



- A field guide for startup CEOs

Your AI works in the demo.

Will it work in production?

Why AI products lose user trust after launch — and **5 design fixes** that win it back. With a 15-point readiness checklist.

By Alex Stepikin · AI Product Designer





EXECUTIVE SUMMARY

The model is rarely the problem. *The experience is.*

1

Demos sell. Production loses trust.

In a demo the input is perfect. Real users bring messy data, edge cases and slow answers — and judge the product by its worst moment.

2

Trust is a design problem.

Users drop AI that sounds sure when it isn't, hides its sources or leaves no way to fix a mistake. All of this is decided in the interface.

3

The fixes take weeks, not quarters.

Five patterns — confidence, human control, guided start, honest waiting and machine-readable data — plus a checklist to score your product today.



WHY NOW

Everyone is building with AI. *Few get it to stick.*

~5%

of enterprise GenAI pilots reached rapid revenue gains; most showed little or no measurable P&L impact.

MIT NANDA, "The GenAI Divide", 2025 (preliminary)

30%

of GenAI projects were forecast to be abandoned after proof of concept by the end of 2025.

Gartner forecast, July 2024

46%

of people are willing to trust AI — while 66% already use it regularly.

KPMG & University of Melbourne, 48,000 people in 47 countries, 2025

The gap between "we have AI" and "people rely on our AI" is where products win or die. That gap is mostly UX.



THE GAP

The demo is a *controlled room*. Production is the *street*.

IN THE DEMO

- Clean, expected input
- Happy path only
- Fast, cached answers
- The founder drives

IN PRODUCTION

- Typos, PDFs, half-thoughts
- Edge cases every hour
- Slow or failed calls
- A tired user at 11 pm

Users don't judge AI by its best answer. They judge it by **the first time it's confidently wrong** — and by what the screen lets them do next.

01

MISTAKE 1

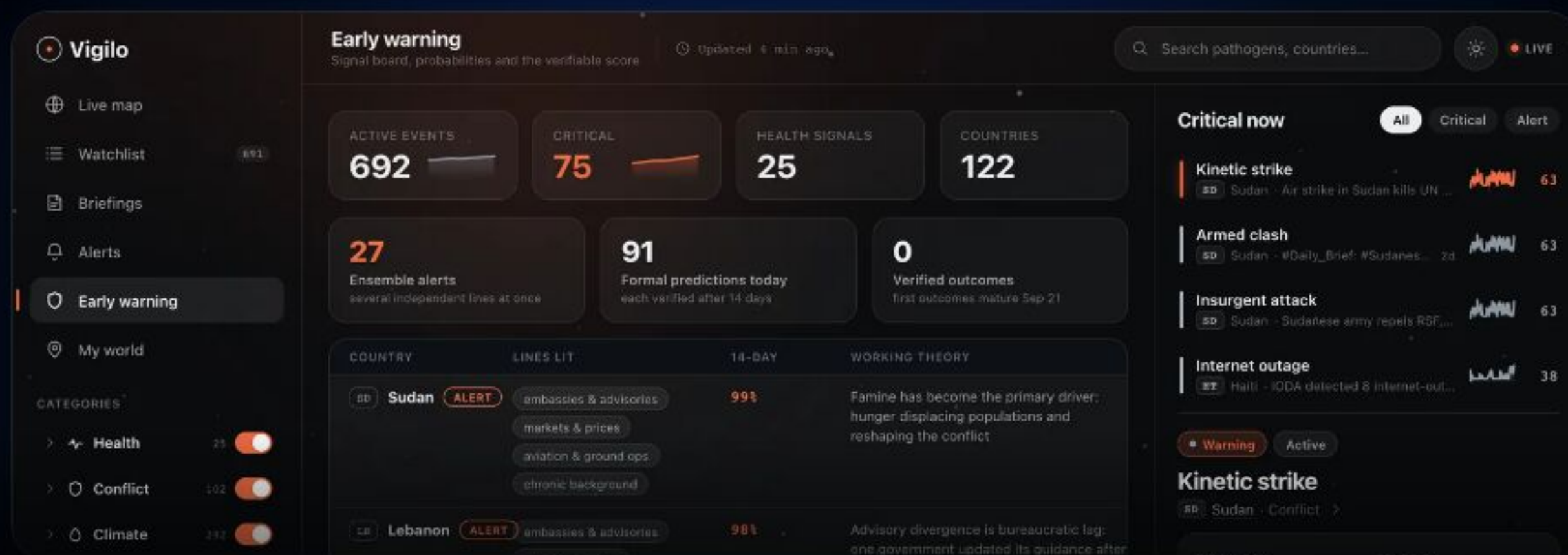
AI sounds sure when it isn't

A wrong answer in a confident tone costs more trust than ten right ones earn. Users can't tell a guess from a fact.

THE FIX

Show confidence and the source next to every answer. Make “not sure” a designed state — and let the user open the evidence in one tap.

EXAMPLE Vigilo: every alert links to the original source, and the misses are published openly, not hidden.





02

MISTAKE 2

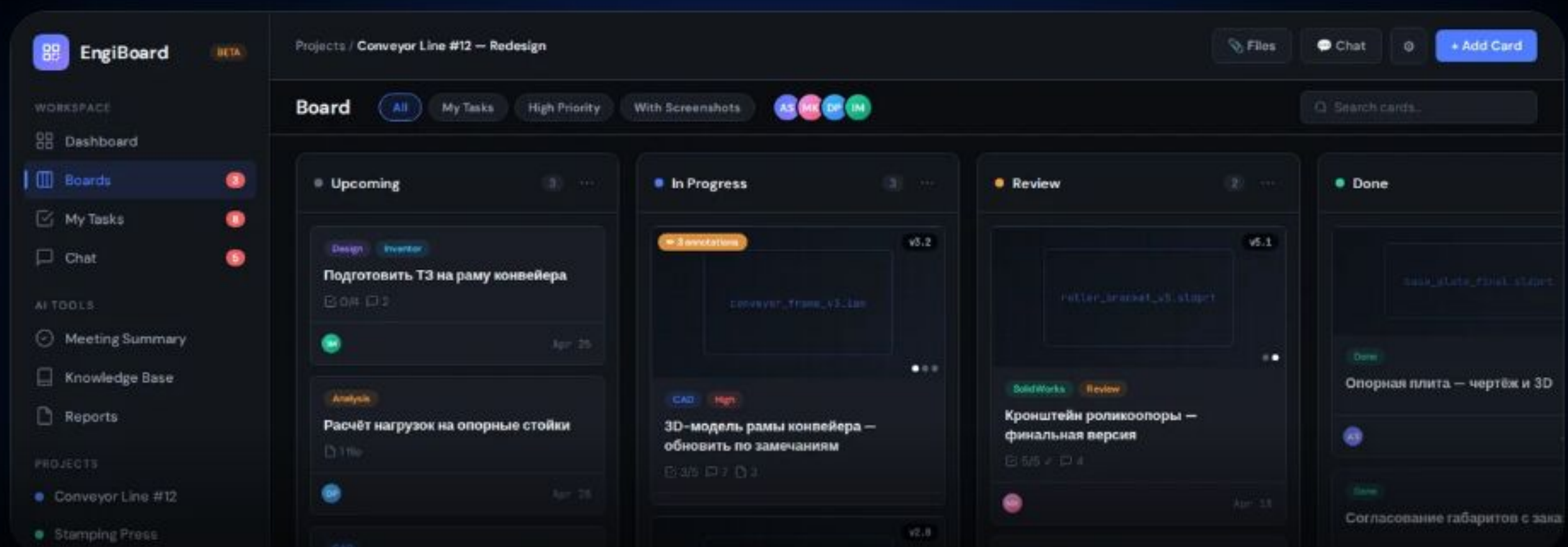
No way for a human to step in

When AI acts and the user can't review, edit or undo, every mistake becomes permanent — so people stop letting it act.

THE FIX

Human-in-the-loop by design: preview before apply, edit the draft, undo any action, explicit confirm for anything risky.

EXAMPLE EngiBoard: undo on every action, so engineers trust the tool enough to move fast. 44 releases in 4 weeks.



03

MISTAKE 3

A blank box and a “magic” button

An empty prompt field asks users to already know what the product can do. Most don't — and leave in the first minute.

THE FIX

Start with examples, templates and suggestions built from the user's own data. The first value should land in the first session.

EXAMPLE VIZI: a guided start around one personal card took the platform from 0 to 1,400 users in 4 months.





04

MISTAKE 4

Waiting with no explanation

A spinner for 20 seconds reads as “broken”. Users refresh, retry and double-charge themselves.

THE FIX

Stream the answer, show the steps the agent is taking, keep partial results, and turn limits and errors into states with a next step.

EXAMPLE Rule of thumb: anything over 2 seconds needs visible progress; anything over 10 needs steps.

- ☒ Reading your files
- ☒ Checking 3 sources
- ☐ Drafting the answer
- ☐ Final check

05

MISTAKE 5

Invisible to other AIs

Buyers now ask ChatGPT, Perplexity and Google's AI first. If your product data isn't machine-readable, you're not in the answer.

THE FIX

Publish structured data (schema.org), an llms.txt and clean public pages, so assistants can find, understand and cite you.

EXAMPLE VIZI: personal cards readable by AI assistants — strictly opt-in, with a counter of agent visits.





HOW TO MOVE FASTER

Design and code, *in one loop.*

Static mockups hide the hard part of AI products: states, latency, failure. I prototype in real code with AI agents, so every flow is tested on live data before the team commits to it.

1

Specs → working build in days

Claude Code writes the code from my specs; I own flows, states and quality.

2

Every change checked on the live service

No “works on my machine” — trust is tested where users are.

3

Proof

EngiBoard: 44 releases in 4 weeks. VIZI: 154-component design system, 85 pages.



THE CHECKLIST

15 questions *before you scale.*

CONFIDENCE & SOURCES

- ☐ Answers show how sure the AI is
- ☐ Every answer links to its source
- ☐ The AI can say "I don't know"

HUMAN CONTROL

- ☐ Users can edit AI output before it's applied
- ☐ Every AI action can be undone
- ☐ Risky actions need an explicit confirm

FIRST SESSION

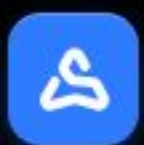
- ☐ The first screen shows examples, not a blank box
- ☐ A new user gets value in under 2 minutes
- ☐ Suggestions use the user's own data

WAITING & FAILURE

- ☐ Anything over 2 s shows progress or steps
- ☐ Errors and limits explain the next step
- ☐ Partial results survive a failure

VISIBILITY & LEARNING

- ☐ Product data is machine-readable
- ☐ You log where users correct the AI
- ☐ Quality is measured with evals, not vibes



YOUR SCORE

Count the “yes” answers.

13–15

Ready to scale

Your AI earns trust. Focus on growth and on measuring quality over time.

8–12

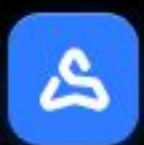
Works in the demo

Users try it, then quietly stop. Fix the gaps in Human control and Waiting first — they’re the fastest wins.

0–7

Trust at risk

The first confident mistake will cost you users. Start with sources, undo and a guided first session.



WHERE THIS COMES FROM

Patterns tested *on real products.*

Diagnocat

HEAD OF DESIGN · AI DIAGNOSTICS

AI for dentists, used in **1,000+ clinics** across North America and Europe. 200+ screens, 3 AI products.

Tango

SENIOR PRODUCT DESIGNER

Live-streaming platform with **500M+ users**.

VIZI

PRODUCT OWNER & DESIGN LEAD

0 → 1,400 users in 4 months, AI agents writing the code.

EngiBoard

DESKTOP APP FOR ENGINEERS

44 releases in 4 weeks, from idea to daily use.

Vigilo

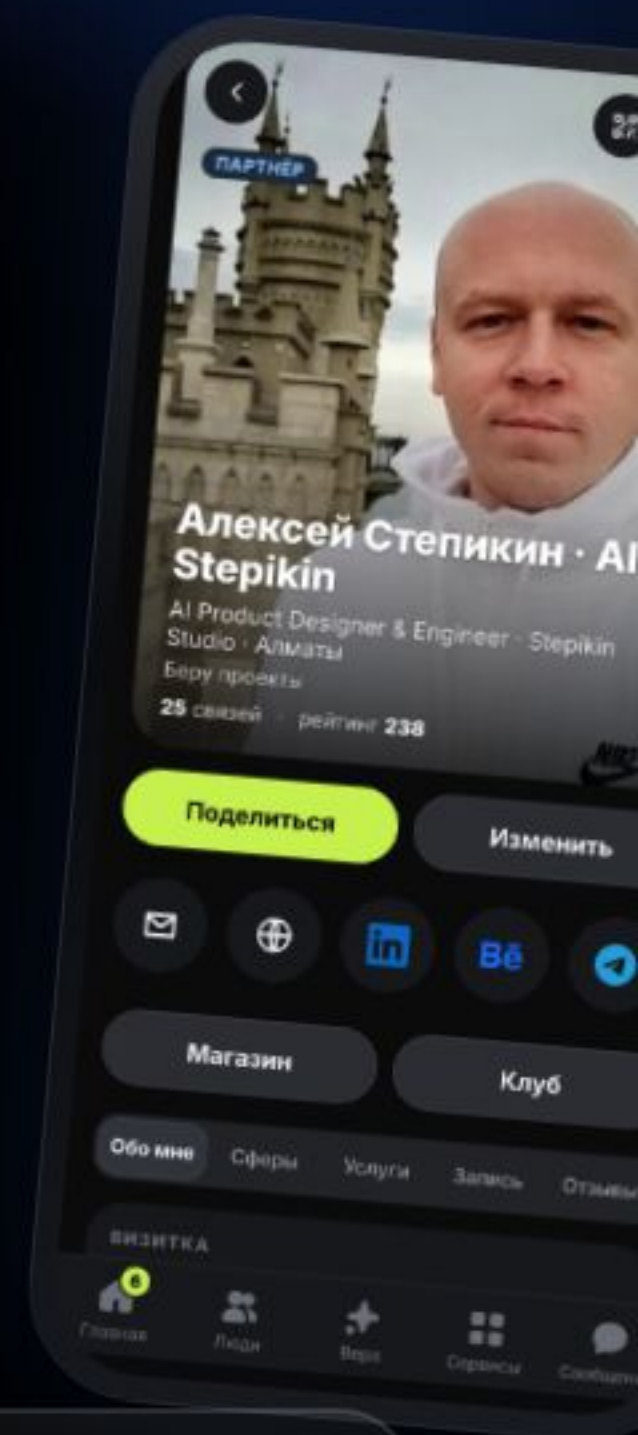
RISK INTELLIGENCE

44 sources, every alert traceable to its origin.

ABOUT THE AUTHOR

Alex Stepikin — *AI Product Designer.*

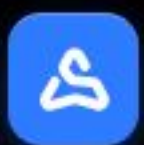
9+ years at international product companies — Tango (500M+ users), Diagnocat (AI diagnostics, Head of Design) and Invitae (US genomics). I design AI products end to end and ship them with AI agents: research, flows, design systems and front-end code, verified on the live service.



1 AI product & UX audits

2 Design of AI features, from first sketch to production

3 Senior / Lead product design roles — remote or relocation



NEXT STEP

Get a free 30-minute *review of your AI product.*

Send me a link or a short demo. I'll score it against this checklist and show the three fixes that will move trust the most — no pitch, no obligation.



stepikin.com/hire

Scan to book

Telegram @uxuialex · aleksey.stepikin@gmail.com

