How to Check Privacy and Reliability Before Using Generative AI Tools

Why a Simple Check Matters

Generative AI tools can produce text, images, and even code in seconds. That speed makes them tempting for professionals who need to draft emails, summarise reports, or brainstorm ideas. But before you paste a prompt into any public chatbot, it pays to stop and ask: where does my data go? Who can see the output? How reliable is the model behind the interface? A few minutes spent on these questions can prevent a privacy breach or a costly mistake.

In a workplace context, the stakes are higher. You might be handling client names, financial figures, or internal strategy notes. A tool that stores your prompts and uses them for training could expose sensitive information. Even if the tool is private, the model itself might produce plausible-sounding but wrong answers. This article walks through the steps you can take to assess privacy and reliability before you rely on any generative AI system.



Step 1: Check What Data the Tool Collects and Stores

Start with the provider's privacy policy and terms of service. Look for explicit statements about data retention. Does the company keep your prompts and outputs on its servers? If so, for how long? Some services use your inputs to improve their models, which means your text becomes part of a training set that other users might indirectly influence. Others promise zero-retention policies, where your data is deleted immediately after the response is generated.

Watch for phrases like "anonymous usage data" or "aggregated trends." These often mean the company collects your prompts but strips personal identifiers before analysis. Still, if your prompt contains enough detail to identify a person or project, anonymisation may not be enough. When in doubt, treat the tool as a public forum and never enter information you would not put on a billboard.

What to Look For in a Privacy Policy

- Retention period: stated in days or months, or described as "not stored"
- Third-party sharing: whether the company shares data with partners or advertisers
- Compliance with regulations: mention of GDPR, Singapore's PDPA, or similar frameworks

If the policy is vague or absent, consider that a red flag. A responsible provider will be transparent about how it handles data. For a quick test, ask the tool itself a neutral question like "Do you store my conversations?" and see what the model says. But remember that the model's answer is generated by language patterns, not by a reliable internal database, so you still need to verify against the official policy.



Step 2: Understand Access Controls and User Authentication

Generative AI tools come in different tiers. A free public chatbot may require no login at all, which means anyone can use it, but also that your session may be linked only to a cookie or IP address. That makes it hard to delete your data later. Enterprise versions often provide single sign-on (SSO), role-based permissions, and audit logs. If you are evaluating a tool for a team, check whether it supports separate accounts for each user and whether an admin can review usage history.

Access control matters most when multiple people will use the same tool. Without it, one person's prompt could accidentally reveal information to another user in a shared workspace. Some platforms let you create isolated projects or workspaces, which helps contain sensitive conversations. For individual use, a tool that requires a personal account and offers a way to delete your chat history is preferable to one that stores everything under a generic session.

Step 3: Evaluate the Model's Reliability and Bias

Privacy is only half the equation. A tool can be secure yet still give you wrong answers. Generative models are trained on vast amounts of text from the internet, and they learn

patterns, not facts. They can produce confident-sounding statements that are entirely fabricated. This is often called hallucination, and it is a known limitation of current technology.

Before you trust a model's output for anything important, run a small test. Give it a question whose answer you already know, and see if it gets it right. Try a question that involves numbers, dates, or specific names, as these are common failure points. For example, ask "What was the revenue of company X in 2023?" and then verify the figure against a reliable source. If the model makes a mistake on a simple fact, assume it will make similar mistakes on more complex topics.

Bias is another concern. Models reflect the biases present in their training data. If you are using generative AI to draft content about people, cultures, or sensitive topics, check whether the output contains stereotypes or unbalanced perspectives. Some providers publish bias evaluation reports or allow you to adjust the model's "temperature" or tone settings to reduce unwanted patterns. Taking a course like the generative AI principles course offered at UCC can help you understand these issues in depth and learn practical mitigation strategies.



Step 4: Set Up Human Oversight for Every Output

No generative AI tool should be the final decision-maker. Always plan for a human review step before you publish, send, or act on any output. This applies to internal memos, customer communications, and especially code or data analysis. A human reviewer can catch factual errors, inappropriate language, and subtle biases that an automated check might miss.

Oversight also means monitoring the tool's behaviour over time. Models are updated, and a tool that worked well last month might behave differently after a new version is deployed. Keep a log of any questionable outputs and report them to the provider if possible. In a team setting, designate one person as the point of contact for AI-related risks. That person can track updates, review policies, and ensure everyone follows the same privacy rules.

Step 5: Test with a Risk Assessment Before Full Deployment

Before rolling out a generative AI tool across your team or organisation, do a structured risk assessment. Start by listing the types of data you plan to input. Sort them into categories: public information, internal non-sensitive data, internal sensitive data, and personal identifiable information (PII). For each category, determine what level of data handling the tool must support. If the tool cannot guarantee zero retention for sensitive data, do not use it for that category.

Also assess the consequences of a wrong output. A tool that suggests a restaurant for a team lunch has low stakes. A tool that drafts a contract clause or calculates a tax estimate has high stakes. For high-stakes use cases, require that the model's output be verified by a qualified professional before it is used. The generative AI principles course at UCC covers how to build such risk frameworks into your workflow, giving you a structured approach that matches the pace of a two-day workshop.

Finally, document your assessment. Write down which tools you evaluated, what you found, and what decisions you made. This record helps you stay consistent and provides a reference if questions arise later. It also demonstrates due diligence if your organisation ever needs to show that it handled data responsibly.



Putting It All Together

Privacy and reliability are not one-time checks. They require ongoing attention as tools evolve and as your own use cases change. Start with the privacy policy, confirm access controls, test the model's accuracy, add human oversight, and run a risk assessment before committing. Each step is simple on its own, but together they form a solid foundation for using generative AI safely in a professional setting.

When you build these habits early, you avoid the scramble that follows a data leak or a public error. You also get more value from the tool, because you know its limits and can work around them. If you are looking for a structured way to learn these principles and apply them in a hands-on environment, consider a classroom course that walks you through building a simple generative AI application over two days. That kind of training gives you direct experience with the decisions that matter most.