skema THINK TANK | PUBLIKA

POLICY PAPER

UNCERTAINITIES

EUROPE'S OPPORTUNITY IN ALTERNATIVE AI

September 2026

Why Europe should compete less on brute-force frontier scale and more on deployable intelligence, sovereign data, and institutionally trusted AI

Engin **IYIDOGAN**

SKEMA PUBLIKA

SKEMA Publika is SKEMA Business School's international think tank. It analyses major social, economic, technological and geopolitical changes to inform public and private decisionmaking. Drawing on the school's research and on scientifically validated external contributions, the think tank fuels public debate and issues recommendations for national and international decision-makers.

SKEMA Publika takes an interdisciplinary and international approach, enriched by SKEMA's global network of campuses and a community of experts from the academic and professional worlds. This international dimension is not simply a network, but a way of thinking. SKEMA Publika therefore connects local dynamics with global transformations to provide a decentralised and multipolar perspective on the major challenges of our time, rather than one centred on a single national lens.

Cover: Blue and yellow flag. François Genon. 2019. Unsplash.com

SKEMA Publika
SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Phone. : +33.1.71.13.39.32
Email : publika@skema.edu
Website : www.publika.skema.edu

COLLECTION

"Exploring the grey areas of geopolitics, digital technology and global risks."

The publications in the 'Uncertainties' series analyse tensions, emerging risks and dynamics of instability globally.

EDITORS

Frédérique Vidal is Director of Development at SKEMA Publika and Director of Strategy and Scientific Impact at SKEMA Business School. A full professor of Biology, she was president of the University of Nice Sophia Antipolis from 2012 to 2017, and subsequently served as minister of higher education, research and innovation in the Philippe and Castex governments from 2017 to 2022. She served as special advisor to the President of EFMD and is currently also the Permanent Representative of the Principality of Monaco to the United Nations Environment Programme and the Whaling Commission.

Sean Scull, Think Tank Project Manager, is a doctoral candidate in Information and Communication Science at Université Paul Valéry – Montpellier III. He holds a degree in Political Science with a specialisation in International Relations from the University of Gothenburg, and a Master's degree in International Politics with a focus on Anglophone politics from the University of Toulon. Sean has lived and worked in Sweden and the United States.

AUTHORS

Engin Iyidogan is a finance professor at SKEMA Business School. He specialises in AI in finance, the crypto economy and the future of money. He obtained his PhD in Finance from Imperial College London.

SUMMARY

In order to develop its Artificial intelligence (AI) champions, Europe should not judge its AI prospects only by whether it can match the largest American or Chinese laboratories in the frontier-model race. A more useful economic question is where Europe could build an advantage in deployable AI. The European Commission's 2025–2026 agenda points towards one answer: AI Factories and proposed AI Gigafactories, the AI Continent Action Plan, the Data Union Strategy, sectoral deployment programmes and the AI Act's governance framework. These initiatives could support a European strategy centred on adoption and deployment, even while public debate remains focused on frontier models. Alternative AI, as used here, is not a rejection of frontier research or an argument for regulatory isolation. It describes a practical combination of multimodal systems, small and medium language models, open-weight models, retrieval-augmented generation, edge and near-edge deployment, model portfolios, and shared data spaces. These approaches fit many of Europe's industrial and regulated sectors. They also reflect the Commission's growing emphasis on compute access, sectoral data and broad adoption rather than benchmark performance alone.

The 2026 technical landscape strengthens the case for this approach. Leading providers now offer model families for different cost, latency and capability requirements. Multimodality, lightweight variants, open weights and on-device deployment are established product categories. Firm adoption is spreading quickly, but evidence of realised gains remains limited and uneven. In surveys of senior executives across the United States, United Kingdom, Germany and Australia, most respondents report no employment or productivity effect in their own firms so far, although they expect larger effects within three years. A Danish study likewise finds no detectable average effect on recorded hours or earnings after two years, while documenting changes in tasks and occupational transitions. These findings do not show that AI has no value; they show that value depends on organisational redesign, workflow fit and governance.

The adoption gap also matters. Workers in the United States use generative AI at work more frequently and intensively than workers in several large European economies. According to the ECB Survey on the Access to Finance of Enterprises, in the euro area 38% of firms describe their adoption as

advanced, while 33% remain infrequent or experimental users[1]. Europe's challenge is therefore not only frontier capability but the slower diffusion of AI through enterprises and public administrations.

Small, multimodal, decentralised and retrieval-grounded systems can reduce latency and infrastructure cost, limit unnecessary cloud transfers, and offer greater local control. These features are relevant in regulated and fragmented markets. The governance question extends beyond large centralised models to distributed systems, orchestration layers and cross-border provenance; it should be addressed in proportion to demonstrated capabilities and risks. Europe's opening is therefore narrower, and more practical than direct imitation of the frontier laboratories. It lies in systems that are auditable, responsive, grounded in domain knowledge, and acceptable to the institutions that use them. Europe could build an advantage here if investment in infrastructure is matched by deployment capacity.

In this policy paper, we present **four policy priorities.** Europe should: (1) turn AI Factories into deployment institutions linked to sectoral data and evaluation capacity; (2) use public procurement to create demand for auditable, portable and interoperable AI; (3) help enterprises move from single-model purchases to governed model portfolios; and (4) build shared validation infrastructure in high-trust sectors.

[1] Ferrando, A., Lamboglia, S., Rariga, J., & Schmidt, M. (2026). Adopting and investing in AI : evidence from euro area firms in the SAFE. European Central Bank. https://www.ecb.europa.eu/press/economic-bulletin/focus/2026/html/ecb.ebbox202602_06~c9f485c478.en.html

TABLE OF CONTENTS

INTRODUCTION

According to the non-profit *Rest of World,* the United States (US) and China dominate the AI market by attracting between 70% and 80% of global investment in AI[2]. The *Boston Consulting Group* makes the same observation. On the one hand, Washington holds a lead in frontier AI models, talent and capital deployment. On the other hand, Beijing stands out with its cost-optimised models and by quickly adopting this new technology across society[3]. The notable absentee seems to be the Old Continent as it has no company that seems able to seriously compete with established tech giants such as Google or Microsoft, but also with prominent AI newcomers such as OpenAI or NVIDIA.

Indeed, Europe faces structural disadvantages in the AI competition with the US and China, especially in the brute-force scale AI playbook. Firstly, European venture capital can't match the investment capacity the American funds and the Chinese state. Secondly, Europe lacks energy and enforces stricter emission targets which make energy-hungry data centers difficult to implement[4]. Thirdly as we highlight it in the SKEMA Publika report ***Artificial intelligence and European sovereignty*** , the European market is fragmented which makes it difficult to implement a coherent strategy and compete globally.

The term alternative AI is used here for a policy and technical shift already visible in Europe. The AI Continent Action Plan links computing infrastructure, data access, sectoral uptake and talent; the Data Union Strategy treats data availability as a condition for competitiveness; and EuroHPC describes AI Factories as ecosystems combining compute, data,

[2] Chandran, R., & Chandran, R. (2026). The Great AI Divide : Navigating U.S. and Chinese dominance. Rest Of World. https://restofworld.org/2026/ai-divide-america-china-world/

[3] Lang, N. et al. (2026). The Great Divide: How the US and China Are Splitting the AI World. Boston Consulting Group. https://www.bcg.com/publications/2026/us-and-china-ai-strategy-causing-global-ai-divide

[4] AI Office publishes frontier AI expert findings on EU competitiveness, sovereignty and security. (2026). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/library/ai-office-publishes-frontier-ai-expert-findings-eu-competitiveness-sovereignty-and-security

support and skills.[5] These initiatives do not yet amount to a settled industrial strategy, but they provide several of its building blocks.

Firstly, multimodal AI can process more than one type of input, including text, images and audio. This is relevant because a lot of enterprise data appears in scans, product images, audio and sensor-linked workflows, not only in text. No single parameter threshold defines a small language model (SLM); one recent survey covers models from roughly 100 million to 5 billion parameters.[6] For this paper, an SLM is a model designed to require materially less computing, latency or hosting capacity than frontier-scale systems, often for specialised or local deployment.

Secondly, deployment architecture matters as much as size. Edge AI processes data on local devices or nearby nodes rather than relying on a continuous connection to cloud infrastructure. Decentralised approaches extend that logic across training or inference systems, although much of the literature remains exploratory.[7] For Europe's industrial sites, hospitals and public administrations, the practical question is which data and workloads need to leave the local environment.

Thirdly, providers increasingly offer model families for different capability, latency and cost requirements.[8] This paper uses model portfolio to mean a deliberately routed organisational stack: a larger model for infrequent complex tasks; a faster model for routine work; a multimodal model for document and vision workflows; a local model for sensitive data; and retrieval layers that keep domain knowledge current. The finance analogy is useful: the objective is to allocate across performance, cost, latency and governance constraints, not to buy the same model for every task.

Fourthly, retrieval-augmented generation (RAG) connects a model at inference time to external knowledge stores, internal documentation or authoritative databases. This is especially relevant in sectors where current,

[5] European Commission, *AI Continent Action Plan* (2025); European Commission, *European Data Union Strategy* (2026); EuroHPC Joint Undertaking, *AI Factories* (2026).

[6] Lu et al., *Small Language Models: Survey, Measurements, and Insights*, arXiv:2409.15790, version 3 (2025).

[7] Dong et al., *Beyond a Single AI Cluster: A Survey of Decentralized LLM Training*, arXiv:2503.11023 (2025), survey preprint.

[8] Official model documentation: OpenAI, *Models*; Anthropic, *Models overview*; Google, *Gemini models*; Mistral AI, *Models overview* (accessed 26 August 2026).

proprietary or regulated information matters more than broad pretraining. Europe's opportunity is therefore less about reproducing the largest training runs than about using trusted institutional knowledge effectively.

EU law governs general-purpose AI models and systemic-risk obligations; it does not define AGI as a regulatory category.[9] AGI remains a contested strategic concept. This paper treats it as a horizon issue and focuses on governable properties of deployed systems rather than assumed thresholds or timelines. The objective is not a single laboratory, benchmark, or sovereign chatbot, but an interoperable ecosystem for trustworthy deployment. In other words, we argue that Europe should not exclusively compete in the race for massive frontier AI models but focus on alternative AI systems that are tailored to the local needs. European businesses do not need a powerful brute-force scale AI but alternative solutions that are less costly and better meet the demand of the European ecosystem, whether it is agriculture, services, finance, industry or the public sector.

[9] European Commission, *Regulatory framework on AI* (updated 2026).

I. EUROPE'S ALTERNATIVE AI MOMENT

A. TECHNICAL FRONTIER AND WHY SMALL MATTERS

Model selection now involves several dimensions such as multimodality, reasoning controls, context length, tool use, deployment flexibility, latency, and cost. Providers differentiate flagship and lighter models to match these requirements. In this context, scale remains important, but deployment of architecture has become part of the technical and commercial frontier.

Open-weight models broaden the options. Meta's Llama 4 Scout combines native multimodality with a mixture-of-experts design; Google's Gemma family covers several sizes and local deployment settings; Mistral offers open models from a European base; IBM publishes Granite models with provenance features; and Hugging Face's SmolLM3 is a 3B multilingual model.[10]

We can identify three developments that are particularly relevant. Firstly, Multimodality is becoming normal, secondly inference efficiency is now a strategic concern, and thirdly, small-model capabilities continue to improve. Reasoning controls, tool-use tiers, and open-weight variants all affect the cost and governance of a system. They should be evaluated as deployment choices rather than treated as secondary engineering details.

However, frontier models remain important, particularly in science, coding and open-ended reasoning. For many other enterprise and public-sector tasks, the practical question is which combination of models meets the required quality at the lowest acceptable cost, delay and governance risk. A firm with strong routing, evaluation and local-deployment capabilities may create more value than one that buys premium remote capacity for every workflow. Europe's strengths in manufacturing, healthcare and public-service delivery provide promising settings for this portfolio approach.

[10] Official model documentation: Meta, *Llama 4*; Google, *Gemma model card*; Mistral AI, *Mistral Small 4*; IBM, *Granite 4.1*; Hugging Face, *SmolLM3* (accessed 26 August 2026).

B. ENTERPRISE AND PUBLIC-SECTOR ECONOMICS

The organisational evidence is more cautious than many headline claims. In a 2026 NBER working paper based on surveys of senior executives, Yotzov et al. report that 69% of firms across the US, UK, Germany and Australia use some AI technology.[11] More than 90% of respondents reported no employment effect in their own firms to date, and 89% reported no labour-productivity effect. These are managers' estimates rather than direct productivity measurements. In Denmark, Humlum and Vestergaard find no detectable average effect on recorded hours or earnings large enough to exceed 2% after two years, while also finding changes in tasks and occupational transitions.[12] Adoption is real, but monetisation often requires organisational change.

Europe's difficulty is not only invention but diffusion. In early 2026, 43% of US workers reported using generative AI for work, compared with 26% in Italy and 36% in the United Kingdom; use was also less intensive in Germany, France and Italy.[13] In the euro area, 38% of firms described their adoption as advanced and 33% as infrequent or experimental.[14] These figures support a practical priority: faster integration into real processes, especially in sectors where European firms already have domain expertise.

Small and multimodal models can be economically attractive when they meet a task's quality requirements at lower cost or latency, or when local control is necessary. Open-weight systems and hybrid RAG architectures can also widen deployment options. None of these features guarantee value; integration burden, evaluation and accountability remain decisive.

Firms should therefore treat AI purchase as an operating model rather than a single-vendor software purchase. The stack typically includes model access, retrieval over proprietary knowledge, guardrails, and governance. Europe's data agenda matters here: the Data Governance Act supports common European data spaces in sectors including health, finance,

[11] Yotzov et al., *Firm Data on AI*, NBER Working Paper 34836 (2026).

[12] Humlum and Vestergaard, *Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI*, NBER Working Paper 33777, revised 2026.

[13] Bick et al., *Differences in AI adoption in Europe and the US* (CEPR, 2026).

[14] Ferrando et al., *Adopting and investing in AI: evidence from euro area firms in the SAFE* (ECB, 2026).

manufacturing, public administration and skills.[15] The Data Union Strategy links those spaces to AI ecosystems through planned data labs. If domain-grounded AI becomes a major source of value, the institutions that make data usable will matter alongside frontier-model capacity.

Manufacturing provides one illustration. For example, Siemens describes its Industrial Copilot as supporting code generation, diagnostics and human-machine collaboration, while Beckhoff markets computer-vision tools for applications such as visual quality control.[16] These are company examples that align with established European strengths in high-value production and complex, bounded workflows. This shows us that factories need reliable, inspectable and vision-enabled systems close to the production line, not the largest available model.

However, finance provides a different case. BNP Paribas has stated the aim of generating EUR 750 million in value from AI integration by the end of 2026 and has partnered with Mistral AI across business lines.[17] This is a company target, not a realised independent estimate. ING's annual report describes an AI governance framework, dedicated risk staff and a qualitative Generative AI Risk Appetite Statement.[18] The Bank of England reports industry support for principles-based regulation alongside skills bottlenecks and caution from second-line risk functions.[19] In financial services, deployment depends not only on capability but on whether decisions can be documented, supervised and defended.

Healthcare is a promising but demanding area because it combines data intensity, clinical risk, cross-border fragmentation, and a large public-sector role. The European Health Data Space Regulation establishes a framework for access, interoperability, and secure reuse of electronic health data.[20] Commission initiatives include Cancer Image Europe and planned AI-powered screening centers, while Google's MedGemma illustrates the

[15] European Commission, *Data Governance Act*; and *European Data Union Strategy*.

[16] Company examples: Siemens, *Artificial intelligence at Siemens*; Beckhoff, *TwinCAT 3 Machine Learning* (2026).

[17] BNP Paribas, *Integrated Report 2024* (2025), company target.

[18] ING Bank N.V., *Annual Report 2025* (2026).

[19] Bank of England, *Summary of AI roundtables* (2026).

[20] European Commission, *European Health Data Space Regulation* (2025).

development of open multimodal models for medical text and images.[21] These initiatives remain at different stages of implementation and validation. Europe's opportunity lies in the infrastructure that connects data, clinical evaluation, and regulation.

Furthermore, public administration is another material opportunity. An OECD brief argues that AI can improve administrative services if governments also develop internal capability, governance and workforce readiness; it reports that Finland's Kela estimates annual savings equivalent to 38 full-time working years from document classification and processing, and that Gladsaxe municipality operates a self-hosted GPT-based model.[22] The Commission has also proposed greater use of open standards and models in public procurement, with sovereignty, interoperability, portability and sustainability as design criteria.[23] Together, they show where public demand could support auditable and interoperable systems. Moreover, as we argue in the SKEMA Publika policy paper *Start-up and SME Growth*, by facilitating access to public procurement for European companies, the State and local authorities can play a decisive role in securing outlets, ramping up activity and scaling up. Used as a strategic tool, public procurement can become a powerful driver of growth.

Across firms and administrations, model access is becoming less scarce than the capacity to redesign processes and assign responsibility. A model can answer a question, but it cannot by itself redesign a claims process, procurement function, clinical pathway or maintenance routine. That work requires people, data, interfaces, and institutional confidence. Europe's deployment strategy should be built around those constraints.

C. GOVERNANCE, SOVEREIGNTY AND ADVANCED AI

The familiar opposition between regulation and innovation is too simple. The AI Act entered into force on 1 August 2024 and became generally applicable on 2 August 2026, subject to specific exceptions. The AI Omnibus, in force since 27 July 2026, extended the deadlines for high-risk systems under Annex III to 2 December 2027 and for AI embedded in regulated products under

[21] European Commission, *Artificial intelligence in healthcare* and *AI-powered advanced screening centres*; Google, *MedGemma model card* (2026).

[22] OECD, *Building an AI-ready public workforce* (2026).

[23] European Commission, *Commission boosts open and interoperable digital ecosystems for public administrations* (2026).

Annex I to 2 August 2028. The Commission has also issued GPAI guidance and codes of practice. The transparency obligations for marking certain AI-generated content apply from August 2026; the related code of practice is voluntary and is intended to support compliance.[24] Together, these measures form an increasingly concrete deployment framework, although implementation remains unfinished.

The framework is also gaining institutional depth. The AI Office supports implementation, while the Scientific Panel advises on general-purpose AI impacts and risks.[25] Europe may derive practical value from common documentation, sectoral standards, scientific advice and more consistent procurement expectations. These mechanisms can reduce uncertainty, but only if they remain proportionate and usable.

Public procurement is especially important. The EU's Public Buyers Community offers resources and standard clauses for trustworthy AI procurement.[26] The Commission has also argued for open standards and models and has used sovereignty as a criterion in its Sovereign Cloud tender. Public procurement can therefore shape demand for auditable, portable and interoperable systems; its effect will depend on how consistently buyers apply those requirements.

The AI Continent Action Plan commits EUR 200 billion to boost AI development, including EUR 20 billion for up to five AI Gigafactories, alongside AI Factories, greater data-centre capacity and the Data Union.[27] EuroHPC presents AI Factories as providing compute access and tailored support to startups and SMEs. These measures will not remove Europe's dependence on non-European accelerators, cloud platforms or software, but they represent a material infrastructure programme. The Data Union Strategy combines greater European control with openness to trusted partners under fair and secure conditions. A workable approach would support trusted cross-border data flows, documented provenance, reusable

[24] European Commission, *AI Omnibus enters into force*; *General-Purpose AI Code of Practice*; *Guidelines for providers of general-purpose AI models*; and *Code of practice on marking and labelling AI-generated content* (2025–2026).

[25] European Commission, *AI Office*; and *AI Scientific Panel*.

[26] Public Buyers Community, *Procurement of AI* (2026).

[27] European Commission, *AI Continent Action Plan* (2025–2026).

open components, local deployment where needed and enough compute capacity to limit critical dependencies.

AGI should remain a horizon issue rather than the organising principle of European policy. A June 2026 scenario paper sketches several possible routes from AGI to more capable systems, including large multi-agent collectives, but it is a forward-looking framework, not evidence that any route is imminent.[28] Its practical lesson is narrower: distributed systems can blur responsibility across model providers, retrievers and orchestrators. Europe should extend provenance, logging and incident-reporting requirements to orchestration layers while calibrating obligations to demonstrated capabilities and risks.

Existing governance tools point in this direction. The GPAI Code includes a model-documentation form. IBM publishes model cards and cryptographic-signing information for Granite, illustrating one approach to provenance.[29] The European Health Data Space provides another sectoral example through certification and interoperability requirements for electronic health-record systems. Trustworthy deployment will require a stack of model provenance, evaluation, sectoral testing, procurement discipline, and cross-border interoperability.

Important counterarguments remain. Europe risks underinvesting in frontier capability; open models do not remove dependence on US or Asian hardware; and a widening quality or ecosystem gap could limit the returns from deployment expertise. These risks justify continued frontier investment. They do not, however, make imitation of US or Chinese laboratories a sufficient industrial strategy.

[28] Genewein et al., *From AGI to ASI*, arXiv:2606.12683 (2026), scenario paper/preprint.
[29] IBM, *Granite 4.1 models* (2026).

II. FIVE STRATEGIC THESES

A. A PORTFOLIO APPROACH CAN LOWER THE AVERAGE COST OF INTELLIGENCE

Public pricing already shows a wide gap between premium and lightweight models.[30] That does not establish that a smaller model is adequate for a particular task. It does mean that evaluation, allocation, and routing can materially affect cost. Firms that match model capability to workflow requirements may outperform those that buy premium capacity by default.

B. IN ENTERPRISE AI, PROPRIETARY DATA IS NOT THE NEW OIL; IT IS THE NEW WORKING CAPITAL

Working capital circulates through processes and supports operations; data becomes valuable when it can do the same. RAG, multimodal retrieval, data spaces, and domain-grounded assistants can turn institutional knowledge into usable inputs. Europe's Data Governance Act and Data Union initiatives matter because they can reduce the friction involved in mobilising data safely, legally and quickly.

C. EUROPE'S STRONGEST OPENING MAY BE SPECIALISED DEPLOYMENT, NOT FRONTIER IMITATION

Frontier AI exhibits strong scale economies, concentrated compute and clustered ecosystems, and Europe begins from a weaker position. Deployment has different economics: institutional trust, integration depth, sectoral expertise and procurement power can matter more. Europe could build an advantage by acting as a coordinated group of strategic adopters, provided it maintains enough frontier capability to evaluate and adapt leading systems.

[30] Public pricing pages: Google, *Gemini Developer API pricing*; Mistral AI, *Pricing* (accessed 26 August 2026).

D. GOVERNANCE CAN BE A PRODUCTION TECHNOLOGY

In high-trust sectors, governance cannot be left as an ex-post legal concern. Documentation, evaluation, certification, and controllability are operational components that determine whether a system can be deployed at scale. Trust architecture can therefore form part of the production process rather than sit outside it.

E. ORCHESTRATION WILL BE A CENTRAL ORGANISATIONAL PROBLEM IN DECENTRALISED AI

As capability moves into local models, multimodal pipelines and agentic systems, responsibility may be divided among the model provider, retriever, orchestrator, buyer and deploying organisation. Europe should address this allocation directly. Clear accountability across a distributed model chain could become a practical source of trust and market access.

III. RECOMMENDATIONS

The previously enumerated theses suggest the four institutional priorities below.

1. Turn AI Factories into deployment institutions: Link AI Factories to sectoral data labs, multilingual evaluation, SME deployment support and applied training. Compute matters, but the relevant test is whether firms and public bodies can turn access into usable workflows.

2. Use public procurement as a trust-building instrument: Make portability, audit logs, model documentation, contractual transparency and retrieval-source records standard requirements where they are relevant to the risk and use case. This would reduce public-sector risk and create demand for systems that are governable and interoperable by design.

3. Move enterprises from single-model procurement to portfolio management: Establish routing rules, retrieval architectures, local and remote model tiers, and criteria for cost, latency, privacy, reliability and

liability. In regulated sectors, this should be supported by model inventories, risk-appetite statements and capable second-line functions.

4. Build shared validation infrastructure in high-trust sectors: Health systems, universities and research funders should treat validation capacity as a shared resource. Sectoral datasets, multilingual benchmarks, retrieval toolchains and edge-inference test beds can support work that no single organisation would build alone. The priority is rigorous, reusable evaluation in healthcare, finance and public services.

CONCLUSION

Europe's AI prospects should not be reduced to ownership of the world's most powerful model. The continent remains weaker in frontier scale, but it has credible assets for deploying auditable, multimodal, cost-efficient and domain-grounded systems across enterprises and public institutions. The policy architecture emerging in Brussels provides useful building blocks, not proof of comparative advantage. Turning them into an advantage will require deployment skills, accessible infrastructure, usable data and proportionate governance. The practical rule is simple: use smaller models where they meet the task, multimodal systems where text is insufficient, local deployment where latency or privacy requires it, and stronger controls where risk is high.

APPENDIX

A. 2026 AI MARKET

Model family	Size / scale	Access	Self-hosting	Deployment fit
OpenAI: GPT-5.4 mini	Undisclosed	Proprietary API	No public weights	Fast enterprise tasks, coding, tools and agent workflows.
Anthropic: Claude Haiku family	Undisclosed	Proprietary API	No public weights	Low-latency Claude workloads; capabilities vary by version.
Google: Gemini Flash-Lite family	Undisclosed	Proprietary API	No public weights	High-volume lightweight multimodal workloads.
Meta: Llama 4 Scout	17B active, 109B total, MoE	Open-weight under Llama 4 Community License	Yes	Local or partner-hosted multimodal workloads, long context.
Google: Gemma family	Several small and medium open-weight variants	Open weights, Apache 2.0	Yes	Edge, laptops, workstations, multilingual multimodal apps.
Mistral: Small 4	119B total / 6.5B active, MoE	Open weights, Apache 2.0	Yes	European multimodal, reasoning, agentic, and customisable deployment.
Hugging Face: SmolLM3	3B	Open weights	Yes	Open multilingual local reasoning and fine-tuning.

Illustrative model-portfolio options (as of 26 August 2026)[31]

[31] Illustrative, not exhaustive. Official documentation: OpenAI, Models; Anthropic, Models overview; Google, Gemini models and Gemma model card; Meta, Llama 4; Mistral AI, Mistral Small 4; Hugging Face, SmolLM3. Models, prices and licences can change.

SKEMA Publika
SKEMA Business School, Campus Grand Paris
5 Quai Marcel Dassault – CS 90067
92156 Suresnes Cedex, France

Tél. : +33.1.71.13.39.32
Courriel : publika@skema.edu
Site Internet : www.publika.skema.edu