

// THE PROMPT STACK

A Field Guide

Getting AI to give you real answers, not confident guesses.

v1.0 · 2026

Contents

This is a working document. Keep it open while you're doing the real thing. The methodology sits in Section 2; the prompts you'll reach for most are in Section 3. If you read nothing else, the page after this one – the field reference card – is the whole guide on one sheet.

00	Field reference – the whole methodology on one page	03
01	Why this exists	04
02	The Prompt Stack – methodology in four stages	05
03	Five worked prompts – copy-paste swipe file	09
04	Two real-world tests	14
05	When to use AI – and when not to	18
06	What's next	19
07	Colophon	20

NOT FINANCIAL ADVICE. This guide documents one private investor's methodology. Nothing here constitutes a recommendation to buy or sell any security. You are responsible for your own decisions.

01

ROLE

Set the stance before the question. The model's default is helpful and quietly bullish. Pick a stance that fights that.

You are a cautious analyst reviewing this position for a sceptical risk committee. You do not assume the case is correct. Surface evidence first, opinion last.

02

FILTER

Separate observable facts from inference before any reasoning starts. If column A is empty, the model doesn't know enough to be useful.

Before any opinion, list the observable facts you're drawing on. Then separately, list the inferences. Anything that is neither, remove. Only then may you proceed.

03

RISK

Make the downside specific and observable. A risk you can't see in advance is a worry; a risk with a tripwire is something you can act on.

Answer three questions: (1) what could go wrong on a 12-month view? (2) what would invalidate the case? (3) what would I see in the next 90 days that would tell me I was wrong?

04

VERDICT

One action, with a confidence level. "Low confidence" is an answer. Hedging into a balanced summary is not.

Give one practical action: hold, add, trim, exit, or wait. State confidence as low / medium / high and one sentence on why that level. Do not hedge into a non-answer.

Why this exists

Last May I asked Gemini to help me think about a covered call on a small US-listed name I was already holding. I told it plainly that I did not have a live options chain in front of it and asked it to reason about the trade-off between two strikes using only the numbers I gave it. It came back with a neatly formatted table containing specific premiums, an implied volatility figure, and a stock price that wasn't the one I'd typed in. None of those numbers were real. The model had pattern-matched what an options table is supposed to look like and produced one.

I am not anti-AI. I use these tools every week. I just don't trust any tool that won't show its working, and a depressing amount of what's marketed as "AI for investors" is exactly that — a tool that won't show its working, dressed up in the language of analysis. The output looks like research. It has a table. It has specific numbers. A reader who didn't know to check would treat those numbers as retrieved market data and act on them.

The Prompt Stack is the method I worked out from the other end of that experience. Four stages, each a checkpoint: set the stance, separate fact from inference, make the downside observable, force a committed verdict. The aim is to leave a paper trail at every step — so that when the answer feels compelling, I can go back and see whether the compelling part was real.

This guide is what I'd hand a friend who told me they'd started using ChatGPT for stock research. Not "here's a list of prompts" — that's a freebie. Here's the method, the failure modes I've documented on real tools at real moments, the five prompts I actually run after I've bought something, and two recent tests where the method either caught the failure or would have. Nothing here is theoretical. Every claim has a date next to it.

The shortest possible version: if you can't see why the AI gave you the answer it gave you, the answer is provisional until you can.

// SECTION 02

The Prompt Stack

Four stages. Each one a checkpoint. The methodology used in every post on dixon.ai, in the form it's actually run – and the failure modes you'll hit at each stage on each tool.

Stage 01 — ROLE

Out of the box, the model leans agreeable, balanced, and quietly bullish. That's not a bug — it's a design choice. It has been trained to be helpful, and helpful in its learned sense means giving you something that sounds plausible and pleases the asker. For stock research, that is exactly the wrong default. So you refuse it. Before the question, set a stance — a specific role with a specific job — that's incompatible with cheerleading.

The role does most of the work the rest of the prompt is meant to do. Spend more time on this than feels natural. A vague "act as a financial analyst" gets you generic. A specific "you are a cautious analyst for a sceptical risk committee, surface evidence first, opinion last" gets you something usable.

// THE STANCE I USE MOST OFTEN

You are a cautious analyst reviewing this position for a sceptical risk committee. You do not assume the case for the position is correct. Your job is to lay out evidence, then list what would change your view. You will not produce a summary or an opinion until you have explicitly listed observable facts.

// WITHOUT ROLE

Asked "is AAPL a reasonable covered-call candidate?", the tool listed pros, listed cons, ended with a hedge. No verdict, no specific evidence the user could check. Sounds balanced; isn't.

// WITH ROLE APPLIED

Same question, cautious-analyst stance set up front. The model produced IV rank (18), IV percentile (12%), flagged AAPL at fresh all-time highs, named the next earnings date, and arrived at a specific HOLD OFF.

// FAILURE MODE AT THIS STAGE

Gemini in particular generates confident-sounding analysis without the cautious-analyst frame, producing outputs that read as balanced but aren't — every paragraph leads with the positive case, the caveats are buried, and the verdict drifts towards "interesting opportunity." If you skip the ROLE stage, you get the cheerleader version by default. The fix is mechanical: write the stance, paste it as the first line of every prompt, never assume a previous turn's framing has carried over.

Stage 02 – FILTER

Once the stance is set, the next move is to force the model to separate what it actually knows from what it is inferring. The instruction is simple: before any opinion, list observable facts; then separately list inferences; remove anything that is neither. What you're after is a column A of numbers, dates, named events and quoted text, a column B of analytical inferences drawn from column A, and nothing else. Filler – adjectives, vague generalisations, "strong management track record" without a specific basis – goes nowhere.

If the model can't fill column A meaningfully on a given question, the answer is: it doesn't know enough to be useful on this one. That is itself a result. Stop there.

// STAGE 02 INSTRUCTION

Before giving any opinion, list the observable facts you're drawing on. Then separately, list the inferences you're making from those facts. Mark anything that is neither – anything you are asserting without a clear basis – and remove it. Only after this step is complete may you proceed.

// WITHOUT FILTER

Asked about MSFT's competitive position, the model returned four paragraphs blending facts, inferences, and "strong moat" filler – no way to check which sentence was based on what.

// WITH FILTER APPLIED

Same question, filter forced. Five observable facts (named products, named contract values, segment growth rates, two cited filings). Three inferences. Two pieces of filler called out and dropped.

// FAILURE MODE AT THIS STAGE

Perplexity runs live web searches even when explicitly told to work from a provided document. I have watched it route to ten external searches on a prompt that began "work only from the document I have pasted below." It imports its own live data and starts reasoning over that instead, which makes the observable-facts-versus-assumptions separation impossible – the "facts" are now a mix of the document and whatever it just retrieved. The fix is to recognise the architectural mismatch and use a different tool for any task where the FILTER discipline matters. Perplexity is built to retrieve, not to read what you give it.

Stage 03 — RISK

The facts and inferences are separated. Now force the model to name the downside in a way you can actually observe. Three questions: what could go wrong on a 12-month view; what would have to be true for the case to be invalidated; and — the one most often missed — what would you see in the next 90 days that would tell you you were wrong.

Question three is where the methodology starts paying back. A risk you can't observe in advance is a worry; a risk with a tripwire — a specific filing, a named metric, a dated event — is something you can act on. Write the tripwire down. Set a calendar reminder. If the model can't give you an observable disconfirming signal, it doesn't have a real view, just pattern-matching.

// STAGE 03 INSTRUCTION

Now that the facts and inferences are separated, answer three questions specifically: (1) what could go wrong on a 12-month timeframe? (2) what would have to be true for the case to be invalidated? (3) what would I see in the next 90 days that would tell me I was wrong? Each answer must be specific and observable — name the filing, the metric, the event.

// WITHOUT RISK

Asked about NVDA at all-time highs, the model returned generic risks: "competition", "regulatory pressure", "AI investment cycle peaking." Nothing observable, nothing dated.

// WITH RISK APPLIED

Same question, three-part RISK forced. 12-month: hyperscaler CapEx slowdown. Invalidation: top-2 customer order revisions. 90-day tripwire: AMZN and MSFT CapEx commentary on the next results calls.

// FAILURE MODE AT THIS STAGE

This is where confabulation is most dangerous. A model asked to name what would invalidate its view will often generate plausible-sounding tripwires that aren't drawn from the document at all. In one of the tests in Section 4 of this guide, Gemini produced specific options strike levels, an implied volatility figure, and a stock price for a covered-call set-up — none of which I had given it, and the stock price wasn't even the one I'd typed. The fix is to require, in the prompt itself, that every tripwire cite its source — a specific filing, paragraph, or number in what you actually provided. Anything that doesn't get treated as provisional until verified.

Stage 04 — VERDICT

One action: hold, add, trim, exit, or wait. One confidence level: low, medium, or high. One sentence on why that level. No hedging. If confidence is low, that is the answer — write down what would have to change to move it up. A low-confidence verdict that names its own gaps is more useful than a balanced summary that says nothing in three paragraphs.

Forcing the confidence label is what makes the stack honest. It gives you something to check against later: was the model right when it was confident, and wrong when it wasn't? Over time that calibration is the only way to tell whether the tool is actually adding anything to your process or whether it's flattering you.

// STAGE 04 INSTRUCTION

Give one practical action: hold, add, trim, exit, or wait. State your confidence level (low / medium / high) and one sentence explaining why that level. Do not hedge into a non-answer; if confidence is low, that is the answer — name what would have to change to move it up.

// WITHOUT VERDICT

Asked whether to trim AAPL after a 22% run, the model offered "you could consider trimming, though there are also reasons to hold." Useful to nobody.

// WITH VERDICT APPLIED

Same question, verdict forced. "Trim 25%, medium confidence — the thesis hasn't broken but position weighting is now beyond my own size rules. Confidence is not high because IV is depressed and a covered-call alternative is on the table."

// FAILURE MODE AT THIS STAGE

ChatGPT on the free plan rate-limits mid-session and quietly downgrades the model partway through a multi-step analysis. On a four-dimensional earnings test I ran in April, GPT-4o's daily allocation ran out between turn three and turn four — the verdict stage. The model that finished the analysis was not the one that started it, and the verdict it produced reflected that downgrade. The verdict stage also gets the most hedging by default — "it depends", "consult a professional", "this is not financial advice as it relates to your situation" — unless you explicitly instruct one practical action and one stated confidence level. Both failures have the same fix: require the verdict structure in the prompt itself, and if a session looks like it's about to run long, run the verdict stage first while the model is still operating at full capacity.

// SECTION 03

Five worked prompts

A copy-paste swipe file for the five most common post-buy moments. Each prompt is complete – orange placeholders mark the slots where your own details go. Replace them before sending.

// PROMPT 01 OF 05

Thesis integrity check

// THE DECISION MOMENT

You bought. Now something has happened — a price move, a news story, a results release. Is the thesis still intact, or has the underlying case actually changed?

// THE DISCIPLINE

Most investors conflate "the stock went down" with "my thesis is wrong." Those are different things. This prompt separates them.

// PROMPT – COPY THIS

ROLE: Act as a sceptical thesis auditor, not a reassurer. Your job is to tell me whether my investment thesis has been challenged by new information – not whether the stock has moved.

FILTER: My original thesis when I bought [COMPANY NAME]: [YOUR THESIS IN 2-3 SENTENCES]. What has happened since then: [WHAT HAPPENED AND WHEN].

Separate: (1) what has factually changed about the business fundamentals, (2) what is market sentiment or price noise, and (3) what remains unchanged about my original thesis. Be specific.

RISK: Name one scenario in which the information above is more serious than I'm treating it – i.e. where I might be rationalising rather than reasoning.

VERDICT: One of three: "Thesis intact", "Thesis weakened – watch X", or "Thesis broken – review your position". One sentence on what would move you between categories.

// WHAT GOOD OUTPUT LOOKS LIKE

Three distinct columns that don't blur into each other. A scenario that specifically describes how you might be rationalising – not generic "the market could be wrong" filler. A categorical verdict and a single sentence on what would shift it.

// COMMON FAILURE IF YOU SKIP ROLE

The model defaults to reassurance. You will get a four-paragraph "here are the reasons to stay calm" response that's structured to make you feel better, not to interrogate the position. Without the sceptical-auditor stance, you're just asking your most agreeable friend whether you should worry.

// PROMPT 02 OF 05

Earnings read (post-results)

// THE DECISION MOMENT

The company has reported. You have the numbers. The harder question is what management committed to, what they hedged, and what they conspicuously didn't mention.

// THE DISCIPLINE

Numbers are the baseline. The forward-looking language is where the asymmetric information lives — and where AI tools are either genuinely useful or visibly out of their depth.

// PROMPT – COPY THIS

ROLE: Act as a cautious analyst reading this results release for what management committed to versus what they avoided committing to. Do not be impressed by revenue beats – focus on forward guidance quality.

FILTER: Here is the key section of the results / earnings call: [PASTE YOUR EXCERPT HERE]. My thesis when I bought centred on: [YOUR KEY THESIS POINT – the thing this result should confirm or challenge].

Identify: (1) specific forward commitments management made (numbers, timelines), (2) language that sounds forward-looking but contains no commitment ("we remain focused on..."), (3) any dimension of the business they did not address that you would have expected them to.

VERDICT: Did this result strengthen, weaken, or leave unchanged the specific thesis element you identified? One sentence. No hedging.

// WHAT GOOD OUTPUT LOOKS LIKE

Two named commitments with specific numbers and timelines. Two pieces of language flagged as forward-looking-but-empty, with the exact phrase quoted back. At least one omission identified — a topic the company would normally address that's missing from this release.

// COMMON FAILURE IF YOU SKIP FILTER

The model summarises the release back at you and calls it analysis. You'll get "revenue beat, margins held, management struck a confident tone" — a paraphrase, not a read. The FILTER discipline forces it to point at the actual words and the actual numbers, separately. Without it, you're paying for a glorified summary of something you already pasted in.

// PROMPT 03 OF 05

Position conviction check

// THE DECISION MOMENT

The stock has moved materially — up or down. You're leaning toward adding, holding, or trimming. Has your conviction actually changed, or are you reacting to the price?

// THE DISCIPLINE

Forces a clean separation between conviction (thesis-based) and comfort (price-based) before any action gets taken.

// PROMPT – COPY THIS

ROLE: Act as a devil's advocate for whatever action I'm leaning toward. If I'm thinking of adding, argue against it. If I'm thinking of trimming, argue against it.

FILTER: I bought [COMPANY NAME] at [PRICE]. It is now at [CURRENT PRICE]. My original conviction level when I bought: [HIGH / MEDIUM / LOW]. What has changed in my view of the fundamentals since then: [WHAT HAS CHANGED, or "nothing"]. What I'm considering doing: [ADD / HOLD / TRIM] and why: [YOUR REASON].

Identify the strongest argument against the action I'm considering. Then tell me what would need to be true for that argument to be wrong.

VERDICT: Is this a conviction-based decision or a price-reaction decision? One sentence. Name the cognitive trap most likely to be at play.

// WHAT GOOD OUTPUT LOOKS LIKE

A specific counter-argument that addresses your actual reasoning — not a generic "the market could keep falling." A named cognitive bias (anchoring, recency, confirmation, sunk cost) tied to the specific decision, not a list of all of them. The verdict commits to one of the two categories.

// COMMON FAILURE IF YOU SKIP ROLE

Without the devil's-advocate framing, the model agrees with whatever action you're leaning toward and finds reasons it's the right one. You'll get sentences that pattern-match to thoughtful analysis but confirm what you walked in believing. The role flip is the only reliable way to break that pattern; "what are the risks of my plan" produces a different output than "argue against my plan."

// PROMPT 04 OF 05

Drawdown discipline

// THE DECISION MOMENT

The stock is down meaningfully from your entry. The honest question — the one most investors avoid — is whether you're holding for a thesis reason or because you don't want to take the loss.

// THE DISCIPLINE

The most useful prompt for avoiding the "I'll wait until it gets back to my entry price" trap. Tests the reason behind the position, not the position itself.

// PROMPT – COPY THIS

ROLE: Act as a loss-aversion auditor. Your job is to surface whether I am holding this position for a thesis reason or a psychological reason. Be direct.

FILTER: I bought [COMPANY NAME] at [PRICE]. It is now at [CURRENT PRICE] – a [X%] move. My original thesis: [YOUR THESIS]. What has changed in the fundamentals since I bought: [WHAT HAS CHANGED, or "nothing visible"]. My honest answer to "why am I still holding": [YOUR HONEST REASON].

Evaluate my reason. Is it a thesis reason (fundamental change would justify the hold) or a psychological reason (anchoring to entry price, sunk cost, hoping for recovery)?

RISK: Name the scenario in which continuing to hold makes the outcome materially worse. What is the cost of being wrong here – not in money terms, but in thesis terms?

VERDICT: "Hold – thesis reason" or "Review – this looks like anchoring." One sentence on what I should look up or decide within the next week.

// WHAT GOOD OUTPUT LOOKS LIKE

A direct read on the reason you wrote, not a polite paraphrase of it. The cost-of-being-wrong is framed in thesis terms (opportunity cost, dead capital, missed reallocation) rather than in pound terms. A specific next-week action — a filing to read, a number to check, a date to mark.

// COMMON FAILURE IF YOU SKIP RISK

Without the RISK stage, the model retreats to "every situation is different, only you can decide" — useless. The RISK stage forces it to commit to one scenario in which holding compounds the error, and to define the cost in observable terms. Skip it and you get a kind, balanced reply that lets you keep doing what you were doing.

// PROMPT 05 OF 05

Exit planning

// THE DECISION MOMENT

At some point you'll sell. Most retail investors have no exit criteria — they wait until emotion forces the decision. This prompt sets the criteria in advance, when you're thinking clearly.

// THE DISCIPLINE

Write the sell thesis when you're calm. Not when you're panicking, not when the stock is running — when you can think.

// PROMPT – COPY THIS

ROLE: Act as an exit-criteria designer. Help me define in advance the conditions under which I would sell this position — before emotion makes that decision for me.

FILTER: I hold [COMPANY NAME]. My thesis: [YOUR THESIS]. My original time horizon when I bought: [YOUR TIME HORIZON]. The current situation: [WHERE THINGS STAND NOW].

Define three specific exit triggers for this position:

1. A fundamental trigger — something that would mean the thesis is broken
2. A valuation trigger — a price or metric level at which the risk/reward no longer makes sense
3. A time trigger — if X hasn't happened by [YOUR TARGET DATE], the thesis has not played out

For each trigger, be specific. "If the stock falls 20%" is not a trigger — it's a price reaction. "If management fails to show margin improvement in two consecutive quarters" is a trigger.

VERDICT: Write the sell thesis in two sentences — the conditions under which you would sell this stock without hesitation, stated clearly enough that you could apply them under pressure.

// WHAT GOOD OUTPUT LOOKS LIKE

Three triggers, each tied to a number or a named event — not "if sentiment shifts." A sell thesis short enough to remember under pressure. Anything longer than two sentences fails the "apply under pressure" test.

// COMMON FAILURE IF YOU SKIP VERDICT

Without forcing the two-sentence sell thesis, the model returns a list of considerations — useful for thinking, useless under pressure. Exit criteria you can't recall in 30 seconds when something is moving don't function as exit criteria. The VERDICT stage compresses the analysis into something you can actually apply.

// SECTION 04

Two real-world tests

Documented findings from live tests on dixon.ai. Each one shows what actually happened, what the methodology read is, and which stage of the Prompt Stack would have caught the failure.

// TEST 01 OF 02

Gemini fabricates an options table

DOCUMENTED 14 MAY 2026 · BMNR · LIVE TEST

// THE PROMPT

I asked Gemini to reason about a covered call on BMNR (BitMine Immersion Technologies) — a name I trade. Real numbers: cost basis \$22.11, current price \$21.50, the strikes I was actually considering (\$26 and \$27). The explicit instruction in the prompt: *"without access to a live options chain"*. Three sub-questions about implied volatility interpretation, the trade-off between the two strikes, and what to verify from the broker before placing the trade.

// WHAT HAPPENED

Gemini produced a formatted comparison table containing specific premium estimates — "\$3.50–\$4.00" for the \$26 strike, "\$2.80–\$3.20" for the \$27 strike. I had not given it any premium data. It generated those ranges. Alongside the table, it stated implied volatility was "currently around 75%" — a figure I never supplied. And it used the wrong stock price (\$28.60 instead of the \$21.50 in my prompt) and the wrong cost basis (\$25.40 instead of \$22.11). It actually noticed the stock-price discrepancy mid-response, flagged it, and proceeded to generate the estimates anyway.

The mechanism is the failure mode the Prompt Stack was designed to prevent: confident-sounding output without underlying evidence. The output looks like research. It has a table. It has specific numbers and ranges. A reader who didn't know to check would treat those premiums as real market data and place a trade against them.

For comparison: Claude, given the same prompt on the same day, returned text only. No table, no premium figures, no IV. Its response was explicit: *"you'll plug in real premiums from the chain."* Conceptual analysis on the trade-off between the strikes, no fictional precision. That is what staying in lane looks like.

The methodology read

This is a Stage 03 (RISK) failure with a Stage 02 (FILTER) precondition. The FILTER stage, run properly, would have refused to proceed: the model had not been given premium data, IV, or even an accurate stock price, so the observable-facts column for any options output was empty. Anything generated past that point was inference dressed as fact. The RISK stage would have caught the rest: a tripwire requirement on every figure ("cite the source — the filing, the paragraph, the number you were given") makes confabulation visible because the model has nowhere to cite from.

What it tells you about the tool: Gemini has a structural willingness to generate plausible-looking numbers when the right answer is "I don't have that information." It is not pattern-matching to retrieved data; it is pattern-matching to *what a response of that shape is supposed to look like*. The table format, the dollar-range notation, the IV figure — they are all consequences of the prompt's shape, not its content.

What it tells you about the method: the Prompt Stack does not stop the model from confabulating. It makes the confabulation visible. Without the FILTER stage, the BMNR options table is unchallengeable — it just looks like research. With it, you have a column A that should contain "premium data given" and "IV given" and doesn't. The gap is the answer.

// THE TAKEAWAY

If you use Gemini for any task that touches options data, don't. It will invent premiums, IV figures, and Greeks with complete confidence, and format them in a way that makes them look retrieved. The wider lesson: anywhere a tool produces a structured number — multiple, ratio, price target — that you did not give it, treat it as provisional until you have located the source. The structure of the response, by itself, is no evidence at all.

Full write-up: dixon.ai/posts/chatgpt-vs-claude-vs-perplexity-stock-research

// TEST 02 OF 02

Meta Q1 — faith-based versus fact-based defence

DOCUMENTED APRIL-MAY 2026 • META • LIVE TEST

// THE PROMPT

Two-turn conversation on Meta's Q1 2026 earnings release. Part A: paste the management commentary excerpt — Zuckerberg on AI spending, "we remain committed", "\$125B–\$145B CapEx", "conviction in the opportunity ahead" — and ask the tool to analyse the language for tone and intent, before knowing what the actual results were. Part B: reveal that Q1 delivered revenue +33%, operating margin held flat at 41%, and the stock fell 6 to 8 per cent on the CapEx raise. Does the language look appropriate in hindsight?

// WHAT HAPPENED

Part A across all three tools produced reasonable surface readings: ChatGPT identified the "we believe" versus "we're seeing" distinction; Claude picked apart the phrasing line by line and flagged "conviction" as a tell-word that surfaces when management wants to project certainty without offering specific metrics. Useful, but the test really hinged on Part B.

Claude's Part B was the standout of the entire battery. Verbatim:

What looks worse in hindsight is the vagueness. With revenue up 33% and margins holding at 41% through an aggressive spending ramp, management actually had a strong empirical case to make and didn't make it. [...] Leaving that unspoken and reaching instead for "conviction in the opportunity ahead" traded a fact-based defence for a faith-based one, and the market priced the faith.

That is the right read. Holding 41 per cent margins through an aggressive CapEx ramp is the hard part — evidence the core business is absorbing the investment rather than buckling under it. Management had numbers to point at. They chose conviction language instead. The market priced the absence of evidence, not the spending. Claude also caught the second-order point: with operating results that strong, management had earned the credibility to give a tighter CapEx range. Leaving the \$20 billion range wide read less like prudent uncertainty and more like wanting permission to spend to the top of it.

The methodology read

This is a Stage 01 (ROLE) and Stage 02 (FILTER) success. The cautious-analyst framing on the cold-read forced the model to treat the commentary as evidence rather than as message — to read it for what it does, not for what it intends. The FILTER stage forced separation of management's specific claims (the numbers they committed to, the dates they named) from their rhetorical moves (the "conviction" language, the relocation of payoff to an unfalsifiable horizon).

Without those two stages running properly on Part A, the Part B observation is impossible. You can't notice that management traded fact-based for faith-based unless you have already mapped what they said factually versus rhetorically — and you can't do that mapping unless you've set the model up to do it. A generic "what do you think of Meta's Q1?" prompt produces a paraphrase of the press release. The structured prompt produces a read of the language as language.

There is a tool dimension as well. Perplexity, on the same prompt, ran 15 external searches and answered Part A based on analyst commentary about the call rather than the language in the excerpt. ChatGPT's Part A was strong and would have made it a clear second on Part B — except the free plan's daily GPT-4o allocation ran out partway through the session, the model got downgraded, and Part B simply didn't run on the same model that started the conversation. Two different failure modes, both addressable: one structural (use a different tool when document-reading discipline matters), one operational (run the verdict stage first if there's any risk of the session running long).

// THE TAKEAWAY

Structured prompting did the work that generic prompting can't. The reading wasn't waiting in the model to be unlocked — it was produced because the stages forced the kind of attention that produced it. The methodology is not a finishing flourish on a good prompt; it is the prompt. The output quality follows from it.

Full write-up: dixon.ai/posts/best-ai-tools-for-earnings-analysis

When to use AI · when not to

The honest version. AI is genuinely useful in narrow, specific situations and reliably bad in others. The dividing line is whether the task requires reasoning over what you give it (where the tool can help) or live, accurate, structured data it has to fetch itself (where it often quietly fails).

// AI IS THE RIGHT TOOL

- Breaking down a 10-K you've already opened — point it at specific sections you've flagged.
- Generating the questions to ask in due diligence you'll then do yourself.
- Stress-testing a thesis you're emotionally attached to — devil's advocate role, fresh conversation.
- Reading transcripts and management commentary for tonal shifts and forward-looking-but-empty language.
- Cross-referencing your own assumptions against a set of named sources you've provided.
- Drafting a pre-earnings checklist for a position you already hold.
- Post-mortem analysis: was the decision a process error or an outcome variance?

// AI IS THE WRONG TOOL

- × Anything requiring live market data — go to the actual provider.
- × Options chain pricing — every tool tested has failed this. Premiums, IV, Greeks are fabricated on demand.
- × Regulatory interpretation. The wrong answer here can have consequences a chat box cannot warn you about.
- × Reacting to breaking news in minutes — the methodology takes longer than the move you'd be chasing.
- × Summarising long filings as a substitute for reading them. The paragraph that mattered is the one most likely to get dropped.
- × Anything where "confident-sounding" looks identical to "correct" without the Prompt Stack applied — which is most things, by default.
- × Position sizing or risk management for your specific portfolio. The model doesn't know what else you own.

The simplest test: if you could not check the answer against a primary source within ten minutes, you should not be acting on it.

The next experiments, when there's something worth writing.

This guide is one version of a working method. It will be wrong about something I haven't found yet, and there will be a v2 once enough of those somethings have stacked up to make a refresh honest. New tool generations, new failure modes, refinements to the four stages — v2 takes the latest evidence into account.

In the meantime, the place I publish the experiments themselves is dixon.ai. Lab reports on real decisions, tool audits against real moments, the Prompt Stack applied and re-applied. The categories are lab-report, agent-build, prompt-stack, tool-audit, market-workflow, decision-process — pick whatever sounds least dull.

If you want to know when there's a new one, the newsletter goes out occasionally — only when something is worth your inbox. No drip campaigns, no "five things you must know this week", no posts written so that another email could be sent next week. The bar is the same as the bar for the site: would a financially literate adult, who is short on time, be grateful for this. If yes, it goes out. If no, it doesn't.

dixon.ai — the site.

dixon.ai/start-here — the curated entry point if you'd rather skip the archive.

dixon.ai/contact — direct line. I read these.

Colophon

Field notes on how this document was made and what's deliberately not in it.

// TYPE

Body text set in Playfair Display, deliberately — Inter was considered and rejected; the slightly old-fashioned serif suits a document that is meant to slow you down rather than be skimmed. Labels, prompts and structural elements in JetBrains Mono.

// METHOD

Written from the same Prompt Stack the guide describes. The methodology section is the result of repeated cold reads of my own writing against the four stages — by the time a paragraph here has earned its place, it has been argued with three times.

// WHAT IS NOT HERE

Generic prompt-engineering tips. Options-specific prompts (covered calls, IV percentile, position sizing). Tool-by-tool comparison chapters. Worked examples on AIM-listed names. All deferred to v2 — v1 is deliberately scoped to core methodology plus the five post-buy prompts plus two recent tests.

// PALETTE

Paper-bg #F4F2EC, paper-dark #EBE8DF, ink-black #0A0A0A, brand-orange #FF5A00. No gradients in light sections, hard 2px borders, no rounded corners. Orange used as punctuation — one accent per page, not a colour scheme.

// PROVENANCE

All real-world tests in Section 04 are linked to the original post on dixon.ai, with the date of the test stamped. Tests are documented evidence "as of May 2026" — they are not current claims about how the tools behave today. v2 will swap in fresh tests.

// WHAT V2 WILL PROBABLY INCLUDE

A covered-calls chapter. A UK-listed worked example. A tool-by-tool comparison once five or more new experiments are documented. A 5th stage to the Prompt Stack if the evidence demands one — and the evidence will be on dixon.ai, not invented for the v2 release.

NOT FINANCIAL ADVICE. Ben Dixon documents his own methodology at dixon.ai. Not FCA-regulated. Nothing in this document is a recommendation to buy or sell any security.



DIXON.AI

Real portfolio. Real prompts. The methodology, with the working shown.

dixon.ai

The Field Guide is updated when the methodology changes. Subscribers get the new version automatically.
v2 will refresh the real-world tests and add the deferred chapters – expected late 2026.

Nothing in this document constitutes financial advice or a recommendation to buy or sell any security.

Ben Dixon writes at dixon.ai. Not FCA-regulated.