

What a pilot gets for free that production has to pay for

A pilot looks inexpensive because it runs on things nobody paid for. None of them is on an invoice, and production sends the bill for each.

Author	Amit Govil, Founder and CEO, XONIK
Published	October 22, 2026
Reading time	8 min read
Online	https://xonik.com/insights/perspectives/what-a-pilot-gets-for-free-that-production-has-to-pay-for/

IN BRIEF

A pilot costs little because it runs on things nobody pays for: a patient audience, a hand-picked extract, a volunteer's time, no audit and no one on call. Production has to pay for each, from a budget sized on the pilot's cost. Writing them down before the pilot ends, with a payer beside each, turns an overrun into a plan.

KEY FINDINGS

1. A pilot's modest bill measures the model, not the work around it.

Each free thing was paid for by someone outside the project, and in production that someone sends it back. The sponsor, seeing a good result at a low cost, sets the production budget by extension.

2. The hand-picked extract becomes a pipeline that someone must build, watch and repair.

In production the data arrives with late entries, changed codes and fields typed at the end of a shift, so the model sees a different population from the one it was shown. The pilot never asked who answers when the source changes.

3. Pilot users forgive faults that production users report or walk away from.

Early complaints are often read as model failure, when much of it is the end of goodwill. The remedy is training, a feedback route and an interface that works for someone with no stake in its success.

4. The volunteers and the pager are decisions, not projects.

Production needs reviewers whose time is funded and a named role, with a rota and the authority to switch the system off. Either the role is funded or the volunteers quietly stop, and the quality of the system goes with them.

5. The approvals the pilot skipped move the go-live date without any change to the build.

Security, privacy, legal and supplier checks are queues held by people outside the project, and they can change the design, for instance by asking for logging the pilot never kept.

Why does a pilot that worked cost more to run than to build?

Because the pilot never paid for the conditions that made it work. A pilot is a controlled experiment: a clean sample, willing people, a narrow task and a safe distance from everything that could go wrong. Someone supplied each of those at no charge to the project. The sponsor sees a modest bill and a good result, and sets the production budget by extension.

What the pilot spent is a fair guide to what the model costs. It is no guide to what the work around the model costs. The same things recur: a patient audience, a hand-picked extract, a volunteer's time, no audit and no one on call. Exhibit 1 lists them. Each is free only because its cost fell on someone outside the project, and in production that someone sends it back.

EXHIBIT 1

What a working pilot received without paying



These are not costs the pilot saved. They are costs it moved onto other people.

What does the pilot assume about its data and its users?

Start with the extract, the first of the loans a pilot takes. For a pilot, someone pulls a period of records, removes the duplicates, repairs the dates and drops the entries that make no sense. The work is real and it is done once, by hand, by a person whose hours are not charged to the project. In production the data arrives continuously, with the late entries, the changed codes and the fields typed by tired people at the end of a shift. The model now sees a population different from the one it was shown. The extract becomes a pipeline that someone must build, watch and repair, and a question nobody asked in the pilot: who answers when the source changes?

A voluntary framework and a regulation both treat this gap as something to be written down. The NIST AI Risk Management Framework 1.0 asks that limitations of generalizability beyond the conditions under which the technology was developed are documented (MEASURE 2.5). A pilot's conditions are narrow by design, so that limit is exactly what the pilot cannot show. Where the EU AI Act applies to a high-risk system, Article 26(4) requires a deployer that controls the input data to ensure it is relevant and sufficiently representative for the system's intended purpose.

The patient audience is the second loan. Pilot users volunteered, were told it was an experiment, and sat close to the team. They report faults kindly, work around errors and do not count the minutes spent correcting the output. In production, users are told to use the tool, carry targets of their own and compare it with the way they worked before. Each fault is either reported, which loads a support route nobody built, or absorbed in silence, which ends the use. The complaints of the first weeks are often read as model failure. Much of it is the end of goodwill, and the remedy is training, a feedback route and an interface that works for someone with no stake in its success.

Whose time was the pilot running on?

The third loan is people. A pilot borrows the people who understand the problem: the senior case handler, the analyst who knows what a bad answer looks like, the clinician who can tell at a glance. They give hours on top of a full job, and their department absorbs the cost without a line in any budget. In production the need does not go away. Someone must review samples, decide the exceptions and tell the team when the world has changed. Either the role is funded or the volunteers quietly stop, and the quality of the system goes with them. Where the AI Act applies, Article 26(2) requires human oversight to be assigned to natural persons with the necessary competence, training and authority, as well as the necessary support. The support is the line the volunteer never had.

The fourth is the pager nobody carried. While nothing depends on the output, nobody needs to answer when it fails at night or at the close of a month. In production someone does, and that means a named role, a rota, a run book and the authority to switch the system off. Article 26(5) expects deployers to monitor the operation of a high-risk system and to suspend its use when it may present a risk. Monitoring and suspension are only words until a person's working week contains them.

What did the pilot skip that a live system has to pass?

A pilot is allowed to say it is only a test, and that sentence waives a great deal: the security review, the privacy assessment, legal sign-off, the checks on a supplier, accessibility, records retention. None of these is a cost in money first. Each is a queue, held by people who were not in the project, who keep their own calendar and who may say no.

What they ask for can change the design: logging the pilot did not keep, access controls it did not need, a description of the model and its limits, a way to explain an outcome to the person it affects. Where the AI Act applies, Article 26(6) expects deployers to keep the logs the system generates, to the extent they control them. Retrofitting that is slower than designing it in. The production plan usually has no time for any of this, because it was drawn from a pilot that had none, and the result is a go-live date that moves with no change to the build.

Who pays for each one in production?

The bills do not land on the project, because the project is over. They land on the budgets of the functions that did not run the pilot. Exhibit 2 sets each free thing against what replaces it, the budget that carries it and the first sign that it was missed.

EXHIBIT 2

What replaces each free thing, and who pays

Free in the pilot	What production needs instead	Who pays	First sign it was missed
Patient audience	Training, a feedback route and support	The business function	Complaints read as model failure
Hand-picked extract	A live data pipeline with monitoring and repair	Technology	Output shifts when a source field changes
Volunteer's time	A funded review role, or a share of other roles	The business function	Reviews stop being done
No audit	Approvals, logs and a description of the model	Risk and security	The go-live date moves and the build does not
No one on call	A rota, a run book and authority to pause	Technology and the business	A user finds the outage first

Each free thing becomes a recurring cost with a payer, and the payer is seldom the sponsor.

What does the table tell you?

Look first for a line with no payer. It is an unfunded cost, and it will either stall the launch or be skipped and become a risk. Then look at who the payers are, which is rarely the sponsor. Technology pays for the pipeline and the rota, the business function for the reviewers and the training, risk and security for the approvals. Each of them was asked for a view on the pilot, and none was asked for a budget.

What does the table predict about the first month?

It predicts the order in which things go wrong. The extract is missed first, because the data begins to behave differently on the first day and the model's output moves with it. The audience is missed next, in the form of complaints that sound like criticism of the model. The volunteer and the pager are missed at the first incident, when it is not clear whose job it is to answer. The audit is missed at the go-live review, when the approvers ask for what the pilot never produced. A team that has read the table can say, before launch, which of these it will meet first and who will be in the room.

A line with no payer is an unfunded cost: it will either stall the launch or be skipped and become a risk.

How do you size the gap before the pilot ends?

Draw the table for your own pilot before the closing review. For each free thing, write what it was in this pilot, what will replace it, and the name of the budget holder who has agreed to pay. A line without a signature is the finding. Then ask the sponsor to review the table and not the demonstration, because the demonstration answers a question the pilot has already settled.

The table also orders the work. The replacement for the extract has the longest lead time, because it depends on other teams' systems. The approvals are the least predictable, because they depend on other teams' calendars. The volunteer and the pager are decisions and not projects, so they can be taken on the day someone is willing to take them.

Sometimes the sum shows that production would cost more than the benefit it brings. That is a legitimate result, and it is better reached in an afternoon, with a table, than after a launch, with a bill. A pilot that ends in a reasoned decision not to proceed has done what a pilot is for.

Sources

The sources below were opened on 20 October 2026. Article 26 binds deployers of high-risk systems only, and this piece does not state when it applies. The free things themselves are patterns seen in delivery work, not accounts of a single engagement.

1. [AI 100-1 from NIST, release 1.0 of the AI Risk Management Framework, January 2023, read for MEASURE 2.5](#)
2. [Regulation \(EU\) 2024/1689, the AI Act, Article 26 \(obligations of deployers of high-risk AI systems\), paragraphs 2, 4, 5 and 6, read in the text reproduced by the AI Act Explorer](#)

USE THIS

[Use the AI strategy decision brief](#)

[Read why pilots stall before production](#)

Write down the subsidies your current pilot receives, and read the list aloud to the people who will be asked to pay for them.

This document was made from the article at

<https://xonik.com/insights/perspectives/what-a-pilot-gets-for-free-that-production-has-to-pay-for/>. The online article is the current version.