

Why AI pilots stall after the demo

Most financial-services AI pilots impress in the demo and stall in production. The usual cause is not the model. It is the missing institutional context that experts add by hand.

01

The familiar pattern

A team picks a promising use case, connects a model to the relevant documents and produces an impressive demo. Leadership approves a pilot. Then progress slows. Experts find that outputs need too much correction, risk asks how decisions can be explained, and users quietly return to the old way of working.

02

What the experts were correcting

When pilot teams look at what reviewers changed, the corrections cluster in a few categories.

Definitions: the model used a general meaning instead of the firm's.

Scope: the right rule applied to the wrong business line, entity or jurisdiction.

Exceptions: a known exception or precedent was ignored.

Authority: the output recommended something only a specific role can decide.

Evidence: the answer could not show which source and rule it relied on.

03

Trust is designed, not assumed

People trust a colleague whose reasoning they can follow and whose mistakes they can correct once. AI earns trust the same way. If a correction made today has to be made again tomorrow, users stop believing the system can learn the institution.

That means the institutional context has to live somewhere outside the prompt: a governed layer where corrections become updates to definitions, rules and precedent that every future output uses, with a visible record of what changed.

04

Designing for production from the first week

Pilots that reach production tend to share three choices made early: they target one consequential decision with clear owners, they capture the experts' corrections as structured knowledge rather than feedback comments, and they show provenance on every output so reviewers can see why an answer was produced. The model matters less than the knowledge it is given and the way that knowledge is kept current.

Bring us one consequential decision.

Penomic makes a financial institution's definitions, policy, precedent and judgment usable by people, models and agents, with provenance on every output.

Request a private briefing at penomic.ai/contact · More research at penomic.ai/insights