

The GDPR checklist for AI projects.

Twenty-five questions, five areas. Answer each one out loud, from memory, without opening a document. That constraint is the point: if you have to go and look something up, the answer is effectively "no", because the enterprise buyer running a security review, the customer who read something, and the authority with a question will not wait either.

Updated August 2026. Includes Art. 50 EU AI Act,
which applies from 2 August 2026.

01 Legal basis Art. 6 GDPR

Every processing operation runs on exactly one of six legal bases. Not "we have a privacy policy." Not "the user agreed to the terms." One named basis, per purpose.

☐ Can you name the legal basis for this processing in one sentence, whether contract, legitimate interest, consent or another, without looking it up?

☐ Does each **separate purpose** have its own basis?

Running the feature, improving the model and analysing usage are three purposes, not one.

☐ If you rely on legitimate interest (Art. 6(1)(f)): does a written balancing test exist, and do you know what happens when someone objects under Art. 21?

Not a paragraph in a policy. An actual document.

☐ Was the data collected for the purpose it is now being used for?

Training a model on data gathered for something else is the most common failure here (Art. 5(1)(b)).

☐ Does any of it touch special categories: health, biometrics, ethnicity, union membership, political or religious views, sex life?

Those need Art. 9 as well, and the bar is much higher.

If a "no" shows up here: this is the expensive one. A missing legal basis is not a formality you bolt on later. It can make the processing unlawful, which means deletion, not correction. Everything built on that data is affected.

The ten-minute version: open a table. One row per purpose, three columns: purpose, legal basis, why. If a row is hard to fill in, you found your problem.

The one nearly everyone skips, because it does not feel like data protection. Anything written to or read from a visitor's device, whether cookies, localStorage, fingerprinting or a pixel, needs consent before it happens, unless it is strictly necessary for the service the user explicitly asked for. It applies whether or not the data is personal, and "strictly necessary" is far narrower than most banner setups assume.

☐ Do you have a list of everything your site or product writes to a device?

Open DevTools, then Application, then Storage. The real list is usually longer than the documented one.

☐ Is analytics or marketing tracking held back until consent is given?

Neither is strictly necessary, whatever the vendor's setup guide implies.

☐ If there is a banner: is rejecting as easy as accepting, on the first layer, at the same visual weight?

☐ Does rejecting actually stop the scripts, rather than setting a flag while the tags keep loading?

☐ Your AI feature, meaning the chat widget, the assistant, session memory: can you say exactly what it stores on the device, and why that is strictly necessary?

If a "no" shows up here: this is the most-enforced item on the list and the easiest for anyone to verify from outside. No access needed, no insider knowledge, just a browser and thirty seconds.

The ten-minute version: open your own site in a fresh incognito window, reject everything, then look at storage. Anything still there that is not strictly necessary is the finding.

03 Processors AVV / DPA

Every external tool that touches personal data on your behalf needs a contract behind it (Art. 28 GDPR, the AVV in German practice). Model APIs, hosting, analytics, mail, CRM, transcription, vector databases, the monitoring tool nobody remembers adding. Not everyone on that list is a processor: some act as separate controllers, and sending them an AVV is wasted motion. Sort that first, because the fix is a different one.

☐ Can you list every third party that sees personal data in this system, including the ones a developer added in a sprint and never mentioned?

☐ Does a signed data processing agreement exist for each processor?
Most vendors have a self-service DPA. Nobody sends it to you unprompted.

☐ For each one: can you say which country the data physically sits in, and which country the **parent company** is in?
Different questions, different consequences.

☐ For transfers outside the EU: can you name the mechanism?
An adequacy decision, standard contractual clauses with a documented transfer impact assessment, binding corporate rules, or one of the narrow Art. 49 derogations. If the answer is the EU-US Data Privacy Framework: is the vendor actually on the certified list, and does the certification cover this data? The decision stands, it is under appeal at the CJEU, and knowing what you would switch to is part of the answer.

☐ Does your model provider train on your inputs by default, and did someone actually check the setting rather than assume the enterprise tier turns it off?

If a "no" shows up here: you are liable for a processor you never vetted, and you cannot answer a deletion request you cannot trace.

The ten-minute version: list every login your team has. That list is close to your processor list, and it is usually the first time anyone sees it written down.

04 Transparency Privacy notice

It has to describe what runs today. Not what ran at launch, not what the template said. If a person reads it and is surprised by what actually happens, it has failed, however thorough it looks. The quickest test is the date: if the last feature you shipped did not change this document, one of the two is wrong.

- ☐ Does it name the actual tools, or does it say "we may use third-party services"?

The vague version is not transparency, it is the absence of it.

- ☐ Does it state how long data is kept, per category, including prompts and logs, in a way you could actually enforce?

- ☐ If AI is involved: does it say so plainly?

What goes in, what the model does with it, whether a human reviews the output.

- ☐ Is there any fully automated decision with a real effect on someone, such as pricing, scoring, rejection or prioritisation?

That triggers Art. 22 and its own duties.

- ☐ Could you answer a deletion request within a month, end to end, including whatever sits in backups and logs?

If a "no" shows up here: this is the cheapest one to fix and the one people judge you on first, because it is the only part of your compliance anyone can read.

The ten-minute version: read your own privacy notice as if you were a customer who dislikes you. Every sentence you would challenge is a to-do.

Not a formality reserved for large companies. It is mandatory wherever processing is likely to result in a high risk, and AI projects hit the triggers more often than not: profiling with real effects, processing at scale, systematic monitoring, new technology. The German authorities publish lists of processing that always requires one, so for a good number of builds the question has already been answered for you.

☐ Did anyone check whether this project needs one, and is that check written down?

A documented "no" is a valid outcome. An undocumented one is not.

☐ Does your processing appear on your supervisory authority's list of operations that always require an assessment?

In Germany that is per Land, and the lists overlap without being identical.

☐ If an assessment exists: does it describe the system as built, including the model, the prompts and what gets logged, or the version that was pitched?

☐ Which risks did you identify to the people in the data, rather than to the company?

Those are two different lists, and only one of them belongs here.

☐ Were the mitigations actually implemented, and does anyone re-check them after a model or vendor change?

If a "no" shows up here: a missing assessment is its own finding, separate from whether the processing was lawful otherwise, and it gets asked for early once an AI system is the subject.

The ten-minute version: write the four triggers on one page and mark which ones your project hits. Two or more, and you are doing the assessment.

+ Also worth checking

EU AI Act, Art. 50. A separate law from the GDPR, with separate obligations, and since 2 August 2026 the transparency duties apply. They bind providers and deployers, so you are in scope for using a system, not only for building one. If your project puts a user in front of a model, three things are live now:

- Does the person know they are dealing with an AI, at the point of interaction, unless that is already obvious?
- If the system generates or materially edits content: is the output machine-readable as AI output? Systems already on the market have until 2 December 2026 for this one duty, and only this one.
- Does anything you run touch emotion recognition, biometric categorisation, or synthetic images, audio or video of real people? Each carries its own disclosure duty, much of it on the deployer.

Fines run to 15 million euros or 3 percent of worldwide turnover. The Commission adopted its guidelines on 20 July 2026, so "the rules are not clear yet" is no longer available.

Records of processing, Art. 30. If you have to produce a legal basis on demand, you need somewhere it is written down. The table from point 01 is most of that document already.

Where you stand

22 to 25 clean. Better shape than most. Re-run this after every significant feature.

15 to 21. Normal, and manageable. Fix point 02 first: visible from the outside, cheap to close.

Under 15. Not a paperwork problem. Something in the build needs to change, and it is cheaper to find that now than after launch.

This checklist is general information, not legal advice, and does not create a client relationship. For a binding assessment of your specific case, consult a qualified data protection lawyer.

Lisa Bodner · Bodner Consulting · bodner-consulting.com · support.bodner@gmail.com