

LLM Deployment Security and Compliance Checklist

A checklist for putting an LLM application into production, built on the OWASP Top 10 for LLM Applications 2025, the NIST AI RMF and ISO/IEC 42001, with EU AI Act transparency duties.

Updated October 29, 2026

Use this checklist before an LLM application goes live. It joins three sources. The OWASP list covers security. The NIST AI RMF covers risk management. ISO/IEC 42001 covers the management system around it. The EU AI Act adds legal duties.

This guide is general information. It is not legal advice.

1. Govern: owners and records

- ☐ The application is in your AI inventory, with an owner and a purpose.
- ☐ The risk tier is recorded. Check if it is high-risk under the EU AI Act ([high-level summary](#)).
- ☐ A named person approves release and can stop the system.
- ☐ Staff who run the system have AI literacy training. Article 4 requires providers and deployers to take measures to support AI literacy ([Article 4](#)).
- ☐ The application sits inside your AI management system, if you run one under ISO/IEC 42001 ([ISO](#)).

NIST makes "govern" a cross-cutting function that informs the other three: map, measure and manage ([NIST AIRC](#)).

2. Map: context and data

- ☐ You know who uses the system and who it affects.
- ☐ You know which data reaches the model: prompts, retrieved documents, tool outputs.
- ☐ Personal data use has a legal basis and, where needed, a data protection impact assessment.
- ☐ You list every model, plugin, dataset and library you depend on.

3. Measure: test against the OWASP Top 10

Test each risk in the [OWASP Top 10 for LLM Applications 2025](#):

ID	Risk	What to check
LLM01	Prompt Injection	Direct and indirect injection through documents, web pages and tool output
LLM02	Sensitive Information Disclosure	The model does not reveal personal data, secrets or other users' data
LLM03	Supply Chain	Models, datasets and packages come from trusted sources and are pinned
LLM04	Data and Model Poisoning	Training, fine-tuning and retrieval data are controlled

LLM05 | Improper Output Handling | Output is validated before it reaches a browser, shell, database or API | | LLM06 | Excessive Agency | Tools and permissions are the minimum the task needs | | LLM07 | System Prompt Leakage | No secrets or access rules live in the system prompt | | LLM08 | Vector and Embedding Weaknesses | Retrieval respects user access rights | | LLM09 | Misinformation | Answers are grounded, and users can check sources | | LLM10 | Unbounded Consumption | Rate limits, quotas and cost alerts are set |

Also measure the NIST trustworthiness characteristics that matter for your use: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed ([NIST AIRC](#)).

- ☐ Red-team tests ran before release. Results are stored.
- ☐ Bias tests ran where the output affects people.
- ☐ Tests run again on every model or prompt change.

4. Manage: controls in production

- ☐ Input and output guardrails are on. You know their false-positive rate.
- ☐ Logs record prompts, responses, tool calls and guardrail actions, within privacy rules.
- ☐ Monitoring alerts on quality drops, abuse, cost spikes and new failure types.
- ☐ A human can review or override high-impact outputs.
- ☐ An incident process exists, with owners and timelines.
- ☐ You can roll back to the last good model and prompt.

5. EU AI Act transparency

[Article 50](#) has applied since 2 August 2026:

- ☐ Users are told they are interacting with an AI system, unless this is obvious.
- ☐ If you provide a system that generates synthetic audio, image, video or text, outputs are marked in a machine-readable format. Systems placed on the market before 2 August 2026 have until 2 December 2026 ([implementation timeline](#)).
- ☐ Deepfake content you publish is disclosed as artificially generated or manipulated.

If the system is high-risk, deployers must also keep logs under their control for at least six months ([Article 26](#)). For Annex III systems, those duties apply from 2 December 2027.

Tools for each step

You can run many checks with open-source tools. See [build vs buy](#). For the layer each tool covers, see [how to choose an AI governance platform](#).

Sources

- [OWASP Top 10 for LLM Applications 2025](#), OWASP GenAI Security Project, accessed 2026-10-29
- [AI Risk Management Framework](#), NIST, accessed 2026-10-29
- [AI RMF Core](#) and [AI RMF trustworthiness characteristics](#), NIST Trustworthy and Responsible AI Resource Center, accessed 2026-10-29

- [ISO/IEC 42001:2023](#), International Organization for Standardization, accessed 2026-10-29
- [Article 4](#), [Article 26](#) and [Article 50](#), [artificialintelligenceact.eu](#), accessed 2026-10-29
- [Implementation Timeline](#) and [High-level summary](#), [artificialintelligenceact.eu](#), accessed 2026-10-29

Frequently asked questions

Is the OWASP Top 10 for LLM Applications a standard?

No. It is a list of the main risks for LLM and generative AI applications, from the OWASP GenAI Security Project. It is a good base for a security test plan. The current list is the 2025 version.

Is the NIST AI RMF mandatory?

No. NIST says it is intended for voluntary use. NIST also published a Generative AI Profile, NIST AI 600-1, on 26 July 2024.

Does the EU AI Act apply to my chatbot?

If people in the EU use it, Article 50 can apply. Providers must design the system so people know they are talking to an AI, unless this is obvious. Article 50 has applied since 2 August 2026.