AI Factories

A European perspective

Structure of this session

- Why are we talking about AI GigaFactories
- AI GigaFactories: what they are, how to build, how much they cost
- Networking for AI GigaFactories

We have digitalized everything…

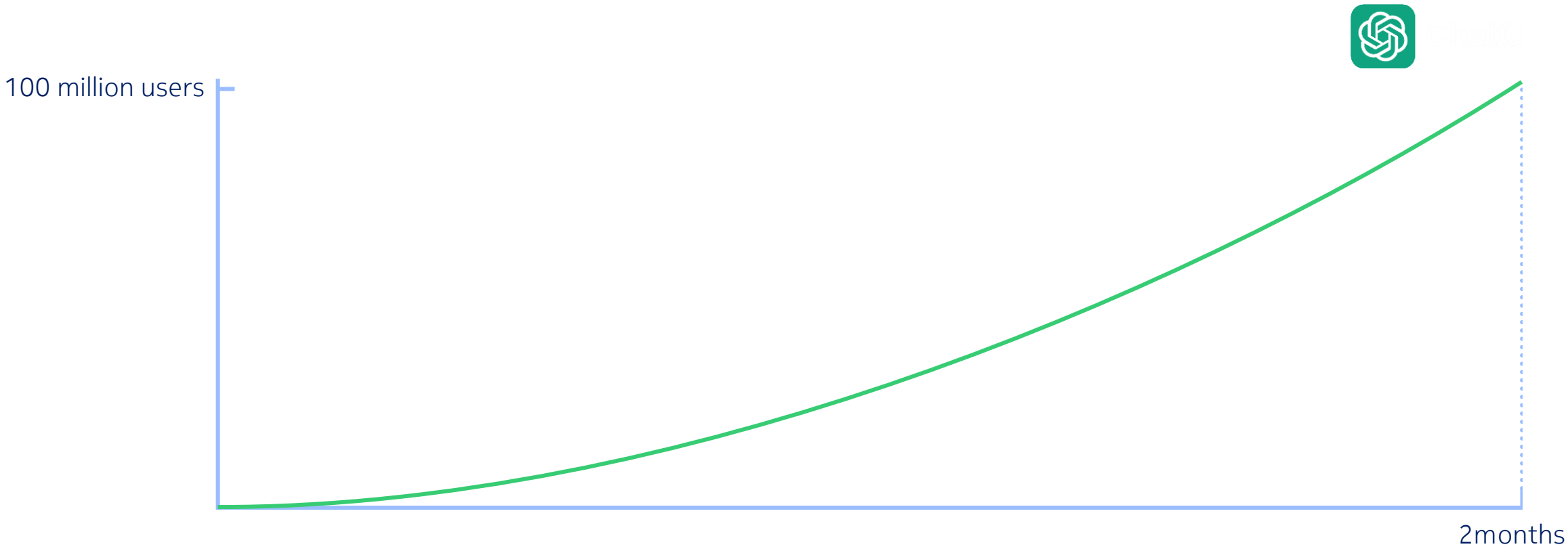
...and sourced it out of Europe

And just like that…

2015	2016	2020	2022 Feb	2022 Nov
	Brexit, Trump, “America First”	Covid-19, Supply chain shocks	Russia invades Ukraine	ChatGPT
We get cheap manufactures from China, affordable gas from Russia, and free Internet stuff from the US!! **Globalization is great**!!!	**Strategic autonomy**: defend European interests in a hostile geopolitical environment	**Strategic autonomy**: mitigate economic dependence on foreign supply chains	**European sovereignty**: reduce dependency on energy (and raw materials, semiconductors, digital technologies, food…)	

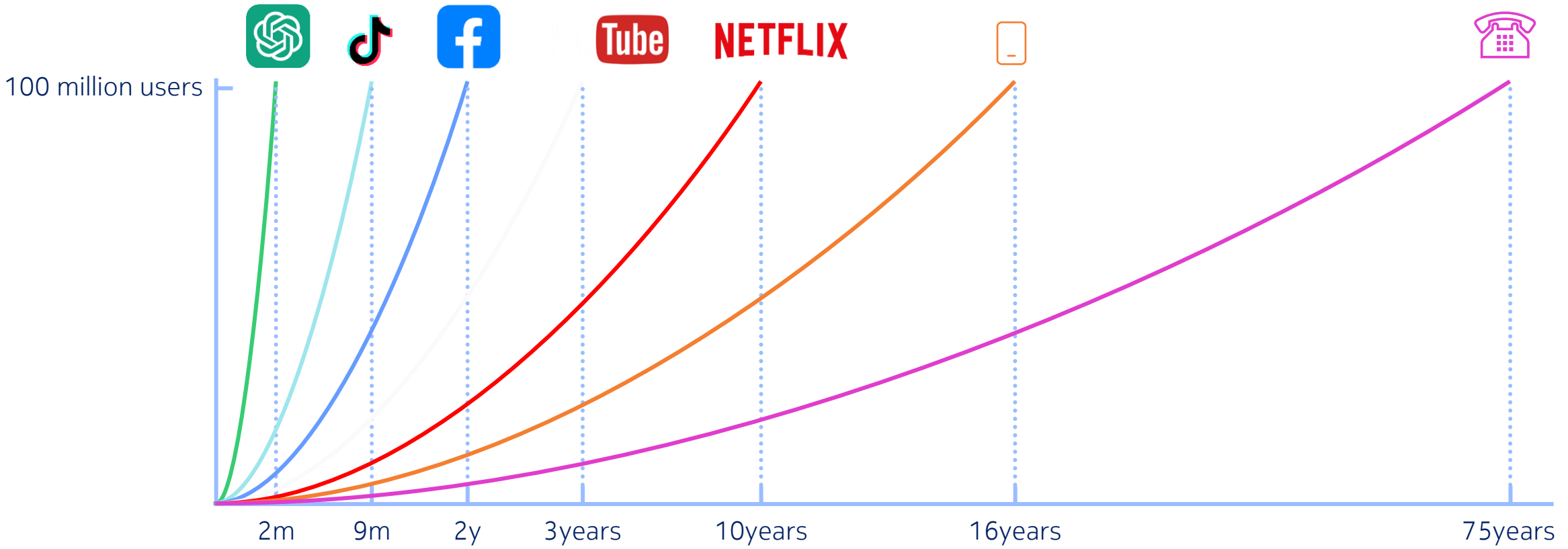
The AI boom

Time to 100 million users



The AI boom

Time to 100 million users



And yet again…

...and this time there's also a Geopolitical drive

THE WHITE HOUSE

Winning the Race

AMERICA'S

JULY 2025

AMERICA'S AI ACTION PLAN

Introduction

The United States is in a race to achieve global dominance in artificial intelligence (AI).

and military benefits. Just like we won the space race, it is imperative that the United States and its allies win this race. President Trump took decisive steps toward achieving this goal

AMERICA'S AI ACTION PLAN

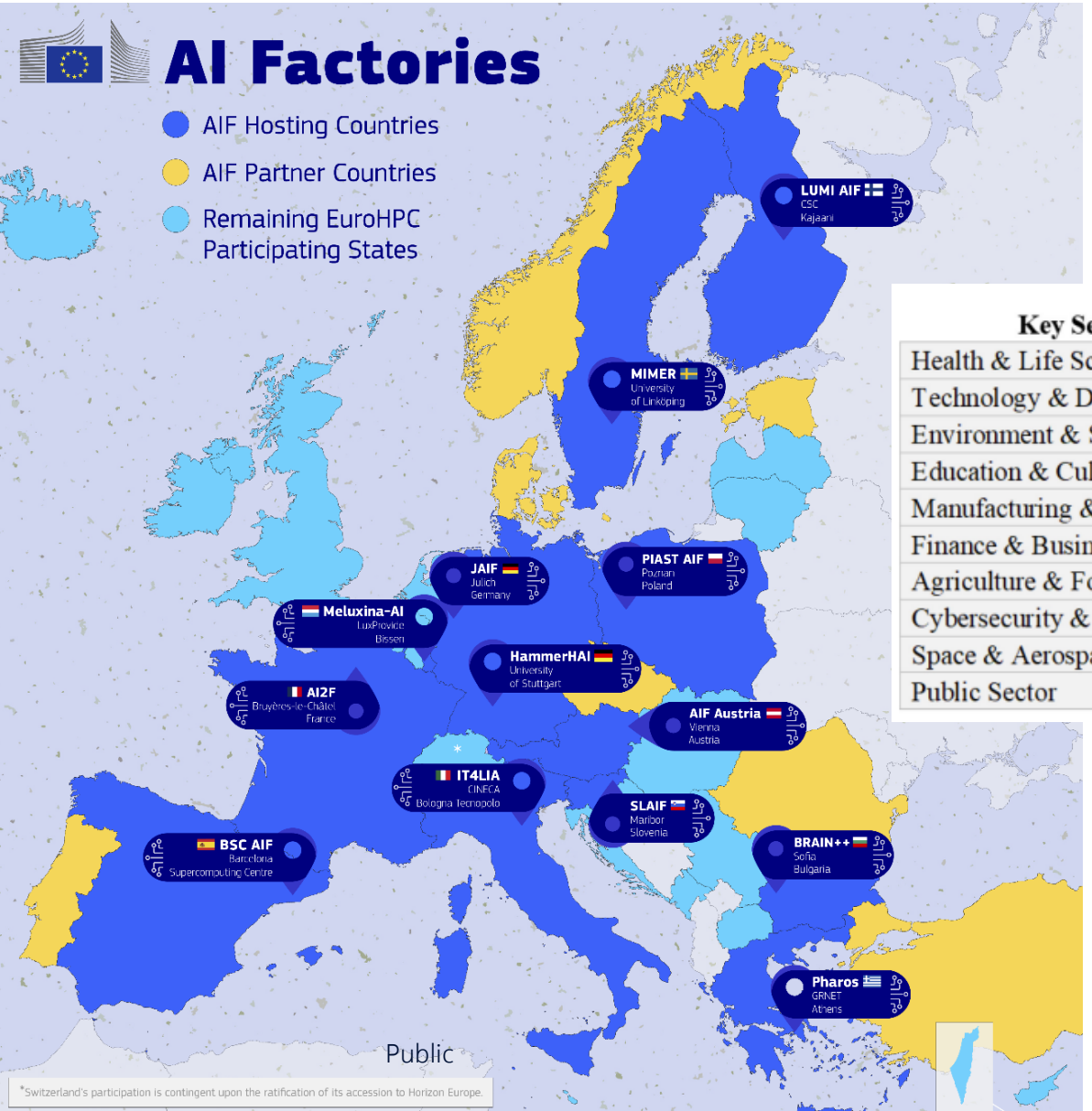
Pillar III: Lead in International AI Diplomacy and Security

To succeed in the global AI competition, America must do more than promote AI within its own borders. The United States must also drive adoption of American AI systems, computing hardware, and standards throughout the world. America currently is the global leader on data

Export American AI to Allies and Partners

The United States must meet global demand for AI by exporting its full AI technology stack—hardware, models, software, applications, and standards—to all countries willing to join

EU had taken some action…



Key Sectors	AT	BG	DE	EL	ES	FI	FR	IT	LU	PL	SE	SI
Health & Life Sciences	●		●	●	●	●	●	●		●	●	●
Technology & Digital		●		●	●	●	●	●	●	●	●	●
Environment & Sustainability		●	●	●	●		●	●	●	●	●	●
Education & Culture	●	●	●	●	●		●	●			●	●
Manufacturing & Engineering	●	●	●			●	●				●	●
Finance & Business	●		●		●		●	●	●		●	
Agriculture & Food	●				●		●	●			●	●
Cybersecurity & Dual use							●	●	●			
Space & Aerospace		●					●		●	●		
Public Sector	●		●		●					●		

...but realized it is falling far behind US and China in the AI race...

VS.

Artificial intelligence *Tech / Big Tech*

Tech war: China poised to expand AI infrastructure to keep pace with US Stargate project

Chinese local governments, telecommunications firms and Big Tech companies have been building new AI facilities over the past few years

Bank of China's AI Industry Development Action Plan

U.S.-China AI arm race is set to escalate

Category	Details
Total Funding	At least ¥1 trillion ($137B) over 5 years, with no less than ¥300B allocated to equity and debt financing.
Core Objectives	Strengthen national technology self-reliance.
	Enhance AI infrastructure (computing power, data supply, and key equipment).
	Accelerate AI innovation and application development.
Key Areas	AI-driven robotics, biomanufacturing, advanced materials, and low-altitude economy.
Strategic Goal	Establish a robust AI financial ecosystem to promote innovation, industrial growth, and sustainable development.

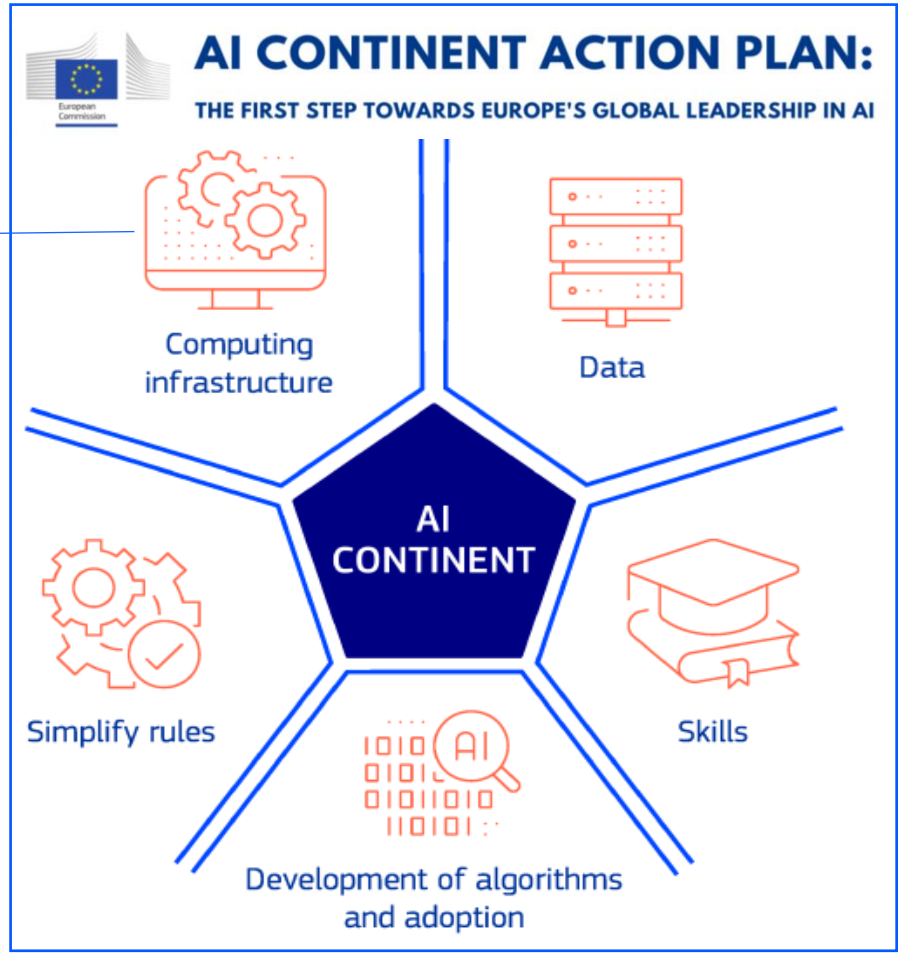
Source: Bank of China • Ray Wang

...and launched an ambitious AI Continent Action Plan...

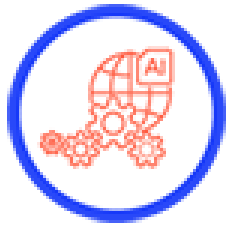
AI Factories
- **Objective**: train and finetune AI models
- Budget: €10 billion from 2021 to 2027
- At least 13 operational AI factories by 2026

AI Gigafactories
- **Objective**: train and develop complex AI models
- 4x more powerful than AI Factories
- €20 billion mobilised by InvestAI
- Deploy up to 5 Gigafactories

Cloud and AI development Act
- **Objective**: boost research in highly sustainable infrastructure
- Encourage investments
- Triple the EU's data centre capacity in the next 5-7 years



...where AI GigaFactories play a key role...



AI Gigafactories

- **Objective**: train and develop complex AI models
- 4x more powerful than AI Factories
- €20 billion mobilised by InvestAI
- Deploy up to 5 Gigafactories

- The AIGF initiative is a once-in-a-generation chance for Europe to establish world-leading AI infrastructure. By building facilities with **100,000+ advanced AI processors**, Europe can match or surpass global leaders, securing **strategic autonomy** and reducing reliance on foreign technology.
- It strengthens Europe's ability to independently develop, deploy, and regulate advanced AI systems, ensuring **alignment with European values** and regulatory frameworks.
- The establishment of a single AI Gigafactory is estimated to require significant investments, encompassing both capital expenditures and operational expenses. Given the magnitude of the necessary investment, these AI Gigafactories will be implemented through **public-private partnerships** and innovative funding mechanisms.

...and Romania is ready to contribute

MINISTERUL ECONOMIEI, DIGITALIZĂRII, ANTREPRENORIATULUI ȘI TURISMULUI

Acasă Despre noi Informați

Ministrul Bogdan Ivan anunță intenția României de a găzdui Black Sea AI Gigafactory, o fabrică AI de ultimă generație care va deservi inclusiv Republica Moldova și întreaga regiune a Mării Negre

23 iunie 2025

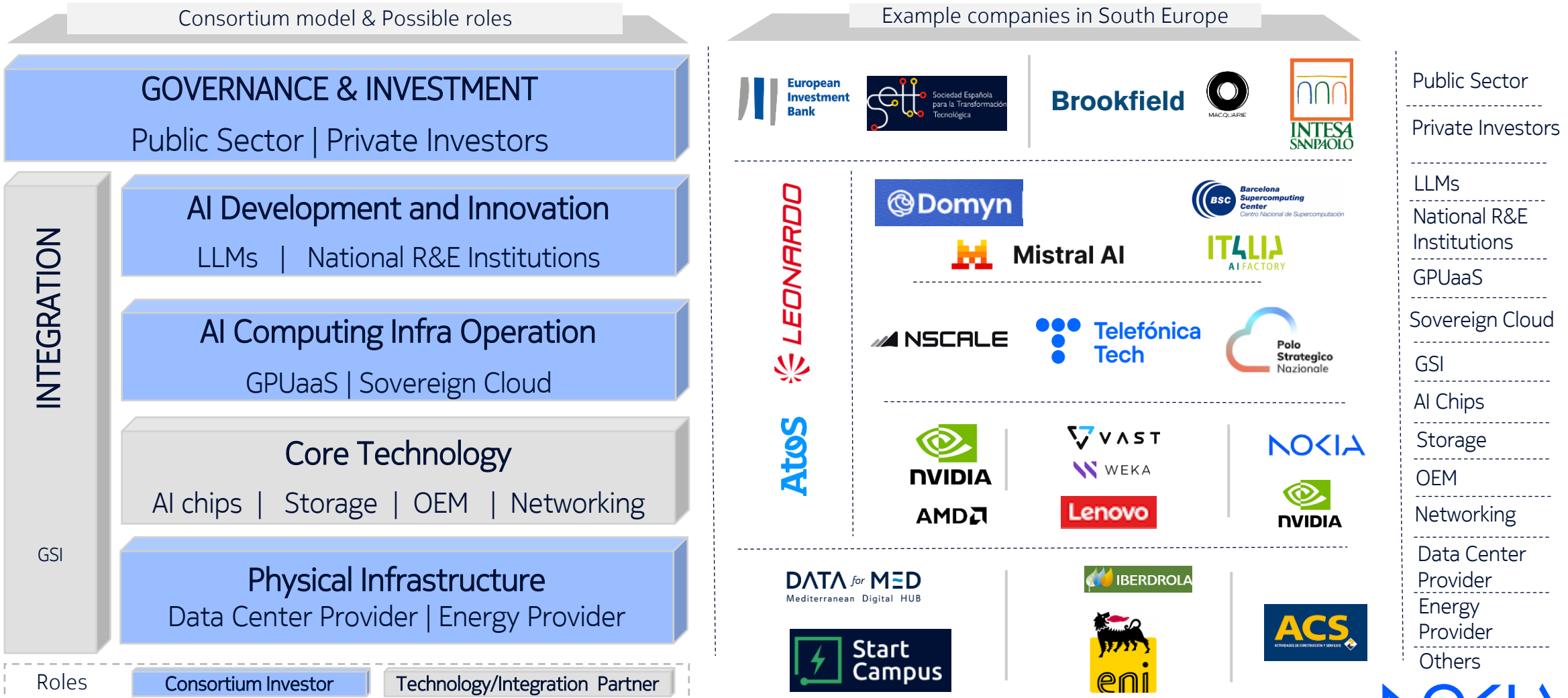
Bogdan Ivan, ministrul Economiei, Digitalizării, Antreprenoriatului și Turismului anunță transmiterea de Autoritatea pentru Digitalizarea României a scrisorii de intenție pentru găzduirea **Black Sea AI Gigafactory**, parte a unei inițiative strategice europene care vizează **consolidarea suveranității tehnologice și a capacităților continentale în domeniul inteligenței artificiale**, în deplină conformitate cu AI Act și obiectivele Uniunii privind suveranitatea tehnologică. Documentul a fost a transmis Comisiei Europene, prin EuroHPC Joint Undertaking.

> „Proiectul depus nu este doar despre construirea unui AI Giga Factory în România, ci reprezintă **accesul real la infrastructură de ultimă generație pentru frații noștri din Republica Moldova, pentru statele din zona Mării Negre – și aici trebuie să menționez Ucraina și Turcia –** care și-au exprimat susținerea fermă față de această nevoie comună. **Această inițiativă este despre aducerea valorilor europene la Marea Neagră: valori precum cooperarea, inovația, suveranitatea tehnologică, securitatea cibernetică, incluziunea și reziliența. Este despre a ne asigura că nicio țară din această regiune nu rămâne în urmă.** Nu construim doar o fabrică de AI, **punem bazele unei regiuni a Mării Negre conectate, sigure și pregătite pentru viitor, ancorate în spiritul unității europene comune**”, a precizat Bogdan Ivan, ministrul Economiei, Digitalizării, Antreprenoriatului și Turismului.

Propunerea României este susținută de un consorțiu național reprezentativ, format din actori publici, privați și academici de mare prestigiu și vizează dezvoltarea unei infrastructuri AI de ultimă generație, cu arhitectură hibridă, capabilă să deservească atât procese complexe de antrenare, cât și de inferență AI, într-un cadru operațional robust, sigur și sustenabil.

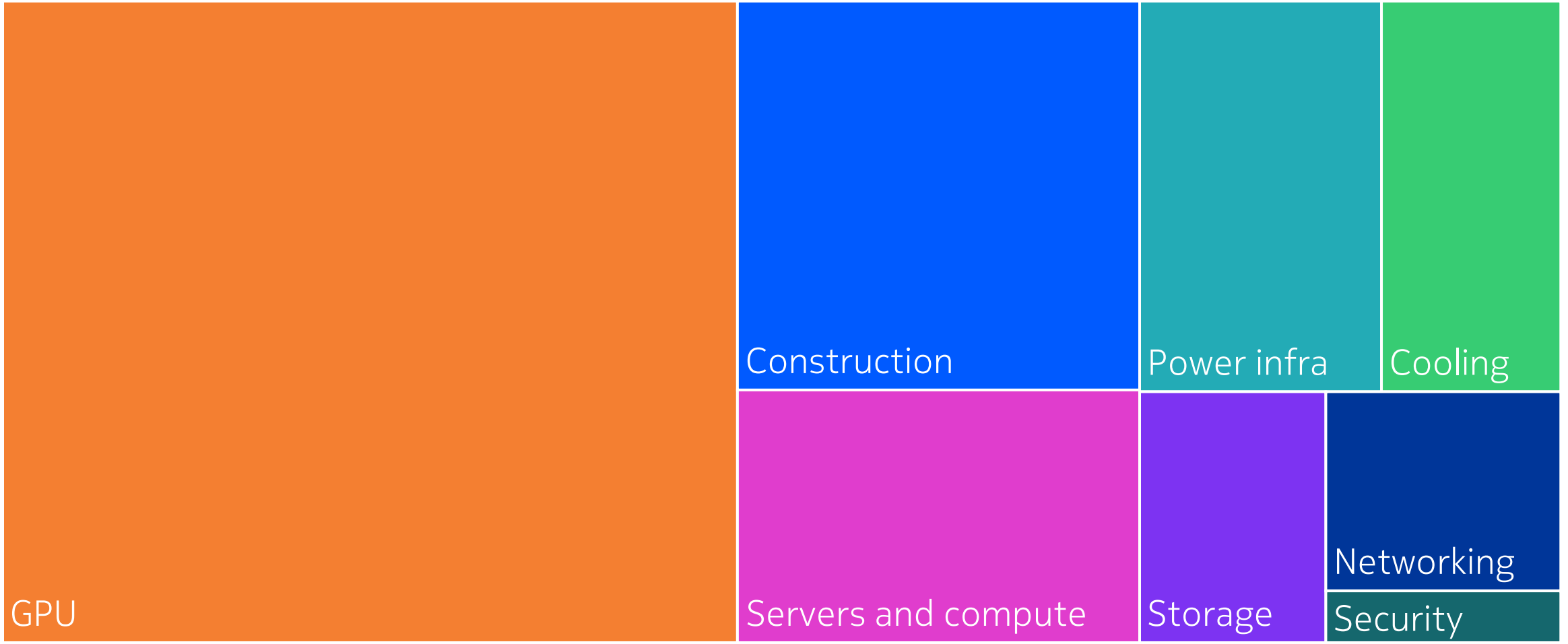
Proiectul prevede instalarea a peste 100.000 de acceleratoare AI în două locații distincte: Cernavodă (Faza I) și Doicești (Faza II), ambele selectate pentru avantajele strategice în materie de energie, infrastructură digitală și acces la conectivitate internațională de înaltă capacitate. Astfel, vorbim de un proiect ce va fi alimentat de un mix energetic de până la până la maxim 1.500 MW, sigur, accesibil și sustenabil, care poziționează România ca un centru strategic pentru calculul de înaltă performanță, respectând mediul, cu energie nucleară fără emisii de carbon.

How do you build an AI Gigafactory



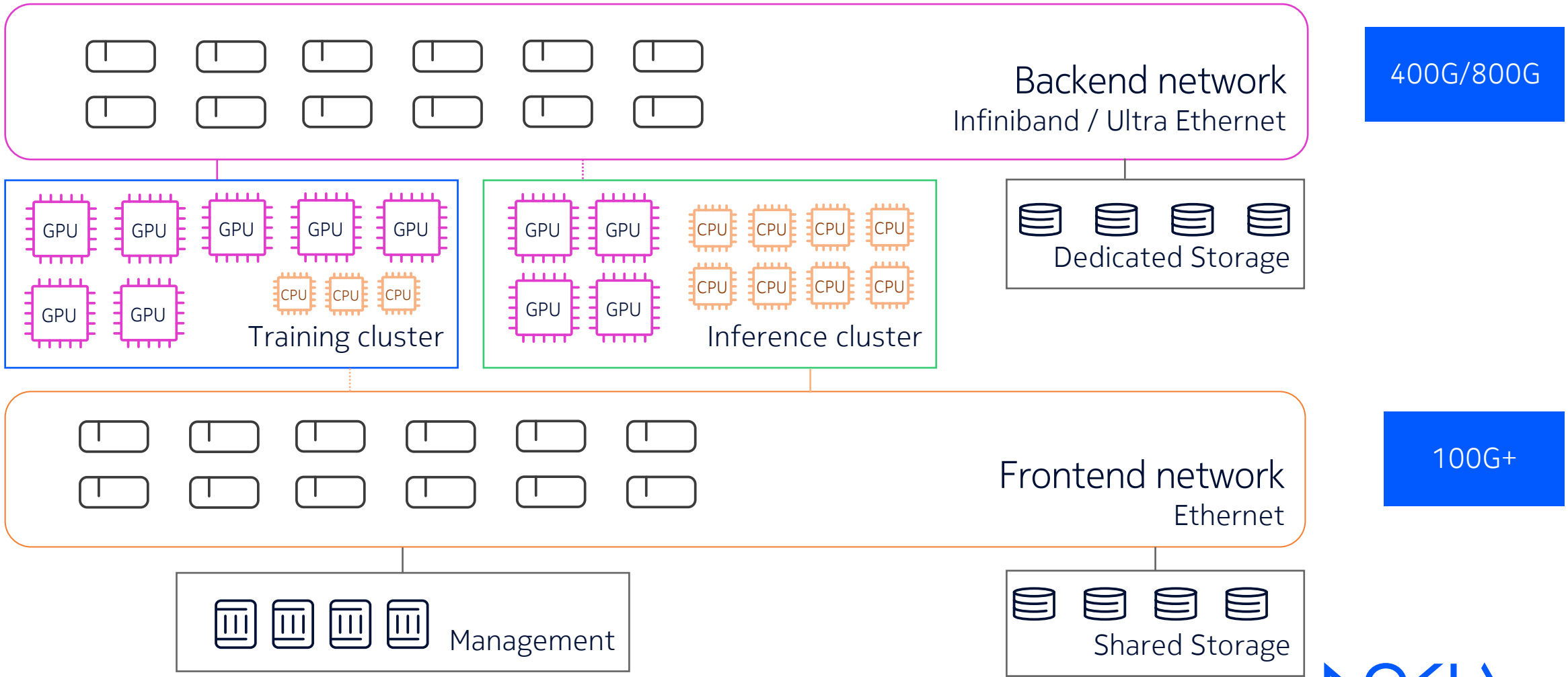
How much it costs to build an AI Gigafactory infrastructure

Estimates: between 5-20B€



Networking in an AI Gigafactory

AI clusters



What's the Problem?

AI and HPC networks are different

- Endpoints are fast, load is high
- Flows are few and high bandwidth
- RTTs are short
- Flows are synchronized
- Completion time determined by slowest flow

Vanilla networking doesn't meet the needs

Ethernet catches up to InfiniBand

How Ethernet closed the gap?

RoCEv2

RDMA performance over L2/L3 Ethernet.

Enables direct memory access between systems.

Eliminates need for TCP/IP networking stack overhead.

Enables low-latency, high-throughput transfers like InfiniBand.

Lossless Ethernet

Stable, predictable performance.

Priority-Based Flow Control: Flow control to prevent packet loss.

Explicit Congestion Notification (ECN): Signals congestion without dropping packets.

Data Center Quantized Congestion Notification (DCQCN): Decreasing the transmission rate when congestion starts.

* Sender to receiver latency

Ethernet's Ecosystem Advantage

XPUs/CPUs

NVIDIA

AMD

intel

...

NICs

NVIDIA AMD BROADCOM

SFPs

Vendor SFPs + 3rd Party SFPs

DACs/AOCs/AECs

Vendor Cables + 3rd Party Cables

Management Layer

Vendors Platforms + Open-Source and 3rd Party Management Tools

Switch

NOKIA ARISTA JUNIPER NETWORKS CISCO NVIDIA DELL Hewlett Packard Enterprise

Toolchain

ANSIBLE Grafana netbox Terraform MuleSoft

NOKIA

Call to action

The AI race is far from over

But Europe needs to take action to deliver on “Thrustworthy AI” promise

We need to invest efficiently

Networking choices will dictate openess

To make Europe successful

By nurturing a European AI ecosystem