@demotoprod · REGULATED AI · CHECKLIST

Air-gapped AI deployment checklist

We ran a tutorial agent stack (a multi-agent framework, a graph framework, a chat-model client, an embedding model, a tokenizer) with every outside connection refused, the way a deny-all network behaves. Before any task ran it made 37 connection attempts to three outside hosts: 12 to a telemetry server, 24 to a model hub with 70 seconds of retries, one for an encoding file. With the offline switches set, the same code made one attempt. Three checks before you ship behind a fence.

1 Pull the cable in staging before the fence does

Frameworks fetch what they need at the first call, not at install. A demo on a connected laptop never runs the path that fails on day one inside the fence. In our run the embedding load retried the hub five times with backoff and took 69.5 seconds to give up; the telemetry thread tried 12 times in the background.

- Run the whole demo in staging with all egress denied (a deny-all rule, or a resolver that answers nothing).
- Keep the list of every host it tried and every step that failed. That list is the migration plan.
- Repeat it on every dependency upgrade; a new version can add a new call.

2 Stage every file behind the fence with a manifest and hashes

Model weights, tokenizers, encoding files, Python wheels and container images are all fetched lazily by default. One of the three hosts in our run was a tokenizer fetching its encoding file on first use; it has no offline switch, so the only fix is to carry the file in.

- List every artefact the demo downloaded (the failure list from check one tells you).
- Transfer them with a manifest and a hash per file; verify the hashes on the inside.
- Point the libraries at the local cache (the hub's cache directory, the tokenizer's cache directory, a local wheel index).

3 Switch off every phone-home in config, not only at the firewall

A firewall rule makes a call fail; a config switch makes it never happen. The difference is retries, timeouts and threads that block the task. With the telemetry SDK disabled, the framework's telemetry flag set and the hub in offline mode, the same code made one attempt instead of 37 and failed fast where it had to.

- Disable the telemetry SDK for every signal (OTEL_SDK_DISABLED=true) and the framework's own telemetry flag.
- Put the model hub and the model library in offline mode so a cache miss is an error in milliseconds, not a 70-second retry.
- Turn off tracing, auto-update and licence checks in config; keep the firewall as the second line.

Nitish Gautam · indica-tech.com · hello@indica-tech.com

INDICA TECH · @DEMOTOPROD

Nitish Gautam

Founder-led AI engineering. Demo to production, under NDA, in regulated industries, on fixed scope.

Where this list comes from

Ten-plus years taking products from zero to one across seven sectors, including defence, fintech, medical and telecom, with AI shipped into regulated and air-gapped environments. The @demotoprod channel takes one production AI failure apart each day. These checklists are the same work written down: every item is a dated failure or a documented default, with the control that stops it and the primary source.

If this list found a gap

1. **Your demo works and production is stuck.**
   Production Readiness Audit. Fixed price, two weeks. A five-dimension scorecard, cost and failure analysis, and a 90-day execution roadmap.

2. **You need senior AI judgement without a full-time hire.**
   Fractional AI Lead. Monthly, 30-day notice either side. Weekly 1:1 with the founder or CTO, architecture and code review, technical hiring support, investor-ready due diligence.

3. **An AI feature has to work, by a date.**
   AI Build Sprint. Fixed price, four to eight weeks. Scope signed off up front, weekly demos, CI/CD and observability included, 30-day production warranty.

4. **A regulator, an auditor or a customer's security team is asking.**
   AI Governance & Compliance. Fixed price, four weeks. ISO 27001 / SOC 2-aligned controls, a model risk framework, human-in-the-loop and audit logging, regulator-ready documentation.

BOOK A CALL
indica-tech.com/#contact

EMAIL
hello@indica-tech.com

WEBSITE
indica-tech.com

ALL CHECKLISTS
indica-tech.com/checklists/

YOUTUBE
youtube.com/@demotoprod

LINKEDIN
linkedin.com/in/startup-founderx

INSTAGRAM
instagram.com/demotoprod

TIKTOK
tiktok.com/@demotoprod

Free to share. The live list, with every source link: indica-tech.com/checklists/#air-gapped-ai