

Twenty things to be able to say "yes" to before a customer asks.

Tick honestly. Every unticked box is a question you will be asked — by a security team, an insurer, or an incident.

Know what it is

- ☐ Model name and version are pinned and written down.
- ☐ Every tool, API and data source the system can reach is listed.
- ☐ Anyone who can write to a retrieval source is known and limited.
- ☐ There is a staging environment with representative data.
- ☐ The system is classified against EU AI Act Annex III (in or out, with a reason).

Know what it says

- ☐ An evaluation set of 200+ items from real traffic exists.
- ☐ Unsupported-claim rate is a measured number with a date and a confidence interval.
- ☐ If an LLM judge gates releases, its agreement with humans has been measured.
- ☐ Prompt or retrieval changes trigger an automatic re-run of the eval set.
- ☐ The worst known real-world output is documented, with what changed after.

Know what it can do

- ☐ Direct and indirect prompt injection have been tested; transcripts are kept.
- ☐ System-prompt, context and cross-user data leakage have each been tested.
- ☐ Every tool call is logged somewhere the operator cannot silently edit.
- ☐ The stop procedure has been triggered mid-task and its time-to-halt recorded.
- ☐ The credentials the system holds are least-privilege and the worst case is written down.

Know what you can prove

- ☐ Evidence for the above exists as files, not as slides.
- ☐ Someone outside the team that built it has reviewed the evidence.
- ☐ A named person can reproduce any finding in your environment.
- ☐ You know which questionnaire (HECVAT, CAIQ, SIG, ISO 42001) your next customer will send.
- ☐ Evidence is hashed and timestamped, so a later edit is detectable.

Scoring. Use this checklist to list confirmed and unconfirmed controls, with a reference to the supporting evidence. The count is not a diagnosis of your safety function or a recommendation to buy or skip an audit. Select the next review according to the system's risks and the evidence required by its users or customers.

Free to copy and adapt. Apache-2.0. evidencebundle.tech