lollms-webui and Settings Documentation

ParisNeo

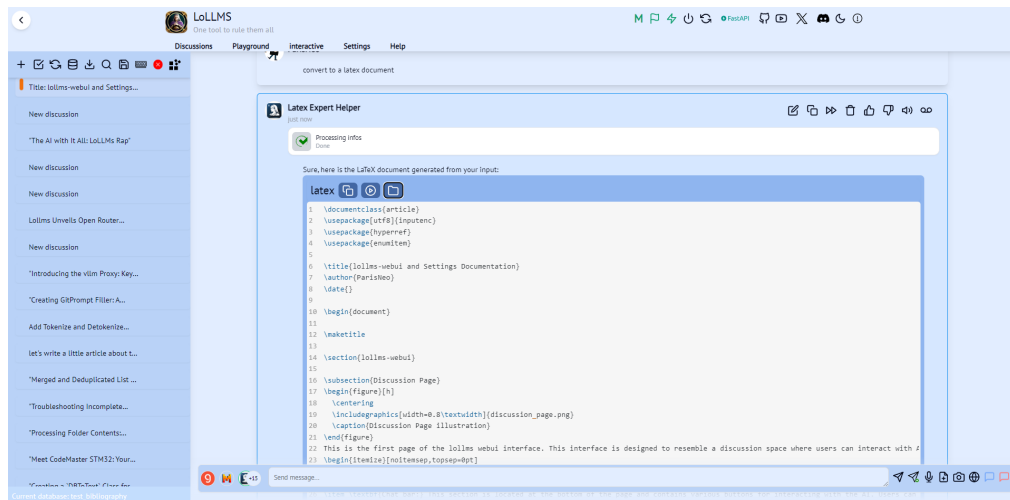# 1 lollms-webui

1.1 Discussion Page

Figure 1: Discussion Page illustration

This is the first page of the lollms webui interface. This interface is designed to resemble a discussion space where users can interact with AI. We can break down this page into three main zones:

- **Left side panel:** This panel contains a list of all the discussions currently stored in the database, along with buttons to manage these discussions. Users can create a new discussion, select multiple discussions to export or delete, reset the database, change the current database, import additional discussions, and search through existing discussions.
- **Message display:** This zone displays all the messages of a selected discussion. Each message contains the name of the user or AI, the date and time of posting, the message text, as well as information about the AI used to respond (if applicable). Users can edit, copy, delete, rate, and listen to any message.

- **Chat bar:** This section is located at the bottom of the page and contains various buttons for interacting with the AI. Users can select the AI model, choose a personality (if available), enter text to send, use additional features such as sending documents, photos, or URLs, and send messages as a user or AI.
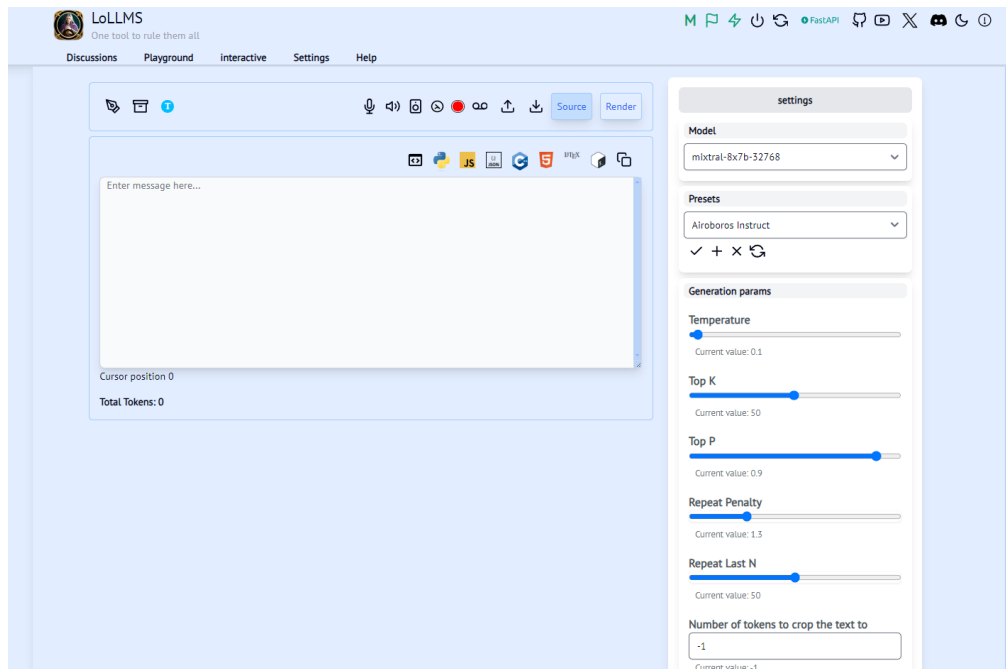
1.2 Playground page

Figure 2: Playground Page illustration

This is the second page of the lollms interface. Here, you can freely experiment with the LLM (Language Models) and improve your text in various ways. You will find a text input field in the center of the page. Simply write your text in this field, then click the pen button to complete your text. You can find the following buttons at the top:

- **Completion button**
- **Completion by keyword (@¡generation$_placeholder$ >)button Show tokens button**
- **Microphone** for dictating your text
- **Speaker** for reading the generated text
- **Upload a voice** for uploading a voice sample to be used for speech synthesis
- **Audio to audio** fo audio to audio tasks like real time translation

- **Start recording** that starts recording and transcribing the speech into text
- **Recorder** that generates an audio file from the entered text

The interface features two tabs and two buttons:

- Two tabs for exporting and importing text
- **Configuration panel** on the right where you can select the model to use and presets (pre-configured texts for specific tasks)
- **Import** and **Export** buttons

You can improve your text using specific keywords:

- **@¡Language:all**$_p$*rogramming*$_l$*anguage*$_o$*ptions* $> @fordisplayingalanguageselectioninterface@$¡put
- **@¡demande¿@** for asking users questions and gathering additional information

Finally, you can adjust the generation parameters such as temperature, top$_p$, *top*$_k$, *etc., andsetaseedforreproducibility.*

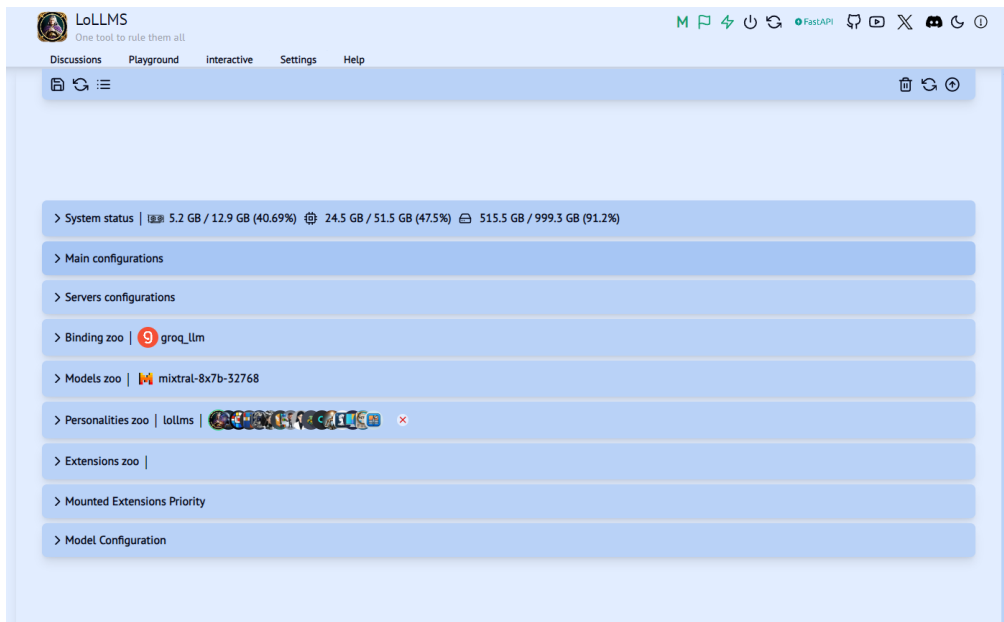# 2 Settings

2.1 Settingspage

Figure 3: SettingsPage illustration

This is the tool setup page, which consists of several sections:

- **System state**
- **Server configuration**

- **Bindings zoo**
- **Models zoo**
- **Personalities zoo**

2.2   System status

First, you will find the system status section, which displays available resources such as memory, disk space, CPU load, and GPU memory.

2.3   Main configurations

This section is divided into several subsections, such as general information where you can specify the type of hardware used (CPU, GPU, etc.), the server address, the current database, automatic tool display options (e.g., whether you want to use it only as a server without a user interface), automatic setting options, automatic update options, and automatic title generation options.

The next subsection concerns user personal information, which can be added for a more personalized experience. You can upload a photo as an avatar, provide information about your preferences that will be used by AI to better meet your expectations, as well as a context management parameter where you can set the minimum number of tokens allocated to the AI to respond.

The following subsection concerns data vectorization, which controls how documents and web pages sent to AI are vectorized, indexed, and stored. You can choose the vectorization method (e.g., tf-idf or embeddings), as well as the number of text pieces to use in the discussion (top k) and their sizes (in tokens).

You will then find a subsection dedicated to LaTeX where you can specify the path to your pdflatex. This allows AI that generates LaTeX to compile their productions into PDFs.

2.4   Boost

The boost subsection allows you to customize the boost text to improve the AI's response. You can, for example, tell the AI that it will receive a reward if it answers correctly or negatively if it returns false information, which can cause damage. You can also force the AI's language response and enable the fun mode to make the AI more fun.

Finally, this last subsection allows you to control the Audio IN and Audio OUT by selecting the AI language and voice.

2.5   Server configuration

This section is dedicated to advanced server configurations, offering various options for installing different types of servers. These servers can be operated by AI, such as Stable Diffusion for generating high-quality images or XTTS for producing high-precision audio.

You can also install text generation servers, such as Ollama or Petals, which allow you to outsource production to other computers. These powerful servers offer fast and efficient text generation for your projects!

2.6   Bindings zoo

This third section is dedicated to bindings, which are modules executing the generation models. Each module can be installed and configured independently by clicking on the **Install** and **Settings** buttons. The configuration settings of each binding may vary, but some common settings include context size or server address. These settings are used when the binding calls a remote service, such as Ollama, OpenAI, Mistral AI, Google, Gemini, or Open Router.

In some cases, an API key is required to access the service. For instance, the Ollama binding requires a key. Depending on the server used, you may be required to provide the key, particularly if an authentication proxy is in place.

2.7   Models zoo

This section allows you to select pre-trained models for text generation. Available models vary based on your needs, ranging from simple to complex. For example, if you need to translate text from one language to another, you can choose a specialized translation model. If you want to generate creative text for a novel or song, you should choose an advanced language model.

2.7.1   Add models for binding

If the pre-trained models do not meet your specific needs, you can add your custom models using this section. Simply provide a link to your local or cloud-based model. This allows you to use models created by other developers or train your own models based on your unique needs.

2.8   Personalities zoo

This section is dedicated to personality selection. Personalities are AI behavior simulation layers that condition AI behavior and direct responses. There are over 500 personalities available in 49 categories (up to version 9.0 of Lollms). To install a personality and use it, click the **Mount** button in the menu located in the personality tab. When you select a category, the personality cards will appear, and you can read the personality description, author, etc. You can then mount a personality to add it to the list of personalities visible in the discussion interface. You can make multiple personalities communicate with each other and collaborate.

There are two types of personalities: unscripted ones, composed only of text describing how they should behave, and scripted ones, containing Python code using logic to perform tasks and provide choices to the AI. For example, Artbot

has a multi-step image generation process where AI makes choices and performs operations.