

# Evaluating Generative AI Tools

A checklist

Nicole Hennig, [nicolehennig.com](https://nicolehennig.com), Feel free to share, with attribution, [CC 4.0 license](https://creativecommons.org/licenses/by/4.0/). Version 1.2, December 23, 2023.

Here are some questions to ask when you are thinking of purchasing a tool or app based on generative AI models.

**Which foundation model is it based on and what API was used?**

Examples: [GPT models](#) from OpenAI, [Claude](#), [Perplexity](#), [Whisper](#), [DALL·E](#), or openly-licensed models like [Llama2 from Meta](#).

**Is it [grounded](#) with sources of factual data?**

Examples: a database of research papers, your documents and web pages, your knowledgebase, access to external tools. For example, [Elicit](#) is grounded with research papers from Semantic Scholar. One method of grounding is [RAG: Retrieval Augmented Generation](#).

**For the underlying model:**

**How was that model trained and with what data?**

Consider this in order to evaluate the model's likely biases and relevance to your specific use-cases. Read this research paper for [details on ChatGPT's \(GPT-3\) training data](#). (OpenAI hasn't released information on the training of GPT4).

**What benchmarks have been used to evaluate the output of the model?**

[This is an area of research with many challenges](#). Benchmarks may be listed, but they may not always be the best way to evaluate the model.

**What guardrails have been put in place?**

Understand the measures that were taken to ensure safe and ethical use. Here's an example: [DALL·E 3 System Card](#). It describes the guardrails they have put in place.

**What privacy settings are available and what is the privacy policy?**

Remember that by default, [OpenAI doesn't use your input for training when a tool is built with their API](#), only with the consumer version of ChatGPT (and that's a setting you can turn off). See also [Enterprise Privacy at OpenAI](#).

**Is it multimodal?**

Can it handle different types of input/output (e.g., text, image, sound)? How do these features contribute to its use?

**And of course, use questions you would ask about any tool, such as:**

- What problems does it solve and what types of tasks is it best used for?
- Does it meet accessibility standards, and does it have a mobile app or website?
- What support and documentation is available?
- What is the pricing model and is there an educational price?
- What future enhancements and improvements are planned?

## **This is just a starting point**

The checklist above is just a starting point for those considering purchasing generative AI tools. You'll think of other questions as you learn more about these tools.

---

Nicole Hennig, [nicolehennig.com](https://nicolehennig.com)

[CC 4.0 license](#). Feel free to share, with attribution. Version 1.2, December 23, 2023.