



OpenLLM-France 

Overview & Ambition

Michel-Marie MAUDET - 
mmaudet@linagora.com

Grenoble - March 15th 2024

 <https://alposs.fr>
Alpes Open Source Software
Le nouvel évènement Isérois de l'écosystème open source

Leader in Open Source

Since 2000, our mission is to invent, develop, implement, and maintain open-source software to preserve the technological independence of our clients and their digital sovereignty, especially when dealing with States. We combine this mission with the desire to have the maximum positive impact for our teams, society, and the planet.

4 

Present on 4 continents, LINAGORA brings together a united community of...

200 

200 employees who strive for excellence in one single area: Open Source..



The only really Open Source workplace

Twake Workplace helps you organize and automate your workflow, collaborate, and communicate easily and securely with your clients, team, and business partners



All in one solution

Feature-rich, multi-platform, SaaS, privately hosted, or on-premises

Open source

Community contribution, ethically by design

Private and secure

RGPD compliant, open and secure code, encrypted

#GoodTech4Good

Positive impact on the world! Contribute to the UN's SDGs!



Michel-Marie MAUDET
General Manager

 mmaudet@linagora.com

 +33 6 60 46 98 52



Schéma Directeur Open Source

Open source software migration



Open Source Software Assurance

Third-party Open Source provider



développement & intégration

Implementation and training

Open Weights ≠ Open Source, Values & Inclusivity

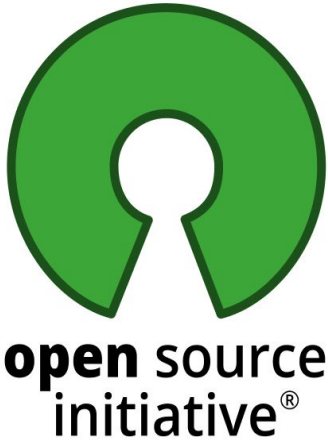
July 20, 2023 Stefano Maffulli OSI opinion

Meta's LLaMa 2 license is not Open Source



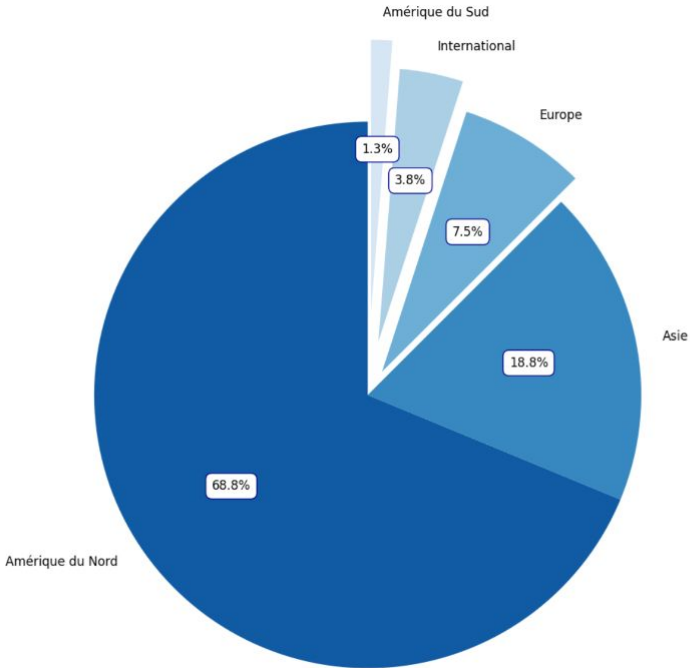
OSI's Open Source Definition

- free redistribution
- source code availability
- derivatives allowed
- no limitations of who may use it or for what
- no additional license in place
- license must not depend on distribution format, technology, presence of other works



<https://github.com/OpenLLM-France/OpenSourceAI-Definition>

Geographical distribution of LLMs with more than one billion parameters since 2018



LLAMA V2 : Language distribution in pretraining data with percentage

Language	Percent	Language	Percent
en	89.70%	uk	0.07%
unknown	8.38%	ko	0.06%
de	0.17%	ca	0.04%
fr	0.16%	sr	0.04%
sv	0.15%	id	0.03%
zh	0.13%	cs	0.03%
es	0.13%	fi	0.03%
ru	0.13%	hu	0.03%
nl	0.12%	no	0.03%
it	0.11%	ro	0.03%
ja	0.10%	bg	0.02%
pl	0.09%	da	0.02%
pt	0.09%	sl	0.01%
vi	0.08%	hr	0.01%

OpenLLM-France

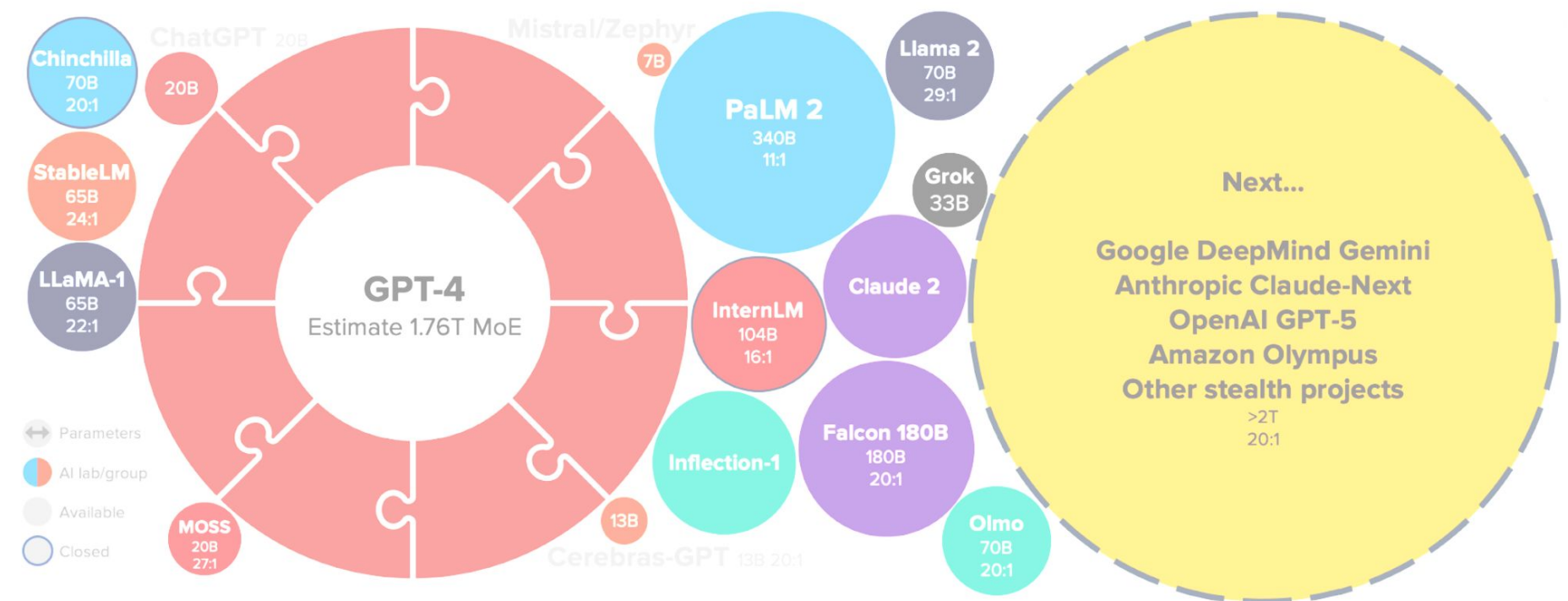


Models labeled as "Open Source," such as LLAMA, or closed LLMs like ChatGPT do not offer these guarantees.

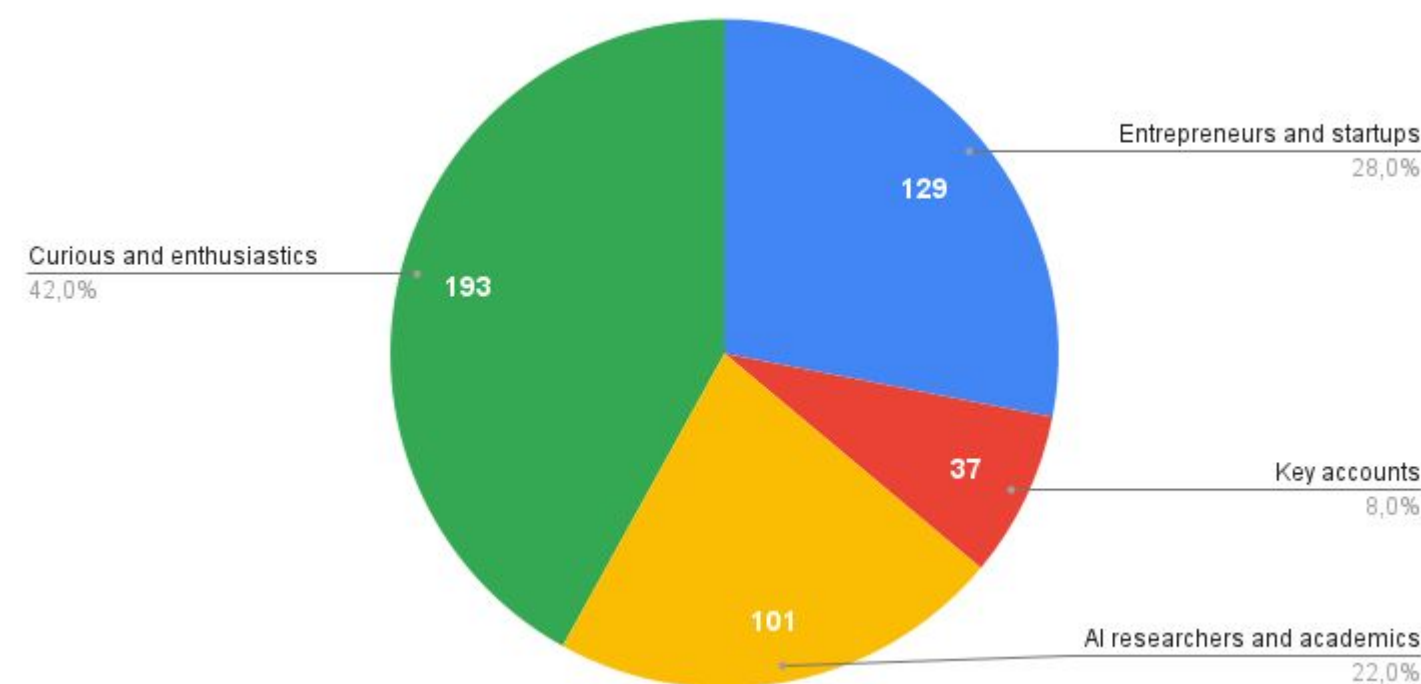
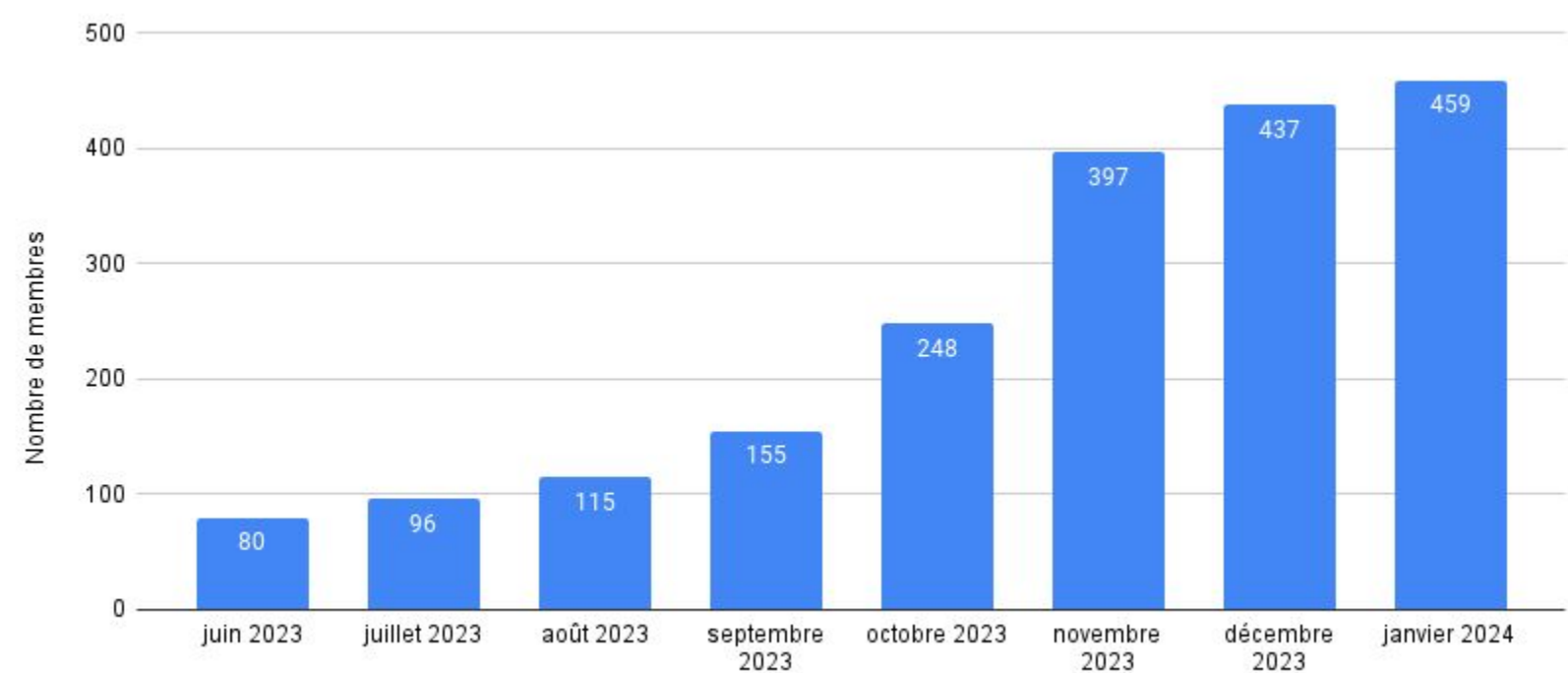
Furthermore, the learning process and data (issues with biases and preferences) are unknown.

Finally, these models are trained 90% on English, with EU languages et Arabic only partially taken into account.

Create **trustworthy, controlled**
and transparent GenAI
commons



OpenLLM France - largest French-speaking GenAI community



Academics / Public research



Corporates



E2E (End to End) Open Source

1 Open Source code

→ Source code of the code model, pre training tool released under OSI compliant Open Source Licence

2 Open Model

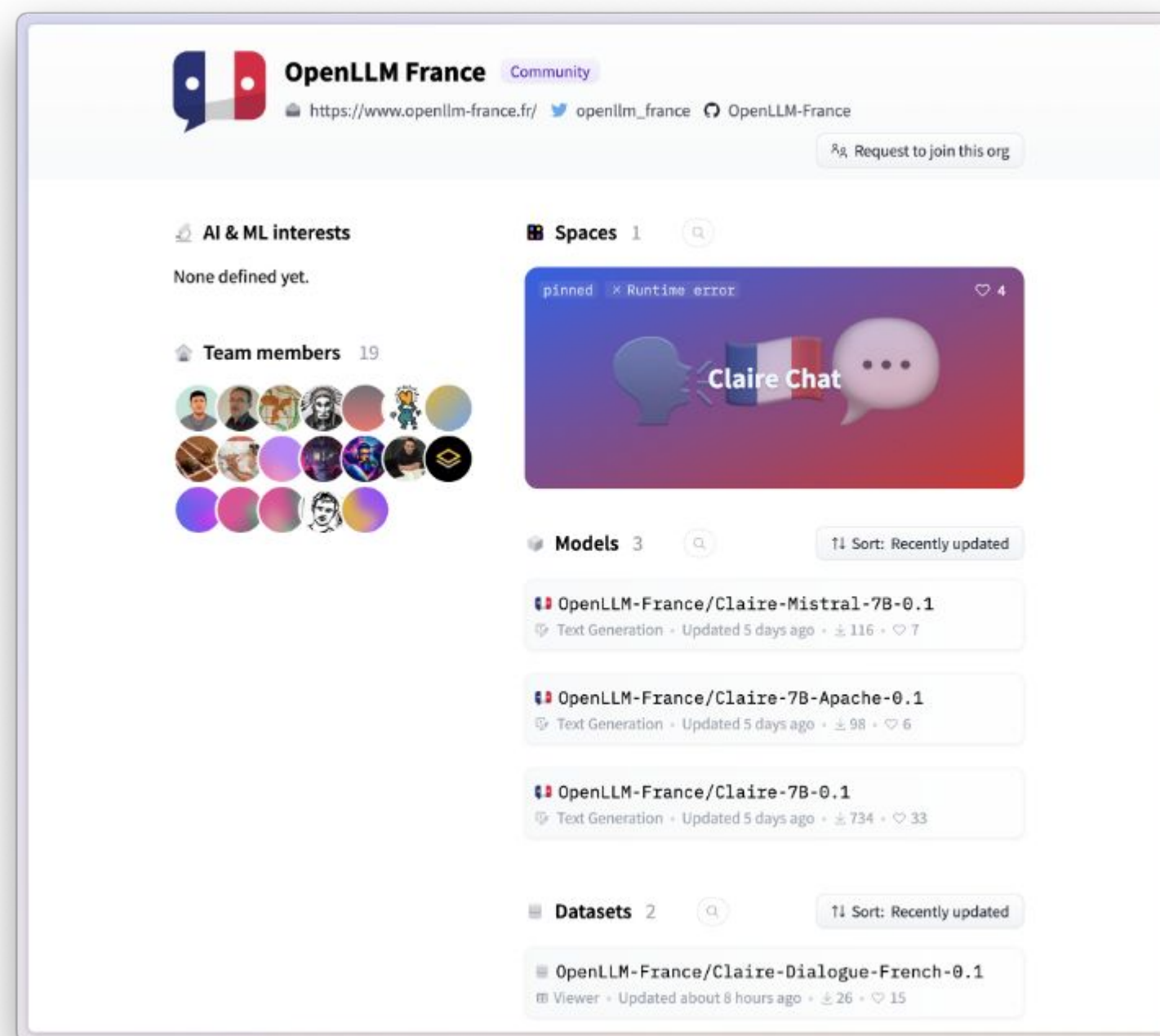
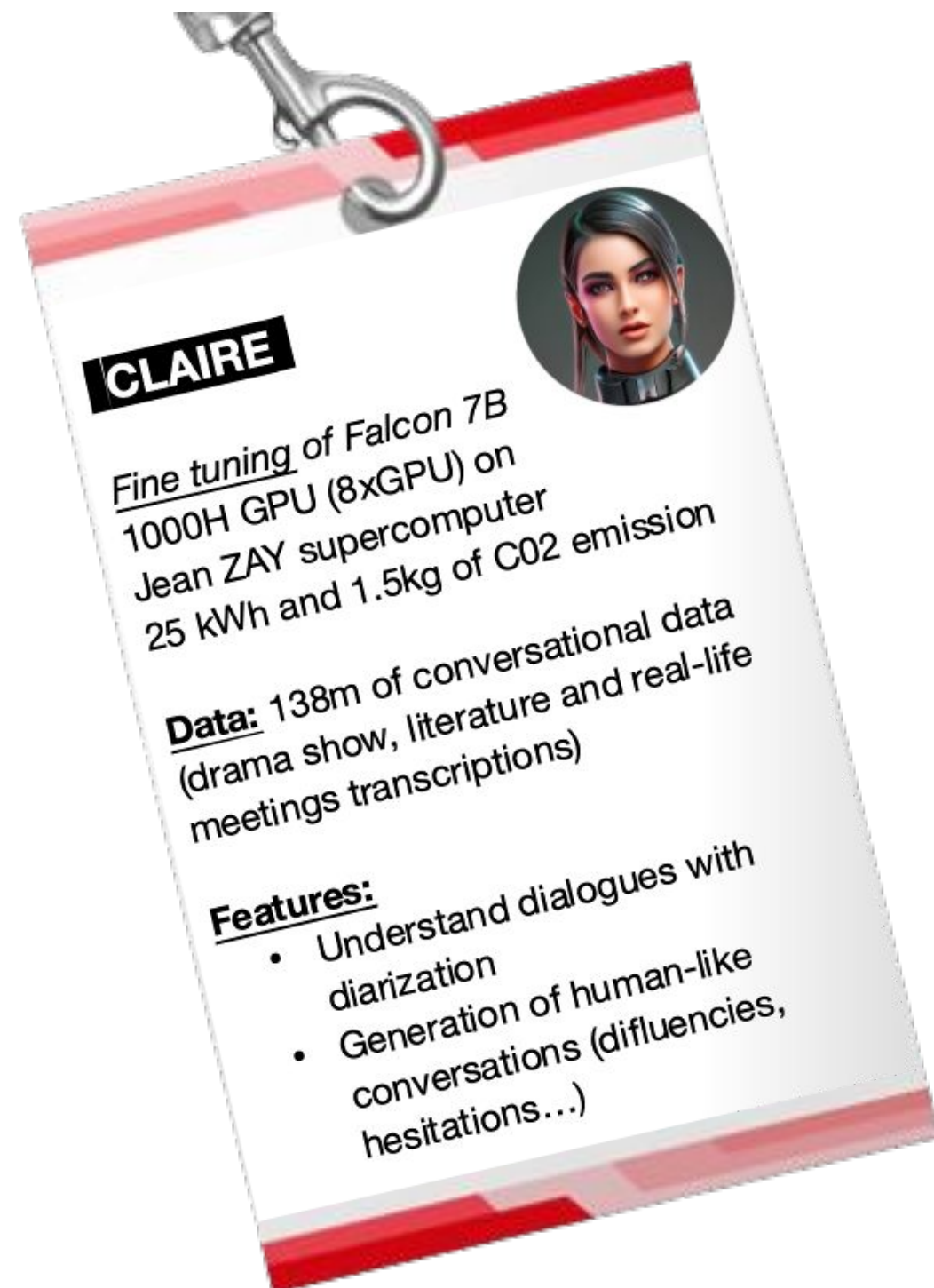
→ Licence and user agreement without any restrictions on who may use it and for what

3 Open Corpus

→ 100% of training data must be publicly available in a format that allows for investigation of the model's biases (preferences) and for retraining

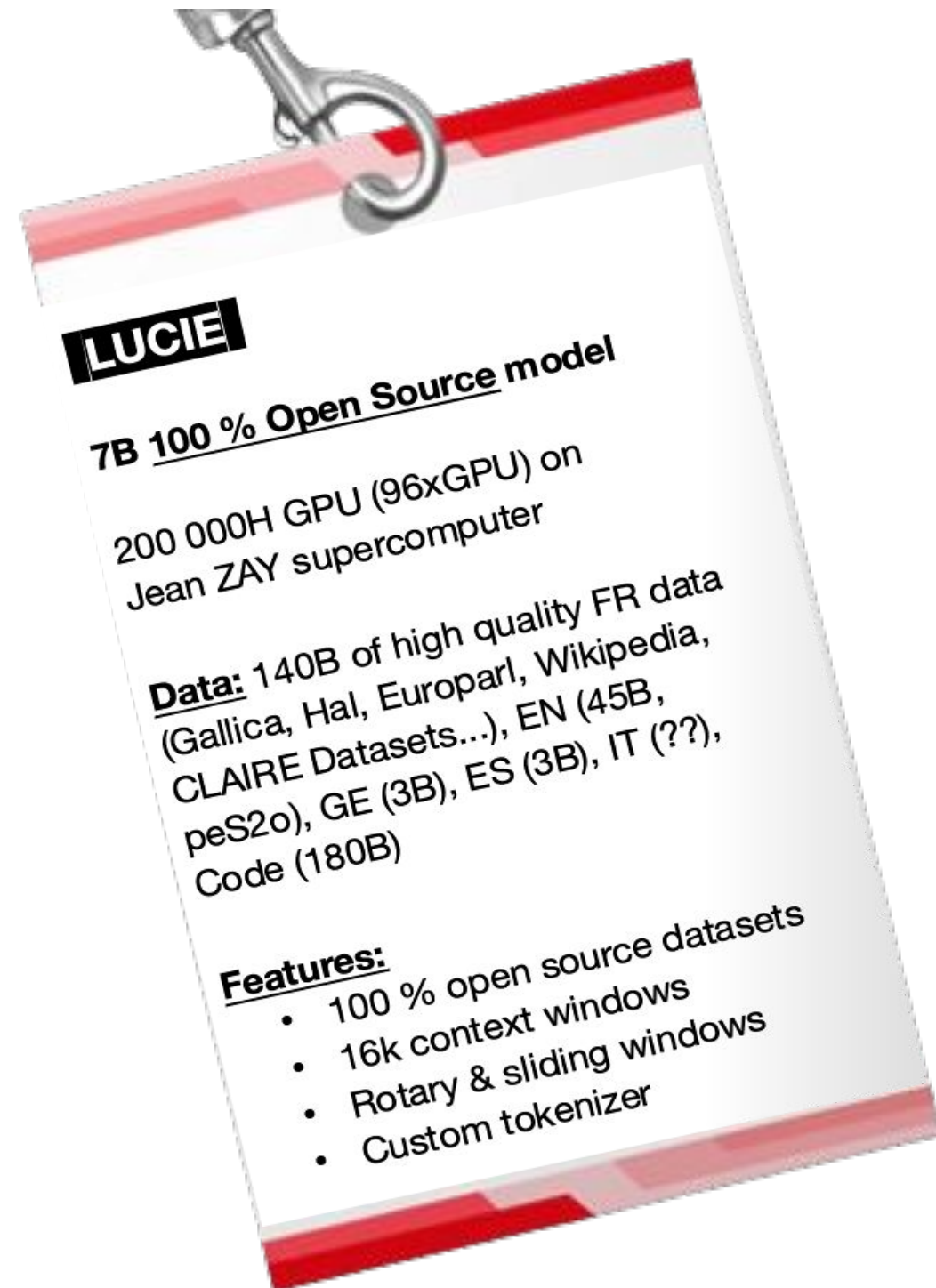
<https://github.com/OpenLLM-France/OpenSourceAI-Definition>

CLAIRE : the 1st Open Source Model (October 2023)



<https://huggingface.co/OpenLLM-France>

LUCIE : Full new pre training (ETA - May 2024)



THE OPEN
SOURCE WAY



LLM4ALL : Model for hospitals and emergency services

LLM4All

Large Language Models for All

Le Projet LLM4All financé par l'**ANR** : un consortium gagnant apportant les dernières avancées **en matière d'IA générative au service de la santé**

ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS

LINAGORA



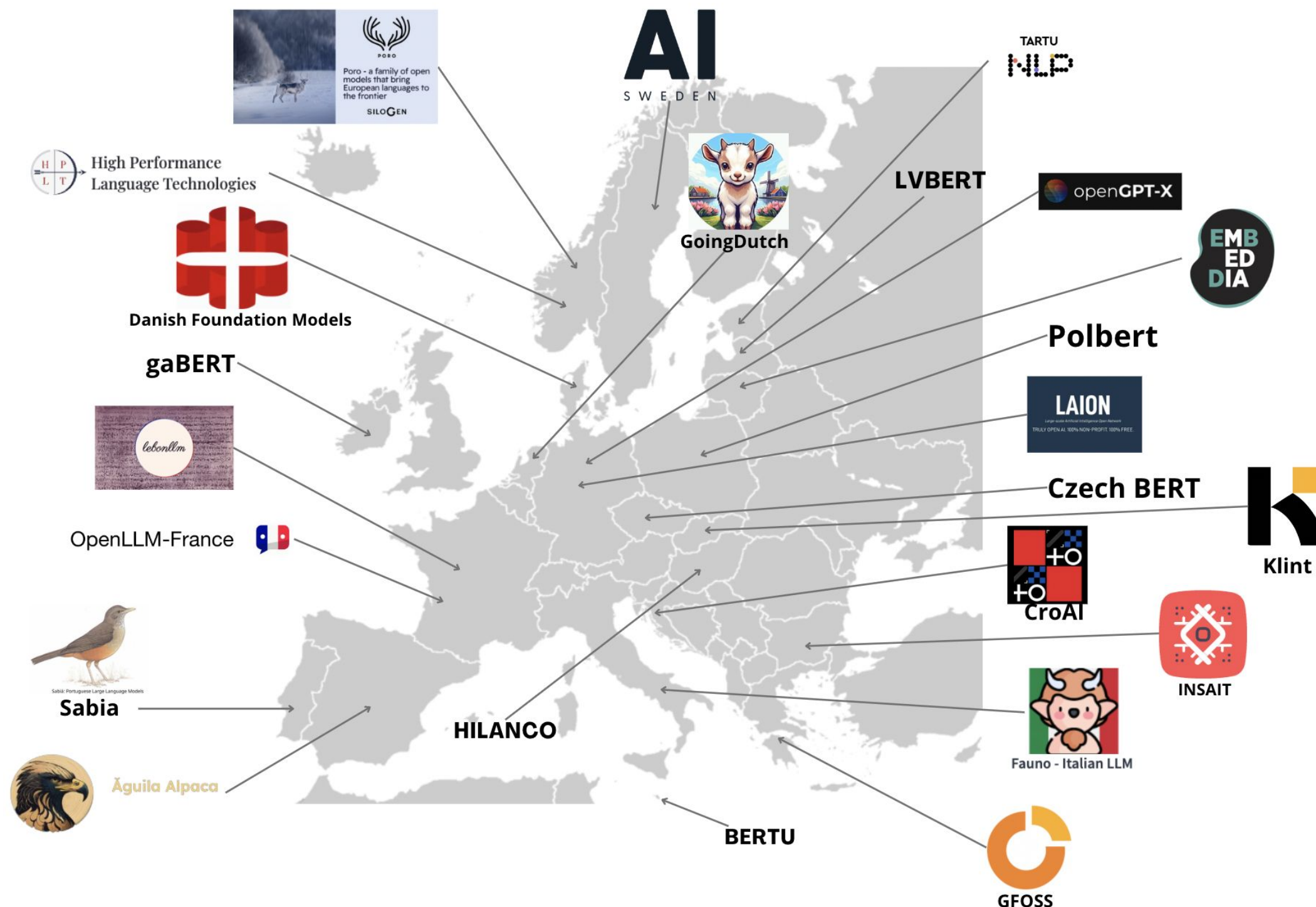
 **loria**
Laboratoire lorrain de recherche
en informatique et ses applications



anr ©
agence nationale
de la recherche

 **Hugging Face**

OpenLLM Europe



<https://github.com/OpenLLM-Europe/European-OpenLLM-Projects/tree/main>

<https://discord.com/invite/b5UQTWQn>

contact@openllm-europe.org

An aerial photograph of a dense urban neighborhood, likely in a city like Istanbul, showing a mix of modern and older buildings, streets, and green spaces. The entire image is overlaid with a semi-transparent red filter. Centered over the image is the text "Thanks for your attention" in a large, white, sans-serif font.

**Thanks for your
attention**