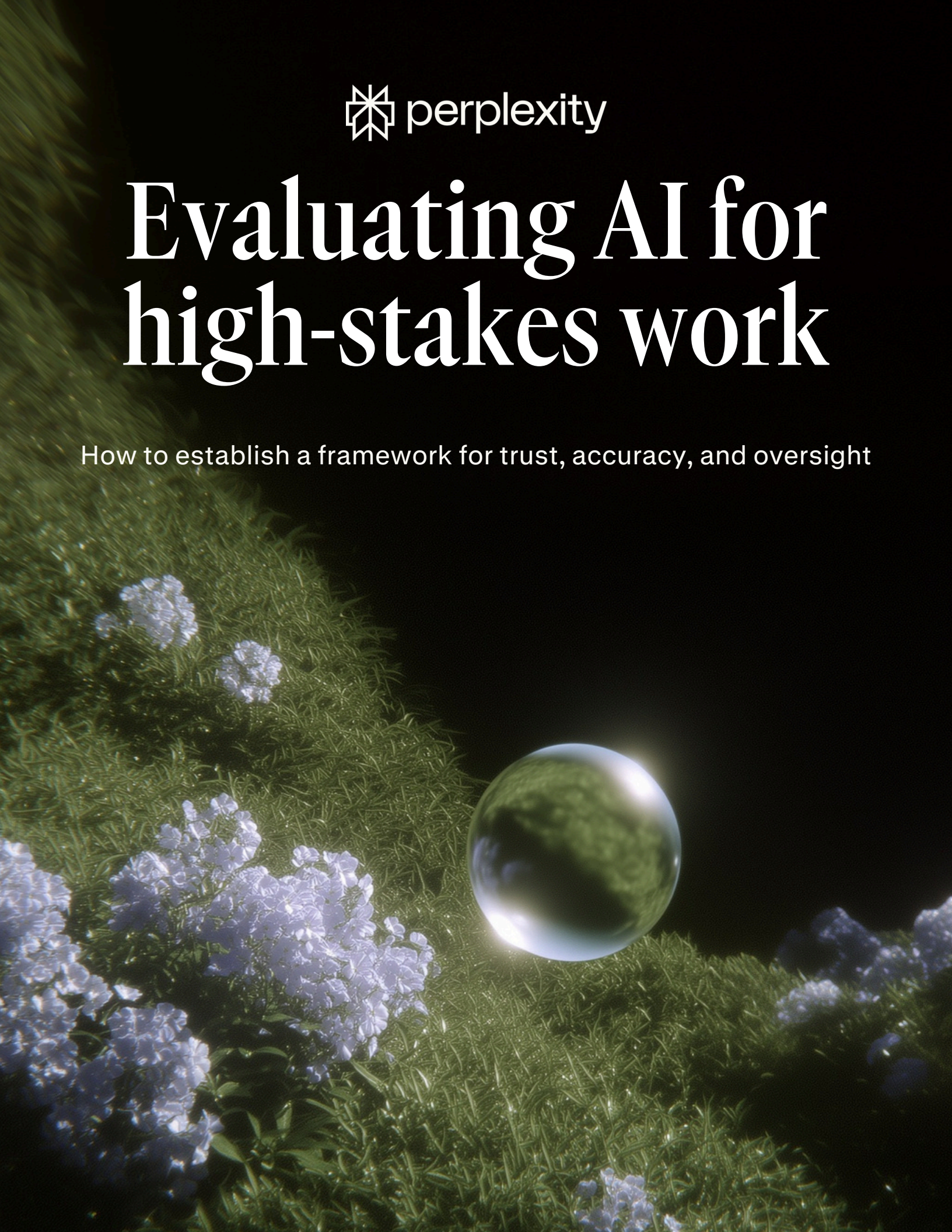
perplexity

Evaluating AI for high-stakes work

How to establish a framework for trust, accuracy, and oversight



perplexity

Identify where trust *blocks adoption.*

Work that shapes important decisions for businesses has to stand up to questions from all manner of stakeholders, including executives, investors, auditors, clients, and legal reviewers. A due diligence report, contract summary, or market analysis needs to show the information it relies on. That standard is especially important when finance, law, and business research teams are using AI, as LLMs pull from a variety of sources before returning a synthesized answer.

Seventy-three percent of IT leaders want end-to-end visibility across AI workflows and systems, according to a 2025 Forrester Consulting study. Teams need a way to verify AI-generated work before it shapes a decision. “Accuracy concerns are the number one barrier to increased AI investment,” states a Thomson Reuters’ Future of Professionals 2025 report. Trust depends on knowing whether the answer is sound, where the information came from, and where human review begins.

By the Numbers

40% of IT leaders said employees do not fully trust AI outcomes.

— FORRESTER CONSULTING

75% of CEOs said trusted AI is impossible without effective governance.

— 2024 IBM STUDY

39% reported strong generative AI governance practices in their organization.

— 2024 IBM STUDY

perplexity

Evaluate AI before it *shapes a decision.*

Teams using AI for work should apply the same rigor to its outputs as to any important piece of work, especially around accuracy and transparency. Before an output informs a decision, reviewers need to answer three questions: Is it accurate? Can the evidence be checked? Where does human review fit in?

Is the answer accurate?

Accuracy looks different depending on the task. At the bare minimum, the output must match the underlying facts, figures, and source materials before style or usefulness matters. More open-ended tasks need standards too. Drafting, editing, summarizing, explaining, and planning can look eloquent even when the answer misses the request. A strong response should follow the instructions, preserve the meaning of the source material, and solve the problem it was asked to handle. Without that check, fluent prose can look finished before anyone knows whether it's useful.

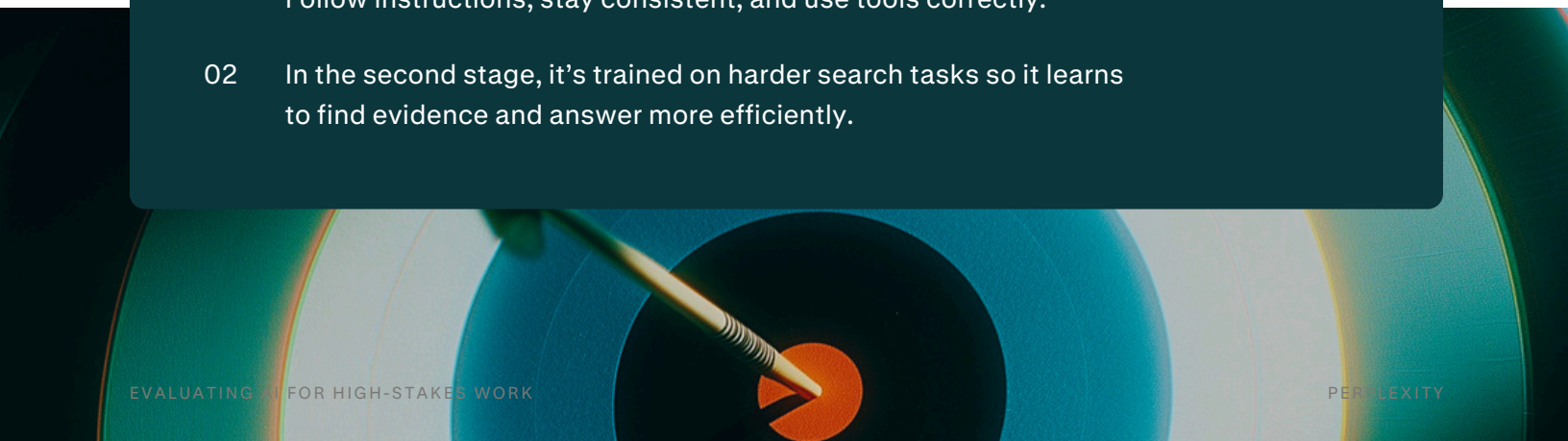
HOW PERPLEXITY TRAINS FOR ACCURACY

Delivering better answers requires constant refinement. No system is 100% accurate, which is why our process pairs training with ongoing testing and measurement.

When post-training models, we prioritize accuracy over polish with a two-stage process:

01 In the first stage, we give the model behavioral instructions: Follow instructions, stay consistent, and use tools correctly.

02 In the second stage, it's trained on harder search tasks so it learns to find evidence and answer more efficiently.



perplexity

Evaluate AI before it *shapes a decision.*

Does the answer have visible reasoning and sources?

A market is easier to evaluate when the supporting evidence is reviewable. More information is not necessarily better information. Additional queries should improve the answer, and longer responses should add value. Having visibility into how an AI system works can help people better evaluate its outputs and correct any errors, notes the National Institute of Standards and Technology (NIST). Researchers studying explainability have also found that people tend to place more trust in systems they can understand.

Where does a human review fit in?

The closer the work gets to an impactful business decision, the more human review matters. A typo in a meeting note is different from an error in a contract summary going to a client, or a research report that will be released to the press. Most organizations assume someone will carefully review important work before it goes out. But with AI-generated outputs, it has to be clear where and how a human stays in the loop.

THE FRAMEWORK

Consistent behavior across these pillars leads to **trustworthy AI.**

Accuracy means the output matches relevant facts, figures, and source materials.

Transparency reveals how a system reached a conclusion.

Oversight designates who's responsible for reviewing AI work before it shapes a decision, a contract, or a filing.

perplexity

Identify sensitive *workflows.*

The common thread in fields like finance, law, and research is accountability. If a deal goes sideways, a contract is challenged, or a recommendation falls apart, someone has to explain how they got there. An investment memo, a contract summary, or a market analysis only holds up when its claims can be traced back to real, reviewable evidence.

Research for financial decisions

Financial decisions often come down to a small set of numbers: revenue, margins, forecasts, valuations. Those figures might appear in investor updates, lender packets, audit reviews, or board decks, where different stakeholders will question how they were produced. It only takes one wrong input in a model or spreadsheet to slip into a presentation and then quietly recur in budgets and valuations. AI can help gather and compare the numbers, but finance teams still need to verify the inputs before the analysis leaves the room.

Diligence in legal matters

Legal teams are already using AI for drafting, review, and research. Professionals report meaningful time savings but remain concerned about transparency, oversight, and data privacy. In one Stanford-affiliated study of leading legal AI tools, nearly one in five queries produced misleading or outright false information, and a large share of the remaining answers were incomplete. In legal work, that kind of error can change how a team reads a clause, cites precedent, or decides whether a contract is ready to sign.

Market and business research

Research often sits behind market analyses, hiring recommendations, competitive assessments, and strategic planning. The work only holds together if the sources are auditable. Someone must be responsible for checking them before it informs a budget decision, hiring plan, or product strategy. NIST notes that documentation can improve transparency, support human review, and strengthen accountability in AI systems.

perplexity

Where to *start using AI.*

The best first AI workflows are not the most complex ones. Work that keeps getting pushed to the end of the week is a good starting point: the diligence tracker that needs updated figures, the vendor agreement nobody compared against the last draft, the research brief with scattered sources.



Financial due diligence

Diligence work often means reading long documents, pulling out key figures, noting what's missing, and tracking issues that need a closer look. A team reviewing an acquisition target might ask AI to build a diligence tracker from financial statements, customer contracts, and board materials. The output could list revenue figures, contract renewal dates, open questions, and source links for each item. The final call still depends on whether the numbers, assumptions, and sources hold up.



Legal document review

Thomson Reuters Institute reports that lawyers can spend up to 60% of their time drafting and reviewing contracts and other legal documents. Teams can compare AI output against existing contracts, prior versions, and approved language. AI might flag a changed indemnity clause, summarize renewal terms, or show where payment language differs from the last draft. Someone responsible for the agreement still needs to review the final language.



Research reports

Employees tasked with a market scan, hiring brief, competitive analysis, or customer research project may have to pull information from articles, reports, transcripts, notes, and internal documents. Using AI, teams can collect findings from multiple sources and turn them into a reviewable summary, such as a competitor brief that links each claim back to a source. The summary still needs to show where the information came from and whether the evidence supports the recommendation.

perplexity

Choose platforms that make *review easier.*

As AI moves deeper into finance, legal, and business research workflows, the design of the system is as important as the model itself. Strong platforms make the path from source to answer easier to inspect. Reviewers should be able to see where information came from, which sources shaped the output, and what still needs human judgment.

That visibility matters more when AI connects to email, calendars, documents, chat platforms, and internal systems. The more quickly information moves between tools, the easier it is to lose track of where a conclusion originated. Platforms should make it easy to trace which information shaped an answer and what still needs human input.

How Perplexity approaches review and control

Perplexity trains models to follow instructions, stay consistent, use tools correctly, and handle search tasks that depend on evidence. Correctness takes priority over polished writing.

Perplexity Computer connects to systems such as Gmail, Slack, Notion, Calendar, Microsoft 365, and hundreds of other tools. When an answer draws from Slack, Gmail, and a contract folder, reviewers need to know which message, file, or thread shaped the result. Perplexity's enterprise security program includes SOC 2 Type 2 certification, SSO with MFA, access controls, and audit-oriented production safeguards.



perplexity

A checklist for evaluating AI for *high-stakes work.*

Before assigning a workflow to AI, test it against a completed example the team already trusts. A manually completed diligence project, contract review, or research brief gives reviewers something concrete to compare it with. The outputs don't need to match word for word, but any differences should be explainable.

Accuracy

- [ ] Do the outputs match the underlying facts and figures from the source material?

Transparency

- [ ] Can reviewers see the sources that support the output?

Oversight

- [ ] Is there a clear point where a person reviews the work before it influences a decision?

Trust

- [ ] Does the system identify uncertainty, or does every answer sound equally confident?
- [ ] Is the result consistent when the same task is repeated?

perplexity

The standard doesn’t change.

AI can help produce many of the documents and analyses that sit upstream of a decision: a diligence tracker, contract review, or research summary. The harder question comes after the output appears. Teams need to know whether the numbers hold up, whether the sources are visible, and whether the reviewer understands what they are signing off on. Faster output still has to survive the same review.



Ready to get started with Computer?
perplexity.ai/products/computer