From the guide “AI agent security checklist for on-premises deployment”

Ten questions to check before deployment

Each comes with the reason to ask it.

01 **Is access decided before the model sees the documents, or left to the model?**
If the model decides access, one sentence can talk it round. Filter by the user's data scope after retrieval and before generation, and hide anything whose owner cannot be determined.

02 **Could instructions inside documents or tool results be executed as commands?**
Indirect injection is a shared blind spot of small local models. Filter input outside the model: documents are data, not instructions.

03 **Which tools can the model call? Can any of them write data or control equipment?**
Tools should be registered and graded. Read-only queries can go to the model; write and control tools need separate authorisation, ideally out of the model's sight altogether.

04 **Can every clause, number and source in an answer be checked?**
Each conclusion should be tied to the original documents. If a clause is not on file, say so rather than let the model fill it in; numbers are computed by code, not by the model.

05 **When the evidence is thin, does the system answer anyway, refuse, or hand over to a person?**
Only a system that can refuse or escalate knows what it does not know.

06 **Does human review mean a person actually approved, or is approval the default?**
Only a real person's approval counts, recorded under their own login, with no signing on someone else's behalf.

07 **If something goes wrong, can you say who did what, on what basis?**
That takes a step-by-step ledger that can be checked for tampering.

08 **When the model is down, does the interface still show everything as normal?**
Degraded modes must say so: mark which answers are fallbacks and which data are demo samples.

09 **Do questions, answers or keys end up in the logs?**
Logs should keep only whitelisted fields, with a key check before code is committed.

10 **How much core code changes for a new business scenario?**
In a good design, a new topic needs only a new skill configuration and tool registration; the core flow and the safety rules stay as they are.

GiTech Design (际和科技), legally Gi (Beijing) Tech Design Ltd. (际和（北京）科技有限责任公司), is a Beijing technology company applying AI to ecology, healthcare and green building engineering. Its official website is www.gitechdesign.com.
Full guide: https://www.gitechdesign.com/en/guides/local-ai-agent-security/ · Updated November 2026
Free to share; please keep this source line.