

AI as a Service



- [Installation](#)
 - [OpenAI integration](#)
 - [IBM watsonx.ai integration](#)
 - [Improve performance](#)
-

At Nextcloud, we focus on creating on-premise AI apps that run fully self-hosted on your own servers in order to preserve your privacy and data sovereignty. However, you can also offload these resource-heavy tasks to an "AI as a Service" provider offering API access in exchange for payment. Examples of such providers are [OpenAI](#), with its ChatGPT APIs providing language model access among other APIs, as well as [Replicate](#) and [IBM watsonx](#).

Installation

In order to use these providers you will need to install the respective app from the app store:

- `integration_openai`
- `integration_replicate`
- `integration_watsonx`

You can then add your account information, set rate limits, and set the providers live in the "Artificial intelligence" section of the administration settings.

Optionally (but recommended), setup background workers for faster pickup of tasks. See [the relevant section in AI Overview](#) for more information.

OpenAI integration

With this application, you can also connect to a self-hosted LocalAI or Ollama instance or to any service that implements an API similar enough to the OpenAI API, for example [IONOS AI Model Hub](#), [Plusserver](#), [Groqcloud](#), [MistralAI](#) or [Together AI](#).

Do note, however, that we test the Assistant tasks that this app implements only with OpenAI models and only against the OpenAI API, we thus cannot guarantee other models and APIs will work. Some APIs claiming to be compatible with OpenAI might not be fully compatible so we cannot guarantee that they will work with this app.

IBM watsonx.ai integration

With this application, you can also connect to a self-hosted cluster running the IBM watsonx.ai software.

Do note, however, that we test the Assistant tasks that this app implements only with the provided foundation models and only against IBM Cloud servers. We thus cannot guarantee that other models or server instances will work.

Improve performance

Prompts from these apps can have a delay of up to 5 minutes. This can be optimized and more information can be found in [the relevant section in AI Overview](#).

📄 #3

📄 cocobook-sync 📄 2026-10-28 16:29:05 UTC

📄 cocobook-sync 📄 2026-10-29 01:09:23 UTC