



Machine Translation Evaluation on Lumigator

Mozilla.ai

Alex Meckes & Hareesh Bahuleyan

28 May 2025



Mozilla.ai is a start-up funded by the Mozilla Foundation

We're on a mission to make AI trustworthy, accessible, and open-source.

Drawing on Mozilla's legacy of advocating for a free and open web, Mozilla.ai build tools for developers to integrate, innovate, and collaborate on responsible AI solutions.

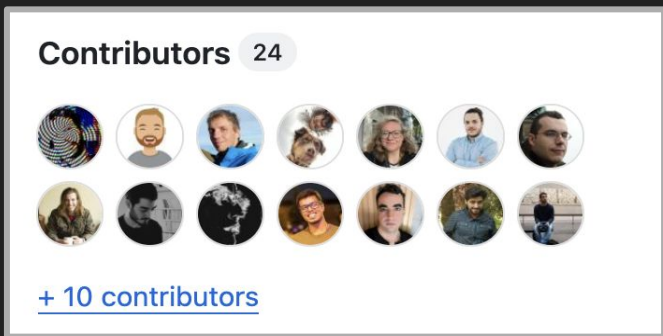
Lumigator is a **platform** that **guides** users through the process of selecting the right language model for their **specific** needs.

Built for users who:

- code
- work with AI and LLMs
- ... but are not necessarily publishing on NeurIPS!
- need to choose LLMs and spend as little time as possible

Lumigator is open-source:

<https://github.com/mozilla-ai/lumigator>



The Challenge: with little time and so many models available, AI developers feel lost on how to evaluate and choose the best model for their task

Lumigator makes model evaluation easier!

Key Features

- Infrastructure agnostic (local/cluster/cloud)
- Relying on existing tools / standards for interoperability
- API + Python SDK + UI
- Extensible: by the community for the community
- Support Eval Use Cases: Summarization and **Translation**

Translation Evaluation on Lumigator


- **Models**


- LLM API-based models
- Self/locally-hosted LLMs via Ollama, Llamafire or vLLM
- Supported models from HuggingFace Model Hub


- **Metrics**


- BLEU
- METEOR
- COMET
- LLM-as-a-judge

Web UI

 Lumigator

 Datasets

 Experiments

 Settings













[GitHub ↗](#)

[Documentation ↗](#)

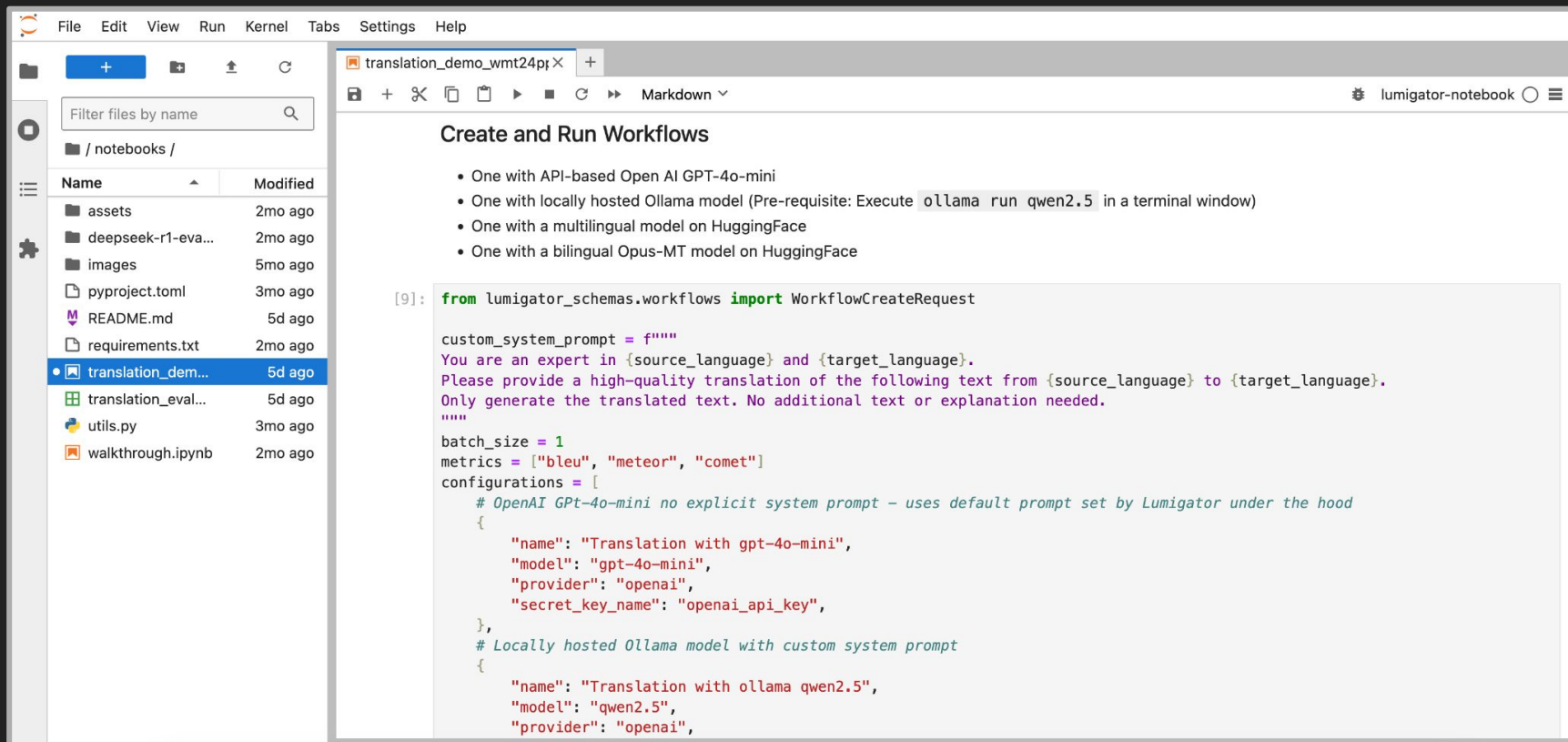
Experiments

Experiments are a logical sequence of inference and evaluation tasks that run sequentially to evaluate an LLM.

+ Create Experiment

NAME ↕	DATASET ↕	USE_CASE ↕	CREATED_AT ↕	STATUS ↕	OPTIONS
WMT24++ Demo Experiment English to German	dataset	translation	20 May 2025, 15:43	SUCCEEDED	  
WMT	wmt14_translate_de-en_test.csv	translation	20 May 2025, 15:37	SUCCEEDED	  
EN2ES	sample_translation_en_es.csv	translation	11 Apr 2025, 14:19	SUCCEEDED	  
COMET	sample_translation_en_de.csv	translation	09 Apr 2025, 09:24	SUCCEEDED	  

Python SDK



The screenshot displays the Lumigator notebook application. On the left is a file explorer sidebar with a search bar and a list of files and folders. The main area on the right is a code editor with a tab titled 'translation_demo_wmt24p'. The editor shows a Python script for creating and running workflows.

File Explorer:

Name	Modified
assets	2mo ago
deepseek-r1-eva...	2mo ago
images	5mo ago
pyproject.toml	3mo ago
README.md	5d ago
requirements.txt	2mo ago
translation_demo...	5d ago
translation_eval...	5d ago
utils.py	3mo ago
walkthrough.ipynb	2mo ago

Code Editor:

Create and Run Workflows

- One with API-based Open AI GPT-4o-mini
- One with locally hosted Ollama model (Pre-requisite: Execute `ollama run qwen2.5` in a terminal window)
- One with a multilingual model on HuggingFace
- One with a bilingual Opus-MT model on HuggingFace

```
[9]: from lumigator_schemas.workflows import WorkflowCreateRequest

custom_system_prompt = f"""
You are an expert in {source_language} and {target_language}.
Please provide a high-quality translation of the following text from {source_language} to {target_language}.
Only generate the translated text. No additional text or explanation needed.
"""

batch_size = 1
metrics = ["bleu", "meteor", "comet"]
configurations = [
    # OpenAI GPT-4o-mini no explicit system prompt - uses default prompt set by Lumigator under the hood
    {
        "name": "Translation with gpt-4o-mini",
        "model": "gpt-4o-mini",
        "provider": "openai",
        "secret_key_name": "openai_api_key",
    },
    # Locally hosted Ollama model with custom system prompt
    {
        "name": "Translation with ollama qwen2.5",
        "model": "qwen2.5",
        "provider": "openai",
```


Demo Time!

Other Projects from Mozilla AI's



Blueprints

<https://blueprints.mozilla.ai/>

Customizable workflows that help developers prototype AI applications using open-source tools and models



<https://github.com/mozilla-ai/any-agent>

any-agent: Python library providing a *single interface* to different agent frameworks (OpenAI, Langchain, Google ADK and more)

Thank you!

Wanna know more?

- [Code](#)
- [Documentation](#)
- [Demo Notebook](#)

