## 1. Context and purpose

To achieve its AI continent objectives, the EU requires its own foundation and specialised AI models, plus a robust data strategy. These are essential for competitiveness and science, providing alternatives to commercially dominant models while delivering reliable, secure, and ethical tools aligned with EU values.

AI strategy cannot be separated from scientific strategy because of their strong reciprocal impact. Social sciences and data science require a special mention here as AI has massive impacts on society (the domain object of social sciences), and data science will be essential for AI development.
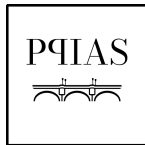
The AI Continent Action Plan addresses five key areas: computing infrastructure, high-quality data, AI algorithms, talent development, and regulation.

This note proposes to strengthen the current and future ecosystem by prioritizing **data as the primary asset**, positioning **regulation as a competitive advantage,** and irrigating the AI continent with a **trusted EU social network platform** coupled with the AI tools developed. This is necessary to develop an AI ecosystem that will sustain science[1].

*Data are more valuable than models, and therefore, investments should take this into account. Models depend on data, not the reverse.*

*Regulation can be played as an asset. The EU's competitiveness and attractiveness to scientists, datasets, and funding in tomorrow's AI landscape may paradoxically be based on its ability to regulate. Regulation and the rule of law provide predictability and trust — key assets for businesses and more generally all users.*

---

[1] This note was sent in June 2025 in response to the European Commission's call for evidence on : "**A European Strategy for AI in science – paving the way for a European AI research council**".

*A trusted European social media platform with specific characteristics can be the lynchpin to build a solid EU data management system and user base that will cement the AI continent.*

## 2. A data foundation model enabling regulated AI models

The main resources necessary for the AI continent are:

- trusted access to large sets of quality data

- computing power to train models

- qualified personnel

These are not enough. A sustainable business model must support the above, with organisational agility to cope with fast evolution.
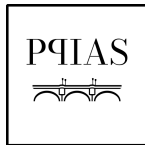
The general goal is to create an ecosystem of the AI continent, which serves EU citizens, society, and economy. As this concerns not only business but also democracy, a not-for-profit element will be required to address specific needs, in addition to a for-profit element to support EU economic actors.

### 2.1. Data:

Data is the real asset in AI. There are already many AI models, there will be even more in the future. Models change quickly as technology advances, and they will be easier to develop as computing power becomes available. However, **a model is only as good as its data and the architecture that powers it ("garbage in, garbage out")**. Most current models are based on large amounts of poorly curated data (e.g. data scraped from the web). **Significant progress can be made by using curated and tagged datasets** for training purposes. Furthermore, data provenance will be necessary for quality and IP issues, and ultimately for access during training or queries — access will depend on trust.

The challenge is therefore how to establish not just data labs, but also trusted data repositories where stakeholders will agree to contribute their data without fear of exploitation. The more labelled and curated these data are, the more valuable they become in terms of provenance, IP, reliability, and history of use. Furthermore, every use of a dataset can improve it with annotations, meaning that data is a form of capital that grows in value with use.

Interestingly, **an initial high-quality dataset can be quickly compiled, with limited IP issues, by EU national libraries and cultural agencies** (e.g. Europeana and its 200 aggregators, and the growing Common European Data Spaces), as well as green open access journals. This European cultural treasure will later grow with curated datasets from research projects, publishers, the press, individual authors, and industrial or public data sources (e.g., scientific or medical data, national observatories, surveys, and industrial datasets). These data are of a high quality: labelled, curated, linked to IP and provenance, securely stored, and often of a high resolution, benefiting from

specialised, dedicated staff. The infrastructure and network of data labs should also be designed with this in mind.
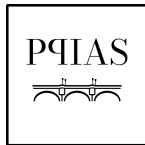
If we want Europe to gain power, **size is an important factor, as is trust. Hence, the suggestion to put all the Data Spaces under the same brand name, and common principles such as those developed in GAIA-X**. While some actors are now realising that it can be risky to store data in countries where the government is invasive, we must be aware that non-EU actors may be reluctant to entrust their data to a repository dependent on the EU as a political entity. Remember how Switzerland once attracted money because it was reputed a safe, protected place. Having the various European entities explicitly under one banner and governance would create the critical mass and trust essential for global influence in a context where the actors are huge.  An option could be to establish an independent EU Data Foundation, responsible for managing the data repositories and access according to EU regulations (GDPR, Data Act, etc.) This foundation could host a federation of data spaces or whatever organisational model.

The repository could archive data for its owners and partners (past, present, and future), while giving them access and, under certain conditions, provide access for training AI models and tools (public or private). The Data Foundation would act as a **trusted third party and data controller**. It would be **a trusted resource for the entire EU ecosystem, research, and beyond, as well as a scientific treasure for decades to come**, since  currently individual projects usually cannot store their data long term, and hence, these data are later lost. Of course, **this curated data repository would become the high-quality training resource for powerful EU AI models; its curated nature opens new avenues for explainability and quality control.**

**Feasibility:** The cost of facilitating access is less than €5 million per national library. Start with public domain data to address IP issues. Later, negotiate copyright with rights holders and their representatives (e.g. SACD in France) for access to data outside the public domain. Include deals with the press and agencies, as some have already done. Then work on regulation (e.g. compulsory legal deposit of works). Many relevant initiatives already exist that can be the basis: Common European Data Spaces, Gaia-X, Creative Commons, Data Labs, etc., and there are already data norms that could contribute to a unified Data Foundation. The current entities above should participate in this foundation.

In short, the approach suggested here is to **establish an independent foundation to host quality data as the solid basis of the AI continent ecosystem,** rather than focusing on transient AI models and clusters**. Primarily invest where the asset is: data.** Such a foundation would serve as a training resource for models (whether for-profit or not-for-profit), it would enable training models in compliance with IP and EU rules and thus building trust. It could provide any organisation with a trusted archival vault (as a paid service) and finally solve the issue of storing valuable scientific data (e.g. medical data) safely for future research, instead of destroying it to comply with GDPR.

Not-for-profit status provides considerable scope for action and agility within the current market system. The Mozilla Foundation and Wikipedia could serve as models for inspiration regarding legal aspects. Not-for-profit associations (e.g. Belgian AISBLs -the status of GAIA-X!-, German VoGs, Dutch VZWs, etc.) could also provide a framework. The Foundation should be explicitly

designed to avoid national rivalries in governance and bureaucracy. Constituency could be based on the provision of datasets (i.e. if you provide data, you have a voice and a vote) rather than financial funding alone.

### 2.2. Computing power:

To train AI, relying on the facilities of companies that own AI tools offers no guarantee against data leakage. Would you entrust your sensitive data to one of these big AI models that train on your data? We need **independent data processing facilities** to control the training of models using the Foundation's data. This would **allow training to be done without exporting the data directly to the model owners**. This limits the possibility of data leakage and enables to **control and certify** the quality of the AI models. The ownership and regulation of these data processor entities matters for trust. In the current roadmap, the exact role of the Data Labs, Gigafactories, etc., in the value and technical chains probably needs to be clarified in this respect. **Trust is essential if we want data owners to agree to their datasets being used for machine learning.**
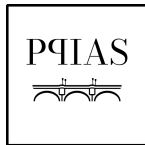
Potential partners are already connected through various initiatives (e.g. EuroHPC JU, OpenEuroLLM). The multiplication of structures with vested interests suggests it is unlikely that an efficient solution will emerge naturally and fast: top-down steering by the EU seems necessary. Commercial models, including current ones (e.g. OpenAI, Anthropic, DeepSeek, xAI, or Mistral AI) would also, under conditions, be able to train their models using data housed by the Data Foundation. In order to be certified, one must abide by EU rules, which could become a standard. Interestingly, **certification could become a business model that funds the data processor.** Unifying data processors under a single brand name would create a critical mass effect.

**Feasibility:** infrastructure components exist, much of it described in the AI continent plan (Gigfafactories, Apply AI plan, etc.), although loosely connected, and possibly requiring tuning. A large EU initiative would likely attract institutional funders: the EU signature and investment is a guarantee against risk in a very uncertain landscape.

### 2.3. Qualified personnel:

One current idea is to attract talent to the EU project by poaching it from competitors. The EU's main selling point is that it is a safe place for you and your family to live a good life. This trumps all the financial incentives offered by non-democratic environments, provided that salaries in the EU are attractive enough. **The EU should leverage its competitive advantage as a place to enjoy a good life, safely.** AI scientists and engineers are also human beings caring for their family. With equally interesting datasets, computing power, and working conditions, the EU lifestyle is more attractive.

A 'Choose Europe' programme including a big pay package and Schengen visas for family, with the perspective of EU citizenship after a while, could be offered alongside staff recruitment programmes for the Data Foundation and Data Processor. This is an attractive proposal for engineers and scientists currently working abroad for the big AI companies. Top European scientists could also, of course, benefit from a similar package.

Moreover, it is paramount to locally train (and retain) the talents in Europe. **It's OK to poach brains, but it's even better to raise them.** This requires training AI engineers and data scientists, and retaining them in the EU. Furthermore, as recommended by the Paris Institute for Advanced Study working group on AI in the Higher Education and Research sector, EU should train *all* students to become apt users of AI. Associations of AI scientists, e.g. CAIRNE (formerly CLAIRE), ELLIS , IEEE, etc., can be leveraged.

Although HR is essential, the scales of investment in talent currently considered are small considered to those planned for machines.

### 2.4.A business model based on a trusted EU social network and trusted services
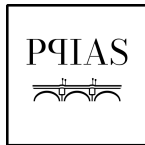
In addition to the resources mentioned above, it is important to consider the conditions needed to create a sustainable business model. The AI tools and infrastructure mentioned above are not a business model, even less a societal model. Those requires a user base, regulation, and funding mechanisms.

First, we need to bring users to adopt these EU AI tools. A suite including a messaging system and a social network would be a powerful way to attract users and make these tools the default solution for daily use. These functions are what the GAFAM and BATX provide, and are their vehicle to market their AI products; but their tools cannot be trusted because of their for-profit nature or potential control by a government. A European trusted and safe alternative would be very attractive, especially if it is free.

So, the other tool suggested here is a **European trusted social network** that provides users with communication capabilities such as email and chat, as well as the ability to have their own webpage. This system would have the competitive edge of being free, safe and not-for-profit. This **EU social networking tool would gain trust by being GDPR-compliant, not-for profit, add-free, and showing guarantee of independence in its governance.** It would **come with a suite of EU AI tools, including search engine**. Finally, it is suggested here **this social network would not be anonymous, which is an added unique selling point.**

Access to this non-anonymous EU social network would be verified (e.g. via smart meter link or digital certificate), therefore enabling trust and secure transactions. **The existence of such a trusted EU social network, with safe and certified access and exchanges, would provide the vehicle for the development of safe data exchange and market that is indispensable to build the Data Strategy of the AI continent.**

The European trusted social network could be made freely available to EU citizens, residents and taxpayers on EU territory. This would be a positive political move, as it would enable European citizens and organisations to become independent, at least for their sensitive data, of current predatory systems based on a for-profit media model (i.e. advertising), which are deliberately addictive and unsustainable in many ways, not to mention scams and other antisocial behaviours which could be policed.

There are several possible approaches to funding this infrastructure. It should not be funded by advertising because this can become a toxic media (ad-based) business model. However, the safe nature of such a network could justify funding from various public organisations. For example, **this trusted network would facilitate administrative and financial transactions**, thereby generating revenue through these intermediary services. Initial funding from philanthropy and EU could be used during the construction phase. In the long term, a more sustainable funding model could be added, involving other components of the ecosystem. For example, data controllers using this network could also generate revenue from their brokerage, archival, and curation services for data owners and users. Also, added-value services could be offered for a fee, and the revenue shared among components of the chain.

A possibility would even be to **offer EU residents this access (EU AI toolkit + EU social network) as a basic right**, therefore partly paid through income tax or VAT.

Other domain-specific ancillary models could be operated by for-profit companies or consortia, as is currently the case.
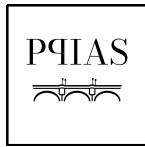
Overall, data processors can have various statuses, such as a foundation, an association, an international agency or a for-profit organisation. However, the different statuses require careful consideration. **The business model is simple in principle: the AI model owners pay for AI training, the data IP rights are negotiated with data owners, the data processor receives a fee**. **The users of the network contribute to its financing through optional paid services. The public service aspect of these tools is funded by member States.**

How can we get there? The idea is to adapt the current ecosystem to the proposed architecture rather than the other way around. This would limit national and corporate gaming, as well as path dependency, in the construction of the new ecosystem. This approach enables the EU (or a coalition of willing states) to steer and nudge the ecosystem towards a collective optimum rather than waiting for the market to impose a fait accompli. The AI continent plan is a great step in this direction.

### 3. The way forward: think big, use science and act together

A key issue is attracting resources (people, data, money). This would be most effective with **large European flagships rather than smaller competing national champions**. While giant size may not be technically necessary (see the DeepSeek example), it is essential for attracting resources, setting standards and reducing uncertainty, from a psychological and institutional perspective. This will also considerably facilitate the creation of a European norm later. Diversity will come from the variety of contributors to the common endeavour.

The current ecosystem suffers from several problems: 1) It was created before the explosion of AI and may be technically inadequate. 2) It is the result of scattered funding and rivalry between States who want their 'champion', resulting in many structures lacking funding and failing to reach the required market scale. 3) The entities have developed an institutional nationalistic spirit as they compete against each other for funding and scientists, which makes it difficult to join forces in a common structure.
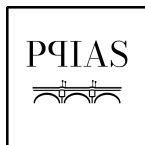
Without guidance at the EU level, the current discussions with and between operators suggest that an efficient solution is unlikely to emerge quickly, if at all, from compromises between parties with competing individual strategies. We may all lose because we are each trying to win individually.

The role of science in society is not only to advance knowledge and to find answers to problems. Its rigorous methods make science uniquely qualified to provide objective evidence - investigations, experiments, simulations, figures, and proofs- that guide decisions under uncertainty and help settle conflicts between competing interests. In the specific domain of AI, there will be many competing interests and significant uncertainty. RAISE could fulfil a critical role by contributing to the research needed to navigate the complex challenges and competing priorities of AI.

As mentioned earlier, developing AI tools is a scientific endeavour that requires collaboration and funding in order to create a critical mass in Europe that is capable of competing with the current leading AI companies that are developing business models which do not align with European ethical and sustainability concerns.

**Foresight is required to guide this process, foster cooperation, and enable research and proofs of concept, all of which require scientific research and innovation. RAISE should be useful in this endeavour.**

PPIAS

**Appendix:**

**A few useful sources**

Group of Chief Scientific Advisors Recommendations Scientific Opinion No. 15 on "Successful and timely uptake of Artificial Intelligence in science in the EU" provides independent policy recommendations on facilitating AI uptake in research and innovation, including the establishment of a European institute for AI in science and development of specialized AI tools for scientific work.

AI in Science Policy Brief: The European Commission's policy brief on harnessing AI's potential in science published in December 2023, advocating for a tailored European Research Area policy to accelerate AI adoption in science and boost Europe's global competitiveness.

Artificial Intelligence in Science: Promises or Perils for Creativity?: European Commission R&I Working Paper WP2025/03 that explores the impact of AI on scientific creativity across 80 scientific fields from 2000 to 2022, showing that AI adoption has surged since the early 2010s and generally enhances scientific creativity, though effects vary by field.

European Research Area Policy Agenda 2025-2027: The Commission's proposal for the next ERA Policy Agenda specifically addresses current challenges including artificial intelligence in science, with envisaged outcomes including a joint roadmap on AI in science and structural cooperation on research security.

CORDIS AI in Science Reports Results pack : ongoing and completed EU-funded projects highlighting various applications of AI in science, demonstrating how AI is expanding scientific boundaries and enhancing innovation across multiple disciplines.
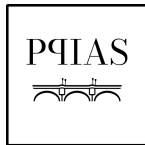
Living Guidelines on Responsible Use of Generative AI: European Research Area Forum guidelines developed by the Commission and European countries to support the research community in responsible use of generative AI, emphasizing research integrity, transparency, and ethical considerations throughout the research process.

ERC Foresight Survey on AI in Science: The European Research Council's comprehensive foresight report examining how ERC-funded researchers are using AI in their scientific processes and their views on future developments by 2030, including opportunities, risks, and the impact of generative AI in science.

Mapping ERC Frontier Research in AI: The European Research Council's comprehensive analysis of ERC-funded projects involved in AI development, application, or study from 2007 to 2022, providing an in-depth portfolio analysis of frontier AI research across European institutions.

**Relevant references to the EU AI Continent Action Plan**

**AI Gigafactories**: The EU AI Continent Action Plan commits to establishing 5 AI gigafactories as "large-scale facilities with massive computing power and data centres" to enable training of complex AI models at unprecedented scale, requiring both public and private investment to secure

EU leadership in frontier AI. The Plan defines gigafactories as facilities housing around 100k advanced AI processors each, designed specifically for training and deploying the most advanced AI models, with the InvestAI facility mobilizing €20 billion for their establishment.

**Data Strategy**: The Plan includes a comprehensive Data Union Strategy (planned for Q3 2025) to foster a true internal market for data, enabling the scaling up of AI development across the EU through data labs within AI factories that gather and organize high-quality data from diverse sources.

**Apply AI Strategy**: The upcoming Apply AI strategy aims to accelerate AI adoption across strategic sectors, noting that currently only 13.5% of EU companies use AI. It aims to bridge this gap through targeted initiatives in healthcare, manufacturing, and public administration.

**Choose Europe Strategy**: The Plan includes talent development initiatives to strengthen AI literacy and skills, including attracting and retaining global AI experts through programs like the Marie Skłodowska-Curie Action 'MSCA Choose Europe' and AI fellowships to encourage the return of European AI researchers living abroad.

**AI Act and Regulation**: The Plan commits to supporting companies in AI Act implementation through guidelines, codes of practice, and the upcoming AI Act Service Desk (launching summer 2025) as a central point of contact for businesses seeking information and guidance on regulatory compliance.
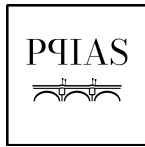
**Regulation as Competitive Advantage**: The Plan emphasizes that "The AI Act raises citizens' trust in technology and provides investors and entrepreneurs with the legal certainty they need to scale up and deploy AI throughout Europe," positioning EU regulation as a trust-building mechanism rather than a barrier.

**RAISE (Resource for AI Science in Europe)**: The proposed dedicated talent and research institution to coordinate and guide the development of AI/Science, also serving as a potential structure to optimize the utilization of Gigafactory-scale computing resources.

**Computing Infrastructure**: €10 billion investments are planned in the AI continent plan over 2021-2027 to establish and enhance AI factories, including procurement of nine new AI-optimized supercomputers through the EuroHPC Joint Undertaking.

**Talent Pool Development**: Comprehensive measures to enlarge the EU's AI talent pool, including significantly increasing European Bachelor's, Master's, and PhD degrees in AI through the AI Skills Academy, with calls for proposals opening in 2025.

**MSCA Choose Europe Programme**: Pilot programme allowing research institutions to attract, develop and retain excellent international AI researchers, co-funding recruitment programmes with links to long-term career prospects including permanent positions.

**Data Labs**: The current Plan specifies that data labs will be integral parts of the AI Factories initiative, federating data from different AI Factories covering the same sectors and linking to Common European Data Spaces, with setup planned by end of 2025.

**Data Governance Framework**: The Plan builds on the Data Governance Act and Data Act, with the upcoming Data Union Strategy focusing on making more data available for AI while ensuring simplified, clear and coherent legal framework for data sharing at scale with high privacy and security standards.

**Common European Data Spaces**: The Plan explicitly references Common European Data Spaces as part of the data infrastructure, connecting them to data labs and ensuring interoperability across sectors for AI development and innovation.

**Sovereign EU Infrastructure**: The Plan emphasizes reducing dependencies on critical technologies and strengthening sovereignty in semiconductors, with the upcoming Cloud and AI Development Act (Q4 2025/Q1 2026) aiming to triple EU data centre capacity within 5-7 years.

**EuroHPC Joint Undertaking**: The European vehicle for procuring and deploying AI-optimized supercomputers (e.g., the energy-efficient JUPITER exascale supercomputer in Jülich).

**Investment Scale**: InvestAI initiative should be mobilizing €200 billion for AI investment in the EU, with specific facilities like the €20 billion InvestAI facility developed with the European Investment Bank Group for establishing AI Gigafactories.