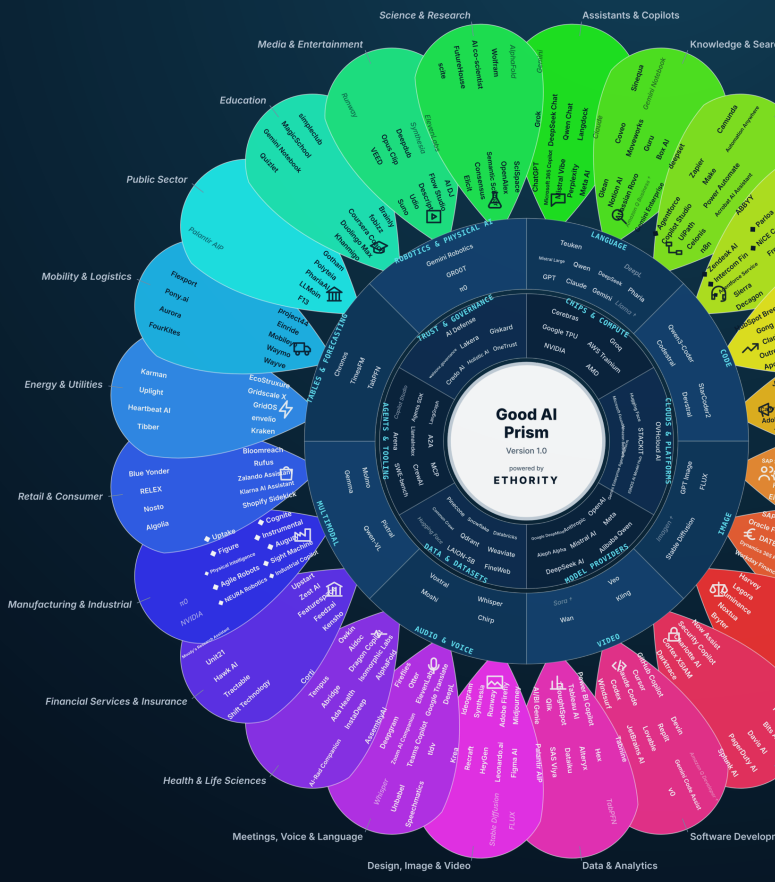
ETHORITY
AI BUSINESS OPERATING SYSTEM



GOOD AI PRISM · TECHNOLOGY BRIEF 2026/27 · CHIPS & COMPUTE · GLOBAL EDITION

Secure the compute your AI runs on

Where AI chips and computing capacity come from, what the market offers, and the rules and energy terms by jurisdiction.

21
AI chip and compute offerings mapped in Chips & Compute

US$89bn
NVIDIA data-centre revenue in one quarter, up 117% year on year (Q2 FY2027)

12 Nov 2026
deadline of the EU call for up to seven AI gigafactories (EuroHPC)

Created by Sten Franke with a people and governance lens by Severin L. Wilson · October 2026

AT A GLANCE

Compute is the scarce input of the AI economy

Every AI use runs on chips, memory, power and data-centre space. Demand for all four grows faster than supply, so where and on what terms an organisation buys compute now shapes its cost, speed and sovereignty.

This brief applies the Good AI Prism Market Study 2026/27 to one technology layer: chips and compute. It combines the study's market map of 21 offerings with current evidence on supply, prices, energy and rules across major jurisdictions. Europe is ETHORITY's hub market and Germany its home and test market, so both get the closest look. The brief ends with the questions to settle before a commitment and with how ETHORITY helps.

+117%
NVIDIA data-centre revenue growth year on year, to US$89.0bn in Q2 FY2027 (NVIDIA, Aug 2026)

+41%
TSMC revenue growth in Q3 2026; 3nm made up a third of wafer revenue (TSMC, Oct 2026)

2027–28
ASML expects its planned supply will not fully meet customer demand (ASML, Oct 2026)

+75%
planned rise in big-tech capital spending in 2026, after more than US$400bn in 2025 (IEA, Apr 2026)

945 TWh
data-centre electricity use projected for 2030, up from 415 TWh in 2024 (IEA)

US$2.53
per H100 GPU-hour at neoclouds, about a third of hyperscaler on-demand rates (Silicon Data index, Jul 2026)

About this brief

Market figures come from the Good AI Prism 2026/27 study dataset: every entry is recorded with a dated source from the provider's own pages. Job groups are ETHORITY's classification. External figures name their source and date and come from company filings, the IEA, EuroHPC, Bitkom and market trackers; company definitions differ and are not combined. This brief is not legal or investment advice.

WHERE THE MARKET STANDS

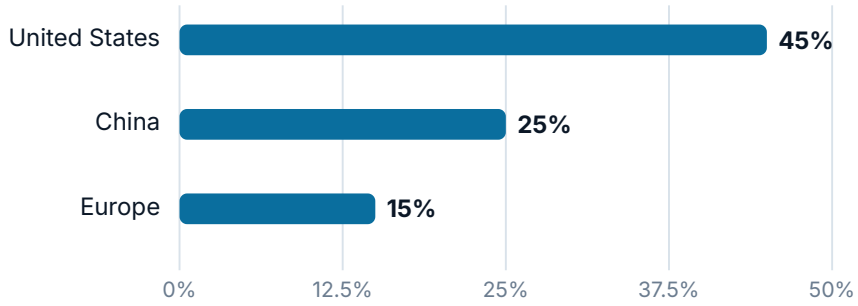
Demand runs ahead of supply

Demand for AI compute grows faster than the supply chain can expand. NVIDIA's data-centre revenue more than doubled within a year to US$89.0bn a quarter, and its Vera Rubin systems are in full production. TSMC's revenue rose 41%, and ASML says its planned 2027–28 supply will not fully meet demand. The head of SK hynix expects memory to stay short until 2030.

Big-tech capital spending passed US$400bn in 2025 and is set to rise a further 75% in 2026 (IEA). Power follows: data centres used 415 TWh in 2024, 45% of it in the United States, and the IEA expects their consumption to double by 2030. For buyers, capacity, region and energy terms now matter as much as chip performance.

Exhibit 1 · Where data-centre electricity is used

Share of global data-centre electricity consumption by region, 2024



IEA, Energy and AI, 2025 · global data-centre electricity use: 415 TWh in 2024

PRICE, POWER AND PLACE

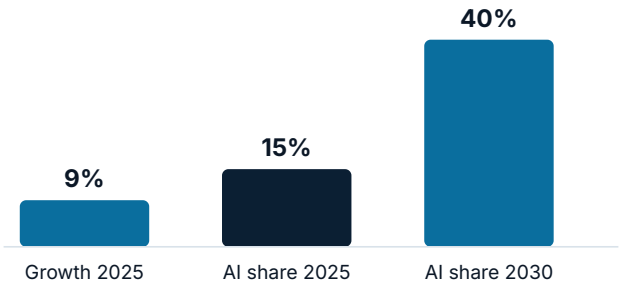
Prices diverge, while power and location decide

Model inference keeps getting cheaper: the price for a given level of capability fell between 9x and 900x a year, depending on the task (Epoch AI). GPU rental is different. H100 hours at neoclouds cost about US$2.53, a third of hyperscaler on-demand rates, and prices have held flat rather than fallen, while memory contract prices rose 15–18% in one quarter (TrendForce).

Germany, ETHORITY's home and test market, shows why location matters. Its data centres had 2,980 MW of IT capacity in 2025, more than a third of it in Hesse. Bitkom expects AI to take about 40% of capacity by 2030, up from about 15%.

Exhibit 2 · Test market Germany: data-centre capacity and AI

Growth of IT capacity in 2025; approximate AI share of capacity, 2025 and 2030 (projection)



Bitkom, Rechenzentren in Deutschland, 10 Nov 2025

21.3 TWh

electricity used by German data centres in 2025, up from 20 TWh in 2024 (Bitkom)

48 GW

data-centre capacity in the United States in 2024, against about 3 GW in Germany (Bitkom)

THE RULES

Export, energy and sovereignty set the terms

ETHORITY reads the rules by jurisdiction. For compute they come from three directions: export controls decide who may buy which chips, energy and data-centre rules decide where capacity is built and at what cost, and the AI Act adds duties for general-purpose AI models. The next pages map the other major jurisdictions.

European Union: the hub market

15 Sep 2024	EED	Data centres from 500 kW report energy and sustainability indicators each year under the Energy Efficiency Directive.
2 Aug 2025	GPAI	Duties for providers of general-purpose AI models apply; the AI Office enforces them directly from 2 Aug 2026.
12 Nov 2026	CALL	EuroHPC AI Gigafactories call closes: up to seven sites, more than €20bn in private investment, selection in early 2027.
From 2028	CAPACITY	New gigafactory capacity can arrive: selected sites are to be operational within 18 months of selection.

AI Act dates as amended by the Digital Omnibus on AI, Regulation (EU) 2026/1744, in force since 27 July 2026. The EuroHPC network already counts 19 AI Factories and 13 AI Factory Antennas (EuroHPC, Jul 2026).

Germany, the home and test market, adds its own building blocks

- Energy Efficiency Act: efficiency, renewable-power and waste-heat requirements for data centres; an amendment is under way.
- JUPITER in Jülich: Europe's first exascale system (1.0 exaflops, 23,752 GH200 superchips), anchor of the EuroHPC AI Factory in Jülich.
- Telekom's Industrial AI Cloud in Munich: up to 10,000 NVIDIA Blackwell GPUs for industry, a €1bn investment.

RULES BY JURISDICTION · EUROPE AND THE AMERICAS

Where chips are made and capacity is built shapes your terms

Europe builds sovereign capacity and sets energy rules; the Netherlands supplies the lithography; the United States sets export rules for advanced chips and hosts the largest projects. Status October 2026.

JURISDICTION	RULE OR PROGRAMME	STATUS	WHAT IT MEANS FOR COMPUTE BUYERS
European Union	EuroHPC AI Factories and AI Gigafactories; AI Act GPAI duties; Energy Efficiency Directive	Call closes 12 Nov 2026; GPAI duties since Aug 2025	Public anchor demand and new sovereign capacity from 2028 at the earliest; data centres from 500 kW report energy indicators.
Germany	Energy Efficiency Act; JUPITER and EuroHPC AI Factory; Industrial AI Cloud	In force; amendment under way	Efficiency and renewable-power duties shape site and price terms; EU-resident GPU capacity from Telekom and Jülich.
Netherlands	Lithography supply (ASML)	Guidance Oct 2026	Planned 2027–28 supply will not fully meet demand; EUV and DUV capacity to rise 30% in 2027.
United Kingdom	Compute Roadmap; AI Growth Zones	Since Jul 2025	6 GW of AI-capable data-centre capacity needed by 2030; £2bn package.
United States	BIS rule on H200-class chips to China	Since 13 Jan 2026	Case-by-case licences, a cap of 50% of US volume and a 25% fee; Blackwell remains barred.
United States	Stargate US (OpenAI, Oracle, SoftBank)	About 0.3 GW operating, Apr 2026	US$500bn and more than 9 GW announced across seven sites; most sites target late 2028.

Asia makes the chips, the Gulf builds capacity

Taiwan and Korea control leading-edge foundry and memory; China builds its own accelerators; Japan and India fund national compute; the Gulf announces large AI campuses.

JURISDICTION	RULE OR PROGRAMME	STATUS	WHAT IT MEANS FOR COMPUTE BUYERS
China	Domestic accelerators; H200 purchases	No H200 shipped by mid-May 2026	Beijing paused orders; Huawei plans Ascend 950 in 2026, 960 in late 2027 and 970 in late 2028.
Taiwan	Leading-edge foundry (TSMC)	Q3 2026 results	Revenue up 41%; 3nm a third of wafer revenue; node and packaging allocation sets accelerator volumes.
South Korea	HBM memory (SK hynix, Samsung)	Shortage expected until 2030	SK hynix held 59% of HBM bit supply in 2025, Samsung 20%; memory prices feed into hardware costs.
Japan	State support for AI and chips; Rapidus 2nm	¥10tn through 2030; 2nm target 2027	Public funding for domestic chip production and AI compute.
India	IndiaAI Mission compute portal	Status Oct 2026	34,381 GPUs listed as capacity, subsidised at ₹100 per GPU-hour.
UAE and Saudi Arabia	Stargate UAE; Humain with Qualcomm	First phases from 2026	Announced: a 1 GW cluster with a first 200 MW in 2026, and 200 MW of Qualcomm AI200/AI250 racks.

Sources: company filings (NVIDIA, TSMC, ASML), IEA, EuroHPC, Bitkom, government programmes and policy reports (CNAS, Epoch AI). Status October 2026; export rules change often. Not legal advice.

Where the 21 providers are headquartered

16
North America

4
Europe, of which 1 in Germany

1
Middle East

Headquarters as of October 2026 from provider pages and public sources. Amazon owns AWS; Deutsche Telekom owns T-Systems; Verda was named DataCrunch until November 2025.

THE MARKET MAP

21 offerings for AI compute, from the chip to the cloud

The Good AI Prism 2026/27 places 21 offerings in Chips & Compute. Five are GPUs and general accelerators and five are GPU clouds and neoclouds that rent capacity. Four are hyperscalers' own AI chips and four are inference chips; two bring AI to the edge, and one links accelerators with light. As of October 2026, Qualcomm's AI200 and AI250 racks, Axelera's Europa and SambaNova's SN50 succeed the recorded flagships, and NVIDIA has licensed Groq's technology (Dec 2025) while GroqCloud continues.

Exhibit 3 · The Chips & Compute map by job

Number of offerings per job group; names alphabetical, not a ranking

5 GPUs and accelerators

AMD · Cerebras · Intel Gaudi · SambaNova RDU · Tenstorrent Blackhole

US export rules decide which chips may go to China; check licence and end-user terms

5 GPU clouds and neoclouds

CoreWeave · Industrial AI Cloud · Lambda · Nscale · Verda

Energy rules for data centres (EED reporting, Germany's Energy Efficiency Act) shape site, price and pass-through terms

4 Hyperscaler AI chips

AWS Inferentia · AWS Trainium · Google TPU · Maia 200

Used in the provider's own cloud regions; check region, residency and portability

4 Inference chips

Corsair · Etched · Groq · Qualcomm Cloud AI 100 Ultra

Check export, end-user and software-portability terms; inference is where volume grows

2 Edge AI chips

Hailo-10H · Metis AIPU

Embedded in devices: AI Act duties follow the product and its use, Annex I from 2 Aug 2028

1 Optical interconnect

Passage M1000

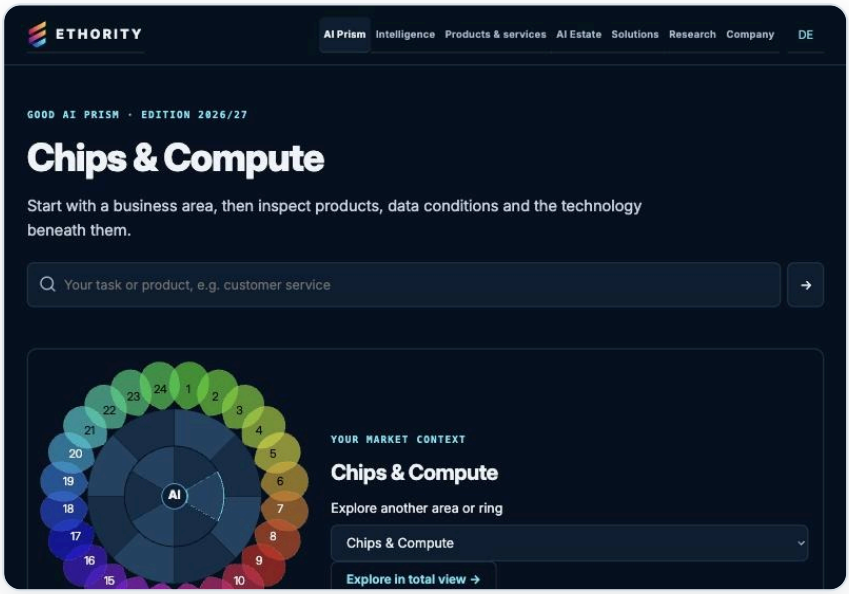
A component of accelerator systems; supply and terms follow the system it is built into

Good AI Prism 2026/27 study dataset · job groups: ETHORITY classification · Rules attach to chips, sites and uses: export rules follow the chip and its buyer, energy rules the data centre, and the AI Act the model and its use.

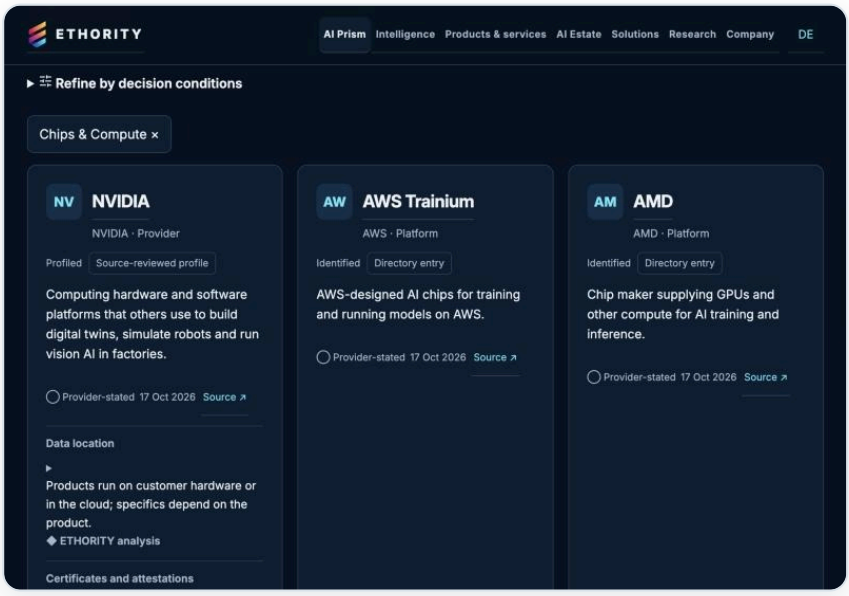
ONLINE VIEW

The Chips & Compute area in the Good AI Prism

The online Good AI Prism places each offering in its business area and shows the source and date behind every entry. The preview below shows the Chips & Compute area as it appears in October 2026. Select an area to see the offerings placed there, then open a profile to check the provider's own statement, its date and the decision facts recorded for it, such as pricing basis, regions and data hosting where documented.



Good AI Prism online preview: Chips & Compute selected, October 2026.



Entries with their source and date. This brief counts all 21 entries recorded for the edition.

NEEDS AND DECISIONS

What compute buyers need, and what to ask before a commitment

What the evidence shows

Capacity
ASML and SK hynix point to supply limits through 2027; memory may stay short until 2030.

Price
H100 hours at neoclouds cost about a third of hyperscaler on-demand rates, while inference prices per task keep falling (Epoch AI).

Energy
About 20% of planned data-centre projects risk grid delays; grid connections, turbines and transformers are bottlenecks (IEA).

Region
New EU gigafactory capacity arrives from 2028; until then EU-resident options are cloud regions, sovereign clouds, neoclouds and EuroHPC AI Factories.

Seven questions before you commit to compute

1. Which workload do we buy for: training, fine-tuning or inference, and how will volume grow over three years?
2. Which region and legal entity host the capacity, and does that meet our residency and sovereignty needs?
3. What is committed: capacity, delivery date, hardware generation and service levels, and what applies on delay?
4. How are price changes handled: reservation terms, memory or energy pass-through, escalator clauses?
5. What is the energy basis: power source, efficiency (PUE) and the reporting we need for our own disclosures?
6. How portable is our software if we change chip or cloud, and what does switching cost?
7. Do export and end-user terms limit where we may deploy, and who signs them off?

WORKFORCE PLAN

Plan the people side with the technology

When agendas change quickly, the people side breaks first: half-adopted tools, skills that lag the system, employees who were not involved, and decisions no one owns. Plan adoption, training, employee participation and accountable oversight as part of every staged commitment, not after it.

Severin L. Wilson, workforce lens in the Good AI Prism Market Study 2026/27

Workforce annex to the pilot charter

Affected roles	Which ML, platform, data-centre, procurement and finance roles the change touches, and how many people
Baseline workload	Utilisation, queue times and cost per training run or per million tokens today, measured before the move
Skills needed	Operating a new chip or cloud stack, porting models, planning capacity and cost; training plan and owner
Staff involvement	Works council and employee representatives where applicable
Oversight owner	The named person accountable for capacity, cost and energy commitments
Escalation	How teams report shortages, outages or cost overruns, and the response time
Capacity plan	Where freed engineering time goes: new models, optimisation or reskilling; who decides
Measured quality	Utilisation, cost and energy per output and incidents, tracked with the outcome measure

Booked capacity is not realised value. Count it when the compute has demonstrably gone into models, products or services in use.

NEXT STEPS

A 90-day start for compute decisions

01
Inventory
List every AI workload with its compute source, region, contract term and monthly cost in one register.

02
Forecast
Estimate training and inference demand for three years, with a low and a high case.

03
Involve
Bring in IT, procurement, finance, sustainability, legal and employee representatives early; in Germany, the works council.

04
Compare
Test at least two chip or cloud options on your own models for cost, speed, energy and portability.

05
Contract
Agree capacity, region, price adjustment, energy reporting and exit terms before you commit.

06
Decide
Scale only on measured cost per output, utilisation, energy and delivery.

ETHORITY recommendation, derived from the rules and evidence in this brief.

CONCLUSIONS

What the evidence adds up to

Compute is now a strategic input. The question is no longer which chip is fastest, but how to secure enough capacity in the right place on terms you can plan. Five conclusions for decisions now.

1 **Plan capacity, not just chips**

Evidence. ASML and SK hynix point to supply limits through 2027; memory may stay short until 2030.

What to do. Forecast demand three years ahead and secure capacity early, with fall-back options. · Where the market stands

2 **Buy on terms, not list prices**

Evidence. H100 hours at neoclouds cost about a third of hyperscaler on-demand rates; memory prices rose 15–18% in a quarter.

What to do. Compare total cost per output and negotiate reservation, price adjustment and exit terms. · Price, power and place

3 **Put energy into the contract**

Evidence. About 20% of planned data-centre projects risk grid delays; data-centre power use is set to double by 2030.

What to do. Ask for power source, efficiency basis and energy pass-through rules in writing. · Price, power and place

4 **Choose the region deliberately**

Evidence. New EU gigafactory capacity arrives from 2028; US export rules decide who may buy which chips.

What to do. Match each workload to a region and provider that meet residency, sovereignty and export terms. · Rules by jurisdiction

5 **Keep your software portable**

Evidence. Hyperscaler chips, AMD, inference chips and neoclouds widen the options; software is the main switching cost.

What to do. Test your models on at least two platforms before long commitments. · The market map

THE PROBLEM

AI is not short of ideas. It is short of compute it can plan, afford and explain.

Organisations must secure enough compute in the right region at a predictable price, with supply and energy risks they understand and can explain to their board.

THE PROBLEM	WHAT WE SEE IN THE EVIDENCE	WHAT IT COSTS
Scarce capacity	Leading-edge chips, memory and packaging are allocated, with long lead times.	**Delayed projects and idle teams.**
Volatile cost	GPU rental, memory and energy prices move differently and are hard to compare.	**Budget overruns and poor contract terms.**
Energy and grid	Grid connections and power supply limit where capacity can be built.	**Higher prices and delayed sites.**
Rules and sovereignty	Export controls, energy rules and data residency apply together.	**Restricted options and rework.**
Lock-in	Software tied to one chip or cloud makes switching expensive.	**Weaker negotiating position.**

The question behind every project

What should improve, which options are credible, what evidence is missing, what implementation requires, and who owns the outcome?

HOW ETHORITY HELPS

From compute demand to decisions you can defend

ETHORITY helps organisations plan, buy and govern AI compute: from independent market evidence to capacity decisions that can be explained to the board and outcomes that can be measured.

01
Discover
Find and qualify the opportunity: options, evidence, constraints and rules.

02
Create
Design and build the solution with your data, systems and people.

03
Run
Operate it responsibly, with owners, controls and support.

04
Improve
Measure against the objective and adjust.

05
Grow
Turn what was learned into new capability and value.

WHERE A PROJECT CAN BEGIN

AI compute case
One workload, such as model training, fine-tuning or a high-volume inference service, assessed against evidence, cost, energy and rules, ending in a decision: build, rent, redesign or stop.

Vendor and evidence due diligence
Good AI Prism research profiles and comparisons of the 21 offerings, with a structured check of performance, availability and price claims.

AI estate audit
A register of all AI workloads and compute sources, mapped to owners, regions, costs, energy use and overlaps.

Governance and contract readiness
Each workload mapped to export, energy, data-residency and AI Act requirements, with proportionate controls and contract terms.

AI literacy for technology teams
Role-based training for IT, procurement, finance and management on compute options, costs and risks.

From pilot to operation
A pilot charter with baseline, threshold and stop conditions, then building and running the solution on ETHORITY's AI Business Operating System.

Who we work with

CIOs and CTOs, IT and cloud procurement, data-centre operators, AI teams in industry, research institutions and public bodies, and investors in AI infrastructure.

Start with one question. Talk to us: sten@ethority.ai · severin@ethority.ai · ethority.ai/ai-prism-study

ABOUT ETHORITY

From understanding markets to building AI capability

ETHORITY connects market intelligence, research and practical delivery to help organisations put AI to work. It is built as an AI1st organisation: humans and AI agents work on one AI Business Operating System, with the research, engineering and operations to build and run AI after the pilot. Its roots lie in the digital intelligence business Sten Franke founded in 2005, which combined monitoring, research and advisory work to help organisations understand changing markets.

The Good AI Prism is the entry point: research profiles, documented relationships and comparisons help users investigate their options. Advisory, governance and transformation work connect those options with an organisation's objectives, data, systems, people and operating conditions.

SELECTED REFERENCES AND PARTNERS SINCE 2005

Bosch · Bayer · BMW · Siemens · Mercedes-Benz · Coca-Cola · Deutsche Bank · T-Mobile · Telefónica o2 · DHL · EDEKA · AIDA Cruises

Selected historical references and partners from the group's market-intelligence and advisory heritage, in projects or through former group companies.

METHOD

Market figures: Good AI Prism 2026/27 study dataset, Chips & Compute area, 21 entries, each with a dated source from the provider's own pages (October 2026). Job groups and rule relevance: ETHORITY classification; notes on newer flagships are dated October 2026. Jurisdictions and market data: status 22 October 2026 from company filings, the IEA, EuroHPC, Bitkom and policy reports; later quarterly results are not included. External figures as cited and not combined. Not legal, investment or procurement advice.

MAIN SOURCES

- NVIDIA, financial results for Q2 fiscal 2027, 26 Aug 2026; NVIDIA GTC 2026 roadmap (DCD, Mar 2026)
- TSMC, Q3 2026 results (Form 6-K), 15 Oct 2026; ASML, Q3 2026 results (Form 6-K), 14 Oct 2026
- SK hynix CEO on memory supply (Reuters via Tom's Hardware), 11 Jul 2026; TrendForce on HBM share (Mar 2026) and DRAM contract prices (Jul 2026)
- IEA, Energy and AI, 2025; IEA, Key Questions on Energy and AI, 16 Apr 2026
- Bitkom, Rechenzentren in Deutschland, 10 Nov 2025; TOP500 (JUPITER), Nov 2025
- EuroHPC JU, AI Gigafactories call, 30 Jul 2026; Deutsche Telekom Industrial AI Cloud (Telecoms.com, DCD, Nov 2025)
- CNAS, H200 export policy, Jan 2026; Epoch AI, Stargate sites (Apr 2026) and inference price trends (Mar 2025); Silicon Data H100 index (Implicator, Jul 2026)
- Microsoft, Maia 200, 26 Jan 2026; AWS Trainium3 (DCD, Dec 2025); Google Ironwood (TechWire Asia, Nov 2025); Jon Peddie Research on Intel, Oct 2026
- UK Compute Roadmap (Jul 2025); Japan AI and chip package (Reuters, Nov 2024); IndiaAI Mission (Oct 2026); Stargate UAE (Oct 2025); Humain and Qualcomm (Oct 2025); Huawei Ascend roadmap (Caixin, Sep 2025)
- Regulation (EU) 2024/1689 (AI Act) and 2026/1744 (Digital Omnibus); Delegated Regulation (EU) 2024/1364 (EED data centres); Energieeffizienzgesetz (EnEfG)
- Good AI Prism Market Study 2026/27 and study dataset (ETHORITY, October 2026)

ETHORITY
AI BUSINESS OPERATING SYSTEM



FROM BUSINESS AMBITION TO AI THAT WORKS

Put AI to work, with evidence.

ETHORITY helps organisations choose, govern and prove AI: from independent market evidence to decisions that can be defended and outcomes that can be measured.

CONTACT



Sten Franke
Founder, Principal Architect & Lead Creator
sten@ethority.ai
linkedin.com/in/stenfranke



Severin L. Wilson
Chief Business & Transformation Officer
severin@ethority.ai
linkedin.com/in/severinwilson

STUDY REPOSITORY AND UPDATES

ethority.ai/ai-prism-study

ethority.ai

THE ETHORITY STUDY SUITE

01	**Put AI to Work: The Good AI Prism Global Market Study 2026/27**	Global edition · EN and DE
02	**Put AI to Work: The Good AI Prism Market Study 2026/27**	Germany–US focus · EN and DE
03	**Good AI Prism sector, function and technology briefs 2026/27**	HR · Health · Financial Services · Public Sector · Education · Energy · Mobility · Finance & Accounting · AI Models · Chips & Compute · Robotics · EN and DE **YOU ARE READING**
04	**AI for Care in Ageing Societies**	Study 2026/27 · EN, DE, US and Global editions
05	**AI in Pharma, Biotech and Longevity**	Study 2026/27 · Global EN and DE
06	**The Social Media Prisma 2009–2025**	Historic paper · EN and DE

© 2026 ETHORITY & STEN.AI LLC · San Francisco HQ · European operations: Hamburg, Germany · ethority.ai
AI Prism for Chips & Compute · Technology brief 2026/27 · Global edition 1.0, 22 October 2026 · All rights reserved. Exhibits may be quoted with source attribution.