PRODUCT REQUIREMENTS · 2024

Ori Search & Ori Optic Deep Research

Written in 2024 as Product Lead for Ori, in my final quarter at Vireon Labs. This is a high-level, abbreviated version, written to show the thinking rather than the whole document.

About a seven-minute read. Pricing, internal dates and confidential detail are left out.

1 Summary

Ori Search answers a market question in about 45 seconds, with every source listed, tiered and linked. Ori Optic Deep Research reads dozens of articles and writes a sectioned briefing that checks every figure across its sources. Both make one promise: the number you read came from somewhere you can check.

This PRD commits two quarters to making that promise airtight before making it louder. Our own measurement found gaps between the claim and the code: retrieval, provenance and arithmetic problems, every one measurable and testable.

Three outcomes, in order:

1. **Trust floor.** A public answer survives a reader who clicks every source.
2. **Research depth.** Optic answers analytical questions (valuation, exposure, thesis) from primary data, not only news.
3. **One plain voice.** Every sentence lands on first read for a smart reader with no finance background.

Nothing here requires a new model, a new vendor or a new price.

2 Where the product stood

	ORI SEARCH	ORI OPTIC DEEP RESEARCH
Who it's for	Every paid plan, plus a free public demo	The top plan, since a briefing is ten times the work of an answer

	ORI SEARCH	ORI OPTIC DEEP RESEARCH
What it produces	A ~250-word briefing with up to 24 numbered sources and a confidence score	A ~1,400-word sectioned report with dated citations and a cross-source check on every figure
Where it looks	About 560 tiered publishers, plus SEC filings and official data releases	The same sources, fanned out: up to 40 scanned, every citable page read in full
How it writes	A model drafts; a rules-based "hardener" scrubs jargon, scores confidence and repairs or drops what fails	One model pulls the facts from each article; a stronger one writes the briefing, word by word
Speed, cold	~15s to the first section, ~45s in total	~60–95s

3 Users and jobs to be done

The buyer is the tool-stacker: a self-directed investor who already pays for a data terminal or two and has been burned by an AI that made up a number. The secondary buyer is the solo adviser or newsletter writer who needs receipts they can forward. Day traders and large firms aren't the target yet.

The reader is deliberately narrower: one smart person with no finance background. One voice everywhere is cheaper to enforce than two.

The top jobs, by how often they arrive:

JOB	EXAMPLE	HOW OFTEN	TODAY
Explain what's happening and why	"Why is the market down today?"	Most common	Works
Check a number I saw	"Did payrolls really fall last month?"	Inside every job	Partly: a figure could still cite the wrong page
This ticker, today	"What is NVDA doing today?"	Common	Works
Compare two companies	"Compare JPM and BAC"	Common	Works as a briefing, not as a valuation
News on a topic	"Latest on the Fed decision"	Common	Works, citing the issuer directly

JOB	EXAMPLE	HOW OFTEN	TODAY
Pull a line from a filing	“What did PayPal guide for next quarter?”	Occasional	An honest refusal that names the filing to read

Two jobs are missing by design: Ori never tells anyone what to buy or sell, and it never guesses at a question it can’t place.

4 The problems

Each was reproduced before it was written down.

1. Retrieval could cite a page about the wrong company.
2. Citation numbers could drift from the source list the reader saw.
3. Provenance was measured; truth wasn’t. Arithmetic slips and dropped revision labels passed every check.
4. Source counts included sources the answer never used.
5. The progress panel showed work that hadn’t happened yet.
6. Optic could treat old pages as current and admit unrated ones.
7. Optic was thin on analytical questions and had no mobile design.
8. The voice was written for professionals; readers new to finance lost the thread.

Not problems: writing quality, the tier gate, or cost at real usage.

5 Goals, non-goals and principles

Goals, in priority order

1. A stranger who clicks every source on a public answer finds nothing to screenshot.
2. Factual accuracy on checkable claims reaches at least 90%, measured monthly against primary sources.
3. Optic answers an analytical question with at least eight fact-bearing, dated sources.
4. Every sentence follows one output contract.
5. Unit economics stay healthy at the 90th percentile of real usage.

Non-goals, refused rather than deferred

- No advice: no buy, sell, trim, hold or hedge.

- Search doesn't read full filing text; that belongs to research mode.
- Optic stays a premium feature.
- No separate beginner and pro voices. Depth adds evidence, never harder language.
- No model change without a bench that tells fresh answers from cached ones.
- No confidence breakdown on public pages; a partial one hides the penalties.

Principles that settle disputes

1. The number came from somewhere you can check. The server builds the sources; the model's own list is thrown away.
2. A deleted claim costs nothing; a wrong one costs the brand.
3. Honest empty beats reaching.
4. Show real work only.
5. Confidence measures provenance, not correctness.
6. One contract, one vocabulary, one registry, kept in step by a parity test.
7. Answers write themselves in as they stream.

6 Requirements, a sample

The full PRD has 24 requirements for Search and 19 for Optic, each with the test that proves it. A few:

ID	REQUIREMENT	ACCEPTANCE TEST
SR-1	Company names resolve on word boundaries from the SEC's own company index. "Delta" finds the airline; "delta hedging" doesn't.	A 50-name test set (delta, oracle, target, apple, visa) and zero false matches on 200 unrelated sentences
SR-4	A page enters an answer only if it names the subject in its title, URL or opening text.	Off-subject sources under 2% on the standing test bench, labeled by hand
SR-11	The source list the reader sees is the one the model numbered against, live, cached or shared.	Every citation marker is in range and points at the page shown at that number
OR-1	Every run sets a recency band from the question; widening a thin search moves to the next band and never drops the date filter.	Tested band by band; "any time" still has a ceiling

ID	REQUIREMENT	ACCEPTANCE TEST
OR-7	The extractor may only return numbers whose digits appear in the article.	Unverified figures are counted: zero on the bench
OR-8	When sources disagree on a figure, both sides are shown; the model never settles it.	Conflicts render both values and both sources

7 Output contract

One reader, one voice, and four beats, in order, in every answer.

BEAT	ANSWERS	RULE
Read	What is happening?	Plain words and one concrete number
Why	Why is it happening?	The mechanism, linked with “because”
So-what	What does it mean for me?	Who gains, who loses; never an instruction
Watch	What would change the read?	A concrete level or dated event: if X crosses Y, then Z

Words fall into three kinds. Shorthand is never used (hike, print, risk-on, the Street). Names are used as they are, with an explanation of six words or fewer the first time: “Brent crude, the world benchmark oil price…”. Concepts are kept and defined on the spot (free cash flow, priced in), at most two per answer.

8 Trust and provenance

The tier letter is the product. A registry of about 560 publishers and official sources gives each one a tier:

A	Primary sources: issuers, regulators, central banks, statistical agencies, wire services
B	Major financial press
C	Secondary press and research shops
D	Retail and aggregator sites

× Excluded: content farms, press-release wires, forums and affiliate sites

Only A and B can anchor a claim; C and D add color. Domains match on dot boundaries, so a lookalike site can't borrow a real publisher's letter.

Confidence combines publisher quality, how many claims and figures more than one outlet backs, and the spread of sources. The server computes it; the model never rates itself. It measures where an answer came from, not whether it's right, so accuracy is reported as a separate number.

9 Success metrics and release gates

METRIC	TARGET	CADENCE
Off-subject sources numbered (Search)	Under 2%	Every release
Dangling citations	0	Every release
Citations pointing at the page the model saw	100%, 30 answers in a row	Every release
Factual accuracy on checkable claims	At least 90%	Monthly
Cited figures found in the cited source	At least 90%	Monthly
Optic: fact-bearing sources on an analytical question	At least 8	Every release
First section, cold	10s or under	Monthly
Full answer, cold	40s or under	Monthly

Release gates: any change to how answers are made clears all seven

1. Unit and contract tests green, including the registry parity test.
2. The standing test bench unmoved, with every row marked fresh or replayed.
3. Deploy from a known base, compared before and after.
4. Independent review through three lenses (correctness, security and cost, product trust), with a chair.
5. For anything visual, QA at two widths with reduced motion, and design sign-off.
6. After deploy, a fresh question in production that proves the change by its behavior.

7. The rollback is named before the deploy.

10 Phased plan

Trust first, then voice, then depth. Each phase ships, and can be rolled back, on its own.

PHASE	WHAT SHIPS	DONE WHEN	WEEKS
0 · Trust floor	Name resolver, relevance floor, honest empty state, aligned citations, cited-only counts	Zero off-subject sources on the test questions	1–2
1 · Output contract	The plain-language contract on every surface at once	Stored answers 100% clean on the vocabulary	3–4
2 · Accuracy gates and trust panel	Four accuracy checks, a labeled question set, the redesigned trust panel	Accuracy 85% in month one, 90% by quarter end	5–6
3 · Optic depth and age	Recency bands, analytical fan-out, age labels, dated source cards	At least 8 fact-bearing sources on the analytical test	7–9
4 · Optic surface	The reading-column design, mobile Optic, streaming kept	A reader finishes a briefing on a phone	10–12
5 · Research mode	A fundamentals and valuation layer with server-owned numbers	Valuation answers with dated, sourced figures only	13–20
6 · Latency and cost	Earlier streaming, a shorter prompt, a monthly cost report	First section in 10s or under, cold	21–24

11 Risks and open decisions

Risks, each with its mitigation

The relevance floor over-filters a page that names only the ticker.

The floor also accepts tickers and brand names, and recall is measured.

An accuracy check repairs a number wrongly.

Repair only when the math is exact; otherwise drop the sentence, and log every change.

Age labels hide history that matters.

Older sources stay readable, marked Historical; they just can't lead or count toward agreement.

One plain voice frustrates professional readers.

Names and concepts stay; depth adds evidence, and churn is tracked.

Open decisions

- Shared pages: split sources into cited and also-retrieved, or hide the uncited ones?
- Investor-relations sites: tier C for all, or A for the largest?
- When the provenance check flags a figure: remove it, cap the score, or both?
- The verdict word: keep "Verified" in the app, or use the softer "Strong signal" everywhere?

Abbreviated from the full PRD. Pricing, internal dates and confidential detail are left out.