# Making European Science Flourish in the AI Age

**Jonathan Wareham**
ESADE Business School, Ramon Llull University, Barcelona, Spain
The Lisbon Council, Brussels, Belgium

Compared with the fast and competitive innovation ecosystems of Asia and the United States, Europe is often portrayed as a "regulatory museum." When it comes to innovation policies, European initiatives are deemed slow and bureaucratic, out of step with the agile culture of digital innovation. The advent of artificial intelligence (AI) has only reinforced this stereotype, with Europe seen as focusing on regulating data and AI usage, and the rest of the world investing hundreds of billions in disruptive models.[1]

The reality, however, tells a different story. Europe remains home to many global leaders in science and innovation. Its competitive edge lies not in copying other regions, but in its deep commitment to scientific excellence and the rule of law. These foundations guarantee the long-term stability and transparent processes that are crucial for competitiveness in general and artificial intelligence in particular.

*'The key question is not whether Europe should aim to be competitive in artificial intelligence, but rather, how it can scale its success stories.'*

The success of AlphaFold – a Nobel-winning AI-driven protein structure prediction system trained on, and built upon the protein database curated over decades by the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) – proves that European public research infrastructures can be the silent engines of global AI innovation. Europe has been laying other such foundations for many years, including long-standing efforts to create data spaces for the secure and effective sharing of data.

1 This policy brief builds on the discussions held at six Science Salons convened by the Lisbon Council in partnership with Google DeepMind. Special thanks to David Osimo, Katarzyna Szkuta, Sebastian Staudt, Maria Chiara Zaccaria, participants at the high-level breakfast roundtable on 04 November 2025 in Copenhagen, and all interviewees listed in annex. Any errors of fact or judgement are the sole responsibility of the Lisbon Council team and the lead author.

The key question, therefore, is not whether Europe should aim to be competitive in artificial intelligence, but rather, how it can scale its success stories. How can Europe replicate AlphaFold-level advances across more fields and sectors? What role can public research infrastructures and data spaces play in supplying high-quality data for AI innovation? What are the main bottlenecks, and how can they be addressed?

This policy brief examines Europe's ambitious policy interventions in the realm of data and AI and identifies the obstacles to greater impact. Drawing on discussions during the high-level breakfast roundtable at the AI in Science Summit in Copenhagen in November 2025, and on a wide range of interviews, it puts forward a set of recommendations:

- **Centralise and harmonise governance:** Establish consistent data and metadata standards across Europe through centralised European agencies to ensure compliance. This includes extending the remit of the Data Spaces Support Centre (DSSC) to provide binding legal interpretations and sector-specific guidance, as well as creating a European Union-wide data stewardship accreditation to formalise professional standards.

- **Optimise high-quality data generation and sharing:** Adapt governance frameworks to accommodate the operational realities of the private sector, enabling companies to share data without compromising trade secrets. This involves providing clearer European Union-level guidance on the use of intellectual property in AI training data and recognising new entities such as "data factories" and brokers as trusted nodes within the common European data spaces (CEDS).

- **Accelerate sustainable and equitable infrastructure:** Address Europe's "compute deficit," i.e., the shortage of specialised hardware, by implementing a European Union-wide allocation system that balances the priorities of public research, small and medium-sized enterprises (SMEs) and deep-tech startups. Additionally, operations must prioritise sustainability by securing low-carbon energy sources and establishing energy management key performance indicators for AI gigafactories.

- **Leverage RAISE as a strategic policy actor:** Use the European Union-funded Resource for AI Science in Europe (RAISE) as a harmonisation layer across different data spaces. Its role should include developing shared compliance resources, championing hybrid scientific AI educational curricula, and defining certification standards for data intermediaries to ensure data remains "evidence-grade" and FAIR+.[2]

## AI in Science: Collaboration and Its Discontents

Artificial intelligence is increasingly transforming basic science by accelerating discovery, enhancing data analysis, and enabling new forms of experimentation. In fields ranging from physics to biology, AI algorithms are now used to sift through massive datasets, identify patterns, and generate hypotheses that would be difficult or impossible for humans accomplish unaided. Machine learning models help researchers simulate complex systems, optimise experimental design, and even automate the interpretation of results. This shift is not only improving the efficiency of scientific workflows but also expanding the boundaries of

2 FAIR principles (standing for findable, accessible, interoperable and reusable) are international guidelines for the management of research data; FAIR+, in addition, implies AI-readiness, with improved provenance and higher reproducibility standards.

what can be explored, enabling scientists to tackle questions of greater complexity and scale than ever before.

For example, physics research now relies heavily on AI across a wide range of applications.[3] At CERN, the European Organization for Nuclear Research, machine learning algorithms scour data from the Large Hadron Collider (LHC), identifying rare particle collisions to precisely measure the Higgs boson and other fundamental particles. In astronomy, deep learning models classify galaxies, detect exoplanets, and isolate gravitational wave signals. In materials science, AI navigates complex datasets from simulations and experiments, accelerating the discovery of novel materials. In energy research, AI helps control fusion reactors to manage the stability of plasma confinement.

AI's increasingly vital role in modern science was acknowledged in 2024, with the Nobel Prizes in both Physics and Chemistry awarded for AI applications. Although the Nobel Prize has historically honoured theoretical contributions or empirical techniques that advance understanding of how natural systems function, it had never been awarded for an exclusively computational technique.[4] More remarkably, both 2024 accolades are essentially void of any theoretical ontology. Surprisingly, science now deploys a wide gambit of theory-agnostic machine learning in tandem with theory-forward physics models.[5] Given the preeminent status of theory throughout most scientific history, these developments are extraordinary.[6]

Perhaps the most high-profile application of AI in science to date is AlphaFold, the basis of the 2024 Nobel Prize in chemistry used to determine protein structures. Proteins are responsible for almost all biological functions, making them critical for drug discovery and life science research. Their structure in three-dimensional space determines their function[7] and dictates what other molecules they can interact with. Proteins, however, are painstakingly difficult to study empirically, often requiring years to model even just a single protein. Hence, without decades of public funding supporting research infrastructures such as the European Synchrotron Radiation Facility (ESRF) or European Molecular Biology Laboratory (EMBL), AlphaFold, heralded as *"The most important achievement in AI – ever,"*[8] would not have occurred. It is only through a sophisticated and costly scientific apparatus that the high-quality proteomic data that trained AlphaFold exist.

> *'AlphaFold… heralded as 'the most important achievement in AI – ever'… would not have occurred without decades of European public investment in scientific infrastructures.'*

This case highlights a challenge for scientific AI: In contrast to mainstream LLMs, which can be trained on a wide range of heterogenous, noisy data types, scientific data must

3 Jonathan Wareham and others, "Regulating AI: Lessons from Scientific Computing," *MIS Quarterly,* 1–20, 2025.

4 Where the differentiation between scientific understanding and empirical technique is blurred, noted examples of Nobel prizes awarded to scientific instrumentation and detection technologies include X-rays (1901), spectroscopy (1930), the cyclotron (1939), chromatography (1952), electron microscopy (1986) and cryo-electron microscopy (2017).

5 George E. Karniadakis and others, "Physics-Informed Machine Learning," *Nature Reviews Physics,* 3.6, 2021, pp.422–440, doi.org/10.1038/s42254-021-00314-5.

6 Mario Krenn and others, "On Scientific Understanding with Artificial Intelligence," *Nature Reviews Physics,* 4.12, 2022, pp.761–769, doi.org/10.1038/s42254-022-00518-3.

7 For example, myoglobin was the first protein structure determined in 1958. It is found in the muscle tissue of most mammals, but sperm whales have it in very high concentrations. Its 3D structure allows it to effectively bind with oxygen molecules, thereby enabling the whales to stay submerged for very long periods.

8 Rob Toews, "AlphaFold Is the Most Important Achievement in AI—Ever," *Forbes,* 03 October 2021.

be generated through disciplined protocols on limited and expensive infrastructures. Surprisingly, then, high-quality scientific data sets suitable for training scientific AI are limited.

In 2021, DeepMind released AlphaFold's source code marked a pivotal moment in the relationship between the scientific community and the private sector, fostering greater collaboration between academic institutions, biotech companies, and pharmaceutical researchers. However, this is a rare case: companies often guard their proprietary technologies closely, driven by commercial incentives, intellectual property concerns, and competitive advantage. Yet, differing goals, funding structures, and cultures can inhibit sharing of competitively sensitive data. Given that some of the best scientific data are generated by industry, this leads to a structural paradox: many of the most informative datasets for scientific AI are often the least shareable.

## How to Unlock the Value of Scientific Data

Europe has invested heavily in public scientific data infrastructure. Since 2021, the European Union has allocated an estimated €1.2 billion to 177 research infrastructure projects under Horizon Europe. A sizable portion of this funding has supported open-science initiatives, including the European Open Science Cloud (EOSC) and data services.[9] New policy initiatives like the common European data spaces (CEDS) and the data act complement these investments. The recent European Union strategy on research and technology infrastructure acknowledges that "soaring research data volumes, particularly in fields with deep AI integration, are outpacing our capacity to put it to use." However, much scientific data is not AI-ready.

*'Much scientific data is not AI-ready. Scientific AI needs evidence-grade data that is physically meaningful, reproducible, validated, contextualised and comparable.'*

To be so, scientific AI needs evidence-grade data that is not only FAIR, but physically meaningful, reproducible and validated, contextualised and comparable. This requires consistent ontologies and metadata with common labelling and measurement conventions. Additionally, scientific data are often generated on sophisticated infrastructures such as satellites, telescopes, particle colliders, synchrotron and free electron x-ray sources. Scientific instrumentation is comprised of leading-edge technologies run by highly trained personnel. In short, AI-ready scientific data (i.e. FAIR+) are scarce for good reason: evidence-grade data are intricate and expensive to generate. And even within the most disciplined scientific organisations, useful scientific data are often stored in institution-specific repositories or governed by project-specific funding rules, leaving them effectively fragmented and siloed.

Deeply coupled with the supply of scientific data is the demand from the private sector. The success of AlphaFold demonstrates this clearly. As protein structure determination is such a critical pain point for the pharmaceutical industry, AlphaFold's level of uptake and integration in drug discovery pipelines has been unprecedented. As industry increasingly integrates evidence grade datasets to provision AI based R&D, operations, and manufacturing, data governance must be attuned to the needs of industry as both a user of— and contributor

9 Goda Naujokaitytė, "Data Corner: Horizon Europe's Six Biggest Research Infrastructure Projects," *Science|Business*, 23 September 2025.

to— scientific data resource pools. This implies that data governance must provide clarity, legal certainty, and predictable access pathways that are congruent with the priorities of the private sector. If inappropriately designed, data governance can create unnecessary compliance burdens, inhibit innovation, and discourage firms from contributing to both common data resources as well as broader scientific and technological ecosystems.

European science has been a historical powerhouse in many domains of basic research. But as we grow into the budding world of scientific AI, an acute risk of a data drought is surfacing. How can Europe foster investments in scientific data infrastructures essential to the use of AI for scientific discovery?

## Europe's Progress to Date

The European digital strategy, launched by the European Commission in 2020, is a comprehensive framework to guide the European Union's digital transformation through 2030. It aims to strengthen Europe's position in strategic technologies such as AI, quantum computing and cybersecurity. Additionally, it seeks to support AI factories and digital innovation hubs, as well as foster open data ecosystems. More recently, the European Union launched the AI in science strategy and the apply AI strategy, both of which aim to position Europe at the forefront of AI-driven scientific innovation. Central to this effort is the creation of RAISE, a virtual institute designed to pool talent, data, and computing resources for scientific AI applications.[10]

These ambitions rely on strengthening the "three pillars of AI," in other words, the foundational components that enable artificial intelligence systems to function effectively: computing power, algorithms and data.

### *Computing Power*

Europe is investing heavily in high-performance computing and AI gigafactories. The InvestAI initiative aims to mobilise €200 billion, including €20 billion specifically for building AI gigafactories. These massive facilities are each equipped with around 100,000 state-of-the-art AI chips.[11] The goal is that these gigafactories will support the training of large-scale AI models for scientific and industrial applications, democratising access to computing power for researchers, startups and SMEs.

However, there is still a significant gap between Europe's computing capacity and that of the rest of the world. As of June 2025, the global graphics processing unit (GPU) cluster performance is heavily concentrated outside Europe: estimates put the United States at 74.5%, China at 14.1% and the European Union at only 4.8%.[12] European researchers can apply for high-performance computing access via EuroHPC AI factories, but demand exceeds supply, and entry-level allocations remain limited.[13]

10 European Commission, "European AI in Science Strategy," *Communication of the European Commission,* COM(2025)724, 08 October 2025.

11 European Commission, "EU Launches InvestAI Initiative to Mobilise €200 Billion of Investment in Artificial Intelligence," [press release] 11 February 2025.

12 Konstantin F. Pilz and others, "The US Hosts the Majority of GPU Cluster Performance, Followed by China," *EPOCH AI,* 05 June 2025.

13 EuroHPC Joint Undertaking, "Open Calls for Proposals" https://access.eurohpc-ju.europa.eu/ [accessed 05 March 2026].

### *Algorithms*

Many AI models used in scientific research are open source, especially those developed by academic institutions or publicly funded initiatives. This reflects the foundational scientific principles of transparency and reproducibility. However, this openness often has its limits. Some models share their weights and code but not the training data or full architecture (e.g., open core), leading to a "gradient of openness."[14]

As such, AI algorithms used in scientific computing exist across a spectrum—from fully open-source to tightly protected proprietary systems.

Private companies and some research labs maintain proprietary AI systems, especially when the models are tied to commercial products, the training data includes sensitive or copyrighted material, or the algorithms offer competitive advantages. Such models are often protected copyright or patents, although patenting AI algorithms can be legally complex due to challenges in proving inventiveness and technical character.

### *Data*

Data remains a critical pillar of AI. Large-scale infrastructures such as the EOSC, ELIXIR, and EMBL-EBI have made substantial progress in building shared repositories and services. Yet, important challenges persist.

Accessibility and standardisation vary widely across European data repositories. For example, in its own study on patent and publication data (2000–2023), the European Union's directorate-general for research and innovation (DG RTD) found that European hubs are significantly less interconnected than their counterparts in the United States, particularly in complex technologies such as AI, biotech, and quantum computing.[15] Additionally, the European Union's Joint Research Centre estimates that only 30–40% of EU scientific datasets meet full FAIR criteria.[16] The European Union further highlights a shortage of trained data stewards and a lack of career incentives for long-term data curation.[17]

## The Importance of Well-Functioning Data Spaces

Arguably, the biggest barriers to accelerating science with AI are data fragmentation and heterogeneity. To address this, the European Union has introduced the common European data spaces (CEDS) initiative, a bold harmonisation strategy built on several legislative foundations, notably the data governance act, the data act and the general data protection regulation (GDPR). Together, these measures mandate clear, harmonised rules for accessing, sharing and reusing data, including data generated by connected devices (IoT).

14 Sylvain Duranton, "What Leaders Need to Know About Open-Source vs. Proprietary Models," *Forbes*, 07 July 2025.

15 European Commission: Directorate-General for Research and Innovation, *Divided We Fall Behind: Why a Fragmented EU Cannot Compete in Complex Technologies* (Publications Office of the European Union, 2025).

16 European Commission: Directorate-General for Research and Innovation and PwC EU Services, *Cost-Benefit Analysis for FAIR Research Data – Cost of Not Having FAIR Research Data* (Publications Office of the European Union, 2018).

17 Laila Güell Paule, "EU Strives to Boost Digital Skills as New Study Highlights Key Gaps," *Digital Skills and Jobs Platform*, 07 March 2025.

The most concrete progress has been made in three common European data spaces:

1. The European Health Data Space (EHDS), implemented in March 2025, provides the legal and technical framework for both primary clinical use and secondary use of health data. It creates a standardised system for researchers to access aggregated, non-identifiable health data.

2. The Cultural Heritage Data Space supports the digital transformation and reuse of Europe's vast cultural data.

3. The European Open Science Cloud (EOSC) aggregates a massive amount of scientific research data, creating a single virtual environment for European researchers.

### *CEDS Criticisms and Hurdles*

The Common European Data Spaces (CEDS) initiative, while ambitious and foundational, faces several significant criticisms and hurdles.

- **Risk of eroding intellectual property (IP) and trade secrets:** The data act mandates that manufacturers and service providers of connected devices (IoT) must share data with the user. Critics argue that this "forced sharing" could require companies to disclose data that is essentially a trade secret or proprietary know-how. Relatedly, while the legislation includes safeguards for trade secrets, their practical implementation is unclear.

- **Implementation and interoperability complexity:** The sheer scale and number of CEDS create enormous practical and technical hurdles. The CEDS rely on a dense web of overlapping legislation. For example, when an AI model is trained on a CEDS dataset containing both personal and non-personal data, determining which set of compliance obligations applies is an extremely difficult task for data custodians. Navigating these different and sometimes conflicting rules, particularly for SMEs, is seen as a significant tax. Secondly, each of the 14 data spaces is being developed by different stakeholders, which can lead to a fragmentation of technical standards and governance models. Finally, the AI Act places stringent requirements on "high-risk" AI systems (e.g., in health, critical infrastructure, law enforcement). These requirements mandate detailed documentation of training data, data provenance, and robust quality checks to minimise bias and ensure accuracy. Meeting these standards for massive, distributed datasets across CEDS requires a level of data governance that few organisations currently possess.

- **Tension with data protection rules:** Some legal experts argue that the framework for data sharing, particularly in sensitive areas like the EHDS, relies on mechanisms that side-step the strict requirements of user consent under the GDPR by creating a "legal obligation" to share.

Relatedly, the secondary use of personal data (e.g., in the EHDS for research) requires anonymisation or strong pseudonymisation. However, critics point out that there is no universally agreed-upon definition of anonymisation, and with large, linkable datasets, the risk of re-identification is always present.

### *Technical Challenges*

AI models thrive on data that is clean, structured and consistently labelled. However, much of the data currently sitting in silos – especially industrial data – is raw, unstructured, or in non-standard formats.

The colossal task of cleansing, annotating and standardising these datasets across 27 member states and multiple sectors is a multi-year, resource-intensive undertaking. National differences compound this problem. For example, being able to share data across national health systems in a way that ensures mutual understanding and clarity of the meaning of that data requires agreement on common data models and APIs that are still under development for many spaces.

# Sector-Specific Opportunities and Challenges

Each of the 14 common European data centres faces distinct data fragmentation challenges, shaped by sector-specific histories and practices. For this reason, while all operate under shared legal foundations — notably the data governance act, the data act and the GDPR – their individual structure has been designed to allow sector-specific barriers to be addressed appropriately.

To illustrate how this works in practice, we focus on three strategically important data spaces: energy, health and agriculture.

## Energy

The Common European Energy Data Space (CEEDS) is a relatively new initiative. Its development began with a blueprint published in March 2024. The INSIEME project, which implements CEEDS, was officially launched on 15 April 2025 under the European Union's digital Europe programme.

### *Purpose and Vision of CEEDS*

The opportunities for AI to advance scientific discovery in the energy sector are immense, with far-reaching societal impact. The most immediate potential benefits lie in real-time monitoring and control of electricity grids, thereby improving stability and reducing outages. Europe's grid management is becoming more complex due to the increase in distributed energy resources such as home solar panels and batteries. Traditional grid infrastructure was designed for one-directional power flows, while today's bidirectional flows can cause local voltage spikes and instability. In addition, solar and wind generation fluctuate with weather conditions. Managing thousands of decentralised units and ensuring predictable supply requires advanced control systems and fast data processing.

CEEDS aims to connect fragmented energy data infrastructures across Europe into a federated system with harmonised data formats and metadata standards to enable data exchange among stakeholders. This federated architecture will link data from transmission system operators (TSOs), distribution system operators (DSOs), energy producers, electric-vehicle charging networks, and consumers. Doing so will give AI systems access to rich, multi-domain datasets, allowing CEEDS to support AI applications like grid balancing, renewable forecasting and demand response. In the long term, CEEDS aims to support the creation of a detailed digital twin of the European energy system. Such a model could improve forecasting, grid optimisation and the integration of renewable energy sources.

Beyond grid management, AI is also playing a growing role in energy-related scientific discovery. In materials science, AI is being used to discover new materials useful for batteries, solar cells and hydrogen storage by simulating molecular interactions and performance.[18] AI is also used in the discovery of new sustainable energy sources, such as biofuels. In fusion research, AI models are used to dynamically adjust magnetic fields in tokamaks to stabilise plasma.[19] AI is also used to support more sustainable buildings by improving design and operational efficiency, integrating building information modelling to simulate energy performance and enabling automated maintenance.[20] As buildings account for 39% of global carbon emissions, these gains are substantial.

### *Criticisms and Challenges of CEEDS*

CEEDS aims to federate multiple national and sectoral platforms, but the energy sector's reliance on heterogeneous data formats and legacy IT systems makes interoperability complex and costly. In addition, a fully defined governance model for data access, monetisation and liability in the sector has yet to be established. Stakeholders are concerned about data sovereignty and intellectual property rights, especially given that energy data related to production, consumption, and pricing is often considered commercially sensitive.

Sustainable business models for CEEDS are also uncertain, particularly for smaller players that may not see immediate returns on investment. As a result, CEEDS risks becoming a policy-driven project without strong interest from the market. Finally, energy systems are still largely regulated at the national level, with each country maintaining its own data governance frameworks, privacy rules and technical standards. This complicates cross-border harmonisation and slows integration.

### *Regulatory Bodies and Frameworks*

Although there are several regulatory bodies supporting energy data standardisation in Europe, enforcement remains complex due to national sovereignty and diverse market structures. The most important is the European Commission's directorate-general for energy (DG ENER). In July 2024, it published guidelines requiring member states to report and align their national electricity metering data practices within a common interoperability repository.[21] The European Network of Transmission Operators for Electricity (ENTSO-E) and DSO Entity, the association for distribution system operators, launched the joint working group data interoperability repository in July 2025.

## Life Science

Europe has invested heavily in public life science data infrastructures. Recent initiatives include the European Commission's life sciences strategy (2025), the forthcoming biotech act (expected 2026) and the life sciences R&D data assembly (2026). Together, these initiatives aim to coordinate data governance, streamline regulatory rules, and ensure AI-readiness for biotech and AI applications.

The centrepiece of these efforts is the European Health Data Space (EHDS), one of the European Union's most ambitious digital health initiatives. The EHDS aims to create a secure,

18 US Department of Energy, Artificial Intelligence for Energy https://www.energy.gov/topics/artificial-intelligence-energy [accessed 05 March 2026].

19 Eirwen Williams, "'It Finally Works': Scientists Use AI to Control Nuclear Fusion (And It Could Unlock Limitless Clean Energy for Humanity)," *Energy Reporters,* 10 June 2025.

20 Andrew R. Chow, "How AI Is Making Buildings More Energy Efficient," *Time Magazine,* 11 December 2024.

21 European Commission: Directorate-General for Energy, *Electricity Metering and Consumption Data Interoperability – Guidance for the Reporting of National Practices in Accordance with Commission Implementing Regulation (EU) 2023/1162* (Publications Office of the European Union, 2024).

interoperable ecosystem for health data exchange across all member states. It has two main pillars: EHDS1 for patient health data and EHDS2 for the secondary use of health data for research, innovation and policymaking.

### *Purpose and Vision of the EHDS*

Life science represents one of the most promising domains for scientific AI, accelerating both computational simulations and traditional empirical research.

AI can be used to integrate multiple data sources (e.g., genomics, transcriptomics, proteomics) to uncover disease mechanisms and identify biomarkers, which supports the shift towards personalised medicine. Precompetitive research platforms like OpenTargets use AI to aggregate both public and private data, aiding in drug target validation across disparate data sources.[22] AI is now also being combined with new technologies in gene editing, single-cell sequencing and cell models to generate novel types of genomic data to understand systems biology.

In drug discovery, AI is now used to optimise clinical trial design, patient recruitment and dosing strategies.[23] An emergent field where AI is critical is synthetic biology and chemistry, where AI can be used to design synthetic proteins, antibodies and pharmacological compounds with desired functions and properties (e.g., binding affinity, low toxicity).[24] On the clinical side, AI is now actively used to diagnose X-ray images as well as tissue samples for more effective radiology and pathology.[25] Finally, in epidemiology, AI models track disease outbreaks, predict health trends and support policy decisions for pandemic response and vaccine distribution planning.

The EHDS aims to unlock these opportunities by enabling access to healthcare data from over 500 million European Union citizens, creating unprecedented potential for AI model development and training. A federated architecture will allow AI models to be trained across multiple countries without transferring raw data, preserving privacy while supporting cross-border collaboration. In addition, automated processes for data access approvals and compliance checks are expected to accelerate research projects and clinical trials.

### *Criticism and Challenges of the EHDS*

Substantial differences across national health systems are likely the greatest barrier to data harmonisation, which requires standardisation of data formats, APIs and protocols. Health data is highly sensitive, and ensuring compliance with the GDPR and other EU privacy laws is complex. This is complicated by unclear definitions and overlapping legal frameworks (e.g., the GDPR, the data act, the medical devices regulation) which can be contradictory, particularly for the secondary use of data in research. Here, there is substantial disagreement about whether the EHDS should use an opt-in or opt-out model for secondary data use. Implementation of the EHDS also requires significant funding, a skilled workforce and infrastructure upgrades. Smaller member states and healthcare providers may struggle to meet these requirements, potentially widening disparities across Europe.

---

22 Laia Pujol Priego and Jonathan Wareham, "Data Commoning in the Life Sciences," *MIS Quarterly,* 48.2 2024, pp.491–520.

23 Gourav Khanna, "Top AI Use Cases in Life Sciences," *appwrk,* 07 July 2025.

24 Cem Dilmegani and Sena Sezar, "Generative AI in Life Sciences: Use Cases and Examples," *AIMultiple,* 29 July 2025.

25 Geng Zang and Zilong Wang, "Top Ten Transformative Impacts of Artificial Intelligence on Life Sciences," *AI in Clinical Medicine, 1,* 2025.

Beyond clinical data, fragmentation is also acute in pre-clinical research. Despite the universality of human biology, pre-clinical life science data often lacks a shared ontology and consistent metadata. Such data is generated using different instruments, sequencing technologies, and analytical pipelines, each producing unique file formats without universal metadata. This limits reproducibility and constrains AI training in all facets of drug discovery, such as genomics, proteomics and pharmacology. Furthermore, research data is often stored in institution-specific repositories or national infrastructures, governed by local privacy laws and funding mandates. Large consortia such as EMBL-EBI and ELIXIR have advanced integration, but smaller labs and regional projects often lack resources to adopt common standards. Rapid technological change adds further complexity, with the fast emergence of new sequencing and proteomics technologies often outpacing politically sluggish standardisation efforts.

### *Regulatory Bodies and Frameworks*

Several European institutions support life science data governance and standardisation. The European Medicines Agency (EMA) launched the European medicines regulatory network data standardisation strategy and invests in AI and big data analytics to support pharmacovigilance and clinical evaluation. ELIXIR coordinates national bioinformatics nodes across 23 countries and supports AI deployment through standardised data resources. Similarly, the European Genomic Data Infrastructure (GDI) is building a federated infrastructure for secure access to genomic and clinical data, with tools for natural language queries and federated AI model training.

## Agriculture

The Common European Agricultural Data Space (CEADS), officially launched in April 2025, is the European Union’s flagship initiative to create a secure, interoperable environment for agricultural data sharing across member states.

### *Purpose and Vision of CEADS*

Agriculture presents significant opportunities for AI-driven scientific discovery and productivity gains. AI systems can be used to analyse soil, weather and crop data to optimise planting, irrigation and fertilisation. When effectively implemented, AI-driven precision farming can boost yields by 15–20% while reducing input costs.[26] Computer vision and deep learning models can detect pests, diseases and nutrient deficiencies from drone or satellite imagery, and simulate climate impacts on crop performance. AI models can also aid in recommending regenerative practices and can optimise carbon sequestration and biodiversity goals.[27] Finally, like in medicine, AI can be used to analyse genetic traits to accelerate plant and animal breeding to develop climate-resilient and high-yield varieties.

The goal of CEADS is to build a secure, interoperable infrastructure using common standards and APIs across a federated architecture, allowing AI models to train across distributed datasets without moving raw data. CEADS provides a catalogue of European Union-wide datasets from farms, machinery, sensors and public administrations, enabling applications such as precision agriculture, crop yield prediction, predictive maintenance for agricultural machinery and supply chain optimisation.

26 Kartik Mittal, “AI Driven Agriculture in 2025: Enhancing Crop Yields, Sustainability, and Workforce Skills,” *IIDE*, 11 December 2025.
27 Ashutosh Kumar Singh and others. “The Role of Artificial Intelligence in Agriculture: A Comprehensive Review,” *Scope*, 14.3, 2024, pp.1203-1215

### *Criticism and Challenges of CEADS*

CEADS faces several major challenges in its implementation. Agricultural data comes from diverse sources such as farm machinery, IoT sensors, drones, satellites and public registries, each using different formats and standards. Integrating these heterogeneous datasets into a secure, interoperable infrastructure is complex and costly, especially for smaller stakeholders. The rapid expansion of digital monitoring technologies further complicates standardisation, with new sensors, imaging systems and AI tools emerging.

Industry dynamics also reinforce fragmentation. As manufacturers of agricultural machinery integrate increasing amounts of telemetry into their equipment, they often adopt closed platform strategies to protect their commercial interests. Proprietary data formats and protocols (e.g., across competing platforms such as John Deere and CNH) can lock agricultural data into vendor-specific ecosystems, limiting interoperability.

Governance and adoption present additional barriers. Designing trusted governance frameworks that balance data sovereignty, privacy and commercial interests remains difficult. The sector encompasses a wide range of stakeholders – including farmers, cooperatives, machinery manufacturers, tech providers and public authorities – whose interests are not always aligned. Building a consensus-based federated system is therefore a major challenge. In addition, farmers and SMEs often lack clarity on the economic benefits of data sharing, which can slow adoption, while low digital readiness makes integration into advanced data spaces difficult.

### *Regulatory Bodies and Frameworks*

The European Commission, through its directorate-general for communications networks, content and technology (DG CONNECT) and the directorate-general for agriculture (DG AGRI), leads CEADS under the digital Europe programme.

However, as the sector lacks widely adopted interoperability standards for key datasets, enforcement of standards is often indirect, relying on legal frameworks and funding conditions rather than strict mandates. Some technical standards exist, such as ISO 11783 (ISOBUS) for tractors and machinery for agriculture and forestry, but adoption remains uneven across the sector.

# Policy Goals and Levers

How can Europe nurture investment in scientific data infrastructures that enable AI-driven scientific discovery? This section explores some of the most recurrent challenges to AI data governance, as well as related issues concerning accessibility, sustainability, and scalability, and outlines potential policy responses.

## Centralise and Harmonise Governance

### *Compliance*

Across scientific and political bodies, there is a broad consensus that harmonising the foundations of AI-ready data is a critical first step. Consistent data semantics and metadata standards are essential for training reliable AI systems. Yet, aligning Europe's diverse scientific data infrastructures remains tactically and politically challenging.

European regulatory bodies can play an important role by embedding data harmonisation into existing regulatory and funding processes. For example, the European medicines agency could establish data standards that are consistent across all phases of drug discovery, regulatory approval and monitoring. This, in turn, would help drive adoption of data standards across the related industries in the pharmaceutical, biotechnology and medical technology sectors.

Likewise, similar European Union agencies – such as DG ENER or ENTSO-E for energy or DG AGRI for agriculture – could define sector-specific "minimum interoperable dataset" (MID) profiles. These profiles would specify required schemas, vocabularies, provenance and quality checks that must accompany publicly funded datasets within the common European data spaces. Embedding data harmonisation into broader funding and regulatory processes would give European bodies strong leverage to influence data harmonisation and compliance.

*'Embedding data harmonisation into broader funding and regulatory processes would give European bodies strong leverage to influence data harmonisation and compliance.'*

### *Legal Interpretation*

Closely linked to compliance is the challenge of legal interpretation. Even the most legally compliant organisations struggle with ambiguous or contradictory readings of the GDPR, the data governance act, the data act and the AI act.

In life science, for example, regulatory provisions differ on how data collected through primary clinical care may be repurposed for secondary scientific use. In the energy sector, issues with interpretation are amplified by a highly layered market structure comprising generators, transmission operators, distribution operators, power exchanges, industrial users and consumers, each with distinct incentives and regulatory obligations.

Following calls for a new, centralised EU "one stop" interpretation facility, the European Union established the Data Spaces Support Centre (DSSC). The DSSC provides common requirements, legal guidance, templates, and governance best practices for organisations building or participating in data spaces. However, while it is indispensable as a coordination hub, its remit is primarily advisory, and it does not offer binding legal interpretations, sector specific adjudication or real time compliance resolution. The DSSC's current remit should be extended to cover the domains facing sector specific regulatory conflicts across the various European Union legislation, heterogeneous technical infrastructures, and diverging commercial incentives that exceed the scope of generic guidance.

### *Education*

Europe has seen rapid growth in the number of educational institutions offering programmes in data science and AI. However, the discipline of data stewardship remains underdeveloped. There is an opportunity to embed data stewardship more systematically in university curricula and professional programmes, with the aim of formalising European Union-compliant data curation practices.

Complementing education with European Union-level data stewardship accreditation could help organisations in need of disciplined data governance (e.g., hospitals, labs and utilities) identify suitably qualified professionals. Establishing European Union-wide data stewardship education and accreditation can cultivate a highly skilled workforce capable of ensuring that data is AI-ready and FAIR throughout its creation, curation, and federation.

## Optimise High-Quality Data Generation and Sharing

### *Private Sector Contributions*

Unlike physical assets, scientific data does not degrade with use. On the contrary, its value compounds with additional contributions, provided the data is of sufficient quality. Today, scientific data is no longer the exclusive domain of research infrastructures. In fact, as industrial processes increasingly rely on scientific insight, scientifically intensive, deep-tech industries are now considerable sources of scientific data in their own right.

Unsurprisingly, requirements for rigorous data reporting can be perceived as a burden or even a risk, particularly when they require disclosing competitively sensitive data. This applies not only to sophisticated organisations in sectors such as life science, energy or manufacturing, but also to less technologically mature industries such as agriculture or small-scale manufacturing.

*'Consequently, data governance must be designed to actively accommodate – and in many cases prioritise – the operational realities and incentives of private industry.'*

Consequently, data governance must be designed to actively accommodate – and in many cases prioritise – the operational realities and incentives of private industry. This means establishing rules that allow companies to share data without compromising trade secrets, intellectual property or commercially sensitive information, while offering safeguards, standardised contractual terms and sector specific usage protocols.

Governance should also make it economically rational for companies to contribute data by ensuring proportional obligations and streamlined compliance and reciprocity mechanisms, enabling them to benefit from shared AI resources. Clearly, the diverse nature of stakeholders across the 14 common European data spaces makes the creation of common data policies and protocols challenging. However, this needs to happen at the CEDS level, with further sub-CEDS differentiation based upon business type and technical maturity. In some instances, the adoption of data standards can be tied to economic subsidies (e.g., in the case of industry and agriculture) along with other levers to enhance digital maturity for smaller organisations with fewer resources.

### *Intellectual Property*

There are broader concerns that current IP regulation is insufficient to delineate appropriate copyright and patentability thresholds for algorithmic methods. Specifically, what specific legal and intellectual property (IP) framework is needed to support collaboration between the public research and the private technology sector that ensure fair returns for public investment? And conversely, what incentives and IP can encourage private companies to share AI tools and data with the scientific community? DeepMind's release of AlphaFold is an exceptional case of the mutual benefits achievable when private and public sector interests align. Consequently, there is a need to publish EU-level guidance on algorithmic and training data IP that balances protection with revelation. This need is particularly acute for AI algorithms, because it is possible to nominally disclose the code and architecture without the training protocols and data. This makes it possible to legally disclose the model while withholding operational knowhow, making the patent more symbolic than socially useful.

We currently see patterns of open-core and hybrid approaches that allow the public use of foundational components (e.g. model architecture and code, weights, training and inference code) while commercially protecting proprietary modifications, training data, or other operational layers. Greater clarity on these boundaries and thresholds will foster the bi-directional sharing of scientific AI across academic and private entities.

### *Managing Heterogeneity*

As industries integrate scientific AI into their core processes, the intersection with IoT data increases. These two data types differ profoundly in origin, purpose, structure and evidentiary requirements, which means they need distinct data governance approaches. Scientific data requires strict metadata, provenance, reproducibility and quality standards to be considered evidence grade. In contrast, IoT data is continuous, high volume, heterogeneous and often noisy, produced by distributed devices for operational monitoring rather than scientific inference.

Effective data governance must therefore impose stronger ontologies, validation protocols and FAIR+ requirements on scientific data, while focusing on interoperability, temporal synchronisation and privacy-preserving access for IoT streams. In short, governance frameworks should avoid a one size fits all model: scientific data spaces require rigorous curation, standardised metadata schemas, and long term stewardship, whereas IoT data spaces must prioritise real time access rules, device level security, and scalable mechanisms for managing industrial confidentiality. Treating these data types according to their intrinsic characteristics ensures that both can meaningfully support AI development without compromising quality, integrity or trust.

### *Data Factories and Brokers*

Despite extensive data collection efforts in natural sciences, scientific AI still suffers from a lack of high-quality training data. Large language models are trained on billions of text, image and video files, but replicating this scale in scientific domains is impossible due to the strict discipline required for data generation, curation, validation and representation. Consequently, and counterintuitively, scientific AI lags behind mainstream AI in access to robust training data.

To address this gap, a new class of scientific infrastructure is emerging: scientific data factories. These entities vary in form, but often combine highly automated laboratory experimentation with *in silico* exploration (i.e., by means of computer modelling or computer simulation) to significantly accelerate scientific cycles. Generative predictions are merged with physics-based models to provide focused, proprietary data sets designed for both broad scientific exploration and focused industrial applications, particularly in life and materials sciences. These systems often operate with partially or fully closed-loop systems that offer epistemic traceability, consistent representations and cross-lab replicability.

Complementing data factories are data brokers, which aggregate datasets from laboratories, research infrastructures, industry partners and public repositories to assemble the high quality training corpora required for scientific AI models. Data brokers harmonise heterogeneous data produced under different protocols, measurement standards and regulatory constraints, ensuring provenance, validation and scientific consistency. Because they operate at the intersection of public research and proprietary industrial data, governance frameworks must address issues of data rights, licensing, traceability and IP boundaries, establishing clear rules for how aggregated datasets can be used, shared and commercialised.

As new players in scientific AI, both data factories and brokers have huge potential to address the current scarcity of high-quality scientific training data. To support their development, European Union funding and governance of scientific AI should encourage the unique attributes of data factories and brokers to feed evidence-grade data into the CEDS. Beyond financial support, such measures should include granting these entities an appropriate legal status as CEDS nodes, implementing effective licensing and intellectual property regimes for scientific intellectual property (in both pre-competitive and industrial applications), establishing interoperability protocols, and defining certifications and evidence-grading standards.

## Accelerate Sustainable Infrastructure

### *Equitable and Appropriate Access*

By any measure, Europe is now experiencing a compute deficit. Traditionally, academic research has been publicly funded, guided by the ethos of scientific understanding as a public good. At the same time, private companies across life science, engineering, energy, transportation and many other industries are increasingly reliant on computational research.

Does it logically follow that scientific AI compute infrastructure should be treated as a purely public service, or should some differentiation be made between public and private research? What are the best mechanisms to govern access to high-performance computing and scientific AI capacity to ensure fair access for both academic researchers and incumbent industries, with sufficient capacity to support SMEs and start-ups founded on scientific AI?

To address this, there is a need for a European Union-wide allocation system that balances public and private computing priorities, with categories ranging from public health, climate science and academic research to private sector SMEs and deep-tech startups.

### *Sustainable Operations*

AI computing demands immense energy resources. Given the high energy costs of operating the massive data centres and AI gigafactories required to train large language models, scientific AI capacity building and operations must account for both economic costs and environmental impacts.

Europe should expand the supply of low-carbon energy sources for high-performance computing and scientific AI gigafactories, alongside actionable energy management key performance indicators that allow full environmental accountability. This will help minimise environmental impacts in the near term, while guiding the development of environmentally sustainable AI infrastructure in subsequent technology generations.

## The Role of RAISE

Scientific AI is now advancing at an extraordinary pace. Many of the most significant challenges are unprecedented, requiring European institutions to develop novel and actionable policy levers. RAISE, the European Union's Resource for Artificial Intelligence Science in Europe, has many unique opportunities to help shape European scientific AI policy.

As most scientific data will remain heterogeneous and locally generated in the intermediate future, it is crucial that such data is FAIR+, evidence-grade, and protected from getting lost in the fog of federation. To mitigate this risk, RAISE can implement common practices, acting as a harmonisation layer across the different CEDS. Specifically, it can develop shared interpretation frameworks, compliance resources and protocols that go beyond conceptual guidance to facilitate common data resource contributions and value appropriation across the broad spectrum of actors.

> *'RAISE should serve as a harmonisation layer across different data spaces… ensuring data remains 'evidence-grade' and FAIR+'*

In the area of education, RAISE can champion data stewardship curricula, along with pan-European training programmes for hybrid scientific AI roles. With the growth of data factories and brokers, RAISE should define certification standards for data factories and intermediaries as trusted CEDS nodes, ensuring interoperability, provenance and FAIR+ compliance.

RAISE can further contribute to a European Union-wide compute allocation system to balance priorities across public health, climate science, academic research and private-sector innovation. Finally, it can provide guidance on future AI policy at the intersection of public, private and scientific interests. As sectors such as life science, energy, manufacturing, aerospace and agriculture grow increasingly reliant on scientific AI, a nuanced understanding of appropriate legal structures and intellectual property considerations is needed to balance public, private and environmental interests.

## Conclusion

Europe stands at a pivotal moment in scientific history. The integration of artificial intelligence into research promises breakthroughs in energy, health, materials, and agriculture – domains critical to both societal well-being and economic competitiveness. Yet, realising this promise will require Europe to address several structural challenges, including fragmented data ecosystems, insufficient compute capacity and unclear governance for emerging infrastructures such as data factories.

These challenges are not only technical, but also political and economic. Without decisive action, Europe risks falling behind in an era where AI-driven discovery is rapidly becoming a key determinant of global scientific leadership. With coordinated investment, governance, and talent development, Europe can transform its proud scientific legacy into a future of innovation – a future where AI amplifies discovery, strengthens sovereignty and delivers solutions to humanity's most pressing challenges.

# Annex 1 – Resources

## Energy Resources

**EUROSTAT Energy Database**
Comprehensive statistical data on energy production, consumption, prices, and market indicators. It also includes data on renewable energy, energy balances and import dependencies. https://ec.europa.eu/eurostat/web/energy/database

**Energy and Industry Geography Lab**
Geospatial data on energy production infrastructure across Europe.
https://energy-industry-geolab.jrc.ec.europa.eu

**European Energy Storage Inventory**
Real-time energy storage deployment across the European Union. It maps technologies like batteries, pumped hydro, hydrogen, and thermal storage, and supports the REPowerEU plan and the strategic energy technology plan. https://ses.jrc.ec.europa.eu/

**International Energy Agency**
Global energy data. https://www.iea.org/

**JRC-IDEES Database**
Integrated database of the European Energy System. Useful for energy modelling, policy analysis and understanding energy system dynamics. https://joint-research-centre.ec.europa.eu/scientific-tools-and-databases-0/potencia-policy-oriented-tool-energy-and-climate-change-impact-assessment/jrc-idees_en

## Life Science Resources

**ELIXIR**
A pan-European research infrastructure that coordinates life science data resources across 23 countries. Enables large-scale model training for genomics, disease prediction and drug response. https://www.elixir-europe.org

**EMBL-EBI**
A global leader in molecular biology data resources, and part of the European Molecular Biology Laboratory (EMBL), that includes:
- Ensembl: Genome browser for vertebrates and other eukaryotes.
- UniProt: Protein sequence and functional information.
- European Nucleotide Archive (ENA): Repository for nucleotide sequence data.
- GWAS Catalog: Genome-wide association studies.

EMBL-EBI hosts the AlphaFold Protein Structure Database in collaboration with DeepMind.
https://www.ebi.ac.uk

**Swiss Institute of Bioinformatics (SIB)**
A leading provider of open bioinformatics resources and tools that includes:

- Expasy: Swiss Bioinformatics Resource Portal.
- SwissRegulon: Regulatory site annotations.
- STRING: Protein-protein interaction networks.

http://www.sib.swiss/

**Protein Data Bank**
A repository for 3D structural data of proteins and nucleic acids. Essential for training models in structural biology, drug docking and molecular dynamics. http://www.rcsb.org/

**Europe PMC**
A free database of life science and biomedical literature. Supports natural language processing and literature mining for biomedical research. https://www.europepmc.org

## Agriculture Resources

**Eurostat Agriculture Database**
Comprehensive EU agricultural statistics, including farm structure, economic accounts, production, prices, organic farming and agri-environmental indicators. Supports predictive models for crop yields, price forecasting, and sustainability analysis. https://ec.europa.eu/eurostat/web/agriculture/database

**ECPGR Central Crop Databases**
Genetic resources for major crops in Europe, including passport data, characterisation and evaluation of germplasm collections. Enables genomic selection, breeding optimisation and biodiversity studies. https://www.ecpgr.org/resources/germplasm-databases/ecpgr-central-crop-databases

**Joint Research Centre (JRC) Data Catalogue**
EU datasets for agriculture, land use, and environment. Crop yield monitoring, nutrient load models, land cover surveys (LUCAS), and climate impact data. Supports precision agriculture, climate risk modelling, and sustainability assessments. https://data.jrc.ec.europa.eu/dataset?keyword=agriculture

**INRAE Research Data Scope**
Agroecology, biodiversity, climate adaptation, and bioeconomy. Experimental data on crops, soils, and ecosystems. https://www.inrae.fr/en

**FAO AGRIS Network**
Global agricultural science and technology literature, widely used in Europe.
https://www.iita.org/knowledge/knowledge-center/research-support/databases/

# Annex 2 - List of Participants at the Copenhagen High-Level Breakfast

| | | |
|---|---|---|
| **Matthias Bethge** | University of Tübingen | Professor for computational neuroscience and machine learning |
| **Ewan Birney** | European Molecular Biology Laboratory (EMBL) | Interim executive director |
| **Martin Brynskov** | University of Copenhagen | Scientific director, AI and digital |
| **Sašo Džeroski** | Jožef Stefan Institute | Head of department, knowledge technologies |
| **Alexander Hammer** | Dunia Innovations | Co-founder and managing director |
| **Agata Laydon** | Google DeepMind Impact Accelerator | Portfolio lead for life sciences |
| **Silke Obst** | European Health and Digital Executive Agency | Head of unit, digital, industry and space |
| **Jan Palmowski** | The Guild of European Research-Intensive Universities | Secretary general |
| | University of Warwick | Full professor, modern history |
| **Liviu Ştirbăţ** | Directorate-General for Research and Innovation European Commission | Head of unit, AI in science and critical technologies |
| **Jean-Philippe Vert** | Bioptimus | Chief executive officer |
| **Max Welling** | CuspAI | Chief technology officer and co-founder |
| | University of Amsterdam | Professor and research chair, machine learning |

# Annex 3 - List of Interviewees

| | | |
|---|---|---|
| **Diego de Alamo** | GSK | Principal scientist |
| **Elena Bou** | InnoEnergy | Co-founder and executive board member |
| **Stavros Chatzipanagiotou** | Fusion4Energy | Senior policy advisor |
| **Simon Dürr** | HES-SO Vallais | Assistant professor |
| **Stefano Gaburro** | Tecniplast | Scientific director |
| **Vadim Kotov** | evotec | Structural biologist |
| **Andrej Ondracka** | | Science and bioinformatics consultant |

# References

Chow, Andrew R. "How AI Is Making Buildings More Energy Efficient," *Time Magazine*, 11 December 2024 https://time.com/7201501/ai-buildings-energy-efficiency/ [accessed 05 March 2026]

Dilmegani, Cem and Sena Sezar. "Generative AI in Life Sciences: Use Cases and Examples," *AIMultiple*, 29 July 2025 https://web.archive.org/web/20251109011626/https://research.aimultiple.com/generative-ai-in-life-sciences/ [accessed 06 March 2026]

Duranton, Sylvain. "What Leaders Need to Know About Open-Source vs. Proprietary Models," *Forbes*, 07 July 2025 https://www.forbes.com/sites/sylvainduranton/2025/07/07/what-leaders-need-to-know-about-open-source-vs-proprietary-models/ [accessed 05 March 2026]

Foss Solbrekk, Katarina. "Three Routes to Protecting AI Systems and Their Algorithms Under IP Law: The Good, the Bad and the Ugly," *Journal of Intellectual Property Law & Practice*, 16.3, 2021, pp.247–258, doi.org/10.1093/jiplp/jpab033

Güell Paule, Laila. "EU Strives to Boost Digital Skills as New Study Highlights Key Gaps," *Digital Skills and Jobs Platform*, 07 March 2025 https://digital-skills-jobs.europa.eu/en/latest/news/eu-strives-boost-digital-skills-new-study-highlights-key-gaps [accessed 05 March 2026]

EuroHPC Joint Undertaking. "Open Calls for Proposals" https://access.eurohpc-ju.europa.eu/ [accessed 05 March 2026]

European Commission. "European AI in Science Strategy," *Communication of the European Commission*, COM(2025)724, 08 October 2025.

– –. "EU Launches InvestAI Initiative to Mobilise €200 Billion of Investment in Artificial Intelligence," [press release] 11 February 2025 https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence [accessed 05 March 2026]

European Commission: Directorate-General for Energy. *Electricity Metering and Consumption Data Interoperability – Guidance for the Reporting of National Practices in Accordance with Commission Implementing Regulation (EU) 2023/1162* (Publications Office of the European Union, 2024)

European Commission: Directorate-General for Research and Innovation. *Divided We Fall Behind: Why a Fragmented EU Cannot Compete in Complex Technologies* (Publications Office of the European Union, 2025)

European Commission: Directorate-General for Research and Innovation and PwC EU Services. *Cost-Benefit Analysis for FAIR Research Data – Cost of Not Having FAIR Research Data* (Publications Office of the European Union, 2018)

Khanna, Gourav. "Top AI Use Cases in Life Sciences," *appwrk*, 07 July 2025 https://appwrk.com/insights/top-ai-use-cases-in-life-sciences [accessed 06 March 2026]

Krenn, Mario and others. "On Scientific Understanding with Artificial Intelligence," *Nature Reviews Physics,* 4.12, 2022, pp.761–769, doi.org/10.1038/s42254-022-00518-3

Kumar Singh, Ashutosh and others. "The Role of Artificial Intelligence in Agriculture: A Comprehensive Review," *Scope,* 14.3, 2024, pp.1203-1215

Mittal, Kartik. "AI Driven Agriculture in 2025: Enhancing Crop Yields, Sustainability, and Workforce Skills," *IIDE,* 11 December 2025 https://iide.co/blog/ai-in-agriculture/ [accessed 06 March 2026]

Naujokaitytė, Goda. "Data Corner: Horizon Europe's Six Biggest Research Infrastructure Projects," *Science|Business,* 23 September 2025 https://sciencebusiness.net/news/r-d-funding/research-infrastructures/data-corner-horizon-europes-six-biggest-research-infrastructure [accessed 05 March 2026]

Pilz, Konstantin F. and others. "The US Hosts the Majority of GPU Cluster Performance, Followed by China," *EPOCH AI,* 05 June 2025 https://epoch.ai/data-insights/ai-supercomputers-performance-share-by-country [accessed 05 March 2026]

Pujol Priego, Laia and Jonathan Wareham. "Data Commoning in the Life Sciences," *MIS Quarterly,* 48.2 2024, pp.491–520, doi.org/10.25300/MISQ/2023/17439

Toews, Rob. "AlphaFold Is the Most Important Achievement in AI—Ever," *Forbes,* 03 October 2021 https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/ [accessed 05 March 2026]

US Department of Energy. *Artificial Intelligence for Energy* https://www.energy.gov/topics/artificial-intelligence-energy [accessed 05 March 2026]

Williams, Eirwen. "'It Finally Works': Scientists Use AI to Control Nuclear Fusion (And It Could Unlock Limitless Clean Energy for Humanity)', *Energy Reporters,* 10 June 2025 https://www.energy-reporters.com/environment/it-finally-works-scientists-use-ai-to-control-nuclear-fusion-and-it-could-unlock-limitless-clean-energy-for-humanity/ [accessed 05 March 2026]

Zang, Geng and Zilong Wang. "Top Ten Transformative Impacts of Artificial Intelligence on Life Sciences," *AI in Clinical Medicine,* 1, 2025, doi.org/10.14740/aicm6

theLisboncouncil
think tank for the 21st century



**The Lisbon Council asbl**
IPC-Résidence Palace
155 Rue de la Loi
1040 Brussels, Belgium
T +32 2 647 9575
www.lisboncouncil.net
info@lisboncouncil.net