

Open & Multilingual LLMs for Europe

André Martins

EuroHPC Summit 2025, Krakow, March 18, 2025



TÉCNICO
LISBOA



instituto de
telecomunicações





- Headquartered in **Lisbon, Portugal**
- Focus on **AI-powered translation** and **multilingual language technologies**
- Strong ties with the research community.

My academic hat:

- Instituto Superior Técnico, University of Lisbon.



Key R&I achievements / awards in 2024

- 🚀 **20+ papers** in top AI conferences (ACL, EMNLP, COLM, ICML, NeurIPS, ...)
- 🚀 Winner of **Large AI Grand Challenge grant** (AI Boost, 2M GPU hours)
- 🚀 Winner of **WMT shared task on General Translation** (also Biomedical and Chat Translation)
- 🚀 Open AI models with **400,000+ downloads** (Tower, EuroLLM, xCOMET, ...)



Map of collaborations in EU projects / EuroHPC

- R&I labs
- Supercomputers



Large-scale AI models using EuroHPC **already** delivered



2023

2024

2025



Large-scale AI models using EuroHPC **already** delivered



CroissantLLM

1.3B @Jean Zay

100K+ downloads



2023

2024

2025

Large-scale AI models using EuroHPC **already** delivered



Tower

2B, 7B, 13B, 70B @MN5

AI & Data Intensive

200K+ downloads



CroissantLLM

1.3B @Jean Zay

100K+ downloads



2023

2024

2025

Large-scale AI models using EuroHPC **already** delivered



Tower

2B, 7B, 13B, 70B @MN5

AI & Data Intensive

200K+ downloads



CroissantLLM

1.3B @Jean Zay

100K+ downloads

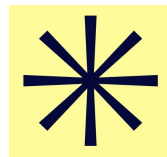


EuroLLM

1.7B & 9B @MN5

Extreme-Scale

300K+ downloads



2023

2024

2025

Large-scale AI models using EuroHPC **already** delivered



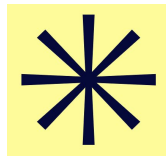
Tower

2B, 7B, 13B, 70B @MN5
AI & Data Intensive
200K+ downloads



EuroLLM

1.7B & 9B @MN5
Extreme-Scale
300K+ downloads



EuroBERT

<1B & 2B @Adastra
Released 1 week ago!

CroissantLLM

1.3B @Jean Zay
100K+ downloads



2023

2024

2025

Large-scale AI models using EuroHPC **already** delivered



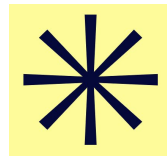
Tower

2B, 7B, 13B, 70B @MN5
AI & Data Intensive
200K+ downloads



EuroLLM

1.7B & 9B @MN5
Extreme-Scale
300K+ downloads



EuroBERT

<1B & 2B @Adastra
Released 1 week ago!

CroissantLLM

1.3B @Jean Zay
100K+ downloads



2023

2024

2025

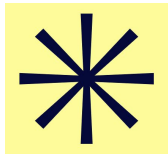
- The success of **Tower** and **EuroLLM** led to a new Unbabel product:



Coming Soon

EuroLLM

22B @MN5
Extreme-Scale



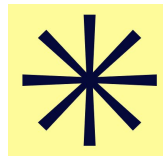
TowerMoE

@Leonardo
AI Boost



**Multimodal
EuroLLM**

@Jean Zay



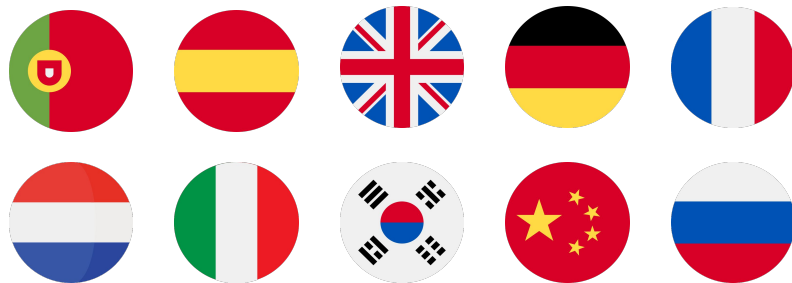


The Tower Project

The Tower Project (started mid 2023)



Goal: Create the best open multilingual LLM optimized for translation-related tasks.



Initial focus on 10 languages, later extended to 15.

The Tower Team



André
Martins



José
Souza



Patrick
Fernandes



Pedro
Martins



Ben Peters



José
Pombal



Amin Farajian



Pierre
Colombo



João Alves



Nuno
Guerreiro



Duarte
Alves



Ricardo
Rei



Sweta
Agrawal

Several streams: data, evaluation, multilingualization, instruction tuning, ...

The Tower recipe

Two steps: **continued pretraining (CPT)** and **supervised finetuning (SFT)**.



Backbone LLM



TowerBase

Base model with
**improved multilingual
performance.**
(20B tokens)



TowerInstruct

Optimized model
(built on top of TowerBase)
for **translation-related
tasks.**

We have released a series of resources related to Tower

1

Models

2

Data

3

Evaluation

4

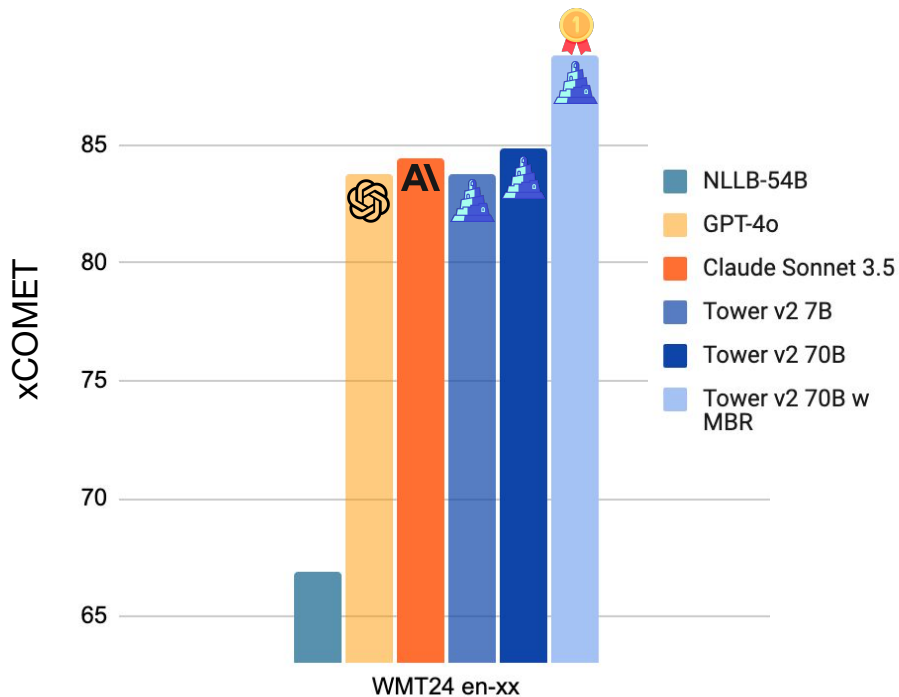
Paper



Check out the
Tower Collection



How does Tower compare to other state-of-the-art LLMs?



Results of the WMT 2024 General MT Shared Task Human evaluation



Tower wins 8 out of 11 language pairs!
(above GPT4, Gemini 1.5 Pro, Google Translate, DeepL)

You can now play with our models!



We are developing even **better versions of Tower** to power widn.ai — your feedback is greatly appreciated!

✦ Add instructions to customize your translation Beta

Personalize your translation with specific instructions... 0 / 150

English ↔

Start typing or paste your text to see Widn in action!

0 / 1000

↶ ↷ Translate

Spanish Glossary

Your translation will appear here.



What next?



Backbone LLM



TowerBase

Base model with
**improved multilingual
performance.**

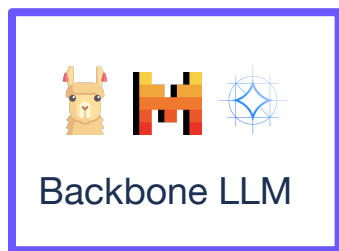
(dozens of billions of
tokens)



TowerInstruct

Optimized model
(built on top of TowerBase)
for **translation-related
tasks.**

What next?



TowerBase

Base model with
**improved multilingual
performance.**
(dozens of billions of
tokens)



TowerInstruct

Optimized model
(built on top of TowerBase)
for **translation-related
tasks.**

Can we get rid of this? Can we train a model from scratch?

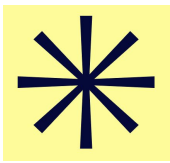
EURO*LLM

The EuroLLM Project



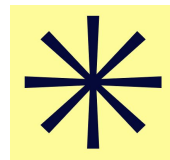
euollm.io

An European collaborative effort



<https://euollm.io/>

EuroLLM



Goal: Develop the best open European LLMs, **trained from scratch**.

3 Model sizes: **1.7B** (*completed*), **9B** (*completed*), **22B** (*next*).

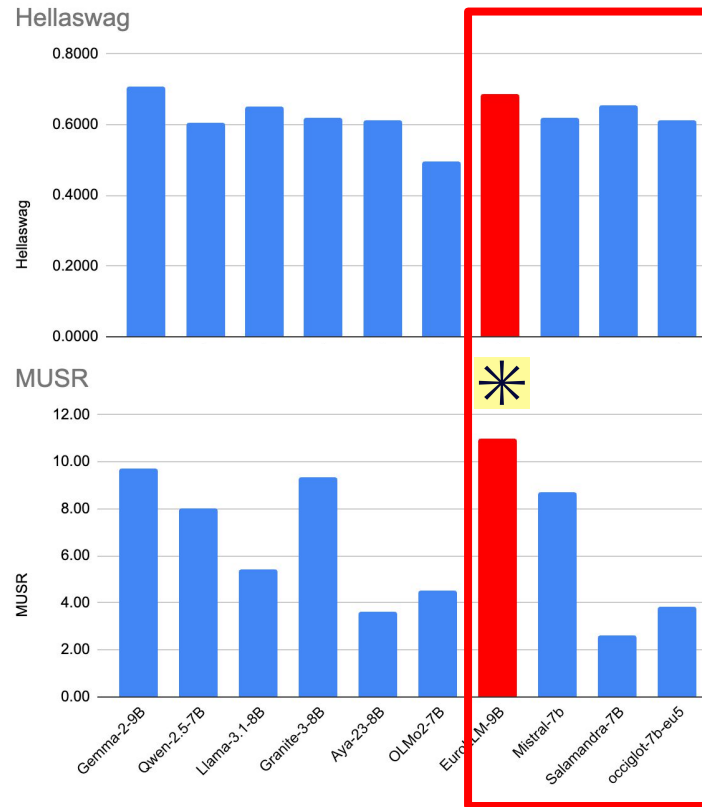
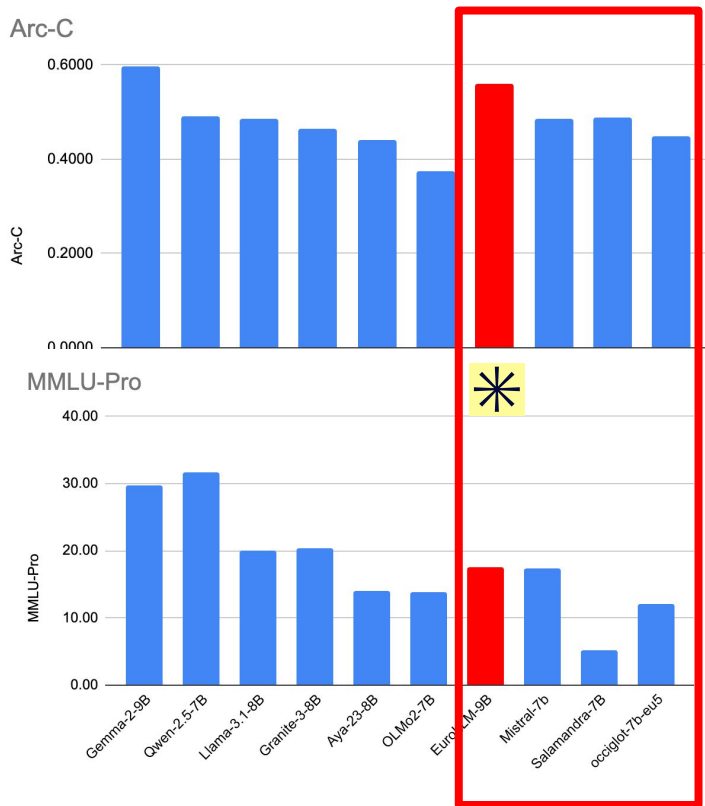
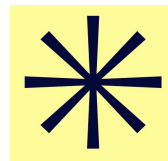
Languages:

- **All 24 EU official languages** + 11 additional languages.

Compute: **1.68M** GPU-hours in MN5 (through an Extreme-Scale EuroHPC grant).

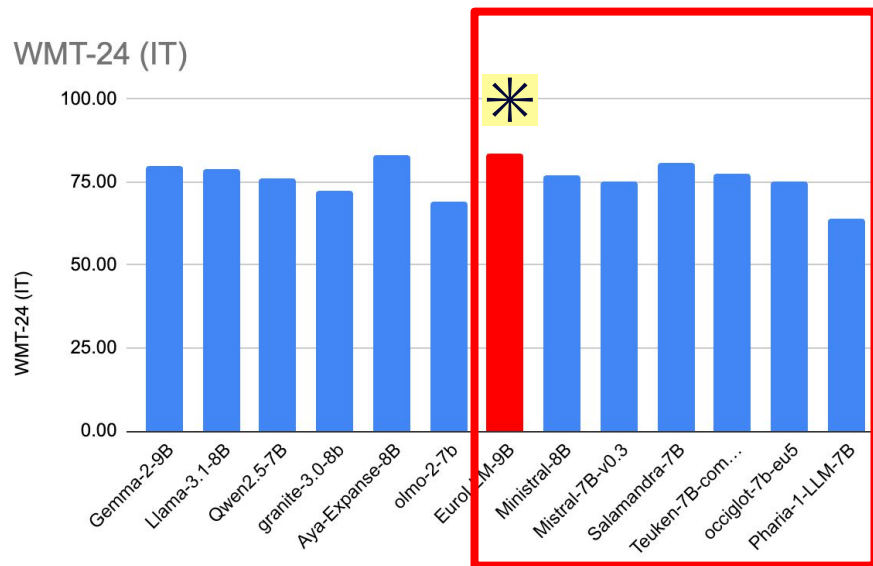
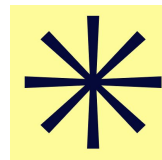
<https://euollm.io/>

Multilingual benchmarks (pretrained models)

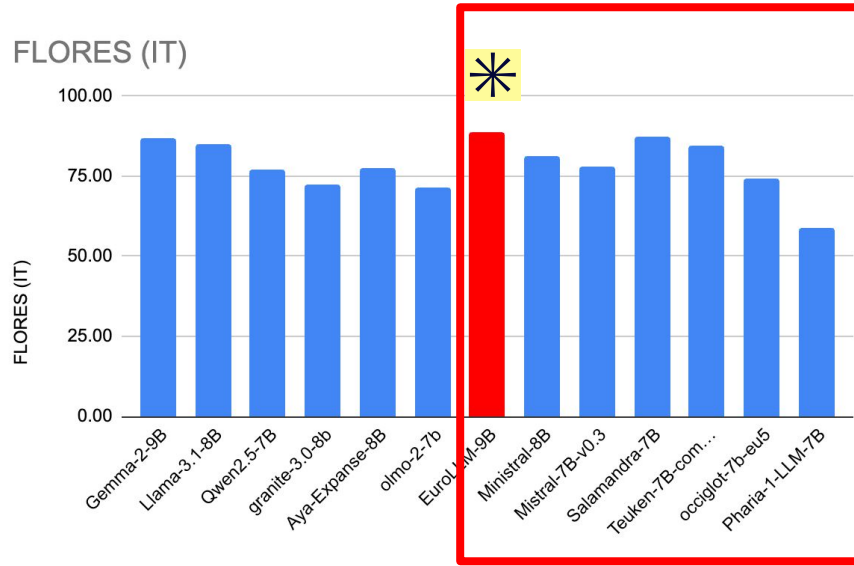


EU-made models

Translation benchmarks (instruction tuned models)



EU-made models



Try out EuroLLM: <https://huggingface.co/spaces/Widn/euollm-chat> (see our demo tomorrow!)

Challenges & Recommendations

Main challenges

Overall no big pain points, but a few hiccups:

- Some delays caused by HW problems in large-scale jobs (communication between nodes)
- The peer review process for grant allocation is too slow (AI moves **really** fast, a wait of 6 months kills EU's competitiveness)
- Large queues (sometimes) and short wall clocks requiring frequent restarts.

Recommendations for AI Factories

Identify success cases and projects that are delivering and give them what they need to keep delivering without interruptions

- Define measure of success
- Faster process for grant allocations (e.g. automatic renewal)

Encourage / require open science (open models, data, code)

Separate training from inference – these are two very different use cases

- Supercomputers for AI model training
- Outsourced cloud providers for inference.



@Widn.AI



Thank you