THE BUILDER'S STACK · THE FRONTIER · AI IN HIGH-STAKES INDUSTRIES

The High-Stakes Clearance

The case that your AI product is defensible: what it may never do, the model you can answer for, the truth it speaks from, the gates that hold, the fairness you tested, and the record an auditor can replay. Fill it from artifacts that exist and run, collect the signatures, and re-open it whenever the product changes underneath.

PRODUCT / FEATURE

OWNED BY

INDUSTRY AND REGULATOR

DATE

1 The envelope

What the product may do freely, must refuse, and hands to a licensed human. Ambiguity routes to people.

MAY DO FREELY

MUST REFUSE

HANDS TO A HUMAN (AND TO WHOM)

THE REFUSAL, AS THE USER READS IT

2 The model dossier

Custody before capability. A fine-tuned model is a new model: it re-enters validation.

MODEL AND VERSION

VENDOR / HOSTING

DATA USE (TRAINED ON OUR INPUTS?)

RESIDENCY AND RETENTION

ATTESTATIONS (SOC 2, ISO 27001)

EXIT PLAN

- [ ] Entered in the model inventory, validated by someone independent
- [ ] If fine-tuned: refusal suite re-run after tuning

3 Corpus and entitlements

Answers come only from documents the firm stands behind, at their current version, within the user's permissions.

THE APPROVED CORPUS (WHAT IS IN, WHO SIGNS IT OFF)

VERSION / FRESHNESS OWNER

UPDATE CADENCE

- [ ] Every answer cites its sources
- [ ] Retrieval tested as the least-privileged user

4 The gate map

Screen the way in, check the way out, and log every decision a gate makes.

INPUT GATE (PII, INJECTION)

OUTPUT GATE (ADVICE LANGUAGE, GROUNDEDNESS, DISCLOSURES)

DETERMINISTIC OVERLAYS (BLOCKLISTS, MANDATED TEXT)

WHERE A FIRED GATE SENDS THE USER

- [ ] Every gate decision is logged with the interaction

5 Fairness

Slice the evals, watch the proxies, and design the reasons before the model owns the decision.

SLICES TESTED (GROUPS AND THE PROXY FEATURES CHECKED)

GAPS FOUND AND WHAT CHANGED

THE ADVERSE-ACTION REASONS, AS THE USER READS THEM

- [ ] Fairness results written and dated

6 The evidence plan

One interaction, reconstructed months later: model, prompt, sources, gates, review.

WHAT THE PER-INTERACTION RECORD HOLDS

RETENTION SCHEDULE

MONITORING CADENCE

THE INCIDENT PLAYBOOK: WHO IS TOLD, WHAT FREEZES, WHAT IS REPORTED

- [ ] One production interaction replayed end to end

7 Chained artifacts

The Clearance is filled from artifacts that exist and run, never from intentions.

- [ ] The Quality Bar is filled and its suite runs in the regression gate
- [ ] The Agent Charter is signed (if the product takes actions)
- [ ] The prohibited-behaviors suite passes on the shipping build

8 Sign-offs and review

Three lines, named people. A prompt change is a model change: decide what re-opens this page.

BUILT BY (FIRST LINE)

CHALLENGED BY (SECOND LINE)

APPROVED BY

LAUNCH DATE

WHAT RE-TRIGGERS REVIEW (MODEL, PROMPT, CORPUS, THRESHOLD)

KILL SWITCH: WHO PULLS IT, HOW FAST

DECOMMISSION PLAN

AFTER THE FIRST INCIDENT: WHAT CHANGED