Building CERN for AI

An institutional blueprint

**centre
for future
generations**

Building CERN for AI

An institutional blueprint

Authors: Alex Petropoulos, Bálint Pataki, Daan Juijn*, David Janků and Max Reddel
* Corresponding author: d.juijn@cfg.eu

This report is available at www.cfg.eu/building-cern-for-ai

Acknowledgements

Table of Contents

Executive Summary

CERN for AI is based on a simple but powerful idea: bring together the brightest minds, give them world-class resources and research autonomy, and point them at one of today's greatest challenges—building AI systems people can trust.

Mario Draghi's [landmark report](#) warns that Europe faces foundational challenges in technology development—from fragmented markets to scattered investments to an underdeveloped venture capital (VC) sector. Draghi rightly calls for both fundamental market reforms and ambitious public catalysts to jumpstart change. European Commission President von der Leyen's [call for a CERN for AI](#) represents such a catalyst—a moonshot initiative that could transform Europe's tech landscape.

Using Draghi's suggestion of an ARPA-type model as our foundation, we synthesized research on ARPA-style institutional design and analyzed best practices across research structures and governance mechanisms. We then adapted and enhanced this model to address the unique challenges of advanced AI development, where issues of safety, transparency, and democratic accountability are paramount alongside the need for rapid innovation.

We propose CERN for AI as a pioneer institution in trustworthy general-purpose AI. It will bridge foundational research with real-world impact across four key areas:

1. Fundamental research into AI interpretability and reliability
2. Scaling promising approaches into cutting-edge systems
3. Developing applications for pressing societal challenges
4. Advancing hardware foundations for responsible AI development

The institution we envisage would operate through a novel two-track structure that combines the best features of ARPA-style agencies with dedicated in-house teams. This hybrid model enables open collaboration with academia and industry, giving Programme Directors extensive freedom to bridge resource gaps that individual institutions cannot manage alone. It also provides in-house, close collaboration that is essential to managing security and fuelling a fast-exchange environment required for integrating breakthrough ideas.

The governance model we propose relies on twin principles of transparency and accountability, implemented through a Member Representative Board that serves as a guardian of the institution's mission. Two independent expert boards—Mission Alignment and Scaling & Deployment—would serve as operational mirrors, ensuring work stays aligned with goals and institutional values through public feedback loops. While these boards provide crucial guidance, ultimate control would remain with member countries, initially comprising EU/EEA states and trusted Horizon Europe partners like the UK, Switzerland, and Canada. A tiered membership structure could enable future broadening while protecting sensitive technology.

A comparative legal analysis we commissioned from [The Good Lobby](#), shows that a Joint Undertaking under [Article 187 of the Treaty on the Functioning of the European Union (TFEU)](#) would be a natural fit for a CERN for AI. This legal basis offers both the necessary accountability and the flexibility to act like a dynamic tech startup while maintaining the benefits of operating under the frameworks of an EU institution.

In our initial report, we [estimated](#) that establishing CERN for AI requires €30–35 billion over the first three years—an investment that would yield cascading benefits for Europe's future. The initiative would be funded primarily through core contributions from the EU, member countries, and strategic private sector partners, supplemented by program-specific industry funding, technology licensing revenue, and compute capacity rental. These resources would position Europe to lead in domains where AI capabilities could be decisive: addressing climate change, enhancing regional cybersecurity, and maintaining global competitiveness and regulatory influence.

The time for action is now. As AI investments accelerate globally, Europe still has a unique opportunity to shape the development of this transformative technology. CERN for AI could not only advance more trustworthy AI development but catalyze a self-sustaining European tech ecosystem that bears fruits for decades to come. With world-class researchers, democratic values, and a tradition of ambitious scientific collaboration, Europe has the essential components for success. What's more, the cost of inaction—in terms of missed economic opportunities, diminished global influence, and potential safety risks—could far outweigh the investment required.

The political will behind a bold, cross-border initiative is growing. Increasingly, leaders are acknowledging that Europe cannot get stuck in complacency or defeatism when it comes to AI.

Instead, Europe needs to build.

1. Introduction

While the EU leads in AI regulation, it hasn't cracked the code for building a thriving AI industry of its own. This isn't for lack of dreams—it's for lack of doing. European AI initiatives have [consistently run short](#) on three essentials: computing power, funding, and concentrated talent.

The EU could be watching from the sidelines – yet again

If the EU does not address these issues, it could be [watching from the sidelines yet again](#). Just as the EU has largely failed to reap the economic benefits from the internet revolution, so too could it miss out on the AI revolution. As Draghi [notes](#), "there is no EU company with a market capitalisation over €100 billion that has been set up from scratch in the last fifty years, while in the US all six companies with a valuation above EUR 1 trillion have been created over this period."

And the opportunity here goes beyond economic impacts. The difficulties of attempting to provide oversight over industries that are largely located outside of the EU has become increasingly acute, from [technology](#) to [climate](#) to [trade](#). European leaders will face an uphill battle attempting to shape the coming AI revolution if the EU remains on the outside looking in.

Public From-Scratch EU Companies Less than 50 Years Old with $10b+ Market Cap
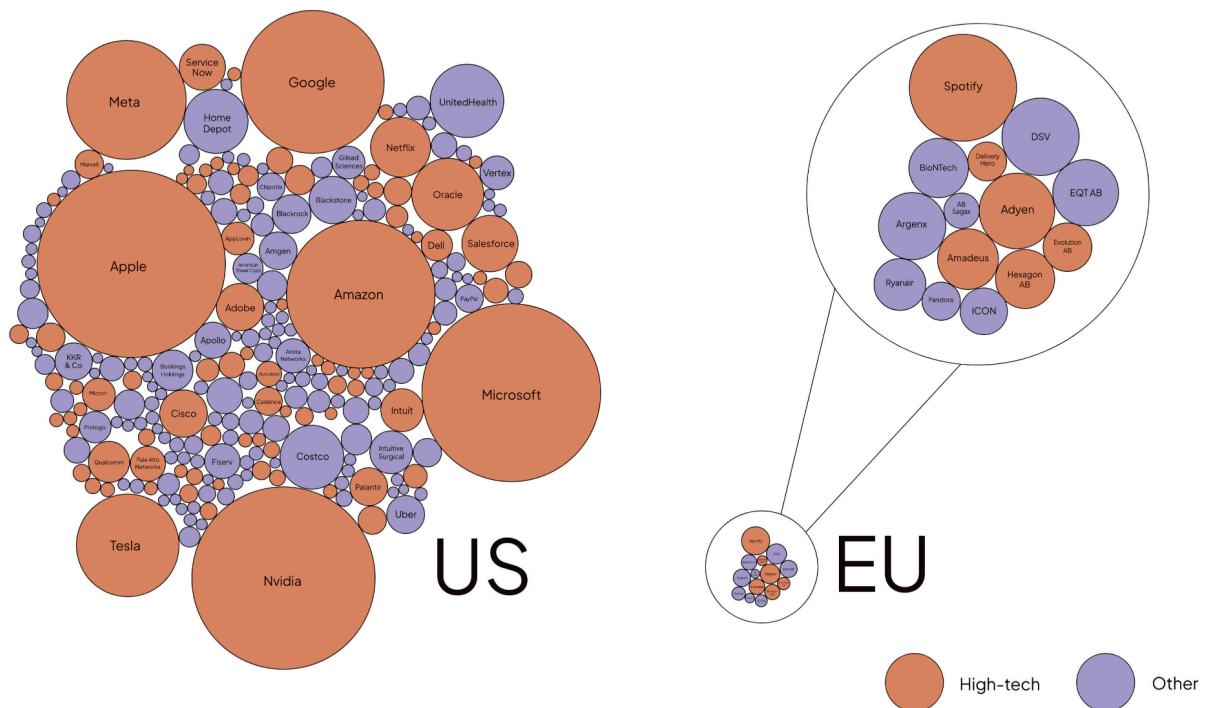
Figure 1: The EU has mostly failed to spur large, new tech companies in the last 50 years.

The current AI gap triggers a troubling chain reaction. Without trusted, European-built AI systems, companies hesitate to adopt the technology. Without widespread AI adoption, Europe's capacity to innovate stalls. The numbers paint a sobering picture: on adoption, European enterprises lag behind their American counterparts by 40% to 70%. This hesitation carries a steep price tag. Goldman Sachs estimates current AI systems can already boost average productivity by 25%, while recent studies show generative AI increases patent filings by 39%. With such gains, every percentage point counts.

Trust lies at the heart of this challenge. European organisations—from hospitals weighing AI diagnostics to manufacturers exploring automated quality control—need systems they can rely on completely. The data is telling: organisations that trust AI are 26% more likely to adopt it, according to Deloitte. Without homegrown, trustworthy solutions, this trust deficit will persist.

The case against defeatism in AI

Meanwhile, overseas investments into AI are accelerating: just a week before the publication of this report, OpenAI, Oracle and Softbank announced the Stargate Project - a joint venture that will invest 100 billion dollars into American AI datacenters, with additional multibillion dollar investments planned. Microsoft also launched a 30 billion dollar AI infrastructure fund with fund manager BlackRock in September 2024, and in January 2025 chief executive Satya Nadella said his company would spend 80 billion dollars on infrastructure this year.

It's easy to respond to such announcements with defeatism: what can Europe really do among such heavy hitters?

This, however, would be the wrong takeaway. Not only has the Stargate Project yet to secure the necessary funding, the AI field has just entered a new paradigm that presents unique opportunities for new entrants. As the recent success of the relatively poorly-resourced Chinese AI company DeepSeek shows, progress has become less reliant on extensive imitation learning from human data, and is now heavily driven by RL-based post-training and inference-time techniques. With the industry exploring novel methods to advance AI capabilities, fundamental research into trustworthiness and reliability could shape the field's direction for years to come. That's not to say money and computational resources are no longer important - but we should not automatically assume the player with the deepest pockets will come out on top.

The EU should view the Stargate announcement not as an excuse for defeatism, but rather as an encouragement to strengthen its own domestic industry. The Draghi report presents a concrete roadmap for Europe to break out of its current 'middle technology trap' and to become

competitive in advanced AI, and tech more generally. As Draghi rightly argues, market reforms and private investment are crucial, but Europe also needs public catalysts. Public leadership can lower barriers, concentrate talent, and show what's possible. As former Minister for Science, Technology and Higher Education of Portugal Manuel Heitor put it, Europe needs "world-class research and technology infrastructures" that serve everyone—researchers, industry, and the general public alike.

European leaders have recognized the urgency. Enrico Letta calls for 'large-scale, cross-border AI projects', and Commission President von der Leyen has recently proposed an even bolder solution: create a CERN-like institution for AI, pooling Europe's resources to tackle this challenge head-on.

The need for a CERN for AI

As our previous report argued, CERN for AI could address three vital European priorities: economic advancement, security enhancement, and the development of trustworthy AI.

Today's technological gap threatens more than market position—it impacts Europe's capacity to fund its climate transition, support strategic security, and sustain essential social systems.

CERN for AI could serve as Europe's foundation for frontier AI development—a hub where trustworthy systems are built and a vibrant tech ecosystem flourishes. By concentrating exceptional talent and resources, it would create the institutional backbone Europe needs to lead in responsible AI innovation and begin tackling its economic woes.

The stakes extend beyond economic advancement, however. As AI becomes increasingly central to cybersecurity, Europe's digital sovereignty hangs in the balance. With geopolitical tensions rising, developing independent technological capabilities isn't optional—it's imperative.

But perhaps the greatest challenge lies ahead: Making advanced AI systems safe and reliable is uncharted scientific territory. As OpenAI admits explicitly, current safety techniques 'won't scale' to much more powerful AI systems. Fortunately, Europe has a proven track record of solving bold scientific challenges. Just as CERN transformed our understanding of physics through bold, focused research, this new institution could revolutionize how humanity builds and integrates AI systems.

A framework for success

Trustworthy AI can be built in Europe, but only with the right conditions and commitments. The 2024 CFG report highlighted nine critical design elements for a CERN for AI.
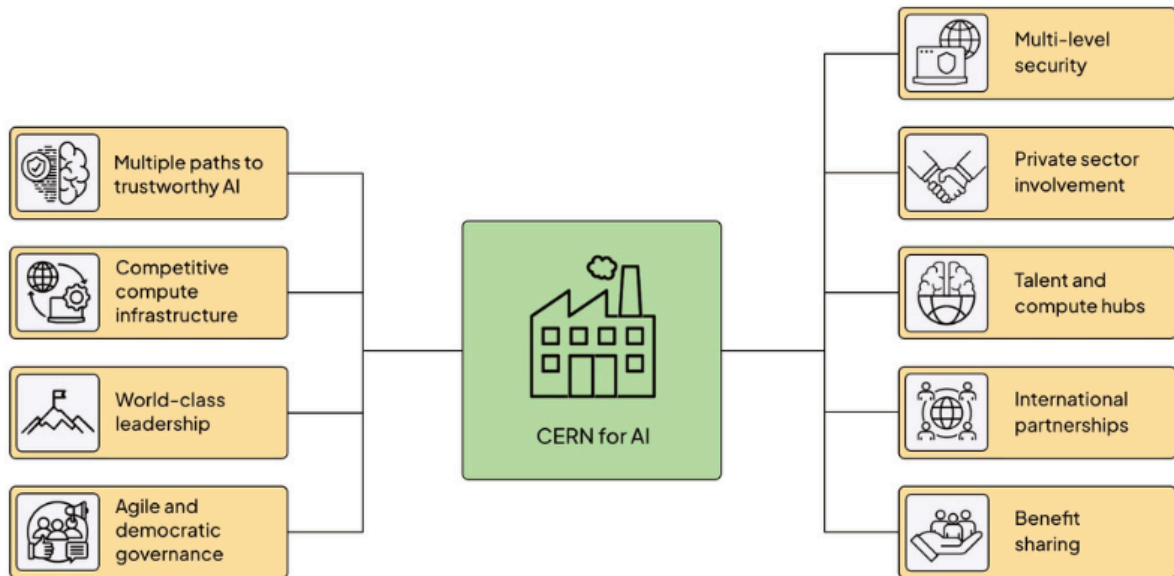
Figure 2 – CFG's previous report identified nine critical design elements for a CERN for AI

These design principles show that CERN for AI can't simply be a scaled-up version of current European AI initiatives. Consider the AI Factories program: while valuable for supporting startups and academic research, it lacks the scale and concentration needed for frontier AI development[1]. In addition, its governance isn't designed for the heightened stakes of increasingly capable, autonomous AI systems, and it doesn't come with a targeted research strategy. To compete at the frontier of AI innovation, Europe needs more than an expansion of existing programs—it needs a purposefully designed institution.

Moving from vision to blueprint

CFG's 2024 report established foundational principles for CERN for AI, but principles alone don't build institutions. After the publication of our research, we had many conversations with European institutions, national governments, civil society organisations and industry

---

[1] The maximum compute investment for a single AI factory is capped at €400 million - enough to buy and connect some 5,000 cutting-edge NVIDIA GB200's . Private actors such as OpenAI are currently building clusters with over 100.000 of the same AI chips.

representatives to assess what they thought were the most important remaining questions. These could be bundled into seven categories:

1. **What should CERN for AI's mission be?** What exactly should the institution aim to solve? AI adoption? Fundamental research problems? Should it become a direct competitor to the American AI companies?

2. **What research areas should CERN for AI focus on?** Should CERN for AI build foundation models? Should it focus on specific niches? Should it build its own hardware?

3. **What kind of research structure could suit CERN for AI's needs?** How could CERN for AI avoid the pitfalls of existing research institutions? Can the EU create a structure that is efficient enough for the fast-moving field of AI?

4. **How should CERN for AI be governed?** Should private actors have a say over its governance? What kinds of organisational guardrails are needed? What should the organigram look like?

5. **What should CERN for AI's membership policies be?** Should it be a purely European institution? If CERN for AI would expand over time, what should the admission procedure be?

6. **How should CERN for AI be funded?** Is it really possible to attract such a large sum of money? In what way would national governments contribute? What should Europe expect from the private sector?

7. **What should CERN for AI's legal basis be?** Does there exist a legal foundation that enables CERN for AI to be set up quickly? How can CERN for AI avoid hiring restrictions that have plagued other European institutions?

This report aims to tackle these questions, providing an institutional blueprint for the organisation's mission, focus areas, research structure, governance, membership policies, funding and legal foundation. In confronting these questions, it also outlines how private sector partners could contribute, how research programs with varying security requirements could operate in parallel, and how the benefits of breakthroughs can be shared among stakeholders.

2. Methodology

Motivation and focus

The [Draghi report](#) highlights significant room for improvement in existing EU research institutions, particularly in rapidly evolving fields like AI. Rather than starting from scratch, we try to build directly on Draghi's recommendations, specifically focusing on what the report identifies as the most promising model for public-led R&D institutions: [ARPA-style research programs](#).

ARPA's track record speaks for itself. Throughout its history (during which it was renamed to [DARPA](#)), it [pioneered](#) GPS technology, the first weather satellite, and the field of materials science. Its Information Processing Techniques office birthed advances in computer graphics, chip hardware, parallel computing, and the field of artificial intelligence itself. ARPA also created ARPANET (the internet's precursor), the first computer mouse, the world's first internet community, modern internet protocols, and the voice-assistant technology underlying SIRI.

However, "let's replicate ARPA" oversimplifies things. ARPA itself has undergone multiple iterations under different leadership, and numerous ARPA-inspired organisations have since emerged. The study of these institutions—and broader questions about how organisational design can accelerate scientific progress—has spawned an entire field: [metascience](#). Given the [high-level support](#) for the ARPA model following the Draghi report and its proven ability to incubate paradigm-shifting innovations, we are confident that the right structure for CERN for AI can be found within this broader metascience literature.

But identifying a research direction is just the beginning. Different organisational structures studied in metascience—from ARPA-style [coordinated research programs](#) to [Focused Research Organizations](#) (FROs), from private companies to national labs—each come with their own requirements and priorities, which often conflict. For instance, ARPA-style programs actively avoid internal full-time hiring to maintain flexibility with a decentralized talent pool, while FROs explicitly pursue full-time hiring to build concentrated teams. To give CERN for AI the strongest foundation, one needs to understand the underlying rationale and benefits of each approach. In this report, we take a deep dive into the metascience literature, and try to find tailored design solutions that are grounded in the specific technical and safety challenges of advanced AI while remaining politically feasible.

Distilled Requirements

While we worked forwards from existing metascience literature, we also worked backwards from a set of pre-existing requirements for CERN for AI, most of which were formulated in CFG's [2024](#)

report. By approaching this research from both directions, we tried to hone in on a set of recommendations for CERN for AI's institutional design.

At the highest level of abstraction, two fundamental requirements shaped our research process and our recommendations. These requirements demand that:

1. CERN for AI is designed with the fundamentals to one day responsibly push the envelope of AI capabilities on the most advanced AI systems in the world, even if it doesn't do so from day one…

2. … while also robustly improving Europe's economic competitiveness and security from day one.

We then turned these high-level requirements into a set of more concrete prerequisites necessary to hit the ground running (short term) while staying on track to cross the finish line (long term). This isn't a neat binary: Some prerequisites are important for both the short and long term, but for clarity they are listed as belonging to either one or the other.

In order to hit the ground running, our previous analysis suggests that CERN for AI would need the following prerequisites:

- **Pursuit of parallel moonshots.** The institution must have the capacity and courage to pursue multiple ambitious research directions simultaneously. While individual initiatives may carry significant risks of failure, their potential transformative impact means that even a small number of successes could justify the entire enterprise. This approach embodies what Draghi terms 'disruptive innovation'—breakthroughs that can reshape entire fields rather than incremental improvements.

- **Streamlined operational structure.** Success demands a radical departure from traditional public institution bureaucracy. Researchers need both the autonomy to pursue promising directions and the administrative agility to quickly assemble teams, acquire resources, and forge partnerships. CERN for AI should aim for the operational efficiency of a tech startup while maintaining the accountability expected of a public institution.

- **Deep private sector integration**. Bridging the gap between research breakthroughs and market impact requires close collaboration with industry from day one. As Draghi highlights, the EU often struggles to translate innovation into commercialization. To avoid this pitfall and maximize its benefit to European prosperity, CERN for AI must actively partner with private companies, creating pathways for its advances to reach practical applications.

- **World-class resource base.** Without substantial capital to offer competitive compensation and secure necessary compute infrastructure, the institution would face

insurmountable barriers. Beyond just enabling large-scale model training and parallel research initiatives, adequate funding is crucial for attracting and retaining top talent in a highly competitive global market.

- **Exceptional founding leadership:** Launching an institution of this scale and ambition requires extraordinary leadership from day one. Only internationally recognized figures with proven track records can drive the necessary momentum, attract world-class talent, and successfully navigate the complex political landscape inherent in launching such a large-scale pan-European initiative. The founding team must combine scientific excellence with political acumen and organisational expertise.

In order to stay on target, CERN for AI would need the following additional prerequisites:

- **Adaptive oversight framework.** Public oversight must strike a delicate balance between ensuring responsible technology development and maintaining research agility. The governance structure should be precise and targeted, avoiding bureaucratic bottlenecks while ensuring AI systems serve European citizens' interests. As AI capabilities advance and pose novel challenges, oversight mechanisms must be able to evolve in parallel—requiring a governance framework that can adapt to emerging realities while maintaining its core principles.

- **Comprehensive security architecture.** As CERN for AI ventures into security-sensitive domains—from cybersecurity to critical infrastructure protection—robust safeguards become paramount. Beyond protecting novel algorithms from industrial espionage, security measures must shield valuable assets like AI model weights. This security architecture should scale with the institution's growing capabilities and evolving threat landscape.

- **Enduring mission alignment.** The institution's long-term success hinges on maintaining unwavering focus on its founding purpose. While many [organisations](#) [drift](#) from their original mission over time, CERN for AI must cultivate a culture where its vision permeates every level of operation. This requires more than stated values—it demands systematic approaches to hiring, training, and decision-making that reinforce the institution's core purpose and ensure its pursuit of trustworthy AI development remains steadfast.

- **Strategic international engagement.** Shaping the trajectory of a general-purpose technology like AI requires influence beyond European borders. Strategic partnerships with trusted non-EU nations can accelerate the adoption of responsible AI practices globally while protecting European citizens from potentially unreliable foreign technologies. These international collaborations offer practical advantages too—expanding talent pools, attracting additional funding, and creating opportunities for knowledge exchange that strengthen Europe's position in the global AI landscape.

- **Catalyst for ecosystem growth.** Perhaps most important of all, CERN for AI must transcend its role as a research institution to become an engine of sustainable innovation. Like a greenhouse nurturing seedlings until they can flourish independently, it should create environments where technological innovation, entrepreneurial talent, and venture capital naturally converge. By fostering self-sustaining innovation hubs, the institution can trigger a cascade of technological advancement that continues long after public intervention ends, maximizing the return on Europe's initial investment.

These high-level requirements naturally led to more detailed specifications for CERN for AI's mission, core focus areas, research structure, organisational design, membership policies, funding sources and legal foundation. While we don't explicitly reference these underlying values in every section, certain lower-level requirements—such as those detailed in Chapter 9 on CERN for AI's legal foundation—are made explicit where their derivation from the core prerequisites is not immediately apparent.

Limitations of this research

The blueprint outlined in the following chapters is not presented as a definitive solution, but rather as a carefully considered starting point. Institutional design at this scale involves complex trade-offs that warrant thorough debate. Moreover, the literature on the intersection of advanced AI and institutional design is scarce, and so are experts who specialise in this area.

The proposals we argue for in this report reflect our findings after five months of research. While we feel like we've made substantial progress during this period and are confident in the broad strokes of our recommendations, it is clear that additional perspectives are necessary.

Yet urgency must balance comprehensiveness: Europe cannot afford to spend years commissioning policy research while falling further behind the US and China. As political momentum builds behind CERN for AI, the conversation must now move from abstract concepts to concrete plans. This report aims to accelerate that crucial transition.

3. Mission Statement

The original CERN's enduring success in bringing about [scientific breakthroughs](#) stems in part from its [ambitious mission](#):

> *"At CERN, our work helps to uncover what the universe is made of and how it works. We do this by providing a unique range of particle accelerator facilities to researchers, to advance the boundaries of human knowledge."*

This mission statement does more than inspire—it provides a compass for decades of research decisions. CERN for AI requires similar clarity of purpose, one that not only guides day-to-day choices but shapes institutional DNA. A mission statement here isn't just words on paper; it must become the shared conviction that drives both researchers and partners toward common goals. It too, must be supported by the EU as a critical priority.

Creating this mission-driven culture demands, [among other things](#), exceptional leadership and robust institutional procedures. Chapter 5 details the specific governance mechanisms we have in mind. Below, we offer an initial proposal for CERN for AI's mission to spark a more detailed conversation.

CERN for AI exists to pioneer the science of trustworthy general-purpose AI and to accelerate AI applications that enrich society.

We pursue this mission through four core pillars:

1. **Diverse research paths.** We place multiple bold bets on different approaches, creating an environment where breakthrough ideas can flourish.

2. **World-class computing power.** We harness cutting-edge computational resources to turn promising ideas into practical innovations. Our infrastructure enables researchers to test, validate, and scale their most ambitious ideas.

3. **Solutions that matter.** We tackle overlooked societal challenges, transforming theoretical insights into real-world impact. The goal isn't just to advance science—it's to solve problems that make a difference in people's lives.

4. **Bringing minds together.** We bring together exceptional researchers from around the world, creating an environment where ideas collide, combine, and evolve. Great minds don't just think alike—they think better together.

Why prioritise general-purpose AI?

This mission centers on trustworthy general-purpose AI—systems that can handle diverse tasks rather than single-purpose tools. It's a deliberate choice that deserves explanation, as it shapes everything from research priorities to potential impact.

General-purpose AI systems are rapidly becoming the Swiss Army knives of the digital world—tools that can handle an ever-expanding range of tasks without specialized retraining. What once required custom-built models is increasingly accomplished by general-purpose systems straight out of the box. We're already seeing this shift in practice: from catching software bugs to translating languages to diagnosing medical conditions, general-purpose models are matching or surpassing their specialized counterparts.

This evolution creates both opportunity and urgency. As the EU pushes for broader AI adoption across the economy, the need for reliable and safe systems becomes paramount. Yet we still lack the scientific foundations for developing truly trustworthy general-purpose AI—to this day, most advanced systems remain de facto black boxes, their learned decision-making processes still too complex for humans to understand. By focusing on this fundamental challenge, CERN for AI can catalyze two crucial developments: a competitive European AI ecosystem and accelerated adoption of more trustworthy AI to improve delivery of public services, like in healthcare, and boost new industrial uses of AI in manufacturing and beyond.

However, a focus on general-purpose AI shouldn't blind the EU to opportunities in specialized domains. Think of it like medical research: while the world needs broad-spectrum advances in human biology, the medical community also pursues targeted treatments for specific conditions. DeepMind's AlphaFold exemplifies the second approach—a specialized system that revolutionized biology by cracking the protein-folding problem. CERN for AI should maintain the flexibility to develop similar specialized models in areas promising exceptional societal returns, from climate science to robotics, from medicine to materials research.

4. Focus areas

Research principles

CERN for AI's success would depend on identifying and pursuing problems of a highly specific shape. To ensure focus and alignment with its mission, programme selection could be guided by high-level principles, including:

1. **Mission Alignment:** programmes should directly support CERN for AI's mission—either by advancing the [science of trustworthy general-purpose AI](#) or by accelerating AI applications that enrich society.

2. **Big bets:** programmes should aim for transformative outcomes. As an example, if only 10% (or some similarly small fraction) of CERN for AI's research programmes meet their goal, that should still be sufficient given their outsized impact.

3. **Additional impact:** programmes should address challenges unlikely to be pursued by academia, industry, or other sectors without CERN for AI's leadership. This ensures the organisation fills critical gaps and catalyses progress in areas that would otherwise remain underexplored.

Focus areas and programme examples

To establish a clear framework for the types of programmes CERN for AI should pursue, research efforts could be organised into four focus areas, as depicted in the figure below. Note that individual programmes may align with one or multiple themes, allowing for flexible, cross-cutting research approaches.

Research Themes

FOUNDATIONAL PROGRAMMES
Programmes focused on making general-purpose AI systems more reliable, interpretable and trustworthy

APPLIED RESEARCH PROGRAMMES
Programmes focused on developing market-ready applications that address critical societal needs

SCALING PROGRAMMES
Programmes focused on scaling up foundational research streams through large-scale compute and data

HARDWARE PROGRAMMES
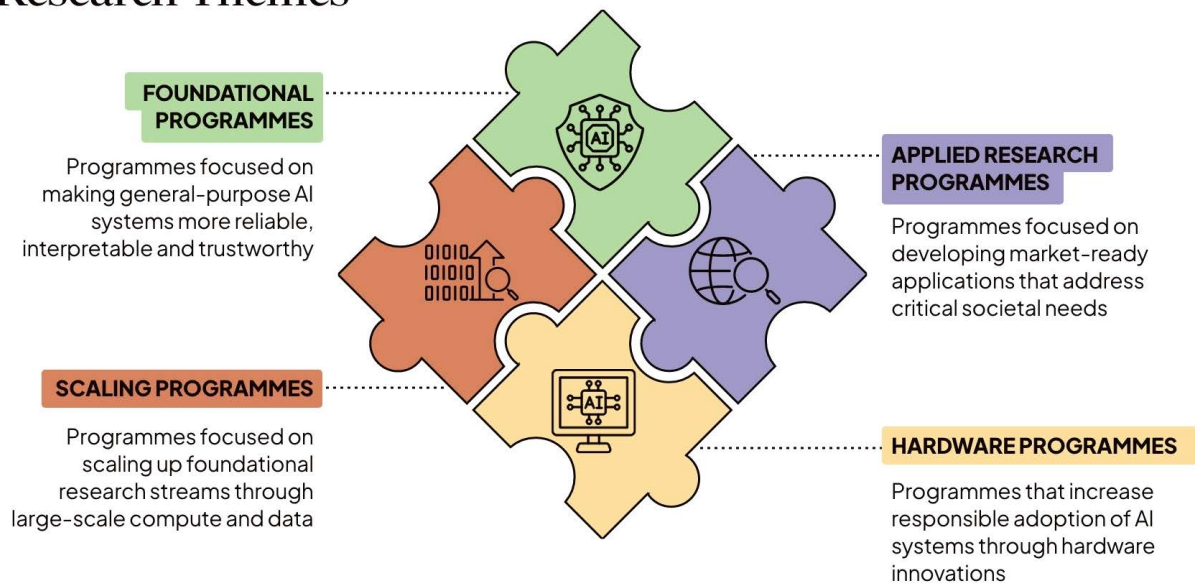Programmes that increase responsible adoption of AI systems through hardware innovations

Figure 3 – CERN for AI's four focus areas would strengthen each other

These four focus areas form a tightly interconnected ecosystem. Advances in hardware infrastructure enable the development and efficient execution of novel algorithms. These algorithms can then be scaled and deployed as foundations for applied research programs. In turn, breakthroughs in applied domains—such as materials science—often catalyze new hardware innovations. While this cyclical relationship may not manifest in every instance, it underscores the strategic importance of the proposed full-stack approach to research and development.

Another common theme in these focus areas is science. Many of the most promising societal applications of AI are in the sciences - think for instance about applying specialised models to help with material science to [invent more efficient](#) membranes and sorbents for direct carbon capture. Or about [speeding up](#) the discovery of new drugs and medicines. Simultaneously, these research themes focus on applying scientific methods to AI itself. CERN for AI has the potential to radically improve our knowledge of what goes on inside AI models - going from trial and error to principled understanding. In short, it's about bringing AI to science, and bringing science to AI.

Foundational programmes

Foundational programmes would focus on the fundamental science behind general-purpose AI, progressing the robustness, reliability, and transparency of AI systems. Examples include[2]:

- **Interpretability:** Developing tools to map and understand neural network behaviours, including fine-grained techniques like mechanistic interpretability and more coarse-grained methods for improving the faithfulness of chain-of-thought reasoning.

- **Bounded and verifiable architectures:** Creating neural architectures with mathematical guarantees, such as provable constraints on outputs and decision-making processes.[3]

- **Reliable and predictable reasoning systems:** Innovating novel architectures and data pipelines to enhance the reasoning robustness of general-purpose AI systems and reduce hallucination rates.

CERN for AI's leadership should maintain the flexibility to allocate resources across various programme themes. However, to reflect that foundational research is at least as important as the other three focus areas, CERN for AI should commit at least 25% of its resources—both compute and staff expenditures—to foundational programmes. Without this commitment, resources might naturally drift toward quick wins and immediate applications. CERN for AI is about making bold bets: programmes with disproportionate pay-offs when they succeed. In this light, foundational research isn't just another program track. It's the bedrock that makes breakthrough

---

[2] For more examples, see [a recent survey by IAPS](#) and discussion on [how to measure progress in making models safe](#).
[3] ARIA's [Safeguarded AI programme](#) would be an example of a programme that could fit this criteria.

innovation possible.

Scaling programmes

Scaling programmes would implement the trustworthiness-enhancing methods at scale, opening the door to societal impact. They could also focus on training domain-specific models from scratch or on curating large, high-quality datasets to train AI systems effectively. Examples include:

- **Trustworthy foundation models:** Building large-scale, general-purpose AI models suitable for sectors where trustworthiness, transparency, and accountability are critical. These models act as versatile platforms for more specialised applications.

- **Automated material science:** Developing large-scale, domain-specific models to address scientific challenges in material science, like predicting the characteristics of novel materials. This would provide the foundation for sector-specific work, for instance to create more efficient batteries.

- **European zero-friction health dataset:** Developing a privacy-preserved, regulatory-cleared health dataset that incorporates anonymized records of large numbers of European citizens, ensuring that personal data cannot be inferred. This dataset would facilitate training AI systems to better understand rare diseases, accelerate clinical trials, and improve healthcare outcomes across the EU.[4]

Applied programmes

Applied programmes would focus on developing AI systems tailored to address critical societal challenges through market-ready solutions. They would often build upon systems created in the scaling programmes. Examples include:

- **AI for cybersecurity:** Leveraging foundational models to create trustworthy cybersecurity solutions that safeguard digital infrastructures.

- **Automated lab robotics:** Designing robotic systems capable of executing intricate laboratory procedures with superhuman precision, such as blood sample preparation, tissue analysis, and other complex diagnostic tasks.

---

[4] This has significant overlaps with the European Health data space, an already existing project. It should therefore be seen more as an example of the sort of programme that CERN for AI should pursue, rather than as a direct recommendation.

- **Net–zero grid managers:** Developing resilient AI systems to efficiently manage 100% renewable power grids on a continental scale. These systems would be able to coordinate millions of distributed energy resources, storage systems, and flexible loads with millisecond-level precision, enhancing both grid efficiency and resiliency.

Hardware programmes

Hardware programmes would focus on creating the physical infrastructure necessary for trustworthy and responsible AI innovation and adoption. Examples include:

- **Energy–efficient computing solutions:** Designing specialised hardware architectures optimised for AI algorithms, significantly reducing energy consumption and improving sustainability.

- **AI–optimised sensor integration:** Developing hardware that embeds AI capabilities directly into sensor technologies, enabling real-time, edge-level data processing to enhance privacy and reduce latency.

- **Hardware–enabled oversight mechanisms:** Creating tamper-resistant, [hardware-based security features](#) that can help verify responsible use of AI hardware.

5. Research structure

CERN for AI's research structure cannot be a copy-paste

Frontier AI development demands an institutional architecture that can deliver both breakthrough innovation and trustworthy systems. While Europe's existing research institutions have served science well, none of its current models—academic, private sector, or public R&D—possess the necessary attributes for this dual challenge.

Limitations of current models

Universities possess extraordinary intellectual capital but face structural obstacles that inhibit transformative research. Their emphasis on publication metrics and departmental divisions creates artificial barriers to the cross-disciplinary collaboration that AI development requires. Moreover, rigid funding mechanisms and administrative overhead constrain the bold, long-term initiatives necessary for advancing trustworthy AI systems.

Meanwhile, European industry is confronted with distinct challenges to drive AI innovation. Market fragmentation and risk-averse investment cultures naturally favor incremental improvements over transformative research[5]. This conservative approach, combined with AI development's winner-take-most dynamics[6], risks further widening the innovation gap between European firms and global leaders.

Finally, a consensus is emerging that Europe's current public research infrastructure requires fundamental reform. As Commissioner Zaharieva emphasized in her confirmation hearing, excessive bureaucracy currently hampers innovation—a point echoed in Mario Draghi's assessment calling for more streamlined R&D structures, like the ARPA-model.

The ARPA model's promise

The success of organisations like ARPA offers valuable lessons for a new approach. The ability of ARPA-type organisations to foster breakthrough innovations stems from a unique combination of ambitious objectives and operational autonomy. However, as Draghi highlights, the difference extends beyond organisational structure to resources: while the EU's European Innovation Council's Pathfinder instrument - the EU's main instrument to support technologies at low

---

[5] As Draghi notes: In 2021, EU companies spent about half as much on R&I as share of GDP as US companies – around EUR 270 billion – a gap driven by much higher investment rates in the US tech sector.
[6] While the relationship between investment and capabilities might have weakened in the new test-time compute paradigm, it is likely that the relationship between capabilities and market share will remain strong.

readiness levels - commands just €256 million for 2024, DARPA operates with $4.1 billion, supplemented by $2 billion for other ARPA agencies.

Some European countries have already begun adapting the ARPA framework to their unique context and have put serious funding behind these efforts. The establishment of the UK's ARIA and Germany's SPRIN-D demonstrates the model's adaptability to European research ecosystems, with well-funded applications spanning climate science, robotics, and beyond[7].

A unique path forward must combine proven models with bold innovation. While the ARPA framework offers valuable lessons in fostering breakthrough research, Europe's AI institution must chart its own course. Drawing inspiration from ARPA's emphasis on operational autonomy and ambitious objectives, while learning from European adaptations like ARIA and SPRIN-D, CERN for AI can pioneer an institutional architecture uniquely suited to advancing both scientific excellence and trustworthy AI development.

Key features of ARPA-type organisations

The formula behind the ARPA-model is elegantly simple: find exceptional talent, set ambitious goals, and remove unnecessary barriers. ARPA-type agencies are renowned for fostering high-risk, high-reward research, accelerating breakthrough technologies with transformative potential in areas of strategic importance, often using open science.

Key features of ARPA-type agencies include:

- **Empowered project leaders:** A focus on picking the best, often atypical, programme directors, who are empowered to enact their bold visions.

- **Time-bound missions:** Fixed 5–7 year limits to projects keeps projects on scope and ensures timely delivery.

- **Mission-driven focus:** A commitment to solving complex, high-priority challenges.

- **Flexibility and agility:** Operating with minimal red tape to allow fast, adaptive decision-making. This includes autonomy over procurement, hiring decisions and regranting/distribution of funding.

- **High-fisk, high-reward investments:** Backing bold ideas with significant transformative potential, but with a large chance of bearing little or no fruit

---

[7] It is still too early to assess the success of either ARIA or SPRIN-D based on outputs. ARIA was founded in 2023 and while SPRIN-D was founded in 2019, it only became "ARPA-type" in Dec 2023, after Germany passed the SPRIND Freedom Act, freeing it from its prior bureaucratic, financial and administrative burdens.

- **Public–private collaboration:** Working closely with academia, industry, and other stakeholders to bring innovations to market.

ARPA-agencies don't optimise for guaranteed wins or neat political victories. This is a purposeful choice. When CERN first set out to understand the universe's building blocks, it promised a vision to stop the post Second World War brain drain to America, and provide a force for unity in post-war Europe: an aspiration and not a definite result to deliver. When DeepMind aimed to "solve intelligence," it couldn't show a detailed roadmap but offered a vision to be brought to life. ARPA-type organisations acknowledge that breakthrough innovation requires such freedom to explore.

What CERN for AI could learn from ARPA-type institutions

A talent-first approach

CERN for AI's success would hinge on deep collaboration with universities, research institutions, and private companies, leveraging their expertise while providing the coordination, resources, and infrastructure necessary for open, large-scale innovation. By adopting an ARPA-style framework, CERN for AI could serve as a research coordinator: uniting diverse stakeholders and empowering researchers to address ambitious challenges that would otherwise remain out of reach.

There are a number of elements from ARPA-type organisations that would be natural for CERN for AI to adopt. For one, ARPAs are relatively flat organisations consisting of loosely structured programmes, each designed to tackle a distinct, highly ambitious problem. Programmes are led by Programme Directors—not your typical managers, but pragmatic dreamers that combine bold ideas with technical brilliance, coordination skills and exceptional drive[8]. They get remarkable freedom: awarding grants, setting challenges, and partnering with anyone who can help crack the problem. This is a perfect fit for the fast-moving field of AI, where competition is fierce. If CERN for AI gets tangled in traditional procurement or hiring limitations - a notable sore spot within existing EU institutions - it won't succeed.

Targeted, front-loaded governance

Such freedoms, however, need to be balanced with robust governance. Therefore, the UK's ARPA-type institution - the Advanced Research and Invention Agency (ARIA) - has introduced a rigorous front-loaded oversight process: only after a thorough selection process and a

---

[8] Nonprofit research organisation Speculative Technologies have a detailed explainer on characteristics that make a good programme director in an ARPA like context.

three-month incubation period, where Programme Director defend their vision before organisational leaders and independent experts, can a programme become fully operational. Programme Directors need to demonstrate a clear strategy, resource requirements, success metrics and programme checkpoints before they gain significant operational flexibility. This flexibility is further balanced by external reviews and transparent reporting practices at agreed milestone checkpoints to ensure responsible innovation. Given the often sensitive nature of AI research, and large resource requirements in the form of compute, it makes sense for CERN for AI to adopt a similar system.

Preventing silos with systems engineers

A critical challenge for CERN for AI is ensuring its programs remain deeply connected to each other and to real-world needs. One concrete way to address this, could be to leverage both connections between Programme Directors and so-called 'Systems Engineers', as they are known in the metascience community.

Systems Engineers, technically trained professionals inspired by the success of institutions like Bell Labs (described here by innovation history researcher, Eric Gilliam), could support CERN for AI by building crucial bridges between research and practical impact. These individuals would maintain a dual focus: tracking emerging AI capabilities while identifying concrete applications and societal needs. Where Programme Directors concentrate on fostering their specific, narrow research networks, Systems Engineers would maintain a holistic, wide view of the organisation's work and its potential impact.

By actively engaging with society, academia, and industry—from major AI companies to European SMEs—Systems Engineers could ensure CERN for AI's research translates into meaningful outcomes for Europe's economic security and prosperity. They'd identify high-potential problems where AI can enhance societal resilience and economic competitiveness, assess technical feasibility of proposed directions, coordinate between research teams and stakeholders, and guide the transformation of research into practical applications.

Systems Engineers could also play a vital role in resolving coordination challenges that would otherwise stall innovation. Consider the development of novel AI algorithms that require fundamentally different hardware architectures[9]: researchers might hesitate to develop such algorithms without available hardware, while hardware engineers lack incentive to create new AI chips without demonstrated algorithmic needs. Systems Engineers can break this deadlock by coordinating across research strands within CERN for AI.

---

[9] For instance, algorithms that are optimised for so-called 'neuromorphic chips'.

The importance of this bridging function would likely have to be reflected in resource allocation. Drawing from DARPA's experience, where [7–9% of R&D spending](#) in 2017 went to Systems Engineers, CERN for AI would need to maintain a substantial Systems Engineering team to maximize its impact and ensure effective integration across programs.

Finetuning the ARPA model

The need for in-house teams

While the distributed ARPA model excels at rapid ground-up innovation, certain research streams inside CERN for AI are better served by a different approach. Some programmes will generate sensitive insights with dual-use potential—advances in AI cybersecurity, for instance, could be repurposed for offensive capabilities. Others require the sustained collaboration and knowledge-building that only comes from dedicated teams working together long-term, such as scaling up foundation models or integrating multiple crowdsourced ideas. For these cases, CERN for AI could employ a second research track based on [Focused Research Organization](#) ([FRO](#)) principles: concentrated in-house teams tackling specific challenges, often in secure environments.

The two tracks could operate distinctly. Distributed, ARPA-style programmes would leverage networks of partner institutions, with Programme Directors acting as network builders who coordinate collaborations across Europe. In-house, FRO-style programmes would focus on tightly integrated teams, with Programme Directors who excel at execution and who would often be supported by dedicated security liaisons.

Security requirements would be determined independently under this framework. While programmes in CERN for AI's lower security tier ([as described in CFG's 2024 report](#)) would typically follow the distributed model, some may require in-house teams based on their specific problem shape. Upper-tier programmes, handling sensitive insights with dual-use potential, would almost always demand the in-house approach with very limited external collaboration.

A tiered approach to **security**

MEDIUM/HIGH SECURITY
In-house teams working with security-sensitive research and information

LOW SECURITY
Open research and information, often through distributed research networks

SL3–4

SL1–2
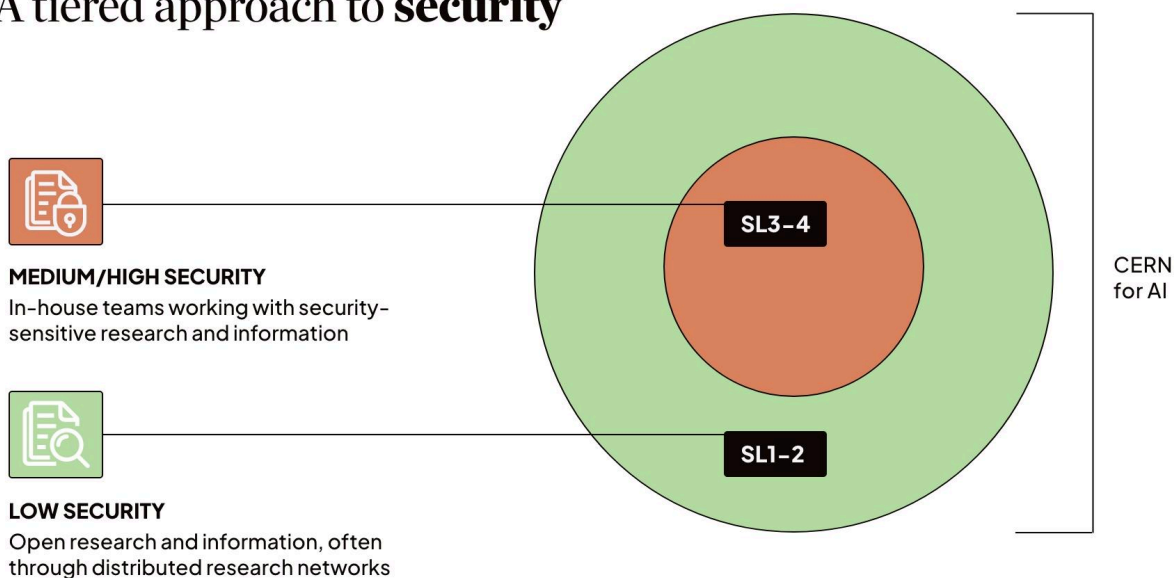
CERN
for AI

Figure 4 - Programmes in CERN for AI's upper security tier would work with in-house teams

This two-track structure could solve two key problems at once. First, it ensures that several member countries can benefit from having CERN for AI-related research within their borders. While sensitive in-house work needs a central facility with very specific locational requirements, distributed programs can spread across Europe through partner networks and multiple compute hubs[10]. This means more countries can benefit from CERN for AI funding without compromising on a single, optimal location.

Second, it provides natural access control. Think of it like a building with public areas and secure zones—partners can engage with lower-tier programs based on their trust level and security clearance (crucial for managing geopolitical sensitivities and industrial espionage concerns), while sensitive research stays protected. Some partners browse the lobby, while only the most trusted get keys to the vault.

Responsible diffusion and commercialisation

CERN for AI's impact must extend beyond scientific breakthroughs to deliver tangible benefits for European society, from democratizing AI capabilities to catalyzing broader technological innovation. By creating an environment that attracts and retains exceptional talent, the initiative

---

[10] Considerations of locational factors for a central hub and distributed compute hubs have been explored in the CFG's previous report (page 17).

can foster thriving AI and deep tech ecosystems across the continent. As AI evolves into the 21st century's foundational general-purpose technology—much like the internet before it—it will become a crucial platform for advancing emerging fields like biotechnology.

In order for Europe to successfully build on such a foundation, CERN for AI's research must diffuse beyond the walls of the institution and out into the real world. ARPA-type organisations often achieve this by using open science and specific IP-sharing agreements with partner or spin-off companies. However, the sensitive nature of certain AI technologies demands that CERN for AI has additional tools at its disposal. This is particularly true for the largest and most advanced foundation models it would create.

Model access and distribution

One way CERN for AI's could achieve responsible diffusion of its most advanced (foundation) models, would be through a flexible, three-stage framework, that is inspired by a [recent report](#) from the Centre for the Governance of AI on the risks, benefits and alternatives to open-sourcing in AI.

Initially, the institution could prioritize open-sourcing its models, enabling smaller firms and startups to build upon its research without the burden of extensive resource requirements. This democratization of access would spark innovation across the ecosystem, particularly benefiting organisations that couldn't otherwise develop advanced AI capabilities.

As model capabilities advance, so do the risks - for instance in the realms of cyber and synthetic biology. For models that pose too grave a risk when being open sourced, distribution could shift to a secure licensing framework. Qualified firms would gain access to model weights in exchange for program funding, enabling them to develop products and services while CERN for AI maintains its research focus. These partners could offer everything from customer-facing applications to specialized fine-tuning services through secure APIs[11].

For models that reach capability thresholds presenting extreme security concerns, CERN for AI could transition again to providing direct inference services, enabled by a new, dedicated product team. This controlled deployment would ensure responsible AI development while generating revenue that could exceed program funding requirements, creating opportunities for new benefit-sharing mechanisms among member countries.

---

[11] API - Application Programming Interfaces - let customers make use of AI models that are run on the cloud by the company providing the service.

Putting it all together

Figure 7 shows how all the elements described in this chapter could work together in practice, during a programme's full life-cycle. A programme would typically start off with either an open call for proposals, or a predefined opportunity space. After a successful application, a probational Programme Director would refine and defend their Programme Thesis during a three-month incubation period. If this incubation period is concluded positively, CERN for AI leadership would place the programme in either the upper or lower security tier. Based on the security level and the programme's other needs, it would then go on to live either as an ARPA-style coordinated research effort, or an FRO-style in-house team. At the end of a programme's 5 to 7-year lifecycle, the focus would shift to technology diffusion, with key breakthroughs typically being open-sourced or licensed to partner or spin-off companies, and in rare cases being offered to consumers and businesses by a dedicated product team inside CERN for AI.

CERN FOR AI

Applies to → 1) Existing Opportunity Spaces (predefined themes/project ideas) or
2) Open Application Call

Aspiring Programme Director

Programme Director Selection

Evaluation based on:
- ☑ Technical vision & expertise
- ☑ Leadership capability
- ☑ Track record of execution
- ☑ Strategic alignment with CERN for AI mission
- ☑ Ability to coordinate diverse stakeholders
- ☑ Understanding of safety/security implications

Approves

CERN for AI Support Staff

CERN for AI Research Lead

Successful application

Probational Programme Director

Creates

Advises

Expert Panel

Advises

CERN for AI leadership team

Refinement of Program Thesis (~3 months)
- Work with support staff to develop detailed program vision
- Map technical risks, resources, and capabilities needed
- Consult domain experts and stakeholders
- Develop milestones and success metrics
- Explore appropriate tier placement
- Define funding structure and transition strategy
- Explore transition/impact pathways

One internal team of CERN for AI employees

Approves

Programme Approval

Key Assessment Criteria:
- ☑ Strategic fit with CERN for AI mission
- ☑ Technical feasibility and innovation
- ☑ Resource requirements justified
- ☑ Clear success metrics established
- ☑ Appropriate tier placement

Hires & Oversees

FRO-style
- Internal research teams within CERN for AI
- Tightly coordinated workstreams
- Clear execution pathways
- Enhanced security protocols
- Builds institutional capabilities

By default

Uncommon

Programme is placed in either Upper or Lower security Tier

ARPA-style
- External research through partner institutions
- Multiple parallel exploration tracks
- Focus on novel approaches and early-stage work
- Build diverse research networks

By default

Very uncommon

Lower Tier / **Upper Tier**

Technology diffusion

Coordinates and integrates

The general public

Multiple external groups pursuing different research pathways

Figure 5 – Schematic representation of a programme's possible lifecycle

6. Governance Structure

CERN for AI has a dual mission when it comes to trust: create AI systems that are genuinely reliable, and build them in a way that the public can verify. The first goal should shape its research; the second should shape its governance - i.e. how decisions are made. When it comes to governance, CERN for AI could work under two guiding principles: transparency and accountability.

Why transparency matters

ARPA-type organisations give unprecedented autonomy to their researchers. As [recognised](#) by ARIA, such autonomy must be balanced by transparency. While some programs may need secrecy for security reasons, these should be exceptions, not the rule.

Transparency would serve CERN for AI at multiple levels. First, it lets management spot and fix problems early. Second, it enables member countries to ensure CERN for AI is on track. And third, it lets the public verify that member countries are staying true to their original goals.

Within CERN for AI, regular team updates and independent internal evaluations (see later sections in this chapter, for instance on the possible 'Mission Alignment Board' for more information) will keep information flowing. But papers and policies only go so far–CERN for AI would need a culture where sharing information is the default, where asking questions is encouraged, and where teams naturally collaborate. The autonomy CERN for AI would grant researchers would have to come with equal responsibility. Hiding information without clear security reasons cannot simply be discouraged—it should be a red flag that could end programs.

CERN for AI can build outward-focused transparency into its DNA through the [concrete tools](#) proposed by prominent AI researchers Dean W. Ball and Daniel Kokotajlo: whistleblower protections, public safety assessments, and clear disclosure of AI capabilities and goals - also of models that are still in training. Public progress reports and public risk communication can further strengthen transparency and help [clarify to the public](#) why and how a CERN for AI needs to take calculated risks. By managing expectations around high-risk investment, CERN for AI can keep the public on board through both successes and failures.

That being said, transparency should be applied where it actually matters to the public. Many public institutions [become ineffective](#) because they have excessive transparency and auditing at the very granular level, and a lack of transparency and accountability at the highest, strategic

levels. In doing this, they end up with poor outcomes: workers are slowed by bureaucracy while oversight is lacking for the most important, strategic decisions. It may be fully clear how taxpayer money was spent within a specific sub-project, but not why the larger programme was started in the first place. Worse, after wrong turns, leadership cannot be properly held accountable. ARPA-type institutions are not immune to that either—in fact, in its later years, DARPA's performance lowered substantially because of added bureaucracy. This highlights the need for clear prioritisation and effective public communication.

More speculatively, AI itself might help reimagine transparency. Just as privacy-preserving analytics lets organisations learn from sensitive data without compromising privacy, new AI oversight tools could help improve transparency while maintaining security[12]. CERN for AI could pioneer not just trustworthy AI systems, but also new ways to make their development visible and accountable to all.

Public accountability should be informed and targeted

If CERN for AI is to build AI systems in a way that the general public can trust and verify, it should be governed by representatives from participating countries to ensure democratic control. Under such a model, private sector actors would still have a role to play in the institution's governance, but in an advisory rather than executive capacity. This preserves public interest and accountability while leveraging crucial industry expertise.

Effective governance requires striking a delicate balance between oversight and operational autonomy. Excessive political interference from member countries could inhibit the institution's functioning, as has been observed within government agencies like the US Environmental Protection Agency and, in Europe, in the European Innovation Council. CERN for AI cannot fall in the same trap. If, for example, countries wielded significant influence over program funding, they might favour projects benefiting their domestic industries over initiatives serving the broader mission - all such incentive structures must be carefully considered.

The solution likely lies in targeted public accountability: member state representatives should shape high-level strategy while maintaining limited influence over day-to-day operations. The public should be able to hold representatives of member countries accountable for high-level strategic decisions, but not for day-to-day decisions. The European Space Agency (ESA) offers an instructive model, with its clear division between the Council's strategic oversight and the Director General's operational authority.

To foster public trust and to make informed decisions on complex technical matters, member country representatives need access to independent expert advisors who can provide impartial

---

[12] See Trask et al (2020) and Rahaman et al (2024) for examples of early work in this direction

analysis of proposed initiatives. This independence prevents overreliance on internal staff and potential conflicts of interest. The success of this advisory system once again depends on operational transparency—external experts must have sufficient visibility into CERN for AI's activities to offer meaningful guidance.

Organisational structure

For CERN for AI's organigram, we again draw inspiration from ARIA's relatively flat organisational structure. Within ARIA, Programme Directors are placed directly below the leadership team, with no intermediate hierarchies. This ensures that Programme Directors retain their autonomy. Further support, such as Directors for hiring, procurement and communications are placed in their own organisational column. Programme panels and other advisors assist both Programme Directors, leadership, and the ARIA's board.

In the context of CERN for AI, many of these structural elements could be adopted seamlessly. However, given the often sensitive nature of advanced AI, we also propose a handful of changes. First and foremost, we argue that CERN for AI's board should be composed of national political representatives, rather than scientists, industry leaders or staff from European institutions (for how collective decisions could be made between these political representatives, see Chapter 7 on Membership Policies). AI is becoming increasingly political – and even geopolitical. It seems better to explicitly reflect this in CERN for AI's structure, than to create a politically neutral face, opening the way for backdoor influence. That said, CERN for AI's leadership should have full autonomy over day-to-day decisions and political influence should be limited to high-stakes strategic matters.

To make sure that political representatives can make informed judgements, we propose that two independent advisory boards under the names Scaling and Dissemination and Mission Alignment report directly to the board. We also argue that CERN for AI should develop two distinct frameworks to ensure the safety of deployed models and to lay the foundation for a governance model that evolves with the technology itself (see the sections on the Safety and Security Framework and the Responsible Governance Framework later in this chapter).
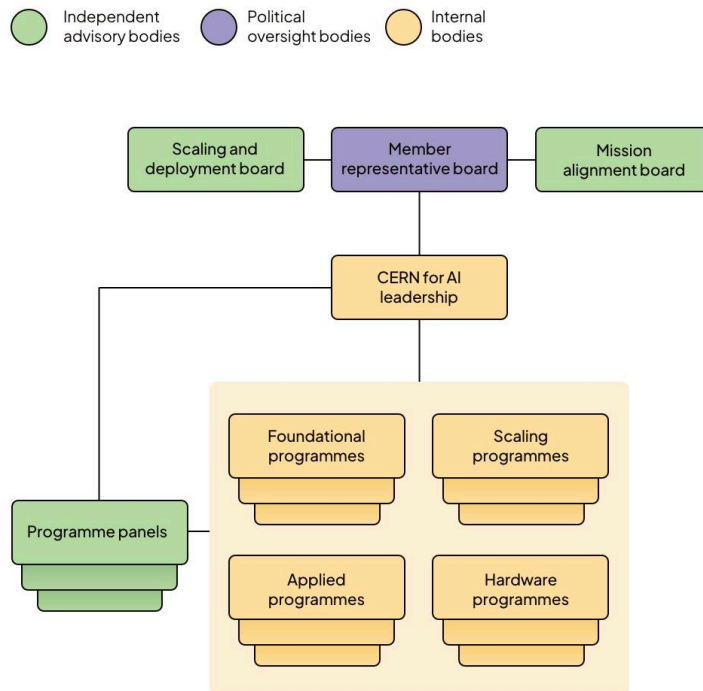
CERN for AI: Organigram

Figure 6 - A proposed organisational structure for CERN for AI

Breakdown of the different bodies and their roles

Below we break down the different bodies presented in the organigram presented in Figure 7 and provide a more detailed description of their possible roles.

Member Representative Board

Under the envisaged structure, the Member Representative Board would comprise political representatives from countries participating in CERN for AI, with a potential additional seat for a European Commission representative that serves as a tie-breaker when needed. Each state would nominate one representative, who must be approved by supermajority vote to ensure both efficiency and broad support.

Voting power within the board could be partly weighted by each country's investment in CERN for AI, with major decisions requiring a weighted supermajority[13]. The Member Representative board would focus on critical high-level strategic matters and carry several key responsibilities. These include appointing and, if necessary, replacing the Executive Director (the head of CERN for AI's leadership team), as well as establishing the initial members of both the Mission Alignment Board and the Scaling and Deployment Board. This board would also be responsible for defining membership terms, overseeing organisational expansion, and approving significant scaling and dissemination decisions proposed by the Scaling and Deployment Board.

While CERN for AI leadership would provide quarterly reports to the board, day-to-day operations would remain outside its purview. Although the Member Representative Board would operate independently without formal external oversight, it should maintain a fundamental obligation to serve the interests of citizens from all participating nations - public accountability of advanced AI models is, after all, a central aim of CERN for AI, so that principle should extend to the organisation itself.

Mission Alignment Board

The Mission Alignment Board's key responsibilities would encompass several crucial areas. It would perform targeted internal audits to enhance organisational transparency (focused on the most strategic matters, not on day-to-day processes), with portions of these evaluations made public and protected from management influence. The feedback loops this creates can enable CERN for AI to quickly respond to changing societal needs, market realities, and policy priorities. The Mission Alignment Board could also develop the Responsible Governance Plan (see the dedicated section below), which would require approval from both the Member Representative Board and Executive Director. Additionally, it would advise the Director of Research on initial Programme selection.

To ensure both stability and appropriate power distribution, Mission Alignment Board membership could follow a staggered rotation system. Here, members would propose their own successors, subject to Member Representative Board approval, allowing for seamless transitions while maintaining institutional knowledge.

Scaling and Deployment Board

The Scaling and Deployment Board would consist of independent domain experts operating outside the Executive Director's authority, complementing the Mission Alignment Board's

---

[13] Similar rules are employed by several European institutions, such as the European Investment Bank, European Commission via the European Stability Mechanism, or Joint Undertakings such as EuroHPC or SESAR for some decisions.

scientific and ethical oversight by focusing on practical implementation of CERN for AI's technologies.

This board would serve several essential functions in CERN for AI's governance structure. It would offer advisory recommendations to both the Member Representative Board and executive leadership regarding program scaling and model proliferation. The Scaling and Deployment board would also collaborate with executive leadership to develop The Safety and Security Framework, while conducting independent audits of the framework's implementation.

While the Mission Alignment Board addresses high-level, long-term strategic concerns, the Scaling and Deployment Board would concentrate on practical oversight and implementation guidance in areas where safety and security needs are high. Members could be appointed by the Member Representative Board through a staggered selection process, ensuring both operational continuity and appropriate distribution of authority.

CERN for AI leadership

CERN for AI's initial leadership team could consist of five key director positions, collectively responsible for all internal decisions: an Executive Director, a Director of Research, a Director of Infrastructure, a Director of Security, and a Director of Operations. Here, the Executive Director would hold overriding power and the authority to appoint and dismiss other leadership positions (for operational efficiency it's tantamount that the Executive Director can assemble their own team). The leadership team structure should allow for future expansion, potentially including roles like a Director of Product, and should probably permit co-director arrangements for particularly demanding positions such as Research and Operations. Each directorship would include dedicated support staff. Crucially, CERN for AI leadership would be restricted from establishing additional hierarchical layers without permission from the Member Representative Board, ensuring that Programme Directors retain their autonomy and that bureaucratic complexity is minimised.

Programme Panels

Each Programme Panel would comprise independent, multidisciplinary domain experts carefully selected to match the specific focus and requirements of its designated programme. The Director of Research would appoint these panels following the selection of a Programme Director.

During the initial three-month refinement period[14], Programme Panels would work closely with Programme Directors to provide expert guidance and support. This crucial phase would ensure programme objectives and methodologies are both clearly defined and achievable. The refinement period would culminate in a formal recommendation to the CERN for AI leadership regarding programme approval, accompanied by detailed suggestions for enhancing the programme plan and addressing potential challenges.

Following programme approval, Programme Panels could maintain an active advisory role throughout the programme's duration. In this capacity, they'd serve the Director of Research by conducting independent evaluations and providing strategic recommendations. This ongoing oversight would help ensure programmes maintain their intended course, align with organisational objectives, and effectively adapt to new insights and changing circumstances.

Forward-looking frameworks

Given the immense pace of progress in AI, we propose that CERN for AI develops a forward-looking framework to guide scaling and deployment decisions, as well as the evolution of the institution's governance itself. This proactive rather than reactive approach can help the organisation prepare for challenges we see looming on the horizon.

The Safety and Security Framework

First of all, CERN for AI should establish a comprehensive Safety and Security Framework as is currently prescribed under the Codes of Practice for the EU AI Act (which is still in the drafting phase). This framework could play a critical role in managing the risks associated with developing increasingly capable general-purpose AI systems by defining deployment and security standards that adapt as AI capabilities surpass predefined thresholds. For example, once systems exceed a specific capability level in synthetic biology, additional real-life monitoring may become necessary, and enhanced security measures may be required to prevent issues such as model theft.

This framework would specify the risk management policies CERN for AI employs to ensure compliance with these deployment and security standards. No standard risk management framework for advanced AI exists yet, so it should build on the existing most advanced frameworks, such as Anthropic's Responsible Scaling Policy, OpenAI's Preparedness Framework and Google Deepmind's Frontier Safety Framework[15]. If any industry-wide standards are

---

[14] This 3–month refinement period is a standard structure for ARPA-type programmes, see e.g. reference to in the similar ARIA programme proposal on page 5

[15] It is telling that the most advanced frameworks that currently exist are still voluntary, and that they are led by industry players.

established – for instance, if [AI companies' voluntary safety commitments](#) are consolidated at the [AI Action Summit](#) – CERN for AI would clearly need to adopt and build on those as well. The goal here should be to create an example for the rest of the AI field, to spur a race to the top, rather than a race to the bottom.

Development of the Safety and Security framework could be led by CERN for AI leadership, together with the Scaling and Deployment Board. Final approval would naturally rest with the Member Representative Board. Once implemented, the Scaling and Deployment Board and Director of Security could be jointly responsible for ensuring the framework is adhered to in daily operations, maintaining rigorous oversight to uphold safety and security as AI capabilities evolve.

The Responsible Governance Plan

We see the Responsible Governance Plan as a forward-looking framework designed to adapt CERN for AI's governance structure itself. While the Safety and Security Framework would focus on deployment and security standards, the Responsible Governance Plan would seek to ensure that the organisation's internal oversight evolves appropriately to maintain accountability, safety, and alignment with its mission. While such a framework may sound superfluous, we think the relentless pace of progress in AI necessitates such a proactive approach to governance.

As AI capabilities improve, new checks and balances may become essential, even if such measures come at the expense of operational efficiency. For instance, the stakes of a large model release may grow so large that additional oversight is necessary. As another example, AI capabilities may eventually reach such extreme dual-use capabilities that CERN for AI's internal security must be completely overhauled, necessitating wide security clearances, increased vetting or more centralisation. Such changes touch on the institution's entire governance, and would likely have to be implemented under tremendous pressure.

The Responsible Governance Plan can provide CERN for AI with a playbook to fall back on during such circumstances by establishing predefined triggers for governance updates, specifying measures that could be implemented when such thresholds are crossed. This if-then approach would allow CERN for AI to function efficiently under its existing structure while avoiding unnecessary bureaucracy. However, when circumstances demand additional checks and balances or new organisational bodies, the plan could ensure that CERN for AI is prepared to act decisively and effectively. This precommitment could strengthen the organisation's resilience and accountability, ensuring its governance remains robust as the field of AI continues to evolve.

Development of the Responsible Governance Plan could be led by the Mission Alignment Board, while final approval could rest with the Member Representative Board.

7. Membership policies

CERN for AI's success depends on both developing and driving adoption of trustworthy AI systems. A strategic approach to improving adoption is to involve partner countries and organisations during the development stage itself. This creates shared ownership and enables knowledge transfer as scientists and engineers bring techniques back to their home institutions. This collaborative model has proven successful historically—it powered both the [Network of AI Safety Institutes](#) and the original CERN, where innovations like the World Wide Web [emerged in 1989](#) to meet physicists' needs for international information sharing. Additionally, expanding membership can bring crucial new funding streams, increasing the resources available to achieve CERN for AI's mission.

However, while a broad, expanding coalition is appealing in theory, implementation presents significant strategic challenges. The institution must carefully balance European competitiveness and security interests when considering non-EU members. Critical considerations include preventing sensitive technology transfer to geopolitical competitors and malicious non-state actors. And perhaps most importantly, CERN for AI must maintain its operational effectiveness and avoid institutional paralysis as its membership grows.

An international institution that advances European goals

In order to justify its long term existence, CERN for AI's should not only advance the science of trustworthy AI, it should simultaneously strengthen the EU's economic performance and security. Including select non-EU members can enhance rather than compromise these objectives. Recent precedents demonstrate this balance: Canada's [€99.3 million contribution](#) to join Horizon Europe boosted EU scientific capabilities, while the [UK's participation](#) in EuroHPC will advance European supercomputing goals.

By developing resilience-enhancing technologies in a controlled environment, CERN for AI can reduce EU dependence on foreign solutions without sacrificing international collaboration. Including trusted non-EU partners can actually strengthen security outcomes by building mutual confidence and enabling deeper cooperation. However, this requires implementing strict controls on which countries can access sensitive workstreams.

The need for founding members outside the EU

CERN for AI's founding members should comprise a select set of trusted non-EU partners, particularly EEA countries and strategic Horizon Europe collaborators. Three non-EEA countries stand out as natural founding members: the UK, Switzerland, and Canada. Each brings unique

strengths—London, Zurich, and Toronto are major AI hubs, while the UK and Canada host well-resourced AI Safety Institutes. Switzerland's central European location and research excellence (hosting, among other, some excellent universities) make it an ideal partner. Furthermore, Canada is home to a world-leading academic institute in AI through Mila.

These partnerships offer strategic advantages beyond technical collaboration. Canada can strengthen Transatlantic ties, while all three nations can provide substantial funding. The UK partnership is particularly valuable in the wake of post-Brexit tensions. The UK and EU share common challenges—both host world-class AI talent but struggle to build competitive domestic industries. Moreover, The UK AI Safety Institute (UK AISI) serves as a great model for CERN for AI when it comes to creating 'a startup within government'. While the EU AI Office has struggled to attract renowned leadership and technical talent, the UK AISI managed to quickly cut through red tape. Finally, with renewed American isolationism on the horizon, CERN for AI presents a timely opportunity to reinvigorate EU-UK cooperation, building on their history of close collaboration.

Tiered membership enables future broadening

While full membership should be limited to trusted allies, CERN for AI's impact could depend on broader international engagement. A tiered membership structure offers a potential solution to further expansion. Full members would receive voting rights and access to upper-tier security models, while partial members would be restricted to lower-security workstreams, such as cloud-run products and non-sensitive foundational research. This approach is similar to the original CERN's membership model that accommodates full members, associate members, and observers. It enables wider collaboration while protecting critical technologies.

Membership at either level would require financial contribution—with full members paying substantially more—and adherence to predetermined criteria. These rules could potentially also cover fundamental values like democracy, human rights, and rule of law, with stricter standards for full members.

The tiered framework can also be extended to private entities: companies from full-member countries could be denied access if they have concerning ties to non-trusted governments or are put on European sanctions lists. Conversely, academic institutions and businesses from non-member countries could still participate in specific programs under strict security protocols.

Expansion requires precedent and careful protocols

CERN for AI's expansion policy must navigate three critical challenges: avoiding overly restrictive barriers that exclude natural partners like the UK and Canada; preventing dilution of mission through overly lenient admission standards; and maintaining efficient decision-making as membership grows.

While robust admission processes are essential, their effectiveness can only be proven through implementation. Including select non-EU founding members serves two crucial purposes: demonstrating the practicality of admission criteria and establishing precedents for future expansion. These early members can showcase membership benefits and encourage broader participation over time.

To facilitate efficient expansion, CERN for AI should adopt streamlined procedures. One promising approach is the champion system, where an existing member serves as the primary contact for aspiring members. This model, successfully employed in Singapore's [Digital Economy Partnership Agreement](#), enables focused bilateral discussions while preserving collective decision-making through final membership votes. Strategic decisions, including operational changes and expansion policies, could require a supermajority approval (with the exact threshold subject to later negotiations among founding members) from both the full membership and CERN for AI's founding members. This structure could preserve the founders' ability to protect the original mission while enabling measured evolution.

Equally important is the ability to address member misconduct. Clear protocols for disciplining and, in extreme cases, expelling members are essential safeguards that are often insufficiently embedded within existing international institutions[16]. These measures could range from temporary suspension of voting rights to restricted access to sensitive technologies, culminating in potential expulsion for serious breaches like unauthorized technology transfer or deliberate obstruction of CERN for AI's mission. This accountability framework would ensure that membership remains contingent on continued alignment with the institution's goals and security requirements, not just initial admission criteria.

---

[16] The most obvious example here would be the UN, where Russian veto power is a thorn in the eye of many.

8. Funding

CFG' 2024 report argued that a successful CERN for AI requires €30 to €35 billion in funding over the first three years, of which approximately €25 billion would be directed toward datacenter investments. Clearly, securing such a substantial sum of funds presents a challenge. This section elaborates on key funding considerations and outlines potential funding sources.

An investment beyond AI

First of all, it is important to realise that funding for CERN for AI would represent more than just an investment in AI products — it would constitute a broader investment in Europe's future. The initiative could advance climate mitigation, enhance European security, and catalyze regional economic growth. On a deeper level, CERN for AI could demonstrate Europe's ability to overcome its historical barriers in technological innovation, implement the Draghi report's vision and signal that Europe's committed to building robust high-tech ecosystems. While the proposed €30–€35 billion investment is large, it represents just a fraction of the annual €34–114 billion in public funding that the EPRS estimates Europe needs to become competitive in high-tech digital innovation.

Breakdown of potential funding sources

CERN for AI's funding strategy would likely have to encompass both core, upfront funding and variable funding streams:

Core, upfront funding

The foundation and majority of CERN for AI's initial funding must be secure, upfront commitments that enable confident investment, grant allocation, and talent acquisition. Primary sources include:

- **The European Union:** The current Multiannual Financial Framework (MFF) still offers opportunities to redirect substantial resources towards a CERN for AI, with substantially more freedom in the next MFF.

- **National governments:** Member countries will likely have to contribute significantly, particularly in CERN for AI's first years. Countries hosting compute hubs or the central talent hub could provide enhanced financial support, for instance by financing their own infrastructure.

- **Private sector partners:** Datacenter providers bring essential expertise for European infrastructure development. Government partnerships are attractive to these providers, potentially enabling favorable long-term pricing agreements. Additionally, a select group of other private sector actors may invest upfront, particularly those poised to benefit from CERN for AI's secure compute facilities, such as semiconductor or pharmaceutical companies requiring protected AI solutions.

Variable funding

Three key mechanisms can generate additional, variable funding streams. While these budget sources are less predictable, necessitating conservative budgeting, they offer potential for growth throughout CERN for AI's lifespan:

- **Private programme-specific funds:** Private sector actors may fund specific programmes aligned with their interests, often through public-private partnerships that can also include staff and/or data contributions in exchange for partial IP rights.

- **Model licensing:** Rather than serving its own models, CERN for AI should aim to license its non-open source models to trusted companies for customer inference operations.

- **Compute provision:** CERN for AI leadership should optimize server utilization by rental of excess compute capacity to external private actors during periods of oversupply.
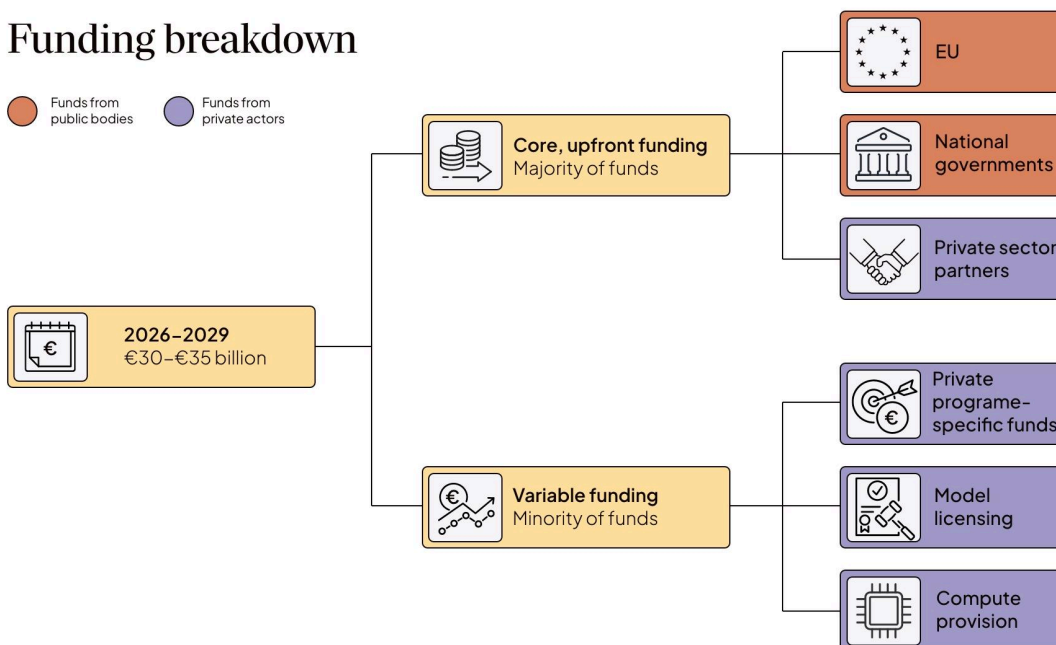
Funding breakdown

Figure 7 - CERN for AI's funding breakdown

Long-term funding is essential

While this funding structure addresses the first three years, CERN for AI requires sustained long-term funding, particularly given AI chips' [four-year average lifespan](#) and increasing competition from [American hyperscalers for talent](#). The initial €35 billion would support the first round of ARPA-style programmes but cannot sustain operations indefinitely. Long-term commitments are thus essential. Inspiration can once again be taken from the original CERN, where national member countries commit to [financing CERN for at least 7 years](#) (after which they are in theory free to withdraw).

9. Legal framework

In the previous chapters, we've laid out a detailed potential design for CERN for AI. CERN for AI's legal foundation would have to accommodate all these elements. In practice, our analysis suggests that this translates to the following legal prerequisites.

1. **Flexibility in hiring to attract top talent.** AI salaries are highly competitive, while public organisations often face rigid salary structures. Though much talent originates from the EU, some roles may require external hiring, as seen when ARIA (UK) [recruited its CEO from the US](#). The EU AI Office's hiring challenges, particularly for its [Lead Scientific Advisor](#), demonstrate these limitations. CERN for AI legal grounding must offer flexibility hiring and compensation to secure top talent.

2. **Accept funding from the European Commission, member countries, and the private sector.** The funding model presented in the previous chapter requires diverse funding streams beyond the EU, including member countries and private sector contributions. CERN for AI's legal foundation should be able to accommodate the diverse funding streams.

3. **Pursue commercial applications in addition to research.** CERN for AI must be able to develop commercial applications, often through private sector partnerships. Revenue from licensing and product development should be reinvestable into programs, with potential for future profitability.

4. **Distribute benefits across state and private actors.** CERN for AI must be able to fairly share benefits—including products, revenue, and other outcomes—among contributing entities, whether through initial funding or research collaboration.

5. **Be publicly controlled while leveraging private sector partnerships.** Despite significant private sector involvement in research and commercialization, CERN for AI should maintain public control to avoid the pitfalls of [private interest capture](#).

6. **Include like-minded non-EU countries as full members.** CERN for AI should be able to welcome non-EU countries sharing its values as equal partners, similar to Horizon Europe's inclusive approach, to maximize impact and success.

7. **Be set up shortly after its formal proposal.** While international organisations typically require years to establish (as with the original CERN), AI's rapid advancement demands swift action. CERN for AI's legal structure must enable quick deployment to keep pace with technological breakthroughs and capture economic opportunities.

Preliminary findings suggest a Joint Undertaking is a strong fit

Given CERN for AI's unique requirements, viable legal entity options are limited. To determine the most suitable structure, CFG asked The Good Lobby to provide an independent comparative legal analysis of existing legal models that could be deployed to establish a CERN for AI[17]. This analysis compared organisational models of the following institutions:

- The original CERN (Intergovernmental organisation)
- Euratom (Intergovernmental organisation)
- The European Research Council (EU legal entity)
- The EuroHPC Joint Undertaking (EU legal entity)
- European Research Infrastructure Consortia (EU legal entity)
- Imec (National nonprofit legal entity)

The Good Lobby found that a Joint Undertaking (JU) under Article 187 of the Treaty on the Functioning of the European Union (TFEU) presents the most fitting model—provided it is customized to CERN for AI's specific conditions.

JUs are purpose-built for public-private partnerships in research and technological innovation. They accommodate funding from both public and private sources, supporting not just research but also commercial applications. Notably, while many JUs reinvest profits into their operations, the legal framework allows profit redistribution among members, subject to governing board approval.

As independent entities, JUs also enable flexibility in employment conditions, allowing CERN for AI to offer competitive salaries and support a growing talent pipeline. That said, previous JUs, such as EuroHPC, voluntarily imposed restrictive conditions on hiring practices and conditions, as well as nationality requirements. CERN for AI cannot afford to do the same.

To avoid governance bottlenecks, these regulations should clearly delineate hiring responsibilities between the Board of Directors and the Executive Director. EuroHPC serves as a

---

[17] For further questions about this report, feel free to contact CFG researcher David Janků (d.janku@cfg.eu).

cautionary tale here, with its [administrative challenges rooted in ambiguities around governance](#)[18].

Flexibility in board composition is another advantage of the JU model. While JUs are established by the European Commission, their governance structures can range from entirely public-sector boards to broader arrangements that include private and international stakeholders. Non-EU participation is also possible, as [demonstrated by EuroHPC](#), which includes non-EU members such as the UK through association with EU research programs like Horizon Europe.

EuroHPC highlights the efficiency of the JU model. The initiative became operational just [10 months](#) after its [proposal](#)—an impressive timeline, and one necessary for the fast-moving AI sector.

Why not other options?

Other legal pathways fall short of meeting CERN for AI's critical needs. Establishing an intergovernmental organisation, akin to the original CERN model, is a protracted process that risks untenable delays. Meanwhile, other existing legal frameworks, like the one underlying the European Research Council (ERC) are equally problematic. The ERC's structure lacks the ability to act as a legal entity, engage in collaborative research with private partners, or support the public-private framework and commercial focus CERN for AI requires.

The Joint Undertaking model stands out for its flexibility in governance, capacity to attract diverse funding sources, and proven ability to be rapidly operational. It aligns well with CERN for AI's ambitious goals and the sector's fast-paced demands. We strongly recommend further exploration of this legal foundation.

---

[18] European Court of Auditors, Annual report on EU Joint Undertakings for the financial year 2023, published on 13 November 2024, available at https://www.eca.europa.eu/en/publications/SAR-JUS-2023, paras 2.24 and 2.25 notes: "[EuroHPC] could only recruit 21 additional staff members, reaching a total of 36 staff. The JU attributes its significant difficulties in meeting the recruitment target to its small number of administrative staff and the challenge in recruiting rare, specialised experts."

10. Conclusion

Europe stands at a crossroads. While it leads in AI regulation, it has failed to develop a thriving AI ecosystem. [CFG's 2024 report](#) sketched out a possible solution—a CERN for AI. Now we've turned that sketch into a blueprint ready for deeper discussion. And while questions remain, one thing shines through: execution is everything.

Our analysis demonstrates that creating a successful CERN for AI requires carefully balanced design choices across multiple dimensions:

First, its research structure must enable both breakthrough innovation and responsible development. A novel two-track approach combining ARPA-style distributed research programs with focused in-house teams could facilitate open collaboration while maintaining security for sensitive research streams.

Second, its governance structure must balance operational autonomy with democratic accountability. Transparent operations and targeted public control will be key here. Our proposal implements these via a Member Representative Board and independent expert bodies that provide guidance without impeding operations.

Third, its institutional architecture must be both inclusive and secure. A tiered membership structure could enable strategic collaboration with trusted non-EU partners while protecting sensitive technologies.

Fourth, its legal and funding framework must enable rapid establishment while ensuring long-term sustainability. A Joint Undertaking under Article 187 TFEU, combined with our proposed €30–35 billion initial funding structure, would provide the necessary flexibility while positioning CERN for AI for sustained success.

CERN for AI's vision extends beyond its immediate research objectives. Think of it as a greenhouse for Europe's tech ecosystem—providing the perfect conditions for innovation to flourish until it can thrive independently. This represents a fundamental departure from Europe's typical innovation strategies. If some readers find themselves uncomfortable with our proposals, it signals that we've successfully conveyed our message: CERN for AI would pioneer very different ways of working, prioritizing transformative impact over quick political wins or universal consensus.

As Draghi rightly noted in [his address at the European Parliament](#), the stakes extend far beyond pure technological advancement. Without developing sovereign AI capabilities, Europe risks not only economic stagnation but also increased dependency on foreign technology providers in critical domains. As AI systems become increasingly powerful, the world needs multiple centers

of innovation working to ensure these technologies develop in alignment with democratic values and societal interests.

The time for action is now. As global AI investments accelerate and capabilities advance, Europe has a narrow window to shape the development of this transformative technology. While many of the details laid out in this report will require further refinement and broader stakeholder involvement, the fundamental architecture we propose offers a viable path forward. Moving from blueprint to reality will require sustained political will and substantial resource commitments, but the cost of inaction—in terms of missed economic opportunities, diminished global influence, and potential safety risks—could far exceed the investment required.

Ideas are crystallising, and the political winds are shifting. Now, what's left is to actually build the institution. One that can move at the speed of AI progress while staying true to European values.

Authors

Alex Petropoulos | Advanced AI Researcher – Policy

Studying the design of AI R&D and governance institutions, Alex analyses what makes them effective and trustworthy. He translates insights from ML research and institutional design into policy guidance.

Bálint Pataki | Senior AAI Researcher – Policy

Bálint researches EU AI innovation policies and shares his findings directly with policymakers. A highlight of his work is bridging the world of the AI research community with EU policymakers.

Daan Juijn | Advanced AI Researcher – Computational Governance

Daan specialises in technical AI policy research. Drawing from his background in scenario planning, he tries to disentangle where and how compute governance can improve future AI outcomes in robust ways.

David Janků | Advanced AI Researcher

David focuses on institutional designs for the strategic development of AI in Europe, disentangling AI hype from reality, mapping expert debates on risks, and advancing the case for a science of safe AI.

Max Reddel | Advanced AI Director

Max Reddel leads CFG's Advanced AI program, merging expertise in AI and governance to deliver rigorous, practical solutions, anticipate challenges, and shape global AI strategies with actionable insights.

About CFG

The Centre for Future Generations is an independent think-and-do tank created to help decision-makers anticipate and govern rapid technological change. We are here to make sure that emerging technologies are used in the best interests of humanity. Find out more on www.cfg.eu.