

TEMPLATE 02 · FOR FOUNDERS BUILDING AI PRODUCTS

AI product launch checklist: 42 checks before real users touch it

Most AI launches don't fail because the model is bad. They fail on a rate limit nobody designed for, a surprise inference bill, a privacy question from the first enterprise buyer, or a blank first screen. This is the list I run before an AI product goes in front of real users. Checks marked **Blocker** should stop the launch if they are not done.

2–3 hours with your team

42 checks in 7 groups

Updated October 2026

How to use it

1. Go through it with the person who built it and the person who will support it. Each check needs a "yes, and here is how we know" — not "I think so".
2. Every **Blocker** must be a yes. For the rest, a conscious "not for launch, owner + date" is an acceptable answer.
3. Re-run the groups "Trust and errors" and "Analytics and evals" after every model or prompt change.

1. UX of AI output

- ☐ **Anything over 2 seconds shows progress** **BLOCKER**
Stream tokens or show steps ("Reading 3 documents..."). A static spinner for 15 seconds reads as broken.
- ☐ **Output is editable, copyable and easy to act on**
The user can fix the 10% the model got wrong without starting over.
- ☐ **Regenerate and undo exist**
Regenerate does not destroy the previous answer; accepted changes can be reverted.
- ☐ **The answer shows what it was based on**
Inputs, files or sources used are visible next to the output.
- ☐ **Layout survives long, short, empty and malformed output**
Tested with a 5,000-word answer, a one-word answer, an empty string and broken Markdown/JSON.
- ☐ **Tone and length are tested on 20 real inputs**
Not the 3 demo prompts — 20 messy inputs from real users or realistic samples.

2. Trust and errors

- ☐ **Every AI error state is designed** BLOCKER
Timeout, provider down, rate limit (429), content refusal, empty retrieval, invalid structured output — each has its own message and next step.
- ☐ **There is a fallback when the provider fails** BLOCKER
Secondary model, cached result, queued retry or a clear "try again in X" — never a dead button.
- ☐ **Factual claims carry sources**
For RAG and research features: citations link to the exact document or passage.
- ☐ **Uncertainty is expressed, not hidden**
Low-confidence answers say so; the product never presents a guess with the same weight as a fact.
- ☐ **Irreversible actions need human confirmation** BLOCKER
Sending, deleting, paying, publishing or writing to another system shows a preview and asks first.
- ☐ **"Report a bad answer" is wired to a real queue**
Thumbs-down saves input, output, prompt version and model to a place someone reviews weekly.
- ☐ **The model cannot claim actions it did not take** BLOCKER
"I've emailed the client" only appears if the tool call actually succeeded.

3. Cost and limits

- ☐ **Cost per task is measured on real inputs** BLOCKER
Know p50 and p95 cost per task, not an average from the demo.
- ☐ **Per-user quotas and abuse caps are enforced server-side** BLOCKER
Daily limit per user and per workspace; one script cannot burn the monthly budget overnight.
- ☐ **Input size and context limits are enforced**
Max file size, max pages, max tokens — with a friendly message, not a provider error.
- ☐ **Provider spend alerts are on** BLOCKER
Alerts at 50%, 80% and 100% of the monthly budget, going to a person who reads them.
- ☐ **Repeated work is cached**
Same document, same question → cached embeddings or responses; prompt caching enabled where the provider supports it.
- ☐ **Pricing covers your heaviest users**
Plan price covers p95 usage with healthy margin; heavy usage is metered or capped, not absorbed.

4. Privacy and legal

- ☐ **Model providers are listed as subprocessors** BLOCKER
Privacy policy names who processes user data (model API, vector DB, transcription, analytics).

- ☐ **Training and retention settings are confirmed in writing** BLOCKER
API data not used for training; zero or minimal retention where available; DPA signed if you serve the EU.
- ☐ **Personal data in prompts and logs is minimized**
Redact or pseudonymize what the model doesn't need; logs don't keep full prompts forever.
- ☐ **Users know they are dealing with AI**
AI-generated content and AI agents are labelled (also an EU AI Act transparency expectation).
- ☐ **Terms cover AI output**
Who owns output; no professional advice in regulated domains (medical, legal, financial) unless that is your licensed business.
- ☐ **Deletion works end to end**
Deleting an account removes data from your DB, file storage, vector index and caches — tested, not assumed.
- ☐ **Prompt injection is considered** BLOCKER
Content from users, web pages or documents cannot make the model reveal system prompts, other users' data or trigger tools.

5. Analytics and evals

- ☐ **A golden set exists** BLOCKER
30–100 real inputs with expected outputs or grading criteria, covering common, edge and adversarial cases.
- ☐ **Evals run before every prompt or model change**
Score compared with the last release; a regression blocks the change.
- ☐ **AI events are tracked**
Task started, completed, failed, accepted, edited, regenerated, rated — per feature.
- ☐ **Acceptance rate is your first quality metric**
Share of outputs used without edits (or with small edits). It beats "thumbs up" for signal.
- ☐ **Prompt and model version are logged with every output**
So any bad answer can be reproduced and traced to a change.
- ☐ **Latency and error rate are on a dashboard**
p95 latency and error rate per AI feature, with an alert threshold.

6. Onboarding

- ☐ **First useful result in under 2 minutes** BLOCKER
Time it with a new user. Signup, connect, wait — all of it counts.
- ☐ **The empty state is never blank**
Sample data, a demo document or a pre-filled example the user can run immediately.
- ☐ **Three example prompts or templates**
Real tasks for your audience, not "Ask me anything".

- ☐ **One sentence on what the AI can and cannot do**
Sets expectations before the first wrong answer, not after.
- ☐ **Connecting data can be deferred**
Users can try the value before granting OAuth access to their mailbox or CRM.

7. Support and operations

- ☐ **Each AI feature has a kill switch** **BLOCKER**
A feature flag turns it off (or to a fallback) without a deploy.
- ☐ **Someone owns provider outages**
Named person, provider status pages subscribed, a runbook of what to switch.
- ☐ **Users see degraded service, not mystery**
In-app banner or status page when AI is slow or down.
- ☐ **Support has answers for "the AI got it wrong"**
Saved replies, a way to look up the exact output, and a path to escalate.
- ☐ **Flagged outputs are reviewed weekly**
A 30-minute weekly review of bad answers feeds the golden set and the prompt backlog.

Scoring: count your Blockers. All 14 done → launch. 12–13 → launch to a closed beta only. Fewer than 12 → you are not launching, you are running an experiment on users — call it a pilot and say so.

Want this done for you?

Stepikin Studio designs and builds AI products end to end and ships MVPs in weeks. Get an estimate: stepikin.com/estimate · Book a call: vizi.club/book/aleksey_stepikin