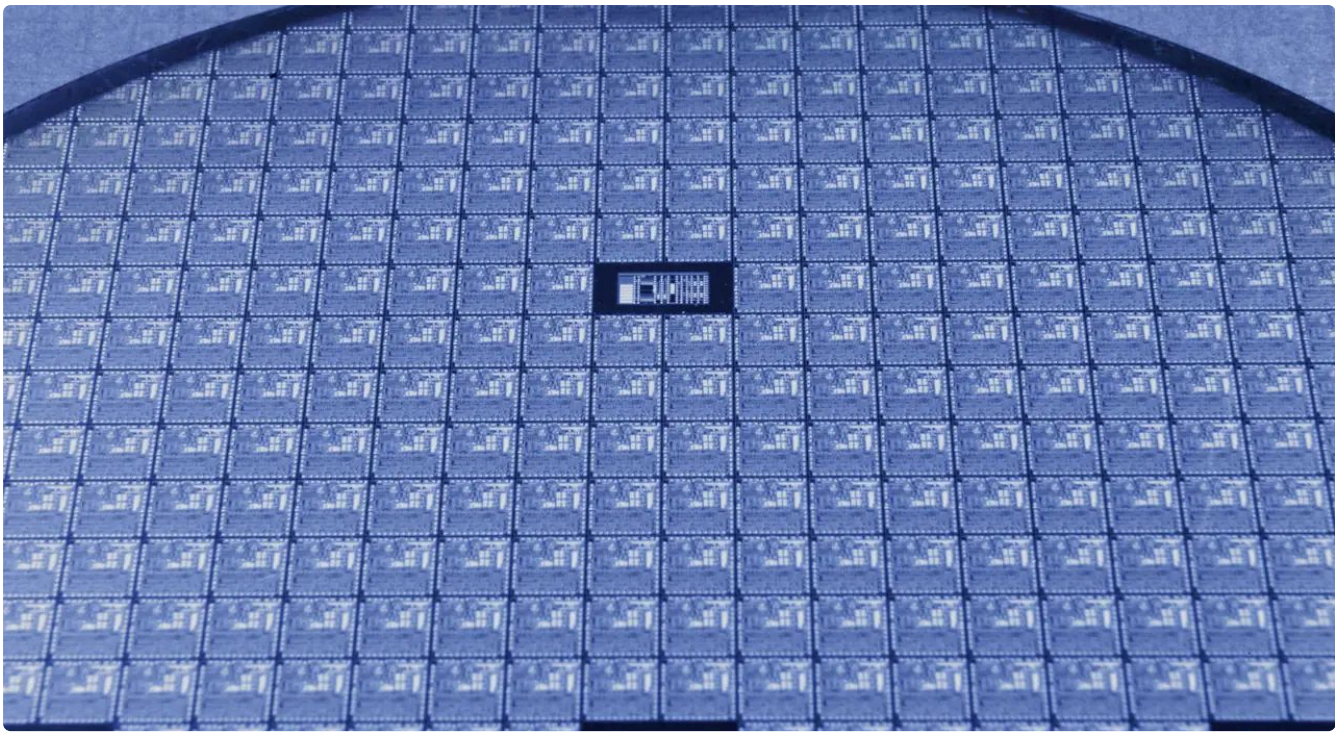


PERSPECTIVE · NOVEMBER 2026 · 6 MIN READ

Prototypes that force decisions

Four rules for experiments that end in a clear decision



Philipp Pöschke
Founder, Paice

AT A GLANCE

- A prototype has done its job when it enables a decision by an agreed date: keep building, adjust or stop.
- Success criteria, measurement and the decision date belong before day one, because any result can be read as interesting after the fact.
- A prototype only becomes meaningful under real conditions, and measurements decide the outcome, because the impressions of those involved can demonstrably mislead.

Prototypes have become routine in many companies. They show that something is technically possible. Whether it's worth it, and what happens next, often remains open afterwards, even though that decision is exactly what they're for.

In July 2024 the analyst firm Gartner predicted that at least 30 per cent of generative AI projects would be abandoned after proof of concept by the end of 2025. It cited poor data quality, inadequate risk controls, escalating costs and unclear business value. We don't see projects ending as a problem. It gets expensive when the decision comes late and nobody has set out in advance what it will be based on.

In the AI delta model, a prototype belongs to exploring: it settles a single question and prepares the next decision. Four rules and an agreement that fits on one page help make that happen.

Rule 1: Start with a question

A good prototype starts with a question you can't answer without it. The question determines what gets built, and that's usually less than expected. Four kinds of question can be distinguished.

FIGURE 1

Four kinds of question a prototype can answer

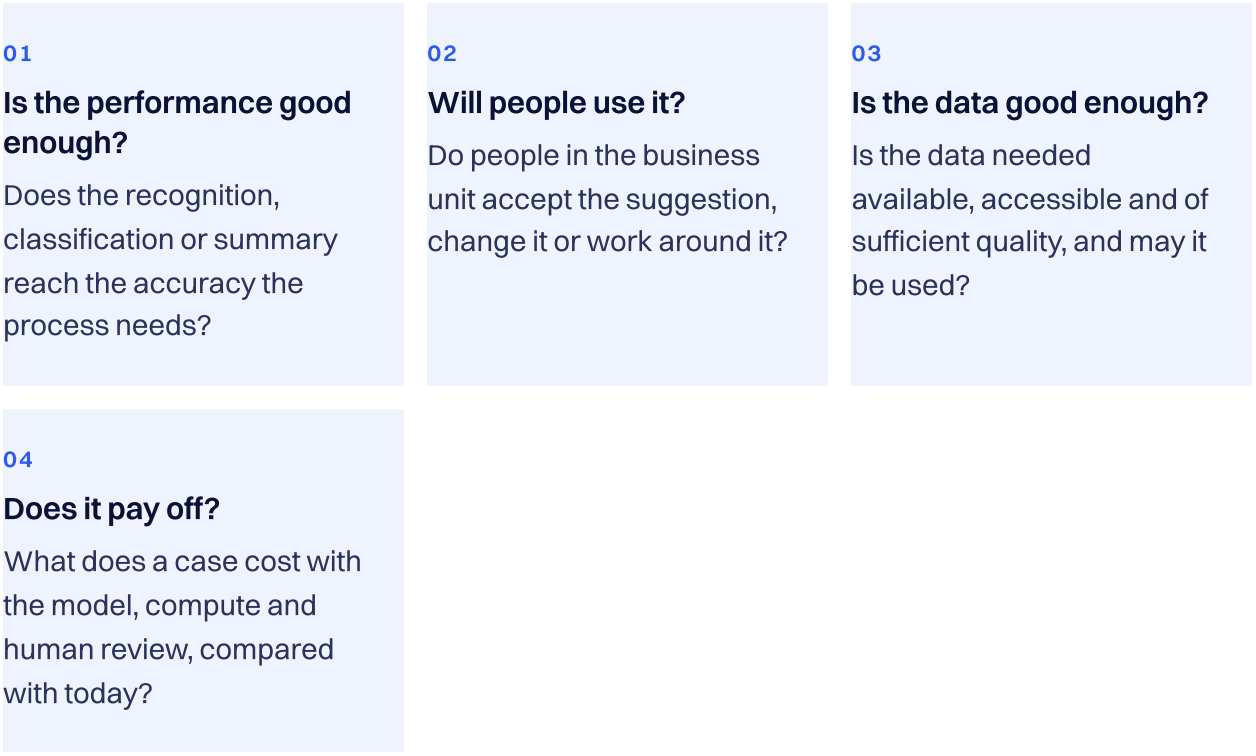


Illustration: Paice

Once the question is fixed, the prototype can be limited to what answers it. A simplified example shows the difference. If the question is whether a model reads the line items on invoices from the most common suppliers reliably enough, the prototype needs neither a user interface nor a connection to accounting. It needs a stack of real invoices, the correct values for them and an evaluation. Only if the answer is positive does the next question follow, such as whether accounting adopts the suggestions in daily work.

Rule 2: Set the criteria before building

Before day one, it's settled when the prototype counts as a success, how that will be measured, who decides and on what date. Without this agreement, almost any result turns into "interesting, let's keep going". With it, there are three clear outcomes: keep building, adjust or stop.

The criteria should come from the business. Accuracy above a certain value is only a meaningful target if someone knows at what value the process can actually change. So the business unit sets the thresholds, and the development team says how they'll be measured.

The duration belongs in the agreement too. It follows from the question and is capped in advance. When the date arrives, a decision is taken, even if the result is incomplete. The decision may then be "adjust", with a new, narrower question and a new date. A prototype with no end date, on the other hand, quietly grows into a project nobody approved.

With generative AI, the same input can also produce different results. A prototype therefore needs a fixed set of test cases with expected results, against which every version is measured. The evaluation also records the type of error, because an invented detail in a contract weighs more than a clumsy phrase.

Rule 3: Test under real conditions

Prototypes on sample data are easily convincing. They only become meaningful with real data, real users and real exceptions. Two studies show why.

Research on the "jagged frontier" of AI shows how misleading examples can be. In an experiment with 758 consultants at a consulting firm, participants using GPT-4 worked considerably faster and better on suitable tasks. On a task that looked similar but lay outside the model's strengths, they were 19 percentage points less likely to produce a correct solution than the group without AI. A prototype tested only on the easy cases tells you nothing about the hard ones.

The second study shows why you have to measure. Experienced open-source developers took 19 per cent longer on their tasks with AI tools. Afterwards they estimated they had been 20 per cent faster. The impression of the people involved is therefore not a result. What gets measured is time, errors and cost, ideally compared with today's process.

Rule 4: Stopping is a result

A prototype that shows something isn't worth pursuing has saved money. For it to be seen that way, stopping needs a form. A short write-up records what was tested, what came out and under what conditions a new attempt would be worthwhile, for example with a better model or new data. That way the learning is kept, and the next team doesn't start from scratch.

The decision should be taken by the person responsible for the budget for the next step. If the team that built the prototype decides, it's ruling on its own work, and that makes stopping needlessly hard.

Companies that accept stopping as a result dare to run more experiments. Many small decisions based on measurements are more robust than a few big bets based on demos.

The common view and ours

The four rules shift what a prototype is measured against. The overview sets the common view against ours.

FIGURE 2

Prototypes: the common view and ours

COMMON VIEW	OUR VIEW
A prototype should show that something is technically possible.	**A prototype should make a decision possible.**
Success means the prototype gets developed further.	**Success means a decision is taken on the agreed date, even if that decision is to stop.**
Tested on sample data, judged by impression.	**Tested on real data and exceptions, judged against measurements set in advance.**
A cancelled prototype is a failure.	**A cancelled prototype has prevented an investment that wouldn't have paid off.**

Illustration: Paice

The difference is clearest in what counts as success. If only a prototype that gets developed further counts, every team has a reason to read results generously and push dates back. If a timely decision counts, it pays to be honest early. For a portfolio of experiments that's the better yardstick, because it steers budget towards where the measurements hold up.

The prototype agreement

The four rules can be captured in an agreement signed before the start by the person who decides and the team that builds. It fits on one page and covers six points:

01 **Question:** Which question does the prototype answer, and which decision depends on it?

02 **Assumption:** What would have to be true for the initiative to pay off?

03 **Criteria:** From which values is the question considered answered, and how is it measured?

04 **Conditions:** With which data, users and cases is it tested, and for how long?

05 **Decision:** Who decides, and on what date?

06 **Outcomes:** What happens if we keep building, adjust or stop, including budget and ownership?

> A prototype is finished when it's been decided what happens to it.

Sources

- Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025, Gartner, 29 Jul 2024
- Navigating the Jagged Technological Frontier, Dell'Acqua et al., Harvard Business School Working Paper 24-013, 22 Sep 2023
- Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, Becker et al., METR, 12 Jul 2025



AUTHOR

Philipp Pöschke

Founder, Paice

Philipp Pöschke founded Paice in Bremen in 2026. He is personally accountable for every engagement and works hands-on on analyses, prototypes and solutions. He is the author of the articles in Insights.