Can the Model Open All 24 Doors?

An analysis of OpenEuroLLM, EuroHPC compute, Gefion, and 24 Doors as a test of whether European AI infrastructure can prove language access rather than only perform multilingual coverage.

2026.07.10 01:44 DENNIS HEDEGREEN ANALYSE V1.1
HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/CAN-THE-MODEL-OPEN-ALL-24-DOORS/

I went into this research with a picture in my head, and the picture came from my feed: Europe regulates, America builds. Europe writes paper, America ships models. The European AI project as a press release with a budget.

The documents corrected me within a week.

What I found was an exascale machine in operation, four world-class supercomputers allocated to one model project, from-scratch pretraining code in public repositories, forty model artifacts you can download today, and a Danish SuperPOD in Copenhagen.

I also found that the promised models still have no shipped delivery behind them, and that the flagship's own coordinator says compute for the final models is not secured.

Both corrections matter. This article is about holding them at the same time — and about a distinction that makes both make sense.

The letter and the receipt

First, why I care at all.

In June I built a tool called 24 Doors — one door for each official EU language. You stamp a letter with up to three languages, pick a threshold, and the tool answers a plain question from weighted Eurobarometer microdata (Eurobarometer 100.1 / ZA8778): how much of the Union can reliably enter this message?

My test case: Irish as the entry door, Danish and Polish as the other stamps. Threshold: "the message can be trusted."

Estimated reach: 10.3% of EU27.

Not because Europeans are not intelligent — the receipt itself lists what it does not measure: not intelligence, not identity, not literacy, not citizenship, not rights, not worth. It measures one thing: whether a public message in given languages can reliably arrive.

Nine out of ten doors stay shut for that letter.

That is not a product gap. Nobody is missing an app. It is an infrastructure gap: the Union has twenty-four official languages, no shared one, and every message that matters either gets translated — at cost, with delay, by institutions — or does not arrive.

Hold that distinction. It is the key to everything that follows.

Product or infrastructure

There are two ways to build AI, and they are judged by different clocks.

AI as product is what my feed shows me: launches, demos, benchmark charts, a new model every quarter, visible the day it ships. Product progress documents itself.

AI as infrastructure is a different animal: the translation layer, the access layer, the thing that decides whether the other ninety percent of doors can open. Infrastructure is invisible until it works — and completely invisible when it works well. Nobody tweets that the bridge held again today.

The European Union, mostly without saying so, has chosen the second kind. OpenEuroLLM — 37.4 million euros total, roughly 20.6 million from the Digital Europe Programme, twenty organisations, coordinated from Charles University in Prague — promises a family of open source models covering all EU languages. Not the smartest model. The widest door.

That choice explains the story gap that fooled me. American labs ship products, and products travel in a feed. Europe ships allocations, reference models and deliverable PDFs on university domains — formats that do not travel at all. The doom narrative is not a lie. It is the product lens applied to a construction site.

But the infrastructure frame cuts both ways, and this article keeps both edges.

What is actually on the construction site

The honest inventory, checked July 8, 2026:

The compute is real, and it is historic in one specific way: OpenEuroLLM is the first AI project granted strategic access across multiple EuroHPC centres — over 10 million GPU hours on four machines: LUMI in Finland, Leonardo in Italy, JUPITER in Germany, MareNostrum 5 in Spain. Named allocations include 1.5 million GPU hours on LUMI for architecture and data-mixture work and 3 million on Leonardo for an open multilingual synthetic dataset.

The code is real: from-scratch pretraining infrastructure, a Megatron-LM fork, a training data catalogue mirrored across three of the machines. The question I carried in — is this just a Llama fine-tune with a European flag? — has a clear answer in the repositories: no.

The artifacts are real: forty models and twenty-four datasets on the public Hugging Face organisation — data-mixture ablations at 0.7, 2 and 9 billion parameters, long-context 9B prelude runs, custom tokenizers built for European languages. And there is precedent that Europe finishes: EuroLLM, an earlier project, trained 1.7, 9 and 22 billion-parameter models from scratch on European supercomputing and released them openly, with a tokenizer deliberately designed so smaller European languages are not tokenised into disadvantage. Language fairness starts in the tokenizer, long before anyone writes a benchmark.

Denmark has its own egg in the basket: Gefion, a SuperPOD of more than 1,500 NVIDIA GPUs run by DCAI in Copenhagen, funded by the Novo Nordisk Foundation and EIFO — capable, sovereign, and notably outside the EuroHPC system entirely.

So no: the basket is not empty. Europe does not have the finished working models yet — but the eggs are in the basket, and they are countable. Forty artifacts. Ten million GPU hours. Four machines.

The seal counted chickens. This article counts eggs.

The real bottleneck has a name

Here is where the infrastructure frame stops being comfortable.

The coordinator, Jan Hajič, has said publicly that significant challenges remain, especially in securing more compute for the final models. Read against the fleet above, that sentence seems strange — until you look at how the compute is shaped.

Ten million GPU hours arrive in slices: separate allocations, on four machines, across two GPU vendors — LUMI runs AMD, JUPITER runs NVIDIA, which means maintaining two software stacks — granted through calls with biannual cutoff dates. The comparison class trains on one homogeneous cluster, occupied continuously for months.

Europe's compute problem is not size. It is contiguity.

The fleet exists. What does not yet exist is the governance move of handing one project one very large block on one machine for as long as the run takes. That is not physics — a single synchronous training run cannot span Copenhagen to Jülich anyway; the latency kills it — it is allocation policy.

One line, and it is the hardest line in this article: Europe has the hardware of a superpower and the allocation model of a research council.

And the fairness note belongs right next to it: slice allocation is also a democratic feature. Many projects share public machines instead of one lab monopolising them. That trade-off — breadth of access versus depth of runs — is a real choice, not an obvious mistake. But it should be chosen consciously, not inherited from how physics grants were always handed out.

The side observation: announcing a product, building infrastructure

One thing in the papers deserves a short section — short, because it is a side observation, not the story.

On February 3, 2025, the Commission announced that OpenEuroLLM had received the first STEP Seal awarded to a Digital Europe Programme project, and described it as the first family of open source large language models covering all EU languages.

The seal is a quality mark for the proposal and its investment profile — a project judged excellent enough to deserve visibility and further investment. Legitimate. But the sentence around it reads in the present tense, and the present tense is not here: the official first-models deliverable is scheduled for December 31, 2026,

with an evaluation-code checkpoint on July 31.

Named with the frame of this article, the mismatch is simple: Europe is building infrastructure and announcing a product. Product language — "the first family", present tense, a seal — laid over a construction site. The error is not dishonesty; it is genre confusion. And the fix costs nothing: talk about it like the public works project it is. The project's own pages mostly do. Say "will be."

Nothing has failed. Nothing has been verified. Both halves of that sentence are true, and this article exists because of the second half.

What the receipt cannot say

Every instrument here has limits, and they belong in the text.

The 24 Doors receipt is a survey-based estimate from weighted microdata, not a census. Hedegreen Research has now checked the article scenario against the local 24 Doors engine, where EU27 plus Irish, Danish, and Polish at the confident-access threshold returns the same 10.3%. That lets the 10.3% travel as a 24 Doors estimate, clearly stamped as such — the tool's own return letter carries that exact caveat, and this article inherits it.

The deliverables timeline proves nothing negative. Early deliverables one year into a three-year project is what a construction site legitimately looks like.

A test run today on the public experimental artifacts documents today's state of play — not the first models, because those do not officially exist yet. An ablation run punished for weak Danish proves nothing about what ships in December.

And I cannot read Maltese, Finnish, or Bulgarian. So the breadth layer of the test to come contains only machine-verifiable gates, and the depth layer exists only where the judge is competent: Danish, my door.

One more limit, and it is the biggest one: both the need measurement and the coming inspection run on instruments I built myself. That is a dependency worth naming before anyone else names it. It is also the open-methodology bet: the tool, the rubric, the letter and the receipts are all public, so you do not have to trust my instruments — you can inspect them, the same way this article inspects the model.

Infrastructure is tested by inspection

If the model is infrastructure, it should be held accountable the way infrastructure is: not by reviews and launch-day takes, but by inspection — against the need it claims to serve.

The inspection method is already on the table, and it is this article's own instrument: the letter.

On July 8, 2026, a letter stamped Irish, Danish, Polish went through 24 Doors, and a return letter came back: 10.3% at the trust threshold, filed, signed, archived. That receipt is the pre-registration. Nobody can say the test was designed after the result, because the result of the first letter is already public.

When the first models arrive, the same letter goes through again — this time with the model as the translation layer. Same stamps, same threshold, same case grammar. A second return letter gets filed next to the first, whichever way it points. Before that run, the pass conditions for each threshold in model mode — what "can be trusted" concretely requires, door by door — will be published openly, so the second letter can be checked, not just believed.

Two dates are on the calendar. July 31, 2026: the evaluation-code package — the project's first testable promise, measured on its own date; I will publish a short note on whether it shipped. December 31, 2026: the first-models window. When they arrive, the letter goes through, and the second return letter gets published.

Three outcomes, registered now. The models open the doors: then Europe built the thing, and the second return letter becomes the independent confirmation nobody else prepared. The doors open unevenly: then "all EU languages" gets measured against its own receipt, door by door — and the project gets a map of exactly where to aim next. The window slips: then the follow-up is about time, compute, contiguity, and what it costs to provision an ambition.

I want the first outcome. You do not build an inspection for something you hope collapses — you build it for something you need to stand, because hope is not a measurement.

And maybe what Europe builds will not be good enough. That is a real possibility, and the second return letter will say so plainly if it happens. But nothing can ever be good enough if nobody begins. Europe began — with machines, allocations, code and forty countable eggs. That is worth more than a seal, and it is exactly why the result deserves to be measured instead of assumed, in either direction.

The letter is stamped. The receipt is filed. The doors are counted.

When the models arrive, we open them like doors — not like press releases.

Open the return letter

— Dennis Hedegreen, trying to see the structure

RELATION MEMORY

NEARBY I Shipped 24 Doors Before the Hackathon · Shares the 24 Doors language-access proof object that this article turns toward model inspection.

NEARBY AI Should Be Infrastructure, Not Authority · Provides the infrastructure frame this article applies to OpenEuroLLM.

NEARBY After Brussels: I Came Home With Requirements, Not a Product · Connects the participation-receipt idea to the return-letter logic used here.

CANONICAL URL: HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/CAN-THE-MODEL-OPEN-ALL-24-DOORS/
PDF VIEW GENERATED: 2026-07-11T15:35:45.764Z