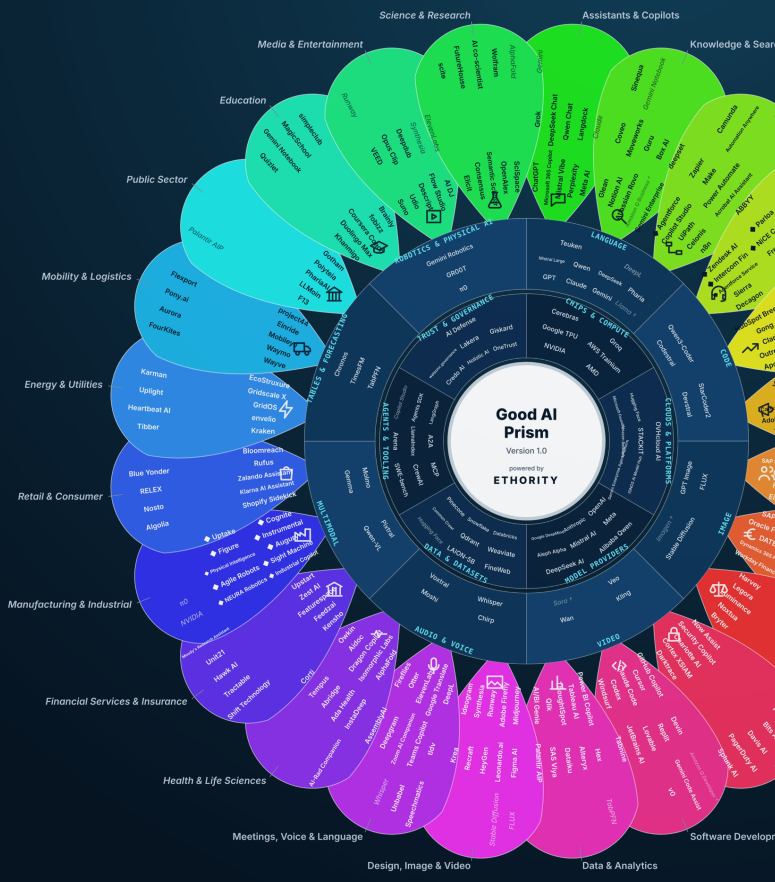
ETHORITY
AI BUSINESS OPERATING SYSTEM



GOOD AI PRISM · TECHNOLOGY BRIEF 2026/27 · AI MODELS · GLOBAL EDITION

Choose the right AI models

Who builds the models behind enterprise AI, what they cost and how long they last, what the market offers, and the rules by jurisdiction.

44
AI model makers, model families and products mapped in the Good AI Prism

2 Aug 2026
EU enforcement of duties for general-purpose AI models begins

81%
of large enterprises use three or more model families (a16z, Jan 2026)

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

AT A GLANCE

Models are the engine room of enterprise AI

A few labs set the frontier, open-weight families from China and Europe follow closely, and small and sovereign models fill the gaps. For buyers, the decision has moved from one vendor to a portfolio of models.

This brief applies the Good AI Prism Market Study 2026/27 to the models layer: the foundation and language models that applications are built on. It combines the study's map of 44 entries with current evidence on capability, price, lifecycle, adoption 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 choosing a model and with how ETHORITY helps.

$285.9bn
US private AI investment in 2025; Europe $20.9bn, China $12.4bn (Stanford AI Index 2026)

2.7%
Arena gap between the best US and the best Chinese model, March 2026 (AI Index 2026)

93 of 102
notable models of 2025 came from industry; 50 from the US, 30 from China (AI Index 2026)

40%
of enterprise LLM API use goes to Anthropic, 27% to OpenAI, 21% to Google (Menlo survey, Dec 2025)

100x
price range of current API models: $0.10 to $10 per million input tokens (provider price lists, Oct 2026)

21
providers have signed the EU Code of Practice for general-purpose AI (European Commission)

About this brief

Market figures come from the Good AI Prism 2026/27 study dataset: every entry is recorded with a dated source. Job groups are ETHORITY's classification. External figures name their source and date; vendor benchmarks are self-reported and not comparable across vendors; survey populations differ and are not combined. This brief is not legal advice.

WHERE MODELS STAND

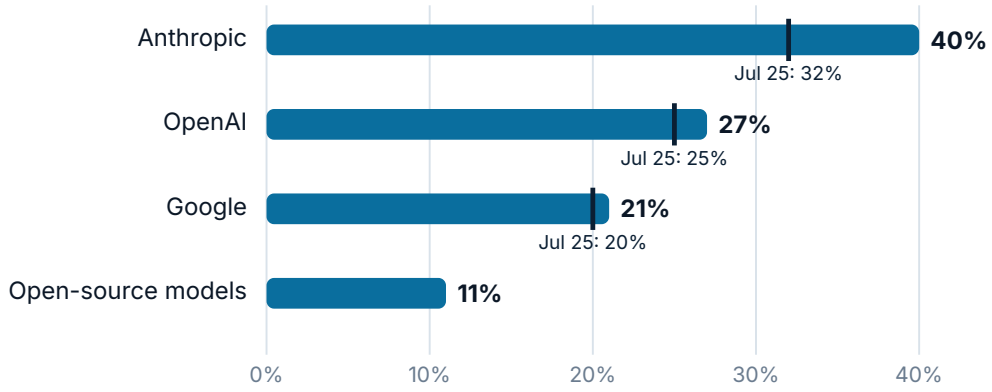
Three flagships at the top, a portfolio below

The frontier now has three closed flagships: OpenAI's GPT-6 Astra, Anthropic's Claude Fable 5.1 and Google's Gemini 4 Argon, all released in September and October 2026. Below them, providers sell workhorse and small tiers: list prices range from $10 to $0.10 per million input tokens, and Anthropic cut its Opus price by 20% between versions.

Enterprises buy accordingly. 81% of large companies use three or more model families (a16z), and the enterprise API market concentrates on three providers. Industry produced 93 of the 102 notable models of 2025, and global AI compute reached about 17 million H100-equivalents (AI Index 2026).

Exhibit 1 · Who enterprises buy model access from

Share of enterprise LLM API use by provider, December 2025; marker: mid-2025



Menlo Ventures, State of Generative AI in the Enterprise, 9 Dec 2025 · n=495; mid-year report, 31 Jul 2025 · n=150

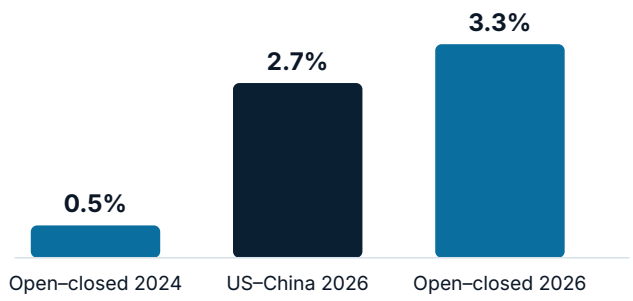
OPEN AND CLOSED

Open weights stay close, but the gap widens again

Open-weight models are a real option. Chinese labs publish very large open models with one-million-token context, Mistral AI previewed Mistral Large 4 with about one trillion parameters and weights announced for October 2026, and Switzerland's Apertus 1.5 is fully open, from data to code. Meta, by contrast, moved its new flagship Muse Spark to closed weights.

At the very top, closed models lead again: the gap between the best closed and the best open model grew from 0.5% to 3.3% on Arena, and six of the top ten models are closed (AI Index 2026). Licences also tighten, with some large open models adding revenue thresholds.

Exhibit 2 · The gaps at the top of the leaderboard
Arena score difference between the top models, percent



Stanford HAI, AI Index 2026, 13 Apr 2026 · Arena data to March 2026

81%

of large enterprises use three or more model families, up from 68% a year earlier (a16z CIO survey, n=100, Jan 2026)

11%

share of open-source models in enterprise LLM use, down from 19% (Menlo Ventures, n=495, Dec 2025)

THE RULES

Europe regulates the models themselves

ETHORITY reads the rules by jurisdiction. Europe is the hub market: the AI Act sets duties for providers of general-purpose AI (GPAI) models, with more for models with systemic risk, and duties for deployers by use. Germany, the home and test market, applies the AI Act directly and is building a sovereign model stack. The next page maps other major jurisdictions.

European Union: the hub market

10 Jul 2025	CODE	Final GPAI Code of Practice published; 21 providers have signed it, and xAI its safety and security chapter.
2 Aug 2025	GPAI	Provider duties apply: technical documentation, information for downstream providers, a copyright policy and a summary of training content. Systemic-risk models add evaluation, incident reporting and cybersecurity.
2 Aug 2026	ENFORCEMENT	The Commission can enforce GPAI duties, including fines; the AI Office gains direct powers over systems built on a provider's own GPAI model. Art. 50 transparency applies.
2 Aug 2027	LEGACY	GPAI models placed on the market before 2 Aug 2025 must comply. High-risk duties for deployers under Annex III apply from 2 Dec 2027.

AI Act dates as amended by the Digital Omnibus on AI, Regulation (EU) 2026/1744, in force since 27 July 2026; GPAI obligations are unchanged. Systemic risk is presumed above 10^{25} FLOP of training compute or set by Commission designation.

Germany, the home and test market, builds a sovereign stack

- Cohere announced on 24 Apr 2026 that it will take over Heidelberg-based Aleph Alpha, subject to approvals; the combined company is expected to run on Schwarz Group's STACKIT cloud.
- Aleph Alpha and Black Forest Labs are among the signatories of the GPAI Code of Practice.
- German entries in the map include Pharia, Teuken from the OpenGPT-X project, DeepL and Jina AI search models.

RULES BY JURISDICTION

Where a model is built and used decides which rules apply

Rules for models combine AI law, transparency and labelling duties. The EU regulates general-purpose models directly; California sets duties for the largest frontier developers; China requires labels on generated content. 47 countries have active AI legislation or policy processes (AI Index 2026). Status October 2026.

JURISDICTION	INSTRUMENT	APPLIES	WHAT IT MEANS FOR MODELS
European Union	AI Act Chapter V (GPAI); Code of Practice; Reg. (EU) 2026/1744	GPAI since 2 Aug 2025; enforcement from 2 Aug 2026; earlier models by 2 Aug 2027	Providers document models and training content; systemic-risk models are evaluated and report incidents; deployers build on that documentation.
Germany	AI Act applies directly	As in the EU	A sovereign stack forms: Cohere with Aleph Alpha on STACKIT; Pharia, Teuken and DeepL in the map.
United States	Executive Order 14365; National Policy Framework for AI	EO since 11 Dec 2025; framework 20 Mar 2026	Federal policy seeks one national framework and challenges state AI laws; the framework is non-binding and Congress has not enacted preemption.
California	SB 53, Transparency in Frontier AI Act	Since 1 Jan 2026	For models trained above 10^{26} FLOP: critical incidents reported within 15 days; large developers (revenue above $500m) publish a frontier AI framework; up to $1m per violation.
China	Measures for Labeling AI-Generated Synthetic Content	Since 1 Sep 2025	Explicit and implicit labels on generated text, images, audio and video; app stores check labelling.

Sources: EUR-Lex, European Commission, California legislature, White House and CAC, with law-firm briefings. Status October 2026. Not legal advice.

Where the 44 entries originate

15
Europe, of which 5 in Germany

14
North America

13
Asia, of which 7 in China

2
Middle East

Home base of the developer as of October 2026. Google owns Google DeepMind, Elastic owns Jina AI and AMD owns Silo AI; Cohere announced its takeover of Aleph Alpha in April 2026.

THE MARKET MAP

44 entries, mapped by the job they do

The Good AI Prism 2026/27 places 44 entries in models and language: providers, model families and products, so OpenAI appears next to GPT. Six belong to US frontier labs, seven each to Chinese and European makers, twelve are sovereign and language models, five enterprise families, five small models, and two serve translation and search.

Exhibit 3 · The AI models map by job

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

6 US frontier labs and flagships

Anthropic · Google DeepMind · GPT · Llama · Meta · OpenAI

GPAI provider duties; systemic-risk duties above 10^{25} FLOP or when designated

7 Chinese model makers

Alibaba Qwen · DeepSeek · DeepSeek AI · GLM · Kimi K2 · MiniMax · Qwen

GPAI duties apply when offered in the EU; non-EU providers appoint an EU representative

7 European model makers

Aleph Alpha · Apertus · LightOn · Mistral AI · Mistral Large · Pharia · Pleias

GPAI provider duties; Aleph Alpha, Mistral AI and Pleias signed the Code of Practice

12 Sovereign and language models

Bielik · EuroLLM · Falcon · HyperCLOVA X · Lucie · PLaMo · Poro · Sarvam (Sarvam-105B) · Solar · Solar LLM · Teuken · tsuzumi

GPAI provider duties; open-source release eases documentation duties

5 Enterprise and cloud families

Amazon Nova · Cohere Command · IBM Granite · Jamba · NVIDIA Nemotron

GPAI provider duties; Amazon, Cohere and IBM signed the Code of Practice

5 Small and efficient models

CompactifAI · Liquid Foundation Models (LFM) · OLMo · Phi · Trinity

GPAI duties where general-purpose; open-source relief may apply

2 Translation and search

DeepL · Jina AI Search Foundation Models

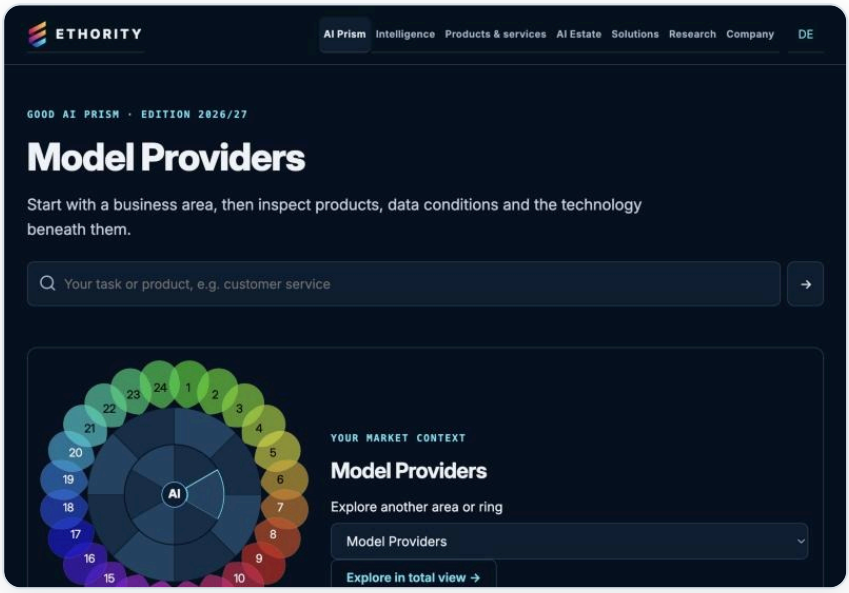
Deployer duties by use: Art. 50 transparency; Annex III only in high-risk uses

Good AI Prism 2026/27 study dataset · job groups: ETHORITY classification · The AI Act regulates models and uses separately: providers carry GPAI duties, deployers duties for their use.

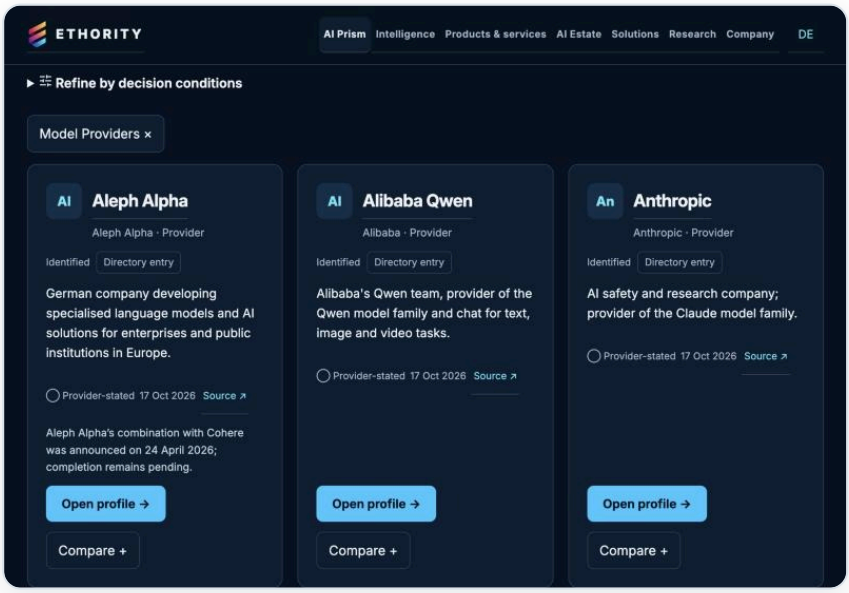
ONLINE VIEW

The AI models area in the Good AI Prism

The online Good AI Prism places each entry in its business area and shows the source and date behind every entry. The preview below shows the models and language area as it appears in October 2026. Select an area to see the entries 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, data hosting and licence where documented.



Good AI Prism online preview: AI models and language selected, October 2026.



Entries with their source and date. This brief counts all 44 entries recorded for the edition.

NEEDS AND DECISIONS

What model buyers need, and what to ask before choosing

What the evidence shows

Choice
81% of large enterprises work with three or more model families (a16z, Jan 2026).

Lifecycle
OpenAI gives at least six months' notice before retiring a model, Anthropic at least 60 days; Anthropic retires Sonnet 4.5 about 14 months after release.

Cost
List prices span 100x, from $0.10 to $10 per million input tokens (Oct 2026).

Rules
GPAI duties are enforceable in the EU from Aug 2026; high-risk duties for deployers follow from Dec 2027.

Seven questions before you commit to a model

1. Which tasks need a frontier model, and which run well on a smaller or open-weight one?
2. Does the provider supply the GPAI documentation we need, and has it signed the Code of Practice?
3. Where are data processed and retained, per model rather than per vendor?
4. What does the licence allow: commercial use, revenue thresholds, fine-tuning?
5. How much notice do we get before retirement, and can our tests prove a replacement?
6. Can we route between models through a gateway to limit lock-in and cost?
7. Which measures decide: quality on our own tasks, cost per task, latency and incidents?

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 developer, analyst, service and content roles the model touches, and how many people
Baseline workload	Handling times, backlogs and error rates today, measured before the pilot
Skills needed	Prompting, evaluating and reviewing outputs, and knowing when not to use the model; training plan and owner
Staff involvement	Works council and employee representatives where applicable
Oversight owner	The named person accountable for each model in production
Escalation and appeal	How staff and customers can challenge an output, and the response time
Capacity plan	Where returned time goes: quality, customer contact, new work or reskilling; who decides
Measured quality	Error rate, incorrect outputs, complaints and audit findings, tracked with the outcome measure

Saved time is not realised value. Count it when returned time has demonstrably gone into better outcomes, quality or new work.

NEXT STEPS

A 90-day start for a model portfolio

01
Inventory
List every model in use, including APIs, models inside software and tools staff bring themselves.

02
Classify
Record provider, version, retirement date, licence, data residency and your AI Act role for each use.

03
Evaluate
Build a test set from your own tasks and compare two or three models on quality, cost and latency.

04
Involve
Bring in IT security, data protection, procurement and employee representatives; in Germany, the works council.

05
Pilot
Run one use behind a gateway with a baseline, a fallback model and the workforce annex.

06
Decide
Scale on measured quality, cost per task and incidents, and plan the next model switch.

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

CONCLUSIONS

What the evidence adds up to

Models improve and change quickly, and buyers already use several. The task is no longer to find the one best model, but to choose, govern and replace models with evidence. Five conclusions for decisions now.

1 **Manage models as a portfolio**

Evidence. 81% of large enterprises use three or more model families; three providers lead enterprise API use.

What to do. Choose by task and tier, not by vendor, and keep a register of every model in use. · Where models stand

2 **Plan for short lifecycles**

Evidence. Providers give between 60 days' and six months' notice; some models retire about 14 months after release.

What to do. Keep evaluation and regression tests ready so a model can be replaced on schedule. · Needs and decisions

3 **Use open weights deliberately**

Evidence. Open models stay within a few percent of the frontier, while licences add revenue thresholds.

What to do. Use open weights where control or residency matter, and read each licence. · Open and closed

4 **Ask providers for GPAI evidence**

Evidence. GPAI duties apply since Aug 2025 and are enforceable from Aug 2026; 21 providers signed the Code.

What to do. Require documentation, training-content summaries and incident contacts in procurement. · The rules

5 **Build sovereign options in Europe**

Evidence. Mistral Large 4 offers European operation; Cohere and Aleph Alpha plan to run on STACKIT.

What to do. Keep at least one European-operated option for sensitive data and critical processes. · Germany; market map

THE PROBLEM

The market is not short of models. It is short of choices buyers can explain and sustain.

Organisations must pick models that fit their tasks, prove that they meet the rules, and keep them running as providers release and retire versions.

THE PROBLEM	WHAT WE SEE IN THE EVIDENCE	WHAT IT COSTS
Many models, unclear evidence	44 entries in this map alone; vendor benchmarks use their own harnesses and are not comparable.	**Wrong choices and overspending on the top tier.**
Fast release and retirement	Models and API features retire within months, often with short notice.	**Broken applications and repeated rework.**
Licences and residency	Open-weight licences and data retention differ by model, not only by vendor.	**Legal exposure and data in the wrong place.**
Rules on two levels	Providers carry GPAI duties; deployers carry duties for each use.	**Gaps in documentation and accountability.**
People and skills	Teams must prompt, evaluate and review outputs and know when not to rely on them.	**Tools unused or trusted too much.**

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 model choice to decisions you can defend

ETHORITY helps organisations choose, govern and prove AI models: from independent market evidence to model decisions that can be explained to boards, auditors and regulators, 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 opportunity case
One use assessed against evidence, cost and rules, ending in a decision: pilot, redesign, expand or stop.

Model and provider due diligence
Good AI Prism research profiles and comparisons of the 44 entries, with a structured check of documentation, licence and data terms.

AI estate audit
A register of all models and AI uses, including shadow AI, mapped to owners, versions, retirement dates, costs and overlaps.

Governance and AI Act readiness
Each use mapped to provider and deployer duties under the AI Act, with proportionate controls.

AI literacy for teams
Role-based training on using, evaluating and overseeing models, with named owners.

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

Enterprises and their IT, data and procurement teams, software makers building on models, public bodies, investors and associations.

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, models and language areas, 44 entries, each with a dated source (October 2026). Job groups and AI Act relevance: ETHORITY classification. External figures from providers, regulators, independent evaluators and surveys, as cited and not combined; vendor benchmarks and press-reported revenues are not used. Jurisdictions: status October 2026. Not legal advice.

MAIN SOURCES

- Stanford HAI, AI Index Report 2026, 13 Apr 2026 (technical performance, R&D, investment)
- OpenAI, GPT-6 Astra and GPT-6.1 Sol announcements and deprecations page; Anthropic, models overview and model deprecations; Google Cloud, Gemini 4 Argon availability, 13 Oct 2026
- Mistral AI, Mistral Large 4, 6 Oct 2026; CSCS, ETH Zurich and EPFL, Apertus 1.5, 24 Jul 2026
- Menlo Ventures, State of Generative AI in the Enterprise, 9 Dec 2025, and mid-year report, 31 Jul 2025
- a16z, enterprise CIO survey, 30 Jan 2026
- TechCrunch, Cohere and Aleph Alpha, 25 Apr 2026; The Register, Meta Muse Spark, 8 Apr 2026
- Regulation (EU) 2024/1689 (AI Act), Chapter V; Regulation (EU) 2026/1744; European Commission, GPAI Code of Practice and signatories
- California SB 53 (TFAIA); Executive Order 14365; White House National Policy Framework for AI, 20 Mar 2026
- CAC, MIIT, MPS, NRTA, Measures for Labeling AI-Generated Synthetic Content, 14 Mar 2025
- 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 AI Models · Technology brief 2026/27 · Global edition 1.0, 22 October 2026 · All rights reserved. Exhibits may be quoted with source attribution.