

NOVA IMS LLMs local infrastructure

2025-05-04

1. To connect, follow the instructions in the document “How to connect” to dah01.novaims.unl.pt”.
2. If you have any issues connecting, please contact the IT Services.
3. After connecting, follow these instructions to use the current two models...

At this moment we have two models available: **IMS-Llama3.3** (the 70-billion-parameter Llama 3.3) for text-to-text, and **Gemma3:12b** (Google Gemma 3, 12 billion parameters) for both text-to-text and image-to-text.

To start using these models, follow these steps:

1. Go to the **onlinev2** site: <https://onlinev2.novaims.unl.pt/>
2. Log in and obtain an **API key** under the “**Data Analytics Hub**” menu.
3. Use the code samples provided.

Example: calling Llama 3.3 with curl

```
curl https://iaapi.novaims.unl.pt/v1/completions \
-H "Content-Type: application/json" \
-H "Authorization: Bearer trincheiras:XXXXXXYourAPIKeyXXXXXXX" \
-d '{"model": "Llama3.3:latest", "prompt": "Say this is a test", "temperature": 0, "max_tokens": 7}'
```

Example: calling Gemma 3 (12 b) with curl

```
curl https://iaapi.novaims.unl.pt/v1/completions \
-H "Content-Type: application/json" \
-H "Authorization: Bearer trincheiras:XXXXXXYourAPIKeyXXXXXXX" \
-d '{"model": "gemma3:12b", "prompt": "Say this is a test", "temperature": 0, "max_tokens": 7}'
```