Memorandum

A Last Window of Opportunity

A Strategy for Europe's Frontier AI Initiative

Submitted to the European AI Office and the Frontier AI Initiative task force.

Executive Summary

In April 2026, Anthropic announced Project Glasswing, a coordinated rollout of its Mythos system, which can autonomously discover and exploit security vulnerabilities in critical-infrastructure software. Glasswing extended priority access to an all-US consortium for defensive patching and vulnerability discovery. Anthropic has separately confirmed it is in ongoing discussions with US government officials about Mythos's offensive and defensive cyber capabilities. No European institution was on the list. EU critical-infrastructure operators, national Computer Emergency Response Teams (CERTs), and defence ministries did not get to use Mythos to harden the software stacks they rely on.

The Frontier AI Initiative was launched to strengthen Europe's capacity to develop and deploy frontier AI. When frontier AI capabilities exist only within a closed US consortium, European institutions cannot use them to protect European infrastructure. The gap in access is a direct consequence of the gap in sovereignty, and closing it is exactly what the Initiative was designed to do.

Six months since its onset, the Frontier AI Forum seems to have produced more disagreement than direction. Several courses of action are currently on the table: training a sovereign European frontier model from scratch, leapfrogging the current paradigm, focusing on vertical integration, and building out compute infrastructure.

This memo makes the case for a single priority: scale and Europeanise a model that already works. The DINUM (Direction interministérielle du numérique) in France operates Albert API, a serving stack for open-weight AI models that runs entirely on EU soil. Its data and inference remain under French legal jurisdiction, outside the reach of foreign law including the US CLOUD Act. Albert API already serves more than 70 public projects across the French administration, processing over 100,000 requests per week. It is built on OpenGateLLM, an open-source gateway developed by the French state and designed for replication. It is hosted on SecNumCloud-qualified infrastructure, the French national security certification for sensitive data. It is the most operationally mature sovereign AI deployment in any member state today. It works. It is underfunded. It has no European equivalent.

This points to three things the Initiative is well-placed to do. First, scale a Sovereign Frontier Access Layer along the Albert API model, accessible to European institutions and Operators of Vital Importance (OIV) across member states. Second, fund the verification, audit, and fine-tuning capacity that makes deploying foreign-origin open-weight models defensible for European critical infrastructure. Third, couple this with an agentic cyber defence programme that uses the same infrastructure to systematically scan and patch European software stacks while the current window for defenders remains open.

The Center for AI Safety (CeSIA) is an official advisor for the European Commission on AI-related manipulation risks evaluations (Lot 4, 2026–2028) and a contributor to the OECD working group on the Hiroshima AI Process (HAIP) reporting framework. An independent expertise center and think tank dedicated to preventing major AI-related risks, CeSIA conducts research, training, and public policy analysis activities to strengthen the safety and governance of advanced AI systems. CeSIA collaborates with national and multilateral institutions including the OECD, UNESCO, the Paris Peace Forum, Sciences Po, the French National Institute for AI Evaluation and Safety (INESIA), and its British counterpart (UK AISI). CeSIA delivers Europe's first university-level training program in AI safety science at ENS Ulm, and is the initiator of the Global Call to Establish Red Lines for AI, launched at the 80th United Nations General Assembly.

Introduction

The Frontier AI Forum surfaced a wide range of strong positions. Several groups have published serious analyses, each capturing a real piece of the problem: the RAND/CFG Europe and the Geopolitics of AGI (December 2025), which identifies Europe's critical gaps in strategic awareness, compute infrastructure, and talent retention; The Future Society's Beware of Geeks Bearing Gifts (April 2026), which decomposes frontier AI sovereignty into five pillars and maps the stack layer by layer; SaferAI's Case for European Investment in High-Risk, High-Reward AI Reliability Research (March 2026), which argues for an ARPA-like reliability mission, potentially as the core of the Initiative itself; and the Oxford/CeSIA Blueprint for Multinational Advanced AI Development (November 2025), co-authored with Yoshua Bengio, which argues that mid-sized "bridge powers" can reach the frontier only by pooling compute, talent, and data.

These are strong contributions, but they are addressed to "Europe" in the broad sense, and they stop short of the decision the FAII actually faces. The FAII cannot pursue every direction at once: it must choose one and execute it. The directions raised at the Forum are largely mutually exclusive in practice. Training a sovereign frontier model, a fundamental-research moonshot, packaging existing capabilities, or scaling compute infrastructure each imply a different Year-1 deliverable and a different funding envelope. We have heard that the Initiative is currently scoped as a single project rather than a portfolio. If that is correct, the choice of which project matters enormously. If it is no longer correct, the prioritisation across parallel tracks matters just as much.

This memo takes a position on that choice, and proposes concrete steps to execute it. Our recommendation converges with parts of this existing work, particularly SaferAI's case that reliability is a structural European advantage, and the Blueprint's argument for pooled sovereign capability. It diverges in one key respect: we argue that an operational serving-and-defence layer must come first, because without it none of the other directions have a substrate to build on. We name those convergences and divergences explicitly in the asks at the end.

Recommendation: Create a European Sovereign Frontier Access Layer

Albert API demonstrates a working model

Albert API is the closest thing Europe has to a working prototype for sovereignty. Operated by France's interministerial digital directorate, the DINUM, it serves more than 70 public projects across the French administration and handles over 100,000 requests a week, running on SecNumCloud-qualified infrastructure with on-premise GPU compute for the most sensitive workloads. A single gateway routes each request to whichever open-weight model best fits the task, with a layer of fine-tuning on top to adapt them to French and EU-language administrative use.

The lineup is deliberately origin-agnostic. The currently hosted models include Mistral's French open-weight models, OpenAI's open-weight gpt-oss, and Chinese open-weight models such as Alibaba's Qwen for code and the Beijing Academy of AI's models for retrieval. All run on French servers, and the platform guarantees that no user data reaches the original model providers. This is the architectural point at the heart of the approach: sovereignty here is not about who trained the weights, but about whose jurisdiction the inference runs under. Because open weights can be downloaded and served locally, France can take the best available capability from any origin and run it inside its own legal perimeter, outside the reach of the US CLOUD Act and beyond any provider's ability to withdraw access.

This matters because unfortunately Europe's own frontier capabilities sit well behind the closed state of the art. Mistral's releases have concentrated in the small and medium class of models rather than at the frontier, leaving a capability gap of roughly twelve to eighteen months. Albert API bridges that gap not by waiting for a European model to catch up, but by serving whatever open weights are strongest, under sovereign control. The one capability this still demands is the ability to audit foreign-origin weights before they touch critical systems, which is exactly what the second component below addresses. And the design is built to scale: Albert API v2 is published as OpenGateLLM, an open-source gateway the French state designed for other administrations to adopt.

This is not a pilot or a prototype. It is a sovereign deployment, sized for real public-sector demand, proven through daily use, and operating under France's doctrine of placing SecNumCloud-qualified infrastructure at the centre of sensitive state systems. No other member state runs anything close.

The Frontier AI Initiative can scale this working model to Europe

The opportunity is to take this national deployment architecture and make it European: a Sovereign Frontier Access Layer available to public institutions, Operators of Vital Importance (OIV), and regulated sectors across member states. The pieces already exist, for example Scaleway, OVHcloud,

Albert API itself. What is missing is the funding, the mandate, and the integration to join them across the EU.

Three components have to be built together.

1. **Sovereign serving**. Contract European cloud operators (Scaleway, OVHcloud, or a consortium) under a public specification: EU-jurisdictional, SecNumCloud-qualified, and audit-ready, with rolling integration of the best globally available open-weight models. The model layer turns over every few months as new open weights are released, so customers subscribe to the access layer, served on EU infrastructure, rather than to any single model.

2. **Safety verification and fine-tuning.** Before any foreign-origin model can be defensibly served to critical infrastructure, it has to be audited: backdoor and trojan detection, capability evaluation, and jailbreak-resistance testing, followed by fine-tuning inside sovereign infrastructure so that weights never leave EU jurisdiction. This is engineering, not open-ended research. The tooling already exists in many published works, including CeSIA's work on how to safeguard benchmarks, findings on the brittleness of misuse detection, and review of safety-measurement methods.

3. **Agentic cyber defence.** Glasswing is a timing problem as much as a sovereignty one. For now, AI finds vulnerabilities faster than attackers can weaponise them, handing defenders a temporary edge that will vanish as the capability spreads. Europe can use the same serving layer as we talked about above (Scaleway, OVHcloud, or a consortium) to scan and patch the software under its critical infrastructure while that edge holds. It does not need Mythos-class access to do so. Public benchmarks from Hacktron and AISLE show that, for the application-layer code most infrastructure actually runs on, well-orchestrated ensembles of cheap open-weight models recover much of what frontier models find at a fraction of the cost. The hardest targets, like browsers and kernels, still favour the frontier, and small models throw more false positives but the moat is the system, not the model. A Sovereign Frontier Access Layer is exactly the system to host it.

Requirements for this project

Roughly €1 billion over three years would carry the Albert API model from French administration to European OIV coverage, fund verification and fine-tuning to a defensible standard, and stand up the cyber defence programme. That is one to two orders of magnitude below the cost of training a sovereign frontier model, and comfortably inside the AI Continent Action Plan envelope. Money is not the binding constraint. Timing is.

The DINUM is being reorganised right now: on 11 May 2026, Prime Minister Lecornu tasked Walter Arnaud with prefiguring "Ariane," its successor, alongside a parallel mandate for a future national digital and AI authority. The window to write a European-scale ambition into that new mandate is open now and will not stay open long. The real constraint is coordination between the AI Office, the French DINUM/Ariane, and equivalent German and Nordic bodies.

Alternative strategies we recommend deferring to the future

Training a sovereign European frontier model on the announced 2030 timeline. This is foreclosed by chip supply, fab capacity, and capital. The compute available in 2028 was largely determined by orders placed in 2024 and 2025, and Europe was not near the front of that queue. A European frontier training programme remains a 2034 option at the earliest, conditional on the next Multiannual Financial Framework and on supply-chain leverage that Europe holds through ASML and Carl Zeiss but does not currently use on its own account. We do not abandon this option. We sequence it. The initial phase is sovereign open-weight deployment now. Phase 2 is fully self-sovereign training in 2034, if the conditions resolve.

Leapfrogging the current paradigm. Wholesale bets on world models, neurosymbolic systems, or neuromorphic hardware all share one structural weakness: the largest historical algorithmic gains have been compute-dependent, so whoever holds frontier compute will validate the next paradigm first, no matter whose research produced the original idea. DeepSeek V3, released in December 2024, was absorbed into the architectures of every frontier US lab within months. Any efficiency Europe discovers will be replicated and scaled, not preserved as a moat. Small research bets in this direction are defensible; betting industrial policy on them is not.

An applications-only strategy focused on vertical integration. Concentrating European effort on industrial AI applications while leaving frontier capability foreign-controlled keeps the input layer in someone else's hands. The applications layer is where most economic value will be captured, but it is fragile to exactly the access withdrawals that Glasswing demonstrated. Vertical integration delivers sovereignty only when it sits on top of the open-weight deployment substrate we recommend. In any case, driving applications strategy is the work of industry with Commission support, not of the AI Office. The AI Office's mandate is the model layer and the safety layer, which is precisely what this memo addresses.

Asks to the Frontier AI Initiative

Five operational asks, in order of urgency.

1. **Make the Sovereign Frontier Access Layer the Initiative's Year-1 deliverable.** Contract European sovereign cloud operators through public tender, against clear technical and jurisdictional specifications, with Albert API as the working prototype. A Phase 1 envelope of €300 to €500 million.
2. **Accelerate SecNumCloud finalisation for Scaleway and equivalent national qualifications across member states.** The DINUM has already proven the model with OVHcloud. Replicating it across Germany (C5), Italy, Spain, and the Nordics is the precondition for European-scale serving.
3. **Stand up the verification and agentic cyber defence programme as the safety-and-security arm of the Initiative.** This is the operational counterpart to the fundamental research that SaferAI and ARIA already fund, and it is the capability the AI Act's high-risk requirements will need in place by August 2026.
4. **Open a supply-chain leverage workstream.** Phase 2, sovereign frontier training, cannot be unlocked by money alone. ASML and Carl Zeiss are EU-jurisdictional chokepoints in the

global compute supply chain. Europe has used that leverage to constrain China but has never used it to negotiate priority allocation for its own capacity. This is a political workstream, not a funding line.

5. **Commission an AGI Preparedness Report.** Along the lines the RAND/CFG report recommended in December 2025. The Bletchley Declaration signatories committed in November 2023 to building independent capacity to evaluate frontier AI risks. France and Germany were among them. Two and a half years on, no European institution has publicly demonstrated an autonomous capability to evaluate a frontier model on CBRN uplift, cyber-offensive capability, autonomy under measurement, or deception. This report would name what European institutions still need to build, and on what timeline.

Conclusion

The Frontier AI Initiative was announced as Europe's response to a strategic gap. Six months later, the gap is no longer in question. Glasswing made the need for sovereignty concrete, and the Forum mapped it in detail. The open question is what Europe builds in the next eighteen months.

The recommendation here is deliberately modest in ambition and concrete in execution. It is scaled to what European infrastructure can deliver now, and it is grounded in a system the French government already runs every day. It is not the most ambitious vision of European sovereignty. It is the floor: the minimum the Initiative has to reach to stay a serious instrument. The more ambitious tiers, sovereign frontier training and full supply-chain onshoring, stay open on longer timelines if the political work gets done.

The Bletchley commitments of November 2023 still await their operational European translation. This is one. The window is closing.