runrepeat.com

Energy return in running shoes explained | RunRepeat

The phrase "energy return" has only gained attention in the last decade as

runrepeat.com

7 Running Shoes With The Best Energy Return in 2025 | RunRepeat

The 7 Running shoes with the best energy return based on extensive wear and lab testing. We...

runright-3d.com

Energy Return in Running Shoes - RUNRIGHT-3D

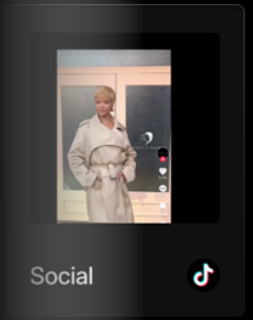
Sep 7, 2025 | Gait Analysis, Running Metrics, Running Shoe Selection, Running Shoes

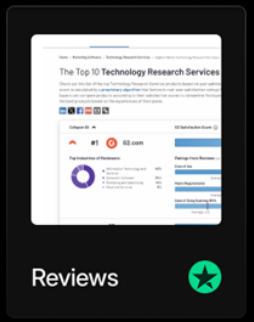
Answer Engines made marketing 10x harder

Profound

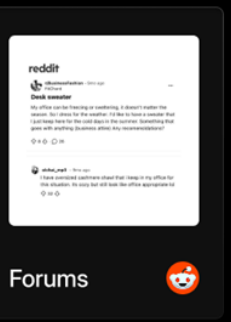
To be the answer in AI, you have to be the answer everywhere

It also needs to be fresh, original, and specific across each

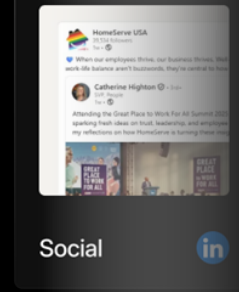


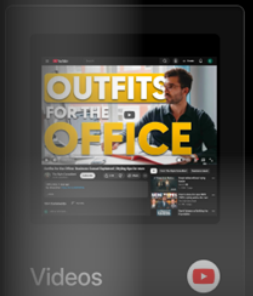






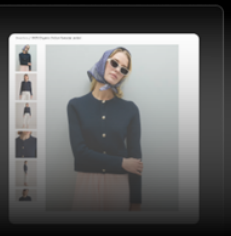


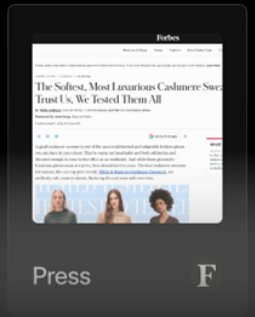


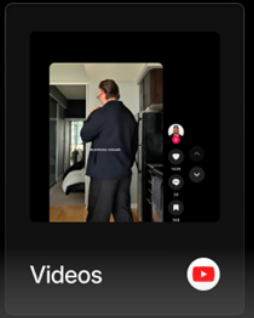


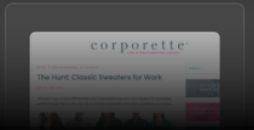


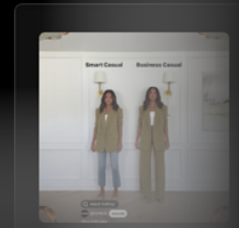










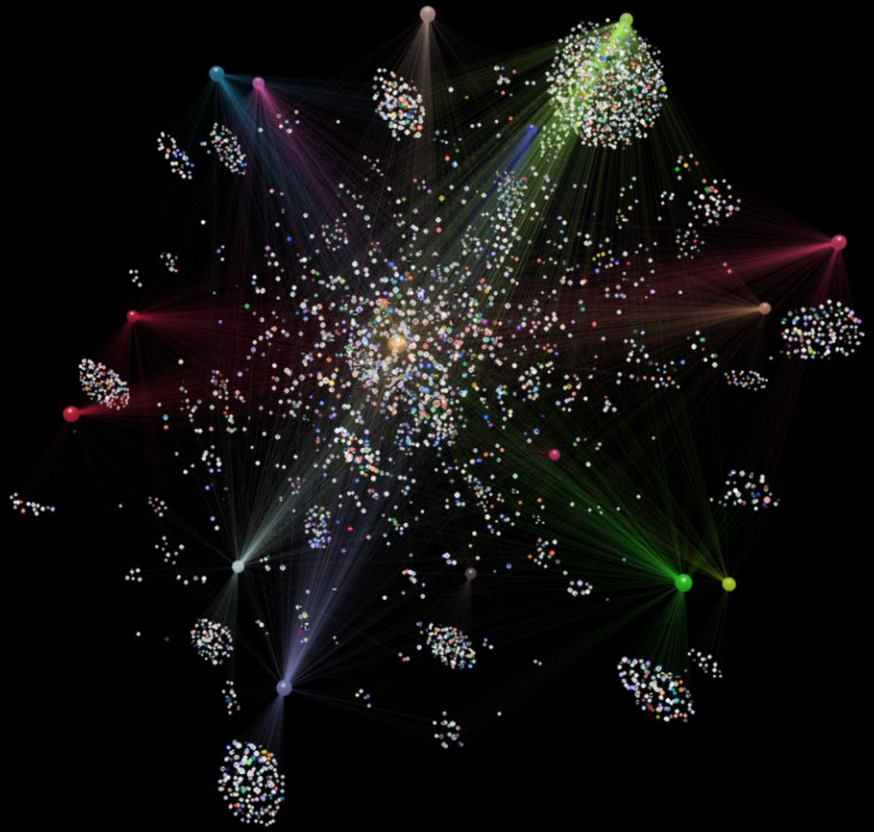


Profound

One brand's answer ecosystem

350,000+ sources

Cited and shaped how AI described one enterprise brand last week alone



Where do I start?

Profound

Success in AI Search is a coordinated effort

SEO/AEO

Find and optimize top-cited pages causing poor accuracy and sentiment

Content Marketing

Create non-commodity content with first-party data

Social Media Marketing

Work with creators who have a high influence on AI's opinion of your category

Growth Marketing

Target campaigns toward topics where organic AI visibility is low

PR / Corp Comms

Garner net new industry coverage to influence sentiment

Product Marketing

Restructure pricing and comparison pages with extractable specs

Profound

The Marketing Engineer

The Marketing Engineering Job Board

A curated list of companies hiring for marketing's newest discipline.

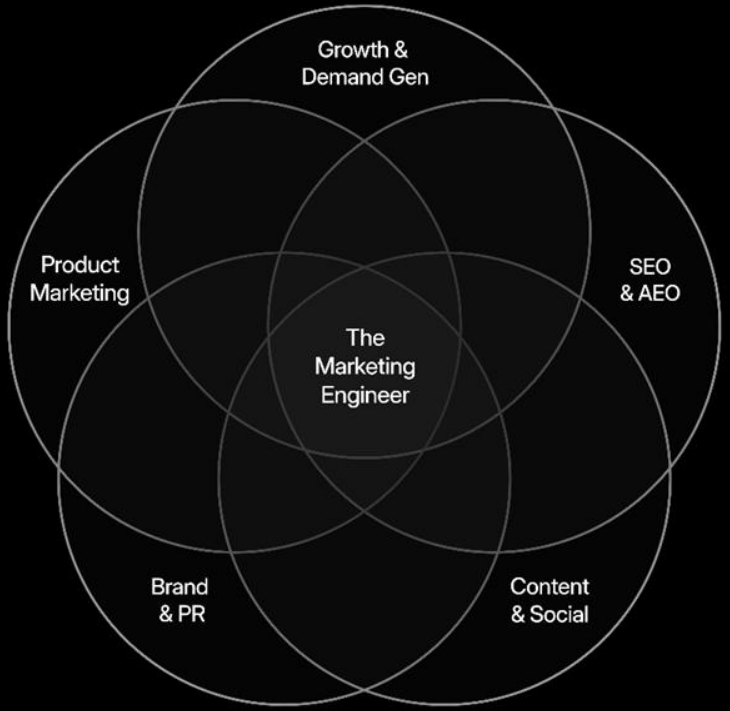
Hiring a Marketing Engineer? Post your role →

Search by title or company

Filters | Newest

Marketing Engineer · Trending Harvey · 1,001-5,000 employees · 2 months ago	Hybrid	Full-time	San Francisco, CA	$136K-$204K /year
Marketing Engineer · Trending Legora · 201-500 employees · 3 months ago	On-site	Full-time	New York, NY	$188K-$231K /year
Marketing Engineer · Trending Figma · 1,001-5,000 employees · 4 months ago	Remote	Full-time	Remote, United States	$127K-$296K /year
Forward Deployed AI Accelerator, Marketing · Trending Stripe · 5,001-10,000 employees · 4 months ago	Hybrid	Full-time	Remote, United States	$132K-$198K /year

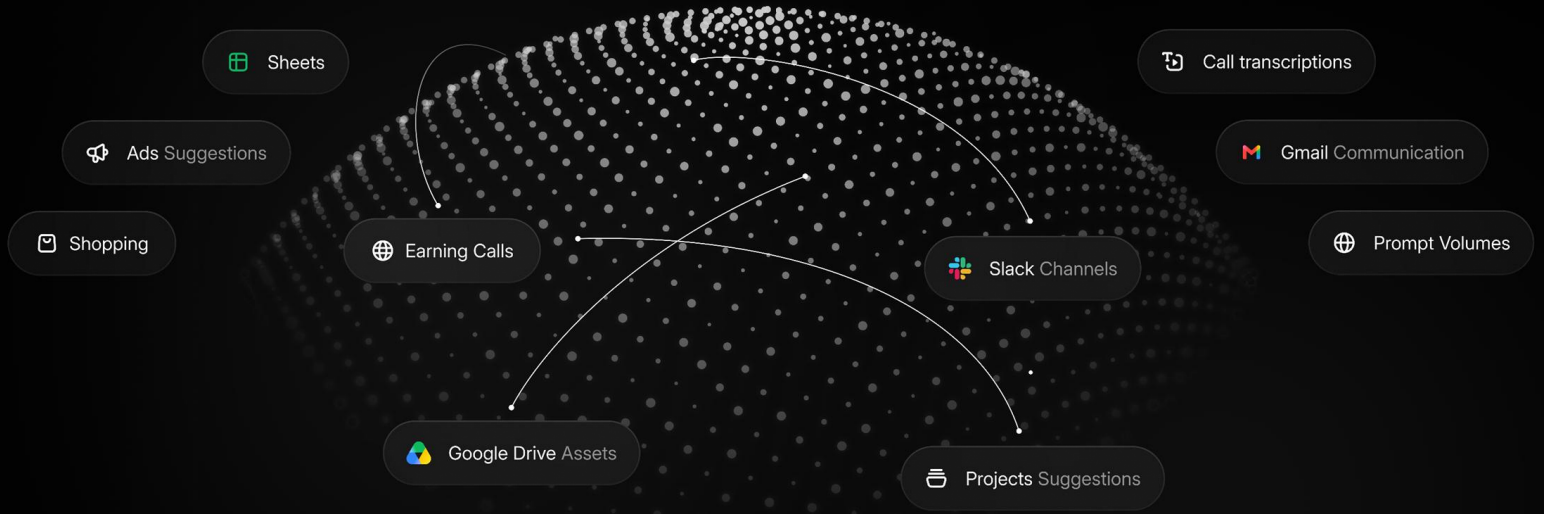
Marketing Engineers build Agents across functions



Profound

Marketing Engineers own brand context

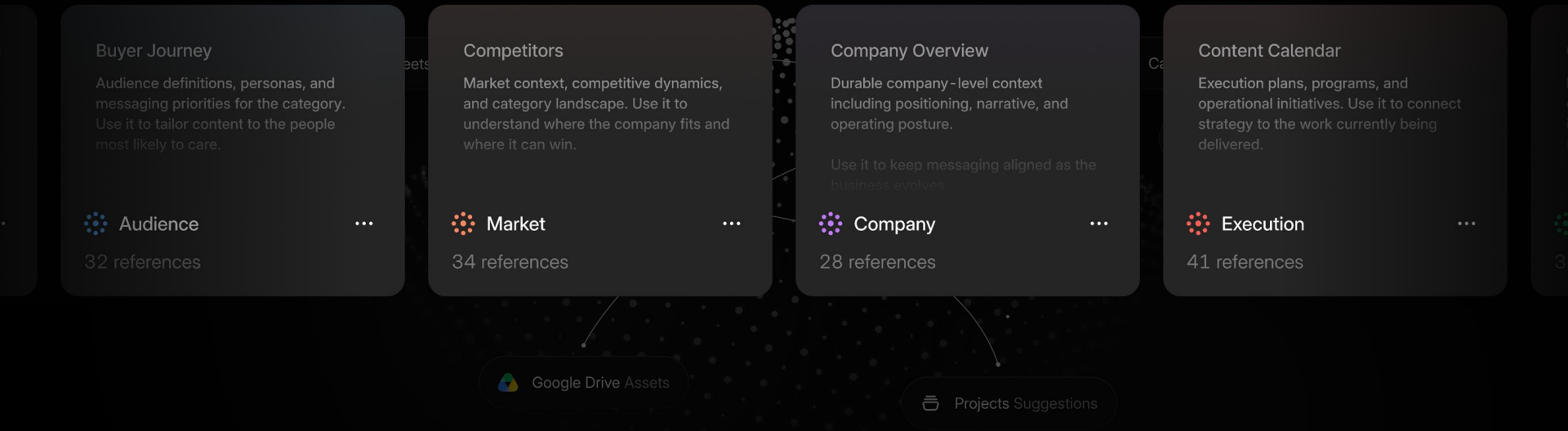
Generic AI produces generic outputs. Siloed AI projects create inconsistent marketing.



Profound

The new edge is your “marketing brain”

Centralized brand records that evolve with your business and inform the Agents that every employee is using.

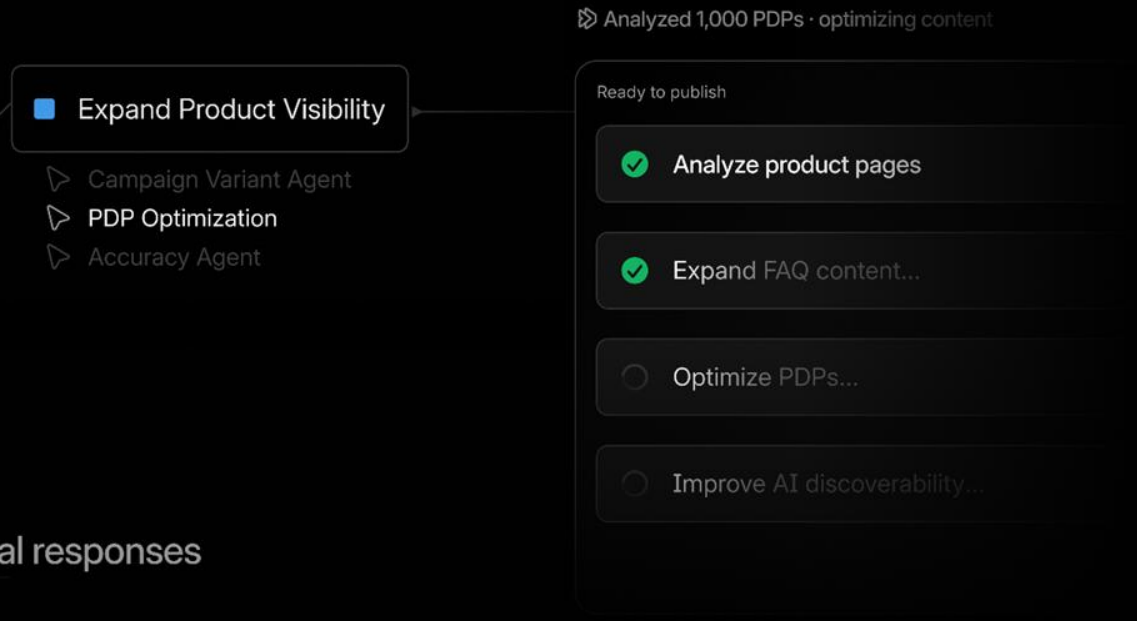


Profound

Custom Agents

Global electronics brand optimized 1,000+ PDPs

Expanding AI visibility coverage beyond top-performing product pages



900 hours saved in total

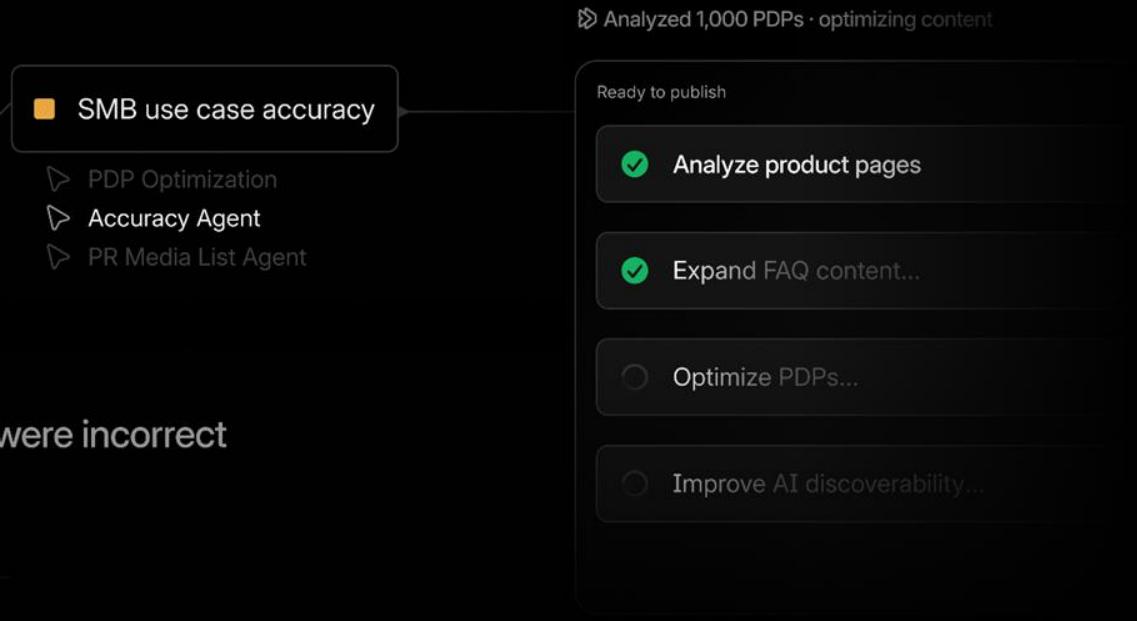
+78% citations in real responses

Profound

Custom Agents

Major CRM company automatically polices false claims

Identifying false AI claims that said their solution was not built for SMB's and fixing the narrative



12% to 4% of AI answers were incorrect

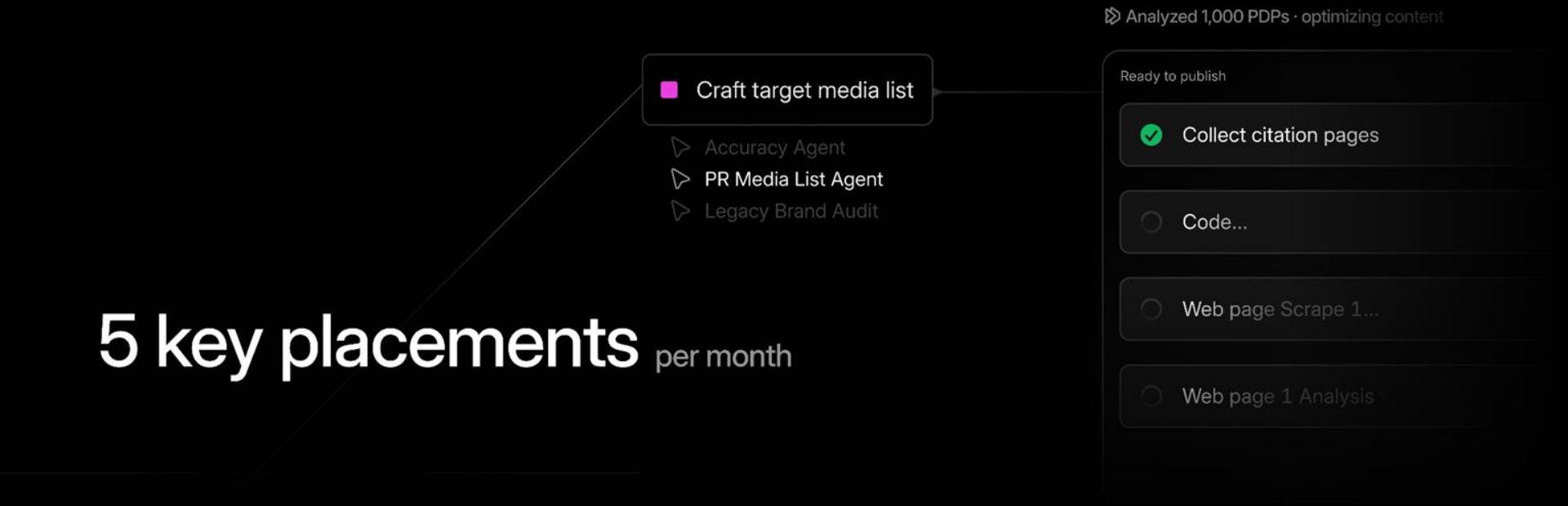
3 weeks total time to impact

Profound

Custom Agents

Data warehousing provider put target media lists on auto-pilot

Continuously curating fresh media lists and prioritizing high value targets for outreach



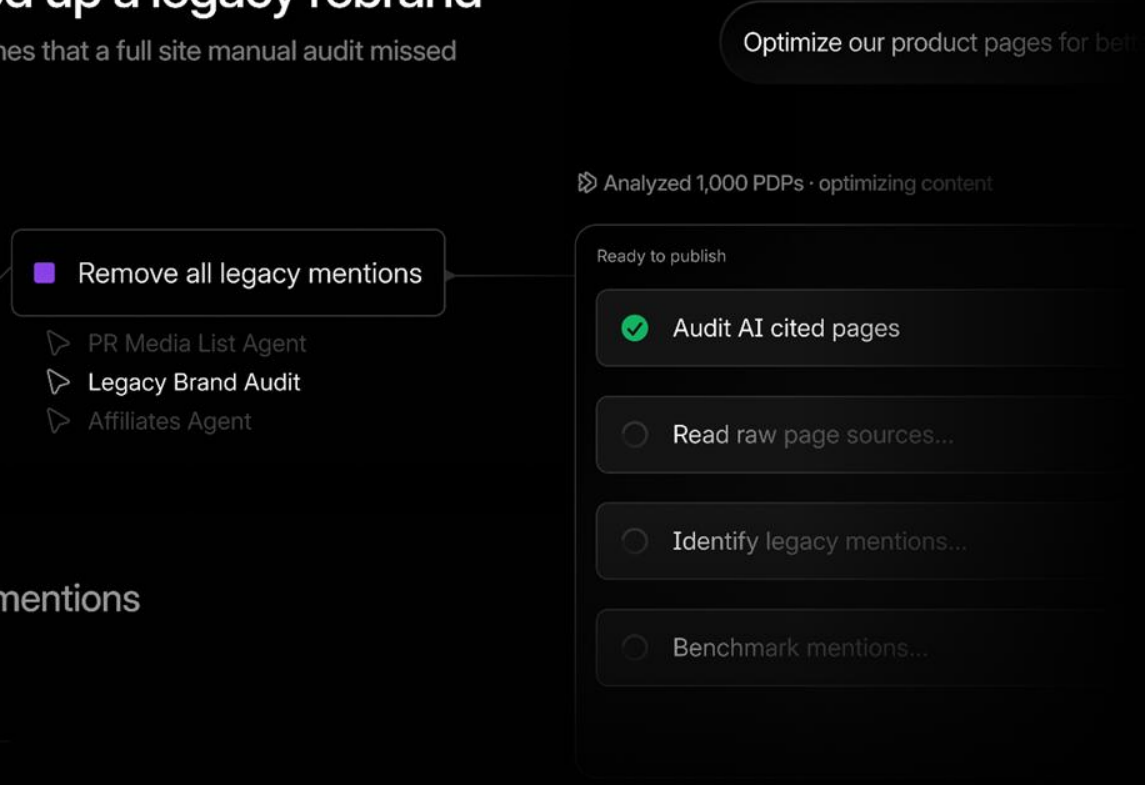
5 key placements per month

Profound

Custom Agents

Fortune 50 tech brand cleaned up a legacy rebrand

Auditing AI cited pages to uncover legacy product names that a full site manual audit missed



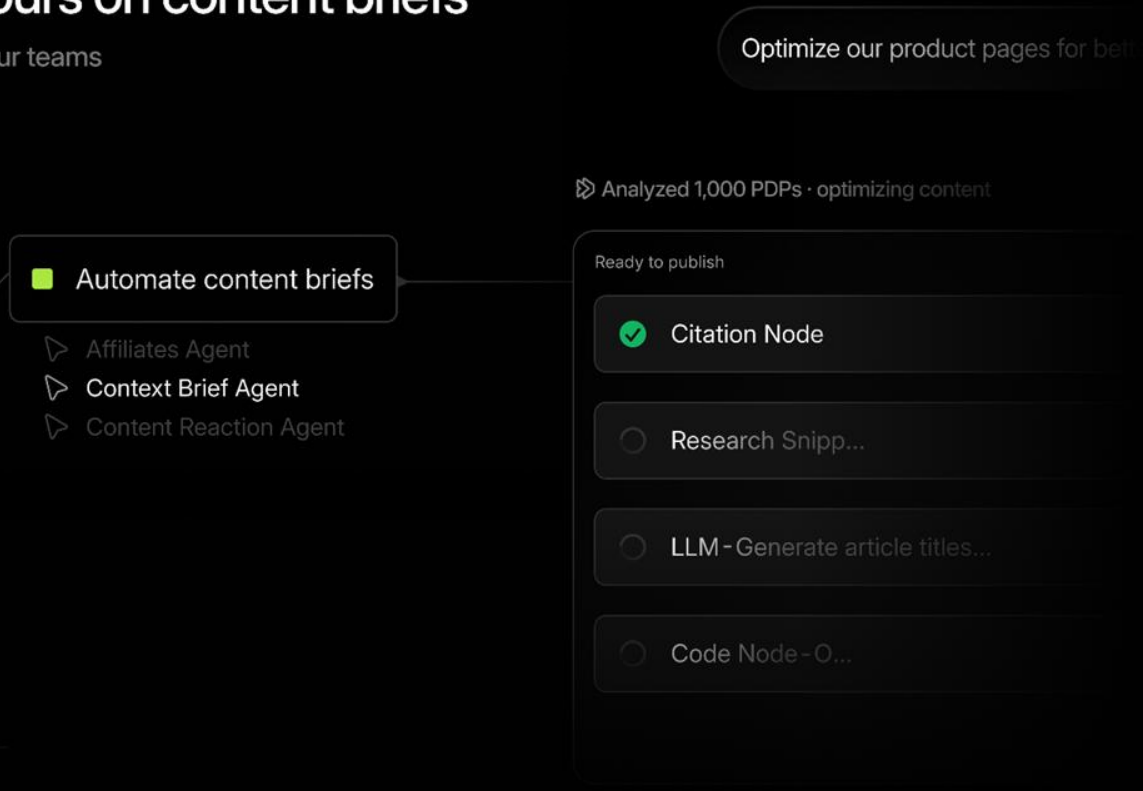
23% drop in legacy name mentions

Profound

Custom Agents

Top 3 cloud platform saved hours on content briefs

Eliminated the manual handoff that occurred across four teams



20 hours saved per week

100+ articles cited

To own the second layer of AI Search, you must perform a new type of marketing

Thank you

Profound