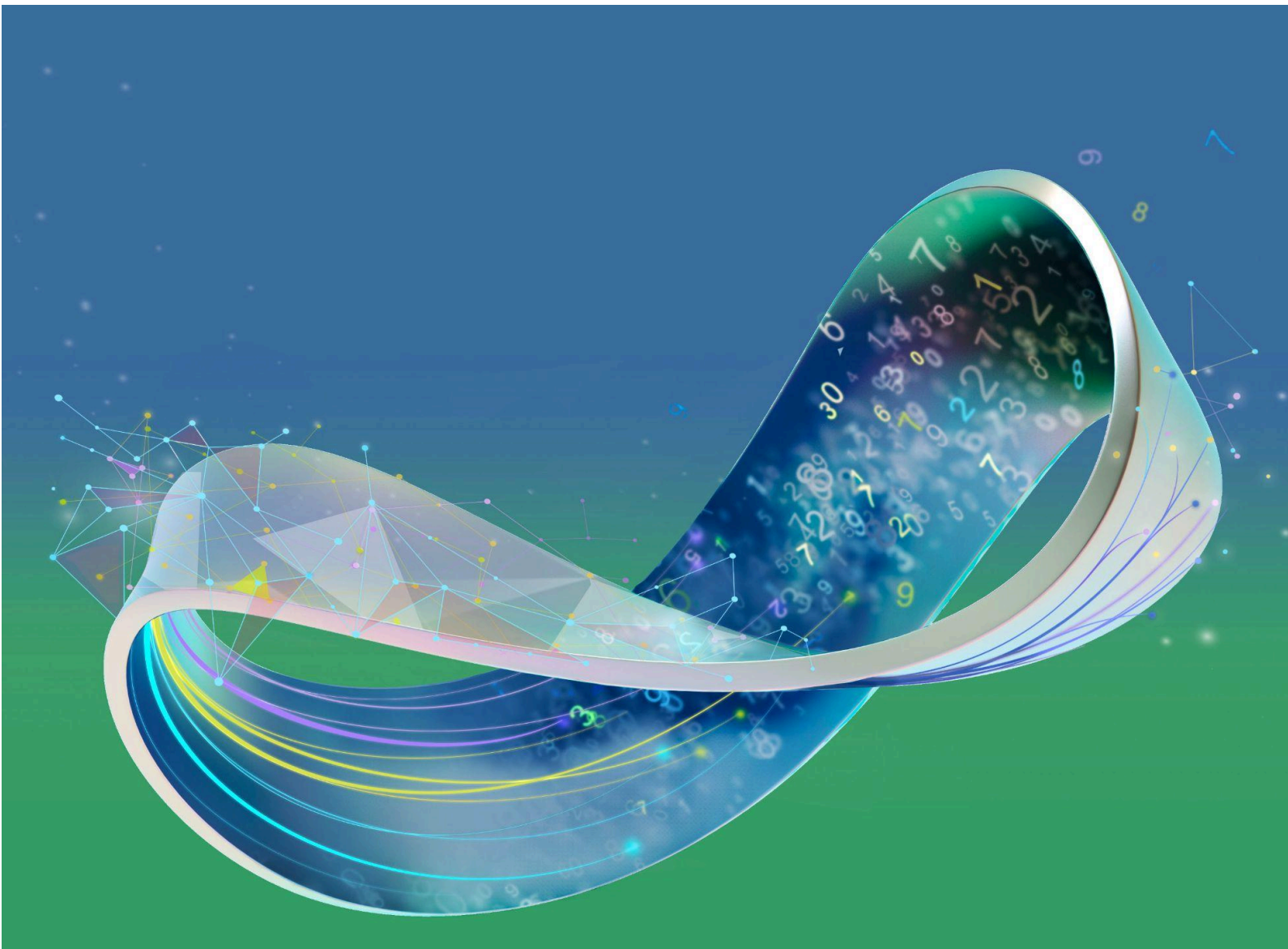
EMBL

A European Strategy for AI in Science

Call For Evidence | European Commission

June 2025



The convergence of life sciences and artificial intelligence is one of the most promising frontiers in research and innovation. *From genomics and drug discovery to imaging, biology is becoming increasingly data intensive, and AI offers unprecedented opportunities to accelerate discovery and unlock new scientific insights.*[1]

The ***European Molecular Biology Laboratory (EMBL)*** *welcomes the European Commission's initiative to develop a comprehensive AI in Science strategy. As Europe's intergovernmental research organisation dedicated to molecular biology, EMBL has long championed the development and integration of emerging technologies to advance biological discovery for the benefit of human and planetary health.* ***A prominent example in the field of AI is the partnership between Google DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI), which led to the development of the Nobel prize-winning AlphaFold*** *Protein Structure Database*[2]*, a landmark that illustrates how AI can redefine the pace and scope of scientific research.*

In this response, EMBL outlines critical priorities for enabling AI to accelerate scientific discovery across disciplines, and scientific opportunities associated with the life sciences. These priorities ***span three key areas****: first, the* ***need to invest in training and talent development*** *to prepare Europe's research workforce for an increasingly data-intensive scientific landscape; second, the* ***importance of strengthening data and computing infrastructure*** *to support the scale, quality, and complexity of AI-driven research; and third, the* ***value of fostering strategic partnerships and European coordination*** *to ensure that the benefits of AI are shared broadly and applied effectively across disciplines and Member States.*

PILLAR I: Training, Skills and Talent Development

AI uptake in science is a human challenge. The transformative potential of AI in science can only be realised through rapid, sustained investment in interdisciplinary training and the development of career pathways that attract, retain, and empower top talent. Europe must act decisively to close the skills gap and prepare its research ecosystem for a data-intensive future.

1. Upskill the Research Workforce

As science becomes increasingly data-driven, there is an urgent need to **equip researchers at all career stages with data literacy and AI fluency**. Tailored training programmes should be developed to support:

[1] EMBL published its AI Science strategy in February 2025: https://www.embl.org/editorhub/wp-content/uploads/2025/02/EMBL_AI-Strategy_Feb2025_Accessible.pdf

[2] Google DeepMind partnered with EMBL-EBI to develop the AlphaFold database and make openly available the results of an AI system which makes state-of-the-art, accurate 3D protein structure predictions, which previously took years, in minutes. https://www.embl.org/news/science-technology/alphafold-wins-nobel-prize-chemistry-2024/

- **Basic AI literacy**, for non-specialists, including bench scientists with no computational background.
- **Advanced AI tool use and development**, for computational biologists and bioinformaticians.
- **Leadership-level awareness**, to help decision-makers understand the strategic potential of AI and the training infrastructure it requires.

These efforts should be supported by a **broad portfolio of accessible training resources**, including **open e-learning platforms**, modular curricula, and on-demand learning opportunities designed for flexibility and scale. A coordinated EU-level initiative could help align these efforts, prevent duplication and **promote equitable access across domains and Member States**.

2. Attract and Retain AI Talent

The **life sciences** offer a wealth of high-impact, societally relevant challenges directly related to health and the environment, and an abundance of open, high-quality data, making them an **attractive domain for AI experts**. However, academic career pathways often fail to compete with industry for top AI talent. A coordinated approach might include efforts to:

- **Reform research assessment frameworks** to ensure that technology-centered contributions are valued, including tool development, data infrastructure, and interdisciplinary research.

- **Incentivise co-funded PhD and postdoctoral positions** where supervision is shared between academic and industry partners in critical sectors such as the life sciences.

- **Support fluid, cross-sectoral mobility, including industry-seconded roles** and short-term positions for AI experts to work in academic environments for a limited period of time. Existing initiatives, such as industry-funded postdocs based at EMBL (e.g. under the GSK-EMBL collaboration agreement[3]), provide scalable models for combining fundamental research and industrial innovation. Similar models should be widely supported and replicated.

- **Launch new or expand existing interdisciplinary PhD training programmes**, designed to foster collaboration across scientific domains, through joint supervision between AI and domain scientists, and ensure **early exposure** of AI and machine learning students **to real-world scientific problems**, through internships, collaborative projects and summer schools. To maximise their visibility and short-term impact at EU level, these programmes should be linked across institutions and build on existing models, such as the ones currently developed in the life sciences.

[3] https://www.embl.org/news/lab-matters/1702-new-collaboration-embl-gsk/

In early 2025, **EMBL and ELLIS signed a strategic agreement[4] to help position Europe at the forefront of AI-driven advancements in life sciences research**. A key focus of the partnership is training and capacity-building to foster closer ties between AI and molecular biology and bring more AI expertise into life sciences.

EMBL and ELLIS already launched a **pilot initiative to connect their respective internationally renowned PhD programmes**. As part of this collaboration, PhD students are jointly supervised by EMBL and ELLIS researchers, creating a new model for interdisciplinary training that combines cutting-edge AI with deep biological expertise.

3. Embed Engineers and Team Science in Research Careers

The deployment and scalability of AI in scientific research **rely heavily on engineers** who build and maintain the necessary computational infrastructure, algorithms, and tools, which are the technological backbone of AI systems. Yet, these professionals are often excluded from traditional academic evaluation and **dedicated career paths are missing**.

- **Establish formal career pathways for engineers** within academic research environments, and **adapt research assessment criteria** to adequately recognise technical contributions, such as infrastructure development, software engineering, and algorithm design. Cross-sectoral mobility notably through **industry-seconded engineer roles** within academia is also essential, and may be supported through a dedicated co-fund programme at EU level.

- **Embrace team-science environments** that enable close collaboration between domain specialists and AI experts, allowing them to remain centred in their own disciplines while acquiring enough cross-disciplinary understanding to work effectively together.

- **Adjust eligibility requirements in EU research training programmes** to allow broader participation, such as removing mandatory PhD requirements for certain roles in fellowship schemes (such as Marie Skłodowska-Curie Actions - MSCA).

One promising model is **ARISE2[5] (2025–2029)**, an MSCA-funded postdoctoral fellowship programme that offers talented STEM (science, technology, engineering and mathematics) fellows the opportunity to develop and improve technologies for life science research, while developing the expertise needed for a career in research infrastructures making them sought after experts both in academia and industry. Such programmes can serve as **scalable and transferrable examples for addressing structural gaps in research engineering careers** and AI specific needs.

[4] EMBL and ELLIS are piloting a joint supervised PhD programme through their collaboration agreement signed early 2025. https://www.embl.org/news/connections/new-strategic-partnership-for-ai-in-the-sciences/
[5] https://www.embl.org/training/arise2/

PILLAR II: Data and Infrastructure

The success of AI in science depends on the availability and quality of data and compute capacity. **Research infrastructures[6] (RIs) play a pivotal role in producing, curating, and maintaining the datasets** upon which AI models depend. For example, decades of public data curated and hosted at EMBL-EBI, through the Protein Data Bank Europe[7], played a central role in the development of Nobel prize winning AlphaFold. In the life sciences and beyond, these infrastructures ensure that AI systems are trained on accurate and real-world information.

Europe is already home to world-leading data infrastructures that support critical research sectors through open data resources. However, the sustainability of many of these data resources is at risk. **Without long-term investment and policy support, Europe's capacity to lead in data-driven research, and to fully capitalise on AI innovation, will be constrained**.

As AI becomes a foundational tool across scientific disciplines, Europe's investments in high performance computing (HPC) and AI Factories[8] must **meet the diverse needs of researchers**. While technically powerful and much needed, existing computing infrastructures could be better aligned with the workflows, data characteristics, and collaboration models of certain scientific domains.

1. Capitalise on Europe's Existing Infrastructure Strengths

Europe has built an impressive and globally unique system of RIs, many of which host open-access data resources that underpin both academic and industrial research, and are actively involved in building the EOSC Federation[9]. These resources, ranging from bioinformatics databases to imaging repositories, form the backbone of AI applications in fields like biomedical research, biodiversity, and climate modelling.

In addition to data services, many RIs are key drivers of technology innovation. **Some**, such as EMBL, **already serve as incubators for novel technologies**, including those powered by AI. Through close partnerships with industry across all technology readiness levels, large scale RIs[10] help translate scientific discoveries into industrial applications, securing Europe's technological sovereignty and its capacity to lead. RIs not only support and enable AI-driven innovation, they can also be transformed and enhanced through the integration of AI technologies into their operations and services.

To strengthen and fully leverage this foundation, the Strategy may:

---

[6] For example, EIROforum members (CERN, ESA, ESO, ILL, EUROfusion, European XFEL, ESRF, EMBL): https://www.eiroforum.org/

[7] https://www.embl.org/news/science/alphafold-using-open-data-and-ai-to-discover-the-3d-protein-universe/

[8] https://digital-strategy.ec.europa.eu/en/policies/ai-factories

[9] https://eosc.eu/eosc-about/building-the-eosc-federation/

[10] See EIROforum's position paper: *Research Infrastructures: Value and Impact for European Science, Industry and Society* (May 2020) - https://www.eiroforum.org/wp-content/uploads/eiroforum-paper-value-of-RIs-may-2020.pdf

- **Recognise existing RIs as strategic digital and technological assets** in the AI for science ecosystem, and leverage their dual role as data and technology providers;
- **Embed AI Factories within the broader European infrastructures ecosystem**, including the EOSC Federation, ensuring coherence and complementarity across Member States and European investments.
- **Accelerate AI-enabled technology development within RIs**, supporting the design and deployment of next generation scientific instruments (for example, smart microscopy) that improve and automate scientific workflows.

2. Safeguard the Long-term Sustainability of Open Data Resources

Unlocking the full potential of AI for Europe's science will depend on access to high quality and highly annotated data at scale, linked to key associated metadata types. Yet, the long-term **sustainability of many open biodata resources**, critical for advancements in health and environmental science, and hosted by European infrastructures that are used daily[11] by the private and public sectors, is **far from secure**. Many are reliant on short-term competitive funding sources, sometimes from a single or very limited number of funders. **This creates operational fragility[12] and long-term risk for services that are now mission-critical for Europe's research and innovation**.

If not addressed, the lack of secure and sustainable support for open AI-ready data resources could be a true roadblock for Europe's resilience and leadership as well as the uptake of world-class data and AI methods in science. The Strategy may aim to:

- Establish a **dedicated European Data Infrastructure Investment Programme**, with stable, long-term financing for open data resources, that are essential to European research and competitiveness.
- **Develop new infrastructure services for AI-generated outputs**, recognising that AI models and their predictions will increasingly complement traditional biodata. These new use cases require infrastructure for archival, curation, and sharing. Future investment should support the development of model repositories and dynamic databases to host pre-computed AI outputs, building on successful examples such as the AlphaFold database.

[11] In 2024, 5.4 billion requests from 6.1 million unique IP addresses to EMBL-EBI data resources came from EU member states.
[12] See Global Biodata Coalition Working Group on Sustainability. (2023). Consultation Paper. Zenodo. https://doi.org/10.5281/zenodo.8384740

3. Align Computing Infrastructure with Disciplinary Research Needs

While Europe has made major strides in developing cutting-edge compute capabilities, further efforts could **enhance accessibility and usability for a broader range of research domains**.

For example, life science applications often rely on data processing patterns and workflows that differ significantly from those used in traditional simulation-heavy fields like physics. For example, **environmental datasets have complex data linkage, while health datasets are increasingly large, complex, and sensitive**, requiring trusted environments and models that support data-intensive, collaborative, and cross-border research.

To maximise uptake, impact and long-term value of those investments, the following priorities could be considered:

- **Lower entry barriers** for researchers, particularly for those with smaller projects, by **simplifying user interfaces** and investing in **infrastructure models (e.g. local GPUs) that bring compute to data**;
- Tailor **computing environments to the specific needs of scientific domains**, supported by technical services that help reduce the burden of integration;
- **Enable federated access to sensitive data across borders** through pre-agreed, trusted environments that comply with relevant ethical and legal frameworks;
- Support **applied use-cases for the life sciences** and **long-term partnerships** between providers, research institutions, and RIs, creating sustainable ecosystems that encourage innovation and ensure long-term return on investment.

AI for the Life Sciences: Scientific Opportunities[13]

AI is particularly impactful in life sciences due to the abundance of structured data, from genomics to imaging, and the growing need for predictive models across scales of biology. Basic research also plays a substantial role, as fundamental scientific advances in biology, chemistry, and data science often lay the groundwork for transformative AI breakthroughs well before their industrial impact becomes visible. Scientific programmes and funding mechanisms need to be adapted to specifically support and enable new opportunities opened up through AI technologies. Examples of those opportunities in the life sciences include:

- **Tailored data generation for AI developments,** thereby filling in the gaps to deliver data required for AI model training and validation. Examples include systematic profiling in areas of biology with the strongest opportunities, such as **foundations of gene regulation, cell atlases or sequencing of the tree of life**. These efforts could go hand in hand with enhancements of RIs, thereby ensuring that data can be generated at scale.

- **Establish community frameworks to advance and systematically benchmark AI**. It will be critical to create and support frameworks whereby scientific communities are generating data, innovate AI and assess progress in a systematic manner. Examples include the successful workflow established in the protein sciences[14], which has been critical to deliver advances such as AlphaFold. This requires **support for benchmarking** definitions of the most impactful biological questions and metrics to assess progress. There is also a strong need for **coordinated meta-data standards**.

- **Advance the scientific method**, and pioneer new approaches to synergistically incorporate AI. The scientific method, standards and structures of recognition and credit allocation will be fundamentally impacted by AI. For example, the formulation of new scientific hypotheses will take big leaps impacting the scientific method as such. New themes and programmes, standards and best practices need to be created.

To maximise the potential and impact of AI on health and the environment, the Strategy should envisage a **dedicated "AI in Life Sciences Roadmap".** This Roadmap should focus on integrating AI tools for **critical areas such as genomics, macromolecular interaction networks and structures, and imaging**, while ensuring that Europe's life science data infrastructure is optimised for AI-ready datasets.

[13] For further details, please refer to EMBL's AI Science Strategy: https://www.embl.org/editorhub/wp-content/uploads/2025/02/EMBL_AI-Strategy_Feb2025_Accessible.pdf
[14] https://en.wikipedia.org/wiki/CASP

PILLAR III: Partnerships and European Coordination

EU-wide collaborations and industry engagement have always been central to scientific progress, and will be even more critical for the successful deployment of AI across research domains. Europe already benefits from a rich landscape of national and European centres of excellence, which should be further leveraged under a coherent AI in science strategy.

Existing European bottom-up AI networks bring together world-class AI expertise across Europe and provide **strong complementarity to domain-focused research organisations**[15]. Deeper collaboration between AI specialists and domain scientists for instance, through co-supervised PhD programmes as mentioned above, shared fellowships, or thematic centres of excellence, can help address complex scientific questions through interdisciplinary approaches. Such collaborations should be supported as **scalable models for future EU coordination**.

To fully realise the benefits of AI across scientific disciplines, the EU's AI in Science Strategy must also support a connected and inclusive research landscape. Programmes that promote **knowledge exchange and mobility between countries or regions with high AI readiness and those with emerging capabilities** will be essential to reduce disparities, **broaden participation across Member States,** and foster a more balanced uptake of AI.

International collaboration remains an essential cornerstone of European research, as it enriches and amplifies its diversity, impact and influence. Additionally, having strong and strategic international links ensures that Europe has access to expertise and unique data unavailable on our continent yet critical for understanding human and planetary life in its full complexity. **This international dimension should be an integral part of the Strategy.**

As the EC is paving the way towards a European AI Research Council, through the proposed **RAISE (Resources for AI in Science in Europe)**, it is essential to **build upon and interconnect existing ecosystems**. Rather than creating entirely new institutions, the emphasis should be on scaling and linking successful initiatives, including **thematic centres of excellence**, established large scale research infrastructures, and robust public-private partnerships. **The scientific leadership, international visibility as well as governance models and interdisciplinary networks, such as the ones of the EIROforum[16] member organisations, could be a solid foundation for Europe's AI future**. These structures can serve as coordination hubs and shared resource platforms, supporting cross-border collaboration, while ensuring sound governance, global visibility, and scientific relevance.

[15] For example, EMBL and ELLIS entered a strategic partnership earlier this year, focusing on collaboration in the fields of AI and life sciences through scientific exchange, training and joint research projects. https://www.embl.org/news/connections/new-strategic-partnership-for-ai-in-the-sciences/

[16] EIROforum members include: CERN, ESA, ESO, ILL, EUROfusion, European XFEL, ESRF, EMBL: https://www.eiroforum.org/

Conclusion

The transformative potential of AI in science can only be fully realised through a Strategy that is technologically ambitious, as well as **ethically grounded, inclusive, and aligned with the values of open science**. Building on its long-standing culture of collaboration, data sharing, and scientific excellence, Europe is **uniquely positioned to shape global AI in science agendas**, leading the way in setting standards and policy frameworks for responsible research, data stewardship, and trustworthy AI development.

Realising this potential will require ensuring that **AI advances are widely accessible across sectors, disciplines, and Member States**, and firmly **embedded in principles of transparency, reproducibility, and community-driven evaluation**. Responsible research practices, robust governance, and diverse participation must guide how AI is developed and deployed in science, with strong safeguards to promote ethical integrity, avoid biases, and protect sensitive data.

With its extensive data ecosystems, impact on human and planetary health, and strong legacy of technological innovation, the life sciences offer a compelling ground to demonstrate how AI can accelerate scientific discovery while advancing values of openness, equity and shared benefit. By investing in the long-term sustainability of data infrastructures, inclusive training ecosystems, and strategic cross-sectoral partnerships, Europe can drive meaningful AI innovation while remaining deeply committed to ethics and scientific responsibility.

———

About EMBL

The **European Molecular Biology Laboratory** (EMBL) is one of the world's leading research institutions, and Europe's flagship laboratory for the life sciences, founded in 1974. With 29 member states, the EMBL is an inter-governmental organisation with more than 110 independent research groups and service teams covering the spectrum of molecular biology across six sites in Germany, France, Spain, the United Kingdom and Italy. EMBL is driving visionary fundamental research, offers vital services to scientists globally, trains Europe's future scientific talent while actively engaging in technology transfer and industry relations, and nurturing policy dialogue in Europe and worldwide. www.embl.org

Contact information

Plamena Markova, Chief of International Relations: int-relations@embl.org

Anne-Charlotte Fauvel, EU Relations Lead: anne-charlotte.fauvel@embl.org