A European AI Powerhouse: Building an ecosystem of excellence and trust

A refresh of ELISE's Strategic Research Agenda

May 2023

elise

Contents

Summary

The European Union's visions for a healthier, wealthier, more sustainable society are intertwined with its ambitions for artificial intelligence (AI). Those ambitions seek to ensure European research excellence in this strategically important technology area, to generate economic and societal benefits through AI adoption across sectors, and to create an ecosystem of trust that aligns AI progress with public interests. Achieving these ambitions is vital in securing European sovereignty in AI; a strategic imperative that invokes the full AI development pipeline, from the manufacture of chips and hardware to the governance of data, to success in research, and the translation of this research to AI-enabled companies that can operate on a global scale. The ELISE network is building a powerhouse of European AI that can deliver these goals. By accelerating the technical capabilities of AI technologies, improving their performance in deployment, and aligning AI development with societal needs, the first phase of ELISE's work has already nurtured a thriving research and innovation community. Building on these successes, this refresh of its Strategic Research Agenda shows how sustained investment can ensure the long-term success of this innovation ecosystem.

ELISE's 2021 Strategic Research Agenda set a roadmap for AI technologies that can deliver European policy priorities. By identifying shared areas of interest between AI research and policy, this agenda provided a framework for creating a powerhouse of European AI. Since publication of this framework, ELISE has been leading the development of an ecosystem of excellence, delivering AI that is 'made in Europe' through a research agenda that reflects technical and policy priorities, and through translation and engagement activities to build a world-leading European AI community. In its first three years of operation, ELISE has already:

- Generated research insights and applications, driving forward a new wave of AI innovation.
- Created a world-leading infrastructure for advanced education, providing a gateway for high-potential researchers from around the globe to access education and research in Europe.
- Supported innovation and entrepreneurship to take AI innovations from research to practice, working across sectors and industries to help companies benefit from access to insights from the frontiers of technology development.
- Enabled mobility between top European labs, increasing connectivity to help build a European AI community.
- Responded to the need to operationalise AI governance principles by developing tools and techniques for trustworthiness by design, for example in the areas of fairness, explainability, and robustness.
- Catalysed new funding and programmes to help grow the European AI ecosystem over the long-term.

Over this period, both policy needs and technological capabilities have continued to evolve. Taking stock of the lessons that emerge from these changes, this Strategic Research Agenda Refresh looks at how technological progress is shaping AI's role in priority policy areas. Research from the network has created new techniques for trustworthy AI; security and privacy; explainability; AI integration into existing systems; and human-centric, responsible AI. The next wave of research progress should ground these new techniques in theory and advance their application, generating powerful new technologies that are supported by responsible innovation practices to deliver real-world benefits for science and society.

The EU's Innovation Missions highlight the benefits at stake. AI could enhance diagnostics and treatments for cancer, improving the lives of patients and their families. It could provide more accurate forecasts of the impacts of climate change for cities and regions, helping city administrations and citizens take action to increase their climate resilience. It could produce management tools to support ocean protection efforts, increasing understanding of marine ecosystems. It could enable monitoring systems to support soil health, while delivering new understandings of this vital natural resource. Its analytical power could provide insights and simulations that help cities transition to climate-neutral, sustainable ways of living and working. Translating these benefits from theory to practice requires action to advance the technical foundations and deployability of AI technologies, and align their design with the needs of human users; to accelerate end-to-end innovation through collaborations across sectors and industries; to connect AI development to the interests of affected communities; and to build an infrastructure for AI research that enables access to data, compute, and skills.

Research from the ELISE network has already contributed new tools for public health policy, for understanding the impact of climate change, for accelerating scientific research, and more. These achievements scratch the surface of what might be accomplished with sustained investment. The rapid technological change that underpins recent high-profile advances in AI capabilities, in particular Large Language Models, signals the progress that could be possible in the coming years. ELISE's research is accelerating progress at these technological frontiers, led by world-leading researchers across Europe. Building on its first phase of work, ELISE will continue to deliver on its agenda for a new generation of European AI, advancing AI technologies that are technically innovative and sophisticated, safe and effective in deployment, and aligned with social needs. Combining this research agenda with its successful initiatives to deliver high-quality advanced education, translate research insights to real-world benefits through collaboration with industry, and foster new research collaborations, ELISE's network will nurture a world-leading European AI community.

1. AI and the rapidly-changing strategic context for AI policy

AI today is intertwined with European policy ambitions for a healthier, wealthier, more sustainable society. Rapid progress in machine learning over the last ten years has raised aspirations for the potential of AI technologies to deliver innovative solutions to major scientific and societal challenges and deliver social and economic benefits. These aspirations are reflected across EU policy agendas, from the Green Deal to the AI Act, which position AI as both a priority policy area itself and an enabler of progress towards a thriving economy, sustainable environment, effective public administration, and healthy society.

National, regional, and international developments in the years since publication of the EU's White Papers on data governance and AI have brought renewed focus to efforts to grow Europe's AI ecosystem. The COVID-19 pandemic demonstrated a demand for AI that can be immediately deployed to solve pressing public policy challenges across a range of domains (Box 1). It also showed how interdependencies across global supply chains have created strategic vulnerabilities. 2022's energy crisis, and continuing pressures on European energy systems, highlighted again the importance of transitioning to sustainable energy sources. This transition requires strategies for the stable delivery of energy from green sources at times of peak demand, raising new questions about both the use of AI to support the green transition and the resources consumed by large-scale AI systems. At the same time, conflict in Ukraine has demonstrated new security vulnerabilities relating to data and digital infrastructure. Across these policy domains – public health, supply chains, energy, security, and more – issues of public interest, technological sovereignty, and AI capabilities interact. In response, governments and parliaments around the world are seeking mechanisms to advance national interests in AI and hold AI leaders accountable for their use of data-enabled technologies. The result is an urgent need to deliver Europe's AI agenda.

Responding to policy ambitions for a European AI ecosystem rooted in excellence and trust, in 2021 ELISE's Strategic Research Agenda set out a vision for a powerhouse of European AI. This Agenda provided a roadmap to connect advances in AI research with policy priorities, building a research ecosystem that secures European technological sovereignty. It focused on three shared goals at the AI research-policy interface:

- To be a world leader in AI and ensure that Europe has capability in a technology of strategic importance, Europe needs to be at the forefront of AI innovation, leading a new wave of innovation to deliver AI that is technically advanced.
- To translate the benefits from these technical advances to positive social and economic outcomes, AI needs to be robust in deployment.
- To ensure all in society benefit from progress in AI research and adoption, AI technologies need to align with societal interests.

ELISE's strategic approach is to accelerate progress at the frontiers of AI development, create methods that ensure AI technologies can be deployed safely and effectively, and connect AI capabilities to areas of societal need. ELISE research programmes provide a mechanism for delivering these goals. Alongside the agendas pursued by these programmes (Annex 1), five cross-cutting themes provide a framework to bridge from specific techniques or applications to the wider strategic landscape for AI development. Those themes focus on AI trustworthiness; security and privacy; explainability and transparency in AI; AI integration into existing systems; and the ethical and societal interests associated with AI development. Across these themes, ELISE's focus is advancing techniques and applications in machine learning. This field has catalysed recent excitement about progress in AI and underpins many of the AI systems in widespread use.

Alongside these research programmes, ELISE initiatives to encourage AI adoption in industry, build skills and attract talent across research and industry, and enhance research collaboration across Europe have translated the network's research to benefits for communities, businesses, and society. With 39 units across 14 countries, collaborations with the ELLIS network have provided an infrastructure to bring together Europe's leading AI researchers.[1]

Two years since the publication of ELISE's Strategic Research Agenda, this document reflects on progress across the network. It highlights ELISE's approach to creating a powerhouse of European AI through action to advance European AI research and nurture a thriving AI ecosystem. Building the bridges between AI research and policy identified by the previous Agenda, it explores what technology needs emerge from current policy priorities, and it considers progress made so far in the areas for action identified in 2021.

1 For an update on the ELLIS network, see: https://ellis.eu/news/four-new-units-join-the-ellis-network

Figure 1. **Summary of ELISE's 2021 Strategic Research Agenda**

Five themes

- Security and privacy
- Explainability, accountability, and decision-making
- Trustworthiness and AI certification
- AI integration across systems
- Embedding AI ethics

Three goals

- To be a world leader in AI, Europe needs to be at the forefront of AI innovation, creating AI that is technically advanced.
- To translate the benefits from these technical advances to positive social and economic outcomes, AI needs to be robust in deployment.
- To ensure all in society benefit from progress in AI research and adoption, AI technologies need to align with societal interests.

Research programmes

- Machine Learning for Health
- Robot Learning: Closing the Reality Gap!
- Geometric Deep Learning
- Human-centric Machine Learning
- Interactive Learning and Interventional Representations
- Machine Learning and Computer Vision
- Machine Learning for Earth and Climate Sciences
- Multimodal Learning Systems
- Natural Intelligence
- Natural Language Processing
- Quantum and Physics-Based Machine Learning
- Robust Machine Learning
- Semantic, Symbolic and Interpretable Machine Learning
- Theory, Algorithms and Computations of Modern Learning Systems

2. Progress towards a European AI Powerhouse

The EU's vision is for a uniquely European approach to AI; an approach that "centers on excellence and trust", delivered through a new generation of AI technologies and applications that are innovative, safe, and aligned with fundamental European rights.[2] This vision reflects strategic ambitions to increase Europe's AI capabilities, making Europe a world-leading hub for AI research and innovation, and to embed the rights and values enshrined in European law into AI development. Implementing it requires action to ensure that Europe is a leader in AI research, that AI advances translate from the lab to the market, and that the resulting AI technologies and applications work for people and society, now more than ever before.

ELISE has been at the forefront of efforts to create an ecosystem for AI that is 'made in Europe'. Its 2021 Strategic Research Agenda set a roadmap for delivering a new wave of progress in AI. Complementing this ambitious research agenda, its research, translation, and education initiatives have: generated new research insights and applications, created a world-leading infrastructure for advanced education, supported innovation and entrepreneurship, increased connectivity across the European research community, and catalysed new funding and collaborations. ELISE has already delivered tangible benefits for Europe's AI ecosystem. Its activities have built a European research community in AI, the development of which has become increasingly significant in the context of a growing need for European technological sovereignty.

Generated new research insights and applications

ELISE's 14 research programmes are advancing research across diverse technical and application domains, including healthcare, climate science, quantum machine learning, machine learning for robotics, human-centric AI, robustness, security, and more (see Annex 1). These programmes have delivered data science-informed insights for policymakers responding to COVID-19 (see Box 1); created machine learning tools to accelerate scientific discovery (see Box 2); enhanced understandings of the Earth and climate systems (see Box 3); created insights to improve diagnostics in healthcare (see Box 11); raised awareness of the importance of human-centred AI (see Box 10); and innovated with machine learning methods and theories to advance the core underlying technologies of AI (see Boxes 7, 8, 9, 10, 11). Some of the research highlights from this work are introduced across this document.

Created a world-leading infrastructure for advanced education

Demand for top AI talent continues to grow, and it is of vital importance that Europe can attract, train, and retain such talent. Convened with the ELLIS network, the ELISE/ELLIS PhD programme is an internationally recognised initiative that attracts global talent to study and research in Europe. Since 2020, the programme has attracted over 3,000 applications from over

2 For an overview of relevant policy initiatives, see: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

80 countries, with over 150 student placements engaging over 100 research institutions and businesses. Alongside these student placements, mobility funds have enabled 245 student exchanges, building links between labs across Europe. By training the next generation of AI researchers in Europe, and providing an entry route to careers in European AI research, this programme lays the foundations for Europe's future AI ecosystem.

Supported innovation and entrepreneurship to take AI from research to practice

ELISE promotes collaboration between industry and research institutes across Europe through its research programmes and use cases (Annex 1) and targeted SME engagement programmes. Launched in 2020, the ELISE SME support scheme offers financial and mentoring support to small businesses that want to implement novel machine learning solutions. This scheme has generated a high volume of interest. Its two open calls have attracted over 800 applications, from over 40 countries. Its first call funded 16 companies to develop new AI tools for varied applications, from civil engineering to machine translation to medical diagnostics.[3] Boxes 7 to 11 introduce some of these projects in more detail. The results of the second call will be announced in Summer 2023, and are expected to double the size of this cohort. Alongside these projects, collaborations with 30 incubators across ELISE partner organisations have also engaged businesses with support in deploying AI in their organisations.

Increased connectivity to build a European AI community

In addition to the 38 world-leading researchers driving ELISE's 14 research programmes, ELISE research programmes connect to over 260 ELLIS fellows working in 23 centres across 10 countries, and a wider network that spans 39 ELLIS Units across Europe. This network of leading machine learning researchers has established a European machine learning community that is driving the development and use of AI. Despite travel challenges during the COVID-19 pandemic, ELISE mobility funds have been instrumental in helping to build this community, facilitating research exchanges and placements across this network.

Catalysed new funding and initiatives to grow the European AI ecosystem

The success of these programmes has helped attract investment in European AI and sparked new collaborations. The newly established ELSA network will drive foundational research in safe and secure machine learning-based AI, building on relationships initiated via the ELLIS network.[4] High-profile examples of the additional funding attracted by these activities include a 100M Euro donation to expand the work of ELLIS.[5]

These foundations provide an opportunity to drive a wave of AI from Europe that aligns technological innovation with human values, that delivers economic and social benefits across Europe, and that enhances EU sovereignty.

3 For further information on each of these projects, see ELISE (2022) How have the 16 champions from ELISE's 1st open call been doing? Available at: www.elise-ai.eu/events/how-have-the-16-champions-from-elises-1st-open-call-been-doing and for insights into the industrial collaborations supported across the network, see ELISE (2023) ELISE Industry Stories, available at: www.elise-ai.eu/events/elise-industry-stories

4 For further information on the ELSA project, see: www.elsa-ai.eu

5 For example, ELLIS (2022) 6 Principal Investigators (m/f/d) as Hector Endowed ELLIS Fellows in Tübingen, available at: https://ellis.eu/PI2023

ELISE's achievements

Generated new research insights and applications

38
world-leading researchers

14
research programmes

Core technologies, such as Natural Language Processing, Deep Learning, Computer Vision, AI for multimodal data, robotics.

Trustworthy AI methods, including robustness, explainability, security, and human-centric AI.

Priority application domains, such as climate science and healthcare.

Created a world-leading infrastructure for advanced education

more than
3,000
PhD applications

245
student exchanges

over
150
student placements

engaging over
100
research institutions and businesses

Supported innovation and entrepreneurship to take AI from research to practice

over
800
applications from over 40 countries

11
use case company collaborations

32
SME collaborators

30
business incubators

Increased connectivity to build a European AI community

over
260
research fellows

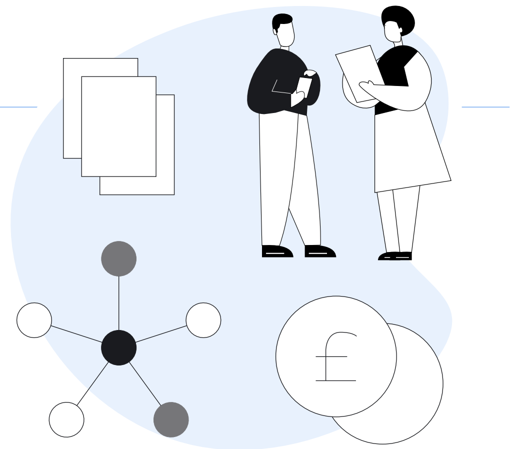
working in
23
centres

across
10
countries

39
ELLIS Units connected across Europe

Catalysed new funding and initiatives to grow the European AI ecosystem

€100m
donation received to expand the work of ELLIS



Box 1. Research highlights: Data science against COVID-19

Today's AI provides a flexible toolkit that can be deployed in response to pressing policy issues. ELISE researchers have driven effective deployment of these tools by combining technical expertise with collaborations that engage policymakers and citizens, delivering trustworthy AI decision-making tools.

Access to up-to-date data and insights about the dynamics of an infectious disease is critical in preventing its spread. By 2019, research using machine learning and data science to analyse outbreaks of the H1N1 virus in Mexico and Ebola in Central Africa had already demonstrated how new modelling tools could help analyse the trajectory of an outbreak and identify potential public health responses. In 2020, as COVID-19 rapidly spread across the world, the challenge for data science and machine learning researchers was to develop analytical tools that could be reliably deployed to help make sense of the pandemic, and to bridge the gap between research and policymaking.

Responding to this challenge, Nuria Oliver, co-Director of the ELLIS/ELISE Human-centric Machine Learning Programme, and her team at the ELLIS Unit Alicante Foundation created and led a coalition of scientists from different research institutions in the Valencian region of Spain with the goal of providing data-driven insights to support decision-making in the regional government. This interdisciplinary team – the Data Science against COVID-19 Taskforce[6] – combined expertise in machine learning, data science, engineering, statistics, epidemiology, and policy, delivering new research in four key areas:[7]

Modelling human mobility

Understanding how people are travelling and interacting is vital in predicting the spread of disease. To help extract insights from large datasets describing people's movements across the Valencian Region of Spain, the team developed analytical tools and accessible interfaces for data visualisation. These tools supported policy discussions about the impact of lockdowns or social containment measures on citizens' mobility, helping identify where such measures were more or less successful, and modelling the spread of COVID-19.

6 Oliver, N. (2022) Data Science against COVID-19: The Valencian Experience. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22).* Association for Computing Machinery, New York, NY, USA, 3–4. https://doi.org/10.1145/3503161.3549913

7 For further information, see: Oliver, N. (2022) Data Science for Social Good: the Valencian Example during the COVID-19 pandemic, *Esade Centre for Economic Policy*, available at: www.esade.edu/ecpol/en/publications/data-science-for-social-good-the-valencian-example-during-the-covid-19-pandemic; and Letouzé, E., Oliver, N., Bravo, M.A. and Shoup, N. (2020) Policy Paper: Using Data to Fight COVID-19 – And Build Back Better, Data Pop Alliance, available at: https://datapopalliance.org/publications/policy-paper-using-data-to-fight-covid-19-and-build-back-better

Building computational epidemiological models

Three different types of model – an SEIR model; an agent-based model; and a deep learning model – helped make predictions about how the pandemic would evolve, according to different scenarios for the use of non-pharmaceutical interventions, vaccination, or other policy measures and at different levels of spatial granularity. The resulting simulations allowed policymakers to explore the implications and trade-offs associated with different policy interventions.

Developing predictive models of healthcare systems

In addition to a suite of epidemiological models, the team also created predictive modelling tools that could translate insights about the pandemic to predictions about the number of cases, hospitalisations, and intensive care patients at a local or regional level. These insights allowed healthcare authorities to plan for waves of new COVID-19 cases, for example as the seasons changed or social confinement periods ended.

Creating a new citizen science initiative

Many aspects of people's behavioural or emotional responses to the pandemic were not well-represented in existing datasets. To help plug these gaps, the team launched a large-scale online survey that captured people's experiences.[8] The survey quickly received widespread attention, with over 140,000 responses in its first 40 hours after launch and over 700,000 answers today. This became an important tool to understand how patterns of social contact changed, the economic effects of policy changes, the prevalence of different symptoms, and more. Rapid analysis and visualisation of results has allowed researchers and policymakers to better understand the practical impact of policy measures.

Strong connections with regional and national governments helped translate these data-driven insights to policy development and implementation. An important focus of the team's work was interpreting, prioritising, and translating the insights from machine learning systems to actionable advice, aided by collaborations with policymakers in the design and use of these AI tools.

The resulting programme demonstrates the value of collaboration across scientific communities, publics, and public administration, leveraging data science and AI to deliver better-informed public policy responses. This work has delivered

8 To explore this work, visit: https://ellisalicante.org/en/covid19impactsurvey

cutting-edge science, demonstrated through a range of publications.[9] It has also directly influenced policy, for example shaping decisions about the deployment of non-pharmaceutical interventions in Valencia, providing an exemplar of data science-informed policymaking that has been recognised nationally and internationally.[10] The AI tools created by this team have also received international recognition: a deep learning tool to forecast infection numbers across the world and identify optimal strategies for containing the spread of disease won a $500,000 XPRIZE Pandemic Response Challenge.[11]

Methods and tools developed through this work also provide a proof-of-concept for AI in policymaking. The team's policy modelling, for example, analyses a range of different policy interventions and their impact on the spread of disease. By allowing policymakers to explore different options and trade-offs, this model connects insights from data to real-world decision-making, opening the door to more effective use of AI for public services in future. In developing these methods, it is vital to understand how to create AI systems that function effectively in deployment, accounting for the limitations of real-world data, designing for the needs of real-world users, and putting people at the centre of AI development. Building on the success of the Data Science against COVID-19 Taskforce, the ELLIS Unit Alicante Foundation continues to advance research that addresses these needs.

9 For example: Martínez-García, M., Sansano-Sansano, E., Castillo-Hornero, A., Femenia, R., Roomp, K., & Oliver, N. (2022).Social Isolation during the COVID-19 Pandemic in Spain: A Population Study. *Nature Scientific Reports*, 12(1), 1–15. https://doi.org/10.1038/s41598-022-16628-y; Lozano, M.A., Garibo, O., Piñol, E., Rebollo, M., Polotskaya, K., Garcia-March, M. A., Conejero, J. A., Escolano, F., & Oliver, N. (2022). Open Data Science to fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge. *International Joint Conference on Artificial Intelligence Organization (IJCAI)* 5304–5308. https://doi.org/10.24963/ijcai.2022/740; Oliver, N., Barber, X., Roomp, K., & Roomp, K. (2020). Assessing the Impact of the COVID-19 Pandemic in Spain: Large-Scale, Online, Self-Reported Population Survey. *Journal of Medical Internet Research*, 22(9), e21319. https://doi.org/10.2196/21319; Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N, Pivetta, F., Saramäki, J., Scarpino, S. V., Tizzoni, M., Verhulstand, S., & Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6(23), eabc0764. https://doi.org/10.1126/sciadv.abc0764

10 Wired (2021) How Valencia crushed Covid with AI, available at: www.wired.co.uk/article/valencia-ai-covid-data

11 ELLIS (2021) Dr. Nuria Oliver and her team take the grand prize in global pandemic response challenge competition, available at: https://ellis.eu/news/dr-nuria-oliver-and-her-team-take-the-grand-prize-in-global-pandemic-response-challenge-competition

Box 2. Research highlights: AI-assisted virtual laboratories

AI can be an engine for innovation, providing adaptable tools that accelerate progress across science and engineering. ELISE researchers are at the forefront of developing next-generation AI tools to support research and innovation, and applying those tools in critical domains.

Scientific progress is vital to develop the innovative solutions that policymakers need in response to today's grand challenges in health, climate, environment, and more. The deployment of AI in science has already demonstrated its potential as a tool for accelerating discovery and innovation. Successes in areas like protein folding,[12] climate modelling,[13] and astrophysics[14] signal the innovation that AI could help deliver, but are also the foothills of the larger transformation of science that AI could unlock.

The laboratory has long provided a physical hub for research. These physical spaces are increasingly digitised, integrating automated devices to support research. Computational tools are also becoming important enablers of scientific practices, allowing researchers to interrogate large datasets and extract novel insights. Extending these practices opens the possibility of new virtual labs, which leverage AI to accelerate discovery, by enhancing researchers' ability to progress their ideas. This vision of AI-enabled innovation through virtual laboratories is the focus of work by Samuel Kaski, ELISE Coordinator and co-Director of the ELISE Robust Machine Learning Programme, and his teams at the Finnish Center for Artificial Intelligence and the University of Manchester.[15]

Achieving this vision requires AI that can facilitate the scientific method. AI techniques already underpin analytical tools such as simulation, emulation, and digital twins that are embedded in today's scientific practices. More sophisticated AI methods could increase the power of these simulations, by combining different data sources, integrating pre-existing domain knowledge to increase the scientific relevance of results, and returning information to human users as actionable, explainable insights.[16]

12 Jumper, J., Evans, R., Pritzel, A. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

13 Requena-Mesa, C., Benson, V., Reichstein, M., Runge, J. and Denzler, J. (2021) EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task. Arxiv https://doi.org/10.48550/arXiv.2104.10066

14 Pultarova, T. (2021) AI discovers over 300 unknown exoplanets in Kepler telescope data, available at: www.space.com/artificial-intelligence-300-exoplanets-in-kepler-data

15 Klami, A., Damoulas, T., Engkvist, O., Rinke, P. and Kaski, S. (2022): Virtual Laboratories: Transforming research with AI. TechRxiv. https://doi.org/10.36227/techrxiv.20412540.v1

16 Berens, P., Cranmer, K., Lawrence, N.D., von Luxburg, U. and Montgomery, J. (2023) AI for science: an emerging agenda, available at: https://acceleratescience.github.io/assets/uploads/ai-for-science-an-emerging-agenda.pdf

Domains such as structural engineering, drug discovery, and graphic design have adopted AI methods to improve decision-making in design.[17] Further progress could extend such tools, creating AI assistants that function as a bridge between virtual analytical environments and the physical lab or research processes. These AI assistants would collaborate with human researchers, providing insights or recommendations that help them accelerate their research. They would combine the functionality of AI for science tools with a human-in-the-loop approach that helps researchers better leverage these analytical capabilities.[18]

To create these AI research assistants, AI agents need to be able to navigate the uncertainty associated with working alongside human researchers. Recognising that human researchers are often operating in contexts where desired outcomes, research questions, and current knowledge are in flux, virtual AI assistants need to be designed to be guided by user behaviours, learning from their users how to best help tackle a research question. They should be able to respond appropriately to problems they have never encountered before, identify relevant actions in situations where the desired end-results might be unspecified or unclear, and be able to generalise these behaviours from one type of research challenge to another.

Recent progress in AI research points to a pathway for developing these AI assistants. For example:

- Researchers and designers are often tasked with creating a solution to a challenge they have not encountered before; they typically know how to approach a task in principle, but lack a specific solution. Many of today's AI tools are developed through training on large datasets with many examples of how to solve a task, but advances in one- or zero-shot learning are creating AI agents that can learn how to perform a task based on only a few examples, or without having seen any data about the task. A prototype published in 2022 leveraged these methods to create an AI assistant that could support a human user in making a series of interrelated decisions, demonstrating the effectiveness of combining AI advice and human decision-making.[19]

- The goals or focus of research projects often change as new knowledge emerges, new problems arise, or scientific interests shift. To keep pace with these changes – without disrupting research activity by retraining an AI system – AI assistants need to be capable of identifying user goals even when they are not clearly specified, and adapting to those goals as they change. This implies a form of social intelligence through which AI agents can communicate with their users to better understand their intent and interests, which in turn invokes a type of inverse reinforcement learning through which AI agents observe a user's activities and use those observations to deduce what they are trying to achieve.[20] A new framework for generative user models creates a basis for such agents, providing a mechanism to model what goals the user might have from their behaviour and plan how best to assist them in response.[21]

- For many researchers or designers, human creativity and autonomy are important aspects of their work. These priorities can be in tension with approaches to AI development that focus on automation of (currently human-performed) tasks. To respond to this desire to have an active role in the design process, AI systems need to be able to work with human users as an active participant, keeping the designer in the loop and providing interfaces that empower the designer in decision-making.[22]

Building on the proof of concept demonstrated by the methodological advances described above, these ideas are being trialled in practice. 2022 saw the first pilot virtual laboratories established in drug design, sustainable mobility, and atmospheric science.[23] Further progress will benefit from broader purpose AI tools that can be developed in one domain and applied in others, alongside further demonstrator projects to show the value of these methods in practice.

17 De Peuter, S., Oulasvirta, A., and Kaski, S. (2023) Toward AI assistants that let designers design. *AI Magazine* 00: 1– 12. https://doi.org/10.1002/aaai.12077

18 Klami, A., Damoulas, T., Engkvist, O., Rinke, P. and Kaski, S. (2022) [n15]

19 De Peuter, S. and Kaski, S. (2022) Zero-shot assistance in sequential decision problems. Arxiv. https://doi.org/10.48550/arXiv.2202.07364

20 Gent, E. (2023) Understanding human intentions will be the next big breakthrough in AI, New Scientist, Available at: www.newscientist.com/article/mg25734260-500-understanding-human-intentions-will-be-the-next-big-breakthrough-in-ai

21 De Peuter, S., Oulasvirta, A., and Kaski, S. (2023) [n17]

22 De Peuter, S., Oulasvirta, A., and Kaski, S. (2023) [n17]

23 FCAI (2022) Solving the complexity of the process industry, available at: https://fcai.fi/news/2022/12/8/solving-the-complexity-of-the-process-industry

Box 3. Research highlights: AI for Earth and climate sciences

Methodological innovations, new applications, and community-building all play a role in developing more powerful AI systems. ELISE's Earth and climate sciences programme demonstrates how technical progress, interdisciplinary collaborations, and pilot projects are contributing to advancing AI R&D in an area of vital strategic importance.

The consequences of climate change can now be seen across Europe. Recent years have seen extreme weather events, including heat waves, droughts, floods, and wildfires, increase in frequency and intensity, affecting people, infrastructure, the environment, and the economy. As the need for accurate information about how the Earth's climate is changing, and how these changes will influence local areas, intensifies, there is growing focus on using AI to help tackle climate change.

The Earth's climate is a complex system, and a core challenge is to model this system in a way that allows researchers to identify changes, understand their causes, and predict their impact. Physics-based approaches construct models from first principles; known laws of physics that describe mechanistic relationships between different parts of the system. Despite much progress in these models, there continue to be uncertainties that limit how they can be used. Data-driven models offer a route to addressing these limitations. Large volumes of observational data available from different sources today – satellite imagery, for example – open the door to machine learning-led approaches to climate modelling. These approaches are the focus of ELISE's Machine Learning for Earth and Climate Sciences Programme, led by Gustau Camps-Valls (Universitat de València) and Markus Reichstein (MPI for Biogeochemistry).[24]

There are already examples of the power of these data-driven techniques in climate science. For example, data-driven quantification of the global carbon cycle has allowed researchers to track how carbon dioxide is taken up and released by different ecosystems, combining in-situ observations of carbon dioxide levels with satellite remote sensing and weather forecast data, producing estimates of carbon dioxide movement across the globe; effectively visualising how the Earth breathes.[25] These estimates can be used to benchmark climate models, helping improve the climate projections that form the basis of policymaking, such as the IPCC reports.

24 See ELLIS (2022) AI is mighty and efficient for this, available at https://ellis.eu/news/ai-is-mighty-and-efficient-for-this
25 For further information, see: https://bg.copernicus.org/articles/17/1343/2020

This work is one of a collection of projects supported by the ELISE Machine Learning for Earth and Climate Sciences Programme, which also includes:

- Understanding and Modelling the Earth System with Machine Learning (USMILE ERC), a project that aims to improve understanding of the feedback loops in the Earth system and use this understanding to create more accurate climate predictions.[26]
- DeepCube, which uses AI to address environmental challenges, focusing on the use of machine learning to examine the impact of drought on vegetation and the drivers of human migration.[27]
- The XAIDA project (eXtreme events: AI for Detection and Attribution) brings together European researchers to examine how climate change contributes to extreme weather events.[28]
- Deep Extremes advances Earth system science through the combination of Earth observation data and AI methods, with a particular focus on explainable AI and physics-informed AI.[29]

Across these projects, collaborative workshops have been influential in advancing scientific research, developing AI methods, and delivering new research tools. One example of such a collaboration is the EarthNet challenge,[30] which brings together machine learning researchers and climate scientists to improve forecasts of the local impacts of climate change. The ability to forecast how climate change will influence seasonal weather locally is important in enabling climate adaption, allowing policymakers to plan interventions to prevent food insecurity, increase infrastructure resilience, and prepare communities for extreme weather events. The challenge is how to apply AI to improve these forecasts. Recognising that the local effect of extreme weather or seasonal change can be seen in satellite imagery – changes in vegetation cover, for example – the EarthNet project seeks to leverage advances in computer vision to deliver improved seasonal weather forecasts.[31] Providing a rallying point for both machine learning and climate researchers, this initiative has already catalysed new research and applications of AI.

26 For further information, see: www.usmile-erc.eu
27 For further information, see: https://deepcube-h2020.eu
28 For further information, see: https://xaida.eu
29 For further information, see: https://eo4society.esa.int/projects/deep-extremes
30 For further information, see: www.earthnet.tech
31 For further information, see: www.earthnet.tech and Requena-Mesa, C., Benson, V., Reichstein, M., Runge, J. and Denzler, J. (2021) EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task. Arxiv. https://doi.org/10.48550/arXiv.2104.10066

A pilot study, for example, has explored how AI can predict changes to vegetation in Africa using satellite data, producing a model that can forecast seasonal changes and the impact of weather anomalies.[32]

These collaborations are also driving advances in AI methods. Causality, for example, has been a long-standing challenge in AI and in climate science. Identifying cause-effect relationships is important for researchers to understand the drivers of change in the climate system and for policymakers to identify appropriate interventions. In complex systems, such as the Earth's climate, which are characterised by dynamic interactions across spatial and temporal scales and emergent phenomena, it is challenging to discern which relationships are causal and which are co-occurring.[33] Differentiating causal connections from a variety of correlations that might exist within a dataset is also a challenge for AI developers. A variety of approaches to causal AI exist: hybrid modelling combines mechanistic models about how a system works with data-driven models, allowing researchers to connect data-derived insights to known laws or principles; explainable AI methods could also help researchers scrutinise how a system works; and the field of causal discovery from observational data and hypothesis is also providing new methods. One challenge that follows is how to benchmark these new methods, validating them against a causal ground truth. In response, the CauseMe platform offers ground truth benchmark datasets that can be used to compare the performance of different causal discovery methods by providing real-world data sets where the causal structure is known and synthetic datasets mimicking real-world challenges.[34]

Such efforts provide a rallying point for a growing community of researchers and practitioners in AI for Earth and climate, supporting a wider wave of action to deploy AI to tackle critical climate and environmental concerns.

32 Robin, C., Requena-Mesa, C., Benson, V., Alonso, L., Poehls, J., Carvalhais, N. and Reichstein, R. (2022) Learning to forecast vegetation greenness at fine resolution over Africa with ConvLSTMs. Arxiv. https://doi.org/10.48550/arXiv.2210.13648

33 Runge, J., Bathiany, S., Bollt, E. et al. (2019) Inferring causation from time series in Earth system sciences. *Nat Commun* 10, 2553. https://doi.org/10.1038/s41467-019-10105-3

34 For further information, see: https://causeme.uv.es

3. AI for European Grand Challenges

AI and the EU's Innovation Missions

As countries across Europe seek to recover from the impacts of COVID-19, policymakers face a constellation of challenges; the need to recover economic growth, to ease pressure on healthcare systems and improve wellbeing, to manage disruption to energy systems, to build resilience to environmental change, and more. With public services in many areas under increasing strain, there is a pressing need to translate policy aspirations for AI to action.

The EU's Innovation Missions offer a vehicle to understand the connections between areas of critical societal need and AI's technical capabilities, and to chart a path for AI that helps address these needs. These five Missions articulate policy priorities and set out a framework for research and innovation to deliver solutions to these challenges. By 2030, their ambition is to:

- Support at least 150 European regions and communities to become climate resilient (Climate Adaptation Mission);
- Improve the lives of more than 3 mission people through cancer prevention and treatment (Cancer Mission);
- Restore European oceans and waters (Oceans Mission);
- Create 100 climate-neutral and smart European cities (Cities Mission);
- Lead a transition towards healthy soils through 100 living labs and lighthouses (Soils Mission).[35]

In early 2023, ELISE convened a series of surveys, interviews and workshops with researchers and policymakers from across Europe, asking how AI could contribute to these Innovation Missions. These dialogues consulted over 150 participants, generating discussion and debate about the role of AI in delivering the Missions. Across the Missions, there are many examples of how AI can contribute to delivering policy goals, by facilitating climate adaptation, improving healthcare outcomes, enhancing marine management, informing urban planning and governance, and monitoring ecosystem health. Figure 2 highlights some of these application areas, which are explored in more detail in the following sections. Together, these diverse application areas showcase the adaptability of AI as a tool for monitoring complex systems, predicting their future state, and delivering data-informed insights to inform decision-making.

ADAPTATION TO CLIMATE CHANGE

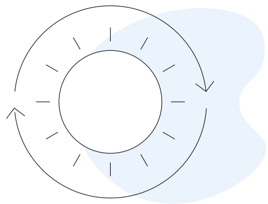
Even with rapid action to reduce greenhouse gas emissions enough to prevent warming beyond 1.5 degrees, new strategies are needed to manage the impact of climate change. The increasing frequency of extreme weather events across Europe, including floods, wildfires and droughts, demonstrates the urgent need to build resilience to the effects of climate change. This requires managing both the impact of these events and the associated social and environmental risks, including threats to biodiversity, human health, and infrastructure.

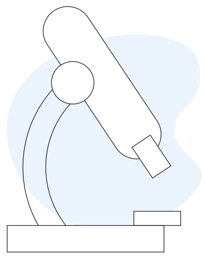
35 For an overview of these missions, see: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

Figure 2. **AI applications for the EU's Innovation Missions**

Climate adaptation

Improve climate models and simulations
Predict extreme weather events
Forecast the local impacts of climate change
Visualise climate risks
Design adaptation measures



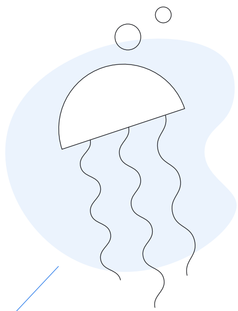


Cancer prevention, treatment and cure

Improve medical imaging for detection, diagnosis, and monitoring
Deliver precision oncology for diagnosis and treatment
Create clinical decision-support tools
Accelerate cancer research

Ocean restoration

Analyse marine ecosystems
Monitor marine wildlife
Track carbon flows
Support efforts to reduce plastic litter
Enhance supply chain sustainability

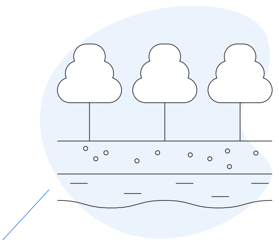


Smart cities

Inform city planning
Analyse building use and management
Optimise transport and service delivery
Support climate adaptation

Healthy soils

Increase understanding of soil ecosystems
Monitor soil health
Inform policy and land management interventions



Climate adaptation is the process by which technical, social, ecological, or economic systems make changes to increase their resilience to climate change. This process can involve a variety of interventions, including policy, infrastructure, environmental, and behavioural changes. With the aim of enabling climate adaptation across Europe, the goal of the EU's Climate Adaptation Mission is to "support at least 150 European regions and communities to become climate resilient by 2030".[36] To deliver this goal, the Mission aims to:

- Help regions and communities plan for climate adaptation, working with local administrations to better understand, prepare for, and manage climate risks.
- Accelerate progress towards climate resilience through collaborations that co-create, develop, and test innovative solutions to climate adaptation challenges.
- Drive systemic transformation that increases climate resilience across Europe, by supporting 75 large-scale demonstrator projects.

In delivering these objectives, AI can support climate adaptation efforts by:

- **Improving climate models and simulations:** Complementing existing, physics-based models of the Earth's climate system, AI can help improve climate predictions by enabling sophisticated simulations of climate sub-systems characterised by high uncertainty. Clouds, for example, affect the Earth's temperature in different ways: bright clouds block sunlight, helping to cool the Earth, while dark clouds have the opposite effect. Processes such as cloud convection can be too computationally costly for physics-based models, but are amenable to AI-enabled analysis.[37] The resulting hybrid models have been shown to increase the accuracy of forecasts, and are helping researchers develop better understanding of the Earth's climate system.
- **Predicting extreme weather events:** Extreme weather events in Europe have caused an estimated half a trillion euros in economic losses and between 85,000 and 145,000 human fatalities over the last 40 years.[38] Predicting when and where such events – such as storms, heatwaves, and flooding – are likely to occur is important in enabling rapid policy responses. While climate models analyse long-term trends and meteorological models provide near-term weather information, seasonal forecasting of the type that can help prepare for extreme weather remains challenging. AI can bridge between weather and climate data, helping increase the accuracy of predictions relating to the location and duration of high-impact weather events and increasing the robustness of local forecasts.[39]
- **Forecasting the impacts of climate change:** Effective climate adaptation requires both the ability to forecast extreme weather events, and the ability to predict the impact of extreme weather events in different areas.

36 European Commission (2021) EU Mission Adaptation to Climate Change Implementation Plan, available at: https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/implementation-plans-eu-missions_en

37 Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Slavin Ross, A., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J. and Bengio, Y. (2023) Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* 55, 2, Article 42. https://doi.org/10.1145/3485128

38 European Environment Agency (2022) Economic losses from weather and climate related extremes in Europe reached around half a trillion euros over past 40 years, available at: www.eea.europa.eu/highlights/economic-losses-from-weather-and

39 Rolnick, D., et al. (2023) [n37] and McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. *Bull. Amer. Meteor. Soc.*, 98, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1

At a local level, the climate, biosphere, land surface and geology, and human behaviours each influence how an extreme weather event will affect a community and its infrastructure.[40] These factors are difficult to model and are not easily captured by physics-based models. AI can support the analysis of these complex scenarios, using various data sources to help develop early-warning systems that highlight where an extreme weather event is likely to translate into dangerous climate impacts.

- **Communicating climate risks:** The pathway from Earth observation to policy decision involves complex interactions between data, analysis, and decision-maker. Through visualisations or simulations that help convey future scenarios, AI can help communicate the impact of climate change to decision-makers and inform decisions about which interventions to take.

- **Monitoring and analysis to help design adaptation measures:** The Climate Adaptation Mission suggests a collection of interventions to support adaptation, including risk assessments, early warning systems, infrastructure management, nature-based solutions, human health protections, and food security measures. AI can play a role in supporting the design of these interventions. Many nature-based solutions, for example, rely on protecting biodiversity to ensure healthy ecosystems; AI-enabled analysis can help monitor biodiversity from satellite imagery.[41]

CANCER PREVENTION, TREATMENT, AND CURE

Each year, approximately 2.7 million people across the EU are diagnosed with cancer, with approximately 1.3 million deaths. A variety of factors, including genetic, lifestyle, and social elements, influence disease development, and the overall prevalence of cancer is expected to increase in the coming decades. Without action to prevent the disease, the number of newly diagnosed people is expected to increase to over 3.2 million by 2040.[42]

The EU's Innovation Mission on Cancer focuses on "improving the lives of more than 3 million people by 2030" through cancer prevention and cure, while helping "those affected by cancer including their families to live longer and better".[43] It aims to:

- **Increase understanding of cancer:** Investigate where, how, and when cancer develops, identifying who is at higher risk of developing cancer and which treatments are more effective for different patients.

- **Enhance prevention and early detection:** Promote awareness of the risk factors that contribute to cancer development, increase public and scientific understanding of those risks, and boost the effectiveness of screening programmes through better access, new methods, and early predictors.

- **Improve diagnosis and treatment:** Increase the speed and effectiveness of diagnosis and treatment by sharing best practices, developing new methods, and delivering innovations in personalised medicine and clinical trials.

40 Bastos, A., Orth, R., Reichstein, M., Ciais, P., Viovy, N., Zaehle, S., Anthoni, P., Arneth, A., Gentine, P., Joetzjer, E., Lienert, S., Loughran, T., McGuire, P. C., O, S., Pongratz, J., and Sitch, S. (2021) Vulnerability of European ecosystems to two compound dry and hot summers in 2018 and 2019, *Earth Syst. Dynam.*, 12, 1015–1035, https://doi.org/10.5194/esd-12-1015-2021

41 Rolnick, D., et al. (2023) [n37]

42 See Mission overview at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/eu-mission-cancer_en

43 See Mission overview at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/eu-mission-cancer_en

- **Increase quality of life for patients and their families:** Integrate patient needs into treatment pathways, by analysing patient perspectives, enabling patient access to data, and building understanding of the impact of childhood cancers.

There are multiple points along the prevention, diagnosis, and treatment pathway where AI-enabled interventions could help deliver better health outcomes for cancer patients. For example:

- **Medical imaging for detection, diagnosis, and monitoring:** Analysis of images such as X-rays or MRIs is central to many cancer screening and diagnosis methods. Leveraging the ability of AI to extract and analyse a large number of features from an image, a variety of AI tools have been developed to help increase the speed of diagnosis and help human clinicians determine the presence or absence of disease. Areas of application include analysis of lung radiography,[44] CT scans,[45] brain MRI scans,[46] and mammograms (see Box 11),[47] amongst others.

- **Precision oncology for diagnosis and treatment:** Precision analysis of individual tumour cells can increase the effectiveness of cancer treatment, through targeted administration of oncology drugs (or other treatments) based on the genetic characteristics of a patient's tumour. AI can enhance this single-cell analysis, helping to profile the genetics of cancer cells[48] and to identify mutational signatures that help tailor treatment plans.[49] By analysing how the proteins in cancer cells interact with each other and with different drugs, AI can help predict which cancer drugs are more likely to be effective for individual patients, helping tackle treatment resistance.[50]

- **Clinical decision-support tools:** By integrating different data types and extracting insights that might not be visible to human decision-makers, AI can contribute to new decision-support tools that help clinicians interrogate patient data and identify appropriate interventions, for example combining data from diagnostics with information captured in clinical reports. The natural language interfaces offered

44 Schultheiss, M., Schober, S.A., Lodde, M. et al. (2020) A robust convolutional neural network for lung nodule detection in the presence of foreign bodies. *Sci Rep* 10, 12987. https://doi.org/10.1038/s41598-020-69789-z

45 El-Regaily, S.A., Salem, M.A.M., Aziz, M.H.A., and Roushdy, M.I (2020) Multi-view Convolutional Neural Network for lung nodule false positive reduction, *Expert Systems with Applications*, Volume 162, https://doi.org/10.1016/j.eswa.2019.113017

46 Zhang, M., Young, G.S., Chen, H., Li, J., Qin, L., McFaline-Figueroa, J.R., Reardon, D.A., Cao, X., Wu, X. and Xu, X. (2020), Deep-Learning Detection of Cancer Metastases to the Brain on MRI. *J Magn Reson Imaging*, 52: 1227–1236. https://doi.org/10.1002/jmri.27129

47 Schaffter T., Buist D.S.M., Lee C.I., et al. (2020) Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw Open*. 3(3):e200265. https://doi.org/10.1001/jamanetworkopen.2020.0265

48 Irmisch A., Bonilla X., Chevrier S., Lehmann K.V., Singer F., Toussaint N.C., Esposito C., Mena J., Milani E.S., Casanova R., Stekhoven D.J., Wegmann R., Jacob F., Sobottka B., Goetze S., Kuipers J., Sarabia Del Castillo J., Prummer M., Tuncel M.A., Menzel U., Jacobs A., Engler S., Sivapatham S., Frei A.L., Gut G., Ficek J., Miglino N.; Tumor Profiler Consortium; Aebersold R., Bacac M., Beerenwinkel N., Beisel C., Bodenmiller B., Dummer R., Heinzelmann-Schwarz V., Koelzer V.H., Manz M.G., Moch H., Pelkmans L., Snijder B., Theocharides A.P.A., Tolnay M., Wicki A., Wollscheid B., Rätsch G., Levesque M.P. (2021) The Tumor Profiler Study: integrated, multi-omic, functional tumor profiling for clinical decision support. *Cancer Cell*. https://doi.org/10.1016/j.ccell.2021.01.004

49 See Cancer Research UK (2022) How "the most advanced machine learning approach" is finding new cancer-causing mutational signatures, available at: https://news.cancerresearchuk.org/2022/10/12/how-the-most-advanced-machine-learning-approach-is-finding-new-cancer-causing-mutational-signatures

50 Coker, E.A., Stewart, A., Ozer, B., Minchom, A., Pickard, L., Ruddle, R., Carreira, S., Popat, S., O'Brien, M., Raynaud, F., de Bono, J., Al-Lazikani, B., Banerji, U. (2022) Individualized Prediction of Drug Response and Rational Combination Therapy in NSCLC Using Artificial Intelligence–Enabled Studies of Acute Phosphoproteomic Changes. *Mol Cancer Ther* 1 June 2022; 21 (6): 1020–1029. https://doi.org/10.1158/1535-7163.MCT-21-0442

by Large Language Models offer a new route to increasing the usability of such tools. AI can also aid the development of robotics technologies used to support surgical interventions.[51]

- **Cancer research:** Underpinning many of these applications is the use of AI in cancer research, advancing understandings of cancer development and progression through analysis of genetic, environmental, lifestyle, and other types of data, and building risk profiles that interrogate the probability of different patients developing different diseases.[52]

OCEAN AND WATER RESTORATION

Oceans cover more than 70% of the Earth's surface.[53] Ocean and aquatic ecosystems are vitally important to the healthy functioning of the Earth system, to national economic activity in many States, and to the wellbeing of local communities. They are also under unprecedented strain. Demands for marine resources, marine pollution, and climate change are affecting species, ecosystems, and services, with implications for the environment and human wellbeing.

In response, the goal of the Oceans Mission is to "restore our ocean and waters by 2030".[54] Connecting across policy agendas on oceans, biodiversity, climate, and pollution, the Mission's three core objectives[55] are to:

- "Protect and restore marine and freshwater ecosystems and biodiversity, in line with the EU Biodiversity Strategy 2030", with interventions to deliver basin-scale restoration projects in the Danube River basin and on the Atlantic and Arctic coast, create a network of protected areas, and develop a 'Blue Parks' platform that promotes innovative ecosystem-based management practices to boost ocean health.[56]
- "Prevent and eliminate pollution of our ocean, seas and waters, in line with the EU Action Plan Towards Zero Pollution for Air, Water and Soil", setting up a demonstrator project in the Mediterranean Sea that scales up effective interventions, and reducing plastic litter at sea by 50%, microplastic releases into the environment by 30%, and nutrient loses by 50% (including the use of chemical pesticides).
- "Make the sustainable blue economy carbon-neutral and circular, in line with the proposed European Climate Law and the holistic vision enshrined in the Sustainable Blue Economy Strategy", reducing greenhouse gas emissions from maritime economic activities, including shipping and aquaculture, and launching an emission-reduction demonstrator in the Baltic and North Sea that acts as a flagship project for sustainable use of maritime resources.

51 Iqbal, M.J., Javed, Z., Sadia, H. et al. (2021) Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int* 21, 270. https://doi.org/10.1186/s12935-021-01981-1

52 Farina, E., Nabhen, J. J., Dacoregio, M. I., Batalini, F., & Moraes, F. Y. (2022). An overview of artificial intelligence in oncology. *Future science OA*, 8(4), FSO787. https://doi.org/10.2144/fsoa-2021-0074 and Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific reports*, 6, 26094. https://doi.org/10.1038/srep26094

53 See work by the IPCC: www.ipcc.ch/srocc

54 See Mission overview at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters_en

55 Cited in the Mission Charter at https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/restore-our-ocean-and-waters_en

56 See, for example, the European Blue Parks initiative: https://maritime-spatial-planning.ec.europa.eu/fundings/european-blue-parks

AI offers a tool to accelerate marine conservation and restoration efforts by:

- **Analysing marine ecosystems:** By enabling data collection from new sources and enhancing data analysis, AI can help researchers better understand the structure and biodiversity of marine ecosystems. AI can support the deployment of autonomous submarine monitoring systems, helping to manoeuvre and target these systems to study areas of interest. It can also help characterise marine ecosystems, analysing data from sensors and images to determine their physical and chemical properties, and the presence or absence of different species, from identifying plankton[57] to tracking whales.[58] These analytical tools can contribute to more effective monitoring, modelling, and managing of marine resources.
- **Monitoring marine wildlife:** Building on these capabilities, AI can help monitor the movements of marine wildlife, tracking the population dynamics of species of interest. For example, in the Southeast Atlantic, Humpback and Southern Right whales are slowly recovering from whaling activities. To monitor this recovery, researchers can use data collected from existing monitoring activities – tracking devices, for example – alongside acoustic data that indicates the presence or absence of whales in an area. Dynamic marine biodiversity maps have also been created to track the movement of fish populations.[59]
- **Tracking carbon flows:** Leveraging the data collected through ocean monitoring activities, AI can help analyse the amount of carbon dioxide absorbed by the oceans and how this relates to carbon flows across the Earth's biospheres. By analysing stores of 'blue carbon', AI can identify pathways to enhancing the carbon storage capacity of the oceans.[60]
- **Supporting efforts to reduce plastic litter:** By analysing images of plastic waste and geospatial data about its location, AI can track how plastic is entering the oceans and the points of intervention that can support clean-up activities.[61]
- **Enhancing supply chain sustainability:** Increasing compliance with environmental and fisheries regulations is an important component of the legislative agendas associated with the Oceans Mission. AI can support compliance and enforcement activities through enhanced data analysis. By analysing the movement of ships using satellite data, for example, AI can help regulators track which vessels are operating in different areas using different fishing methods. This can contribute to increased transparency in seafood supply chains.[62]

CLIMATE NEUTRAL AND SMART CITIES

Cities are responsible for more than 70% of global carbon dioxide emissions.[63] Achieving the European Green Deal's ambitions of net zero

57 See, for example: https://sci.vision

58 British Antarctic Survey (2022) Using AI to track whales from space, available at: www.bas.ac.uk/media-post/using-ai-to-track-whales-from-space

59 Farrell, L. (2022) AI is benefitting our oceans: here's how, available at: https://revolutionized.com/ai-for-oceans

60 AltaSea (2019) Artificial intelligence in the ocean: what it is and how it facilitates ocean conservation, available at: https://altasea.org/artificial-intelligence-in-the-ocean-what-it-is-and-how-it-facilitates-ocean-conservation

61 ITU (2022) How AI is helping protect our ocean, available at: https://aiforgood.itu.int/how-ai-is-helping-protect-our-ocean and Microsoft (2022) The Ocean Clean Up, available at: www.microsoft.com/en-us/ai/ai-for-earth-the-ocean-cleanup

62 European Parliament (2022) AI and the fisheries sector, available at: www.europarl.europa.eu/RegData/etudes/ATAG/2022/699.644/IPOL_ATA(2022)699644_EN.pdf

63 See Mission overview, at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/climate-neutral-and-smart-cities_en

carbon emissions by 2050 requires rapid action to embed sustainability across European cities. However, this transition to climate neutrality will require a transformation in how these complex systems are designed and managed.

This transformation is the focus of the EU's Cities Innovation Mission, which aims to deliver 100 climate-neutral and smart European cities by 2030, and to leverage these cities as innovation hubs that lead the way for all European cities to become climate neutral by 2050.[64] In support of this ambition, the Mission is working to:[65]

- Create a programme of Climate City Contracts through which 100 selected cities can implement investment and management plans to deliver climate-neutral initiatives.
- Support research and innovation demonstrators that help scale successful solutions and share best practices.
- Build links and networks across existing city-relevant European policy initiatives, helping ensure the effectiveness of different programmes to promote climate neutrality.
- Connect city administrators and local stakeholders to EU-wide expertise in city sustainability, accelerating the spread of best practices and skills in implementing climate-neutral projects.
- Increase capacity for city administrations to deliver innovative sustainability solutions through effective governance.
- Establish monitoring and evaluation frameworks to track progress towards the goal of climate neutrality.
- Leverage investment from local, regional, and national authorities and private sector stakeholders to help scale the impact of the Mission.

Across these areas for action, the Cities Mission identifies a role for AI in delivering smart city solutions that reduce carbon emissions and increase environmental sustainability across sectors including transport, buildings, water and waste management, and air quality.[66] Examples of potential application areas include:

- **City planning:** Inhabitants live in cities for different reasons and have diverse – often competing – requirements, which include employment opportunities, access to transport, energy, and waste services, environmental quality, and social inclusion and community. These requirements shape city planning and development, as decision-makers try to balance the needs of different communities alongside sustainability goals that might include reducing air pollution, reducing traffic, increasing the capacity of energy networks, or other services. Digital twin projects can help explore different scenarios for city planning, by simulating outcomes under different types of policy intervention.[67]
- **Building management:** By monitoring activity in buildings – through data generated by smart meters and other sensors – AI-enabled building management systems can predict patterns of demand and optimise energy use in support of sustainability goals.[68] By providing information about the energy intensity of daily activities, AI could play a role in encouraging low-carbon behaviours by building users. By analysing building features and environmental conditions, AI systems also offer an opportunity to target retrofitting interventions to increase the energy efficiency of existing buildings.[69]
- **Transport and services:** City inhabitants rely on various utilities, including waste, water, and energy infrastructures, which are subject to increasing demand as populations increase and the effects of climate change are felt at a local level. AI can optimise the management of these services, increasing their efficiency and resilience.[70] In transport, for example, AI can help develop traffic management systems that adapt to changing traffic flows, reducing emissions by decreasing the number of vehicles in traffic.[71]
- **Climate adaptation:** In developing digital twins to support urban planning, AI can contribute enhanced weather forecasting tools (see Box 3) that support local administrations to adapt to the changing climate.

HEALTHY SOILS

Soils deliver ecosystem services that are vital for human wellbeing. They enable food production, support biodiversity, manage water flow, store carbon, and provide a foundation for the living landscapes that are an important part of human cultural heritage. Maintaining healthy soils requires careful environmental stewardship. This natural resource is under increasing pressure from a range of human activities, including pollution, intensive land management, and climate change. The result is degradation of their vital ecosystem services. Together, the cost of these environmental impacts in the EU is estimated at 50 billion Euros per year.

The EU's Healthy Soils Mission aims to "establish 100 living labs and lighthouses to lead the transition towards healthy soils by 2030".[72] Its 8 Mission objectives set out its ambitions to: "(1) reduce desertification (2) conserve soil organic carbon stocks (3) stop soil sealing and increase re-use of urban soils (4) reduce soil pollution and enhance restoration (5) prevent erosion (6) improve soil structure to enhance soil biodiversity (7) reduce the EU global footprint on soils (8) improve soil literacy in society".[73]

In support of these objectives, the Mission plans to increase understanding of soil stewardship through research into the processes that affect soil health, to support innovative approaches to increasing soil health, to create a monitoring system that allows tracking and evaluation of soil

64 See Mission overview, at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/climate-neutral-and-smart-cities_en

65 See Mission implementation plan, at: https://research-and-innovation.ec.europa.eu/system/files/2021-09/cities_mission_implementation_plan.pdf

66 See Mission implementation plan, at: https://research-and-innovation.ec.europa.eu/system/files/2021-09/cities_mission_implementation_plan.pdf

67 See, for example: Forbes (2022) How to fight climate change using AI, available at: www.forbes.com/sites/markminevich/2022/07/08/how-to-fight-climate-change-using-ai/?sh=3ed2c0512a83 and Connecting Cambridgeshire (2022) Digital twins to design a better city infrastructure, available at: www.connectingcambridgeshire.co.uk/smart/better-use-of-data/digital-twins

68 Rolnick, D. et al (2023) [n37]

69 Centre for Sustainable Energy (2022) What is retrofit? Available at: www.cse.org.uk/news/view/2687#:~:text=Retrofit%20is%20the%20latest%20and,fossil%20fuels%20with%20renewable%20energy

70 AI Magazine (2021) 10 ways AI can be used in Smart Cities, available at: https://aimagazine.com/top10/10-ways-ai-can-be-used-smart-cities

71 AI Magazine (2021) [n70]

72 See Mission overview, at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/soil-health-and-food_en

73 See Mission overview, at: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/soil-health-and-food_en

health across Europe, and to increase soil literacy and citizen engagement in this area.[74] AI can support activities across these areas for action, for example by:

- **Increasing understanding of soil ecosystems:** Soil has been studied for over 100 years, but while its physics and chemistry have been well-characterised, scientific understanding of its biology is still in flux. Soil biology is in the early stages of the sequencing revolution, with genetic studies offering new insights into how these complex ecosystems work, for example in terms of the relationships between soil, microbes and invertebrates, and the metabolic pathways that soils support. Marrying new sequencing data with existing understandings of soil physics and chemistry could unlock new insights into what contributes to healthy (or unhealthy) soils, and what happens when soil ecosystems are perturbed. By drawing connections between different data types – or previously unidentified links across large datasets – AI could support a systems-level understanding of soils.
- **Monitoring soil health:** A variety of sensors and monitoring systems already collect data about indicators of soil health, including soil quality, pest prevalence, and chemical or physical characteristics.[75] AI can enhance these monitoring efforts, helping to map soil characteristics, simulate the processes that shape ecosystem services, predict the state of soil in different areas and how this might change over time, and enable new monitoring strategies by using data from remote sensing and Earth observation.[76]
- **Planning policy and land management interventions:** Drawing from these increased analytical capabilities, AI could deliver insights that inform policy strategies to increase soil health and sustainability. For example, predictive models of water demand can help land managers monitor water resources and take action to prevent droughts;[77] chemical sensors that monitor weather and soil conditions can predict soil nitrogen levels and provide advice on fertiliser use to help avoid over-fertilisation;[78] AI-enabled image analysis can automate pest identification, supporting farmers to control their spread;[79] and automated systems can help minimise water waste from irrigation.[80]

Connecting AI to policy priorities

The potential of AI stems from its pervasiveness: a variety of AI tools now exist that can be adapted to help tackle areas of pressing need across different application domains (see Box 1, for

74 See Mission implementation plan, at: https://research-and-innovation.ec.europa.eu/system/files/2021-09/soil_mission_implementation_plan_final_for_publication.pdf

75 Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard,L. (2018) Soil quality – A critical review, *Soil Biology and Biochemistry*, Volume 120, 2018, 105–125, https://doi.org/10.1016/j.soilbio.2018.01.030

76 Examples of this work can be found from the work of the AI4soil initiative, with information available at: https://ai4soilhealth.eu

77 Dedezade, E. (2019) Combating drought with AI and the cloud, available at: https://news.microsoft.com/europe/features/combating-drought-with-ai-and-the-cloud

78 Open Access Government (2022) AI soil sensors could minimise fertiliser use in agriculture, available at: www.openaccessgovernment.org/soil-sensors/128613/#:~:text=AI%20benefits%20both%20the%20environment%20and%20farmers&text=The%20data%20is%20then%20used,farmers%20to%20not%20over%2Dfertilise

79 Loudjani, P., Devos, W., Baruth, B. and Lemoine, G. (2020) AI and EU agriculture, report from the Joint Research Centre, available at: https://marswiki.jrc.ec.europa.eu/wikicap/images/c/c8/JRC-Report_AIA_120221a.pdf

80 Varatharajalu, K. and Ramprabu, J. (2018) Wireless irrigation system via phone call and SMS, *International Journal of Engineering and Advanced Technology*, Vol 8, December 2018, ISSN: 2249-8958 (Online) www.ijeat.org/wp-content/uploads/papers/v8i2s/B10821282S18.pdf

example). The illustrations above of how AI can be deployed to help deliver the Innovation Missions are diverse in their purpose, focus, and the disciplinary expertise they engage. Together, they point to a pipeline of AI interventions for mission delivery. AI provides an adaptable tool to:

- Monitor and make sense of the complex systems that influence human wellbeing.
- Predict changes or events in those systems for which preparatory action is needed.
- Enhance human decision-making, leveraging insights from data to create strategies for action.

The demands that these applications of AI in priority policy areas make of AI technologies share similar characteristics. They require sophisticated technical methods that can analyse complex datasets to deliver actionable insights; they need AI methods that can be deployed into real-world systems; and they demand AI technologies that work effectively alongside people, aligning with the needs, interests, and concerns of human users.

ADVANCED ANALYTICAL TECHNIQUES

Each of the Mission domains represents a complex system. Climate adaptation and ecosystem management require an understanding of the climate system, and the interactions between physical, biological, and human processes that influence the behaviour of its sub-systems over space and time. Cities are built from complex interconnections between people, technology, and infrastructure. The emergence and progress of cancer in different patients are influenced by complex genetic, economic, lifestyle, and social factors. These systems are characterised by interacting physical, environmental, social, and technological influences across different scales, resulting in unpredictable or emergent properties, where the localised or individual impacts of policy interventions are hard to predict, and where the long-term impact of those interventions are difficult to anticipate.

In response, advanced analytical tools are needed that can work with real-world data, extracting insights from that data and combining those insights with existing domain knowledge, allowing users to interrogate the workings of complex systems. Each of these needs represents active areas of research.

Analyse real-world data

By combining different types of data from different sources, the hope is that AI can identify opportunities for intervention that are otherwise not visible to humans. In cancer prevention, for example, a combination of genetic data, lifestyle data from devices that monitor daily activities, and medical records could help identify signals that indicate emerging health issues. Combining and analysing these diverse data types presents several methodological challenges. Such real-world data is often not collected with a specific research question in mind; it is typically 'messy', with areas where data is missing; it might come in different forms, including text, images, or audio; and different data sources may be working to different data standards. To respond to these challenges, AI techniques that can work with dynamic, multimodal data are needed, alongside methods that allow AI systems to learn from few examples, or transfer learning from one domain to another.

Bridge between data-driven and domain knowledge

Existing domain knowledge – about the physical laws that underpin climate change, or the interactions between genes, for example – provides important insights that can enhance the analytical power of AI systems. In soil science, for example, by marrying data-driven insights from sequencing studies with existing knowledge of soil physics and chemistry, AI could help unlock new understandings of the drivers of soil health. This type of 'digital twin of the soil' could help researchers take a systems approach to the study of soil health, based on combining biological, chemical, and physical processes at an appropriate level of abstraction. In climate science, researchers can more effectively identify the cause-effect relationships that help explain climate phenomena by combining physics-based models of the Earth system with data-driven modelling.[81] Achieving these results requires AI techniques that can encode domain knowledge in system design, for example by specifying known scientific laws or principles, or by embedding known symmetries.[82] Advances in causal AI can also help link data-driven insights and domain knowledge, by identifying cause-effect relationships from data.

Interrogate workings of complex systems

Researchers and policymakers are often interested in the interaction between the physical, biological, and social elements of complex systems. Each of these elements brings different data types, underlying processes, and scales of operation. By meshing together models representing different sub-systems, AI offers a route to building enhanced simulations that researchers can use to better understand system dynamics, explore different scenarios for action, and make predictions about the impact of different perturbations. Building these simulations brings a collection of engineering and methodological challenges. Data pipelines that can smoothly integrate different data types are needed, enabling researchers to combine models working at different levels of granularity. Implementing such pipelines at scale requires progress in research software engineering to enable access to data held across multiple locations in different forms. In climate science, for example, to deliver accurate forecasts of climate change and its impacts, machine learning tools must be able to integrate real-time, multimodal data, across multiple scales of time and space, synthesising these inputs to identify local impacts of climate change. In marine science, digital twins of the ocean need to combine different physical, natural, and social elements, such as ocean circulation, fish movements, and economic activities. Delivering these sophisticated simulations requires advances in techniques for simulation and emulation, alongside the ability to encode existing domain knowledge.

DEPLOYABLE METHODS AND TOOLS

These advanced analytical tools need to be deployable into real-world environments, which are often dynamic or uncertain and where their interactions with human users influence their effectiveness. Despite much progress in the capabilities of AI, many attempts to deploy these technologies in complex, real-world systems fail. A recent review of the use of AI in COVID-19 diagnosis, for example, highlighted that many AI systems developed for use in pandemic response were unsuitable for deployment.[83] More robust AI methods are needed to tackle this 'deployability' issue; such methods should be safe and effective in real-world systems, which are typically characterised by changing conditions, unexpected events, and complex incentive structures. To effectively serve human users, AI techniques also need to function as decision-support tools, requiring communication and collaboration with human users.

Robust under dynamic conditions

Deployed environments, particularly in complex socio-technical systems, often bring challenges that were not anticipated when training AI systems for a particular task. These dynamics pose technical challenges: not only must AI technologies demonstrate the characteristics of trustworthy

81 Reichstein, M., Camps-Valls, G., Stevens, B. et al. (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

82 Berens, P., Cranmer, K., Lawrence, N.D., von Luxburg, U. and Montgomery, J. (2023) AI for science: an emerging agenda, available at: https://acceleratescience.github.io/assets/uploads/ai-for-science-an-emerging-agenda.pdf

83 Roberts, M., Driggs, D., Thorpe, M. et al. (2021) Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell* 3, 199–217. https://doi.org/10.1038/s42256-021-00307-0

AI, as articulated by policy frameworks,[84] but they must also be able to maintain these characteristics over time and in changing conditions. AI must be able to respond to circumstances not represented in training data, be resilient to adversarial challenge or security issues, and be maintainable in deployment. Provably robust AI methods are needed, supported by mechanisms for verification and validation, and certification procedures that reflect the demands of real-world systems. Technical methods to ensure performance is maintained over time are also needed, such as AutoAI or other active learning methods, to automate performance monitoring and management.

Effective as decision-support

To deploy AI effectively as a decision-support tool – for doctors designing treatment plans; land managers deciding how to use fertilisers; city administrators investing in climate adaptation; and elsewhere – interfaces to enable knowledge exchange between AI systems and human decision-maker are needed. AI agents need to be able to identify and respond to the needs of human users, providing insights tailored to their goals, while system users need to be able to explore different ways of achieving a goal. This type of collaboration requires new approaches to human-agent teaming, based on new learning strategies, ways of combining insights from humans and data, and effective user interfaces. Box 2 explores research directions to achieve these outcomes in the context of the use of AI for scientific discovery.

Explainable and interpretable in decision-making

Transparency and accountability are important contributors to the trustworthiness of AI in decision-making. While their delivery requires a collection of governance mechanisms, including organisational or procedural interventions, technical approaches also play a role. AI systems deployed in decision-making need to be explainable: their users need to understand how the system works in ways that resonate with their needs. This might include understanding how an output was produced, what influenced that output, or what alternatives might be possible.

AI THAT WORKS FOR PEOPLE

Who benefits from advances in AI, and who bears the risks associated with its use, is shaped by a range of digital divides. How to manage the development and deployment of AI without reinforcing patterns of exclusion – enabling all in society to benefit from its use – remains a major challenge for researchers, practitioners, and policymakers. The EU's AI policy frameworks highlight the importance of ensuring that AI reflects the rights and values set out in European law, through ethical principles to guide its development and use, alongside regulatory proposals to govern high-risk applications. Human-centric AI seeks to embed these ideas in research and development. Applications in the context of the Innovation Missions highlight the importance of AI that demonstrates trustworthiness, that is integrated effectively into decision-making, and that represents diverse perspectives and needs.

Trustworthy

The ethical principles and characteristics of trustworthiness in AI that have been set out in policy frameworks translate into a variety of design features for AI systems. These include human oversight, robustness and safety, privacy and security, explainability and transparency, fairness, and accountability. Each of these areas is the focus of active research, with the aim of developing AI tools that are trustworthy by design.

84 See work by the EU High Level Group (2019) Ethics guidelines for trustworthy AI, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Integrated into practice

Delivering real-world benefits, and managing the risks associated with mis-use, requires effective integration of AI into practice. This in turn requires an understanding of the deployed environment, the design of AI tools to work alongside users in that environment, and organisational interventions to manage its use. In cancer treatment, for example, the use of AI in clinical-decision making requires research practices that centre patient and clinical needs, technical systems that are trustworthy by design, user interfaces that allow doctors to interact with AI-enabled data analysis, and organisational processes that embed careful stewardship of data and AI across practices from hospital procurement to staff training.

Represent diverse perspectives

To ensure AI systems are effective, inclusive, and aligned with societal interests, effective engagement with affected communities is needed across the design and implementation pathway. In smart cities projects, for example, the process of creating new AI-enabled sustainability initiatives offers opportunities for public discussion about where, how, and why technology should be used in city governance. Participative design can help bring diverse stakeholder perspectives to the development of AI systems, giving people a voice in discussions about how data and AI are used to shape their services and environments. Individual initiatives have experimented with different approaches to research co-design. Building on existing examples of such practices, there is an opportunity for a new wave of responsible research and innovation that helps create AI that serves diverse communities and centres public interests.

An AI for Grand Challenges policy agenda

The technical agenda that emerges from this review of the application of AI to deliver the Innovation Missions highlights the importance of developing AI systems that are technically advanced, robust in deployment, and aligned with societal interests. Alongside this agenda for AI technology emerges a policy agenda to accelerate the use of AI in areas of critical societal need.

To support further progress in the deployment of AI to address the goals of the Innovation Missions, policy interventions are needed to:

Advance the technical foundations and deployability of AI in real-world systems

Progress in core underlying technologies continues to be of critical importance in driving forward the application of AI in areas of need. Further investment in AI research is needed to allow this progress to continue.

Accelerate end-to-end innovation through challenge-led programmes and pathfinder projects in areas of need

Pathfinder projects in priority areas can help build understanding of the pipelines from data collection to AI development, to deployment, to decision-making. By convening projects that close the loop from data collection to decision-maker – from field to Ministry – it is possible to identify and resolve the interface issues that arise when communicating across communities or between organisations.

Connect to local innovation ecosystems and citizen interests

To ensure that communities and organisations across Europe benefit from advances in AI, AI research and translation activities need to be deeply connected to local innovation ecosystems. Local connections between researchers, industry, and the public play an important role in bringing user perspectives into AI design, and in helping drive adoption to create social and economic benefits.

Build an infrastructure for AI R&D

- Access to data is central to the development of current AI technologies. Data needs to be available, shareable, usable, and governed to ensure their use adheres to regulatory standards and public expectations. Wider data sharing and re-use require standards and procedures that are agreed and adopted across different communities of use.
- Today's AI methods – in particular large foundation models – require access to powerful computer systems for training and experimentation, alongside software and hardware systems to enable their use. A pathway to delivering enhanced compute capability in Europe is considered in Box 4. Increased uptake of AI across disciplines can accelerate innovation, producing new solutions to the challenges set out in the Innovation Missions. Easily available toolkits that can be adopted 'off the shelf', alongside access to AI training and know-how, can help increase the use of AI across science.

Attract talent and build skills

To remain at the forefront of AI innovation, Europe needs to be able to attract leading research talent and build capacity across sectors and organisations in the use of AI.

Box 4. Research highlights: Building a compute infrastructure for European AI

Access to compute is vital to continuing progress in AI. Europe already has the building blocks for a compute infrastructure that can support the development of sophisticated AI systems. In the long-term, next-generation compute methods can deliver a step-change in the performance of this infrastructure.

Increasing compute power has been an important enabler of recent advances in AI technologies. Following a phenomenon known as Moore's Law, over the last six decades the number of transistors that can be built into a computer chip has roughly doubled every two years. Engineering innovations have also delivered chips specialised for use in developing AI systems. The resulting hardware and software applications support rapid information processing, allowing researchers to train sophisticated AI systems, including large foundation models – for example, Large Language Models such as ChatGPT – that enable AI inference by using standard language ("prompting") and that have generated renewed excitement about the potential of AI.

This current rate of hardware development is not sustainable. Engineering design is approaching the limits of circuit miniaturisation on traditional silicon chips. Today's chips have become so small that dissipating heat-cased errors significantly corrupt signals during operation. Quantum computing is currently facing similar challenges. Increasing energy demands arising from the use of large-scale compute have also highlighted the environmental impact of AI development, demonstrating the need for both energy-efficient AI development methods and sustainable energy sources for compute facilities. In response, a new approach to building compute infrastructure is needed. To chart a path towards this new infrastructure, ELISE researchers Petr Taborsky and Lars Kai Hansen at Technical University of Denmark (DTU) have created a strategic roadmap for compute in Europe.

Demonstrating how the EU can transform the compute landscape in the long-term while widening access to high-performance computing today, this roadmap highlights three important technical directions.

- **HPC virtualization:** In the near future, a new generation of middleware, based on improved interconnections between CPUs, GPUs, and nodes, is set to emerge. This, together with task-specific accelerators, such as Transformer engines,[85] will enable HPC virtualization, offering user- and task-specific HPC environments, which are more powerful and more energy efficient than today's systems. Virtualization, instead of relying on a single computer to perform a computation task, relies on distributed computing and effectively leverages multiple software and hardware – for example, memory – components across multiple systems to complete a shared task. This federated architecture may also be required from a privacy perspective, if the datasets being processed are multiple and localised, and cannot be copied or moved to a central location.

- **Quantum computing:** Quantum computing develops computing methods based on quantum principles. The resulting systems should be able to solve problems that no classical computer could solve in a feasible amount of time. While there has been exciting recent progress in this field, whether or not so-called quantum supremacy – the creation of a quantum device that can perform calculations that could not be performed by a classical computer – has been achieved is the subject of active debate. Despite a few successful real-world quantum applications,[86] further work is needed to take these prototype systems to commercial scalability.[87]

- **Neuromorphic computing:** Taking inspiration from the highly energy-efficient information processing of the human brain, neuromorphic computing seeks to develop orders of magnitude more energy-efficient computing systems. Neuromorphic computing methods and processors are in development; while at present there are few commercial offerings, there are a variety of applications where neuromorphic computing could play a role.[88]

85 For example, Transformer engines from NVIDIA.

86 The quantum encryption link has been developed under the FIRE-Q project – supported by Innovation Fund Denmark – and comprises academic and industrial partners that are now ready to commercialize the technologies. The technology is the result of 20 years of basic research supported by the Danish National Research Foundation through the research centres Silicon Photonics for Optical Communication (SPOC) and Hybrid Quantum Networks (Hy-Q).

87 One example of an EU initiative in quantum computing is the EuroHPC JU project High Performance Computer – Quantum Simulator (HPCQS): www.hpcqs.eu

88 One example of an EU initiative in neuromorphic computing is EBRAINS: http://ebrains.eu

While research in these areas continues and moves towards commercialisation, the ELISE compute roadmap identifies steps that can be taken today to maximise the effectiveness of Europe's existing High-Performance Computing infrastructure. Created in 2018, the European High Performance Computing Joint Undertaking (EuroHPC JU) provides a framework for EU countries to coordinate and pool their compute resources, creating a federated infrastructure that can support EU leadership in high-performance computing.[89] This network already provides access to eight supercomputers located across Europe, alongside a further seven systems managed by the PRACE network.[90] As a result, researchers can access compute resources more powerful than those typically available at universities, enabling delivery of robust experimental results in an environment that adheres to regulatory requirements regarding data governance and security.

By reducing the procedural complexity of accessing this network, facilitating the discussion between AI community and EuroHPC (workshops) in support of virtualizing the HPC user environments, and explaining how to use the systems to wide AI user bases, including the governance requirements, the roadmap sets out a route to expanded, powerful, while energy conscious, access to this compute resource across Europe.

This snapshot summarises the findings of ELISE deliverable D-4.6 Report on recommendations for infrastructure roadmap, by Petr Taborsky and Lars Kai Hansen.

89 For further information, see: https://eurohpc-ju.europa.eu/index_en

90 For further information, see: https://prace-ri.eu

4. An evolving research-policy agenda

ELISE's Strategic Research Agenda and trends in AI

ELISE's 2021 Strategic Research Agenda set out the research challenges that need to be addressed to strengthen the technical capabilities of AI; improve its performance in deployment; and align AI development with societal interests. This Agenda sought to bridge between the frontiers of technology development and the EU's AI policy agendas, recognising that the success of those policy agendas would depend on Europe's ability to pursue excellent research that both advances foundational AI technologies and applies those technologies to areas of critical social and scientific need.

The research themes and areas of research interest explored in 2021's Strategic Research Agenda are considered in the following sections. As technical capabilities continue to shift, the intention here is not to provide a comprehensive review of the state of the art under each theme, but to convey a sense of key issues and how the field has progressed in recent years. To illustrate how the research topics explored relate to practical applications, a selection of use cases introduces how ELISE industrial collaborators have deployed machine learning to enhance their work.

Over the two years since the publication of the initial ELISE Strategic Research Agenda, research, practice, and policy have changed at pace. Headline-grabbing advances in technologies such as Large Language Models have re-ignited debates about the opportunities and risks associated with AI, highlighting the potential for rapid advances in technical capabilities. Understandings of how to deploy both known and state-of-the-art technologies have continued to evolve. Legislative and regulatory proposals have grappled with the challenge of stewarding a technology that is dynamic, pervasive across sectors, and associated with both beneficial and harmful uses. From this shifting landscape emerge ten trends in technology and regulation that are shaping the development of AI today (Box 5).

Box 5. Trends in AI[91]

Today's trends in AI suggest that continued technological advances will bring discontinuities in AI capabilities and disruptions for society and the economy; policy will play an important role in shaping the impact of these changes.

AI is achieving better-than-human levels of performance for certain very specific, or closed, tasks; broader application of these successful tools in real-world environments requires progress in robustness.

The benchmark of achieving better-than-human performance has been a catalyst for public conversations about the potential of AI technologies. There are already some tasks where AI's performance against human benchmarks has been well-documented: in object detection from static images, for example, AI tools have performed at higher-than-human levels since 2017.[92] In other areas, how AI compares to current human practices is less clear. In image analysis systems for cancer diagnosis, for example, the performance of AI in deployment in comparison to human radiologists has been mixed. Performance for restricted tasks continues to improve, and the range of tasks for which AI could be used continues to grow.[93] However, these restricted tasks are not often representative of real-world challenges; even small deviations from these constrained environments might derail performance. Adding snow to landscape images, for example, drastically decreases the performance of current systems for image recognition. However, large, foundation models are opening the possibility of creating broader-purpose AI tools that can be trained on one task and applied for another. One approach to increasing AI performance on a range of tasks is transfer learning – the ability to take learning from one task and apply it to another – which, despite being close to human standards, remains challenging. The creation of AI that delivers high performance in dynamic environments will require progress in robustness and deployability, creating tools that are suitable for broader applications.

New learning strategies mean AI can use less data and function effectively; advancing methods to elicit and integrate knowledge from domain experts can deliver further progress in 'low-data' AI.

Rapid progress in AI has been facilitated by the availability of well-curated data that can be used to train machine learning models. While highly effective, these data-intensive methods rely on access to such data, or synthetic equivalents, which may be difficult or undesirable in some domains, for example where privacy

91 This box summarises trends reported in more detail in a forthcoming ELISE White Paper: Smeulders, AWM., Montgomery, J. (2023) A digital society with AI built with trust: technology with regulation and regulation with technology.

92 Stanford University (2021) AI Index, available at: https://aiindex.stanford.edu/report

93 Freeman, K., Geppert, J., Stinton, C., Todkill, D., Johnson, S., Clarke, A., et al. (2021) Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy *BMJ*; https://doi.org/10.1136/bmj.n1872

is an important concern. To overcome this limitation, researchers have developed a suite of approaches that are allowing AI to learn from less data. These include training methods that bootstrap from similar datasets and learning strategies such as zero- or one-shot learning that generalise from a small number of data points. Strategies to introduce structure or domain knowledge into an AI system can help drive further progress, through encoding domain knowledge in the form of laws or principles, developing causal AI methods, or eliciting knowledge from expert users. The resulting systems would combine the ability to derive insights from data with pre-existing knowledge about the system and interactions with its users. These low-data approaches bring with them their own tensions: prior knowledge added in the place of data can translate into prejudice, while prior assumptions added to the model can translate into losing precision for the problem at hand.

Large models are delivering impressive results and will continue to improve; the next wave of progress in foundation models will come from combining different types of foundation models.

Impressive outputs from Large Language Models, which can generate convincingly human text in response to questions from users, have sparked new conversations about progress in AI and its implications for society. Progress in this area has been driven by the creation of very large models. Large foundation models are also being developed for other core AI functions, such as computer vision[94] and image generation.[95] These models potentially unlock broader applications of AI tools. However, they also face a variety of deployability issues that are shared across AI technologies. Building on the successes of these models, further progress will come from integrating different types of foundation models, enabling broader problem-solving. Large models intrinsically suffer from limits in transparency. When larger datasets are used to create a system, it is almost impossible to manage or even assess bias, for example. The associated risks are increasingly clear as progress in these models continues.[96]

Approaches to performance evaluation are shifting, with increasing attention to robustness, explainability, and fairness; better integration of these approaches can connect performance to real-world needs.

While accuracy benchmarks have been helpful in driving progress in some areas of AI technology, these benchmarks are not typically representative of the demands AI systems face when deployed in real-world contexts. The risk that follows is that

94 Such as the CLIP model ft30 Florence model from Microsoft.

95 Such as DALL.E.

96 Bender, E.M., Gebru, T., McMillan-Major, A. and Mitchell, M. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

achieving a benchmark creates a misleading impression of performance or progress in the field.[97] In response, AI research has been generating new approaches to benchmarking, for example measuring performance under changing conditions.[98] Further research in this area can help develop benchmarks that better reflect the characteristics needed to deliver trustworthy AI in practice, while also helping raise awareness amongst developers and users on the system side about the limits of methods for robustness, explainability, and fairness. In the long term, systems should be able to assess their own limits, indicating when they are not able to give a reliable answer; such systems will require significant progress to deliver.

Software and hardware tools are helping make AI easier to deploy; low-energy approaches can connect the benefits of these tools to the desire to achieve sustainable AI.

The availability of software tools that align with modern hardware designs has been an important enabler of AI deployment, reducing the time taken for software development and allowing easier access to GPUs. Open-source development has been central to the success of this approach, with software published under flexible BSD-licenses. Maintaining this culture of open-source software development and code sharing will be vital for the continuing success of the field. In the context of a trend for larger models, effective management of access to compute will become increasingly important, balancing the demand for increased compute with the need to ensure energy efficiency in AI. To chart a path towards delivering both these outcomes, ELISE has created a roadmap for compute infrastructure in Europe (see Box 4).

Increasing attention on explainable AI has driven a proliferation of methods; the gap between these methods and implementation needs to be bridged.

European policymakers have identified transparency as a core component of trustworthy AI. One technical response to this policy demand is the development of explainable AI; AI tools whose workings or outputs can be understood and scrutinised by human users. Growing interest across the AI community in the challenge of how to create explainable AI methods has resulted in a proliferation of approaches to explainable AI. While technically feasible, the extent to which many of these methods address the needs of real-world users is not clear. Further progress to deliver trustworthy AI in practice will require action to bridge this gap between technical capabilities and the needs of different communities affected by AI. Countervailing trends in the development of large models that are less interpretable and a lack of methods for assessing performance in practice are likely to influence progress in addressing these challenges.

97 Raji, I.D., Bender, E.M., Paullada, A., Denton, E. and Hanna, A. (2021) AI and the everything in the whole wide world benchmark. Arxiv, https://doi.org/10.48550/arXiv.2111.15366

98 See, for example, benchmarking work by: https://shift-happens-benchmark.github.io

The sophistication of the fake text, images, audio, and video that AI can generate is increasing; what follows is a demand for AI-enabled tools to detect these fakes.

While manipulation of images or text is not a new phenomenon in the information environment, AI has increased the sophistication and availability of tools that can generate convincing fake images, audio, text, or video. The resulting capabilities increase the challenges policymakers face in building a trustworthy online information environment, with implications for science and society. In response, technical, regulatory, and societal interventions are needed to allow better detection of fake content, prevent harm arising from such content, and support users to interrogate the trustworthiness of different information sources. AI must play a role in supporting these interventions, for example through services to help detect online fakes. The scale of distribution and ease of creation of fake media requires automated tools in response, akin to the domain of cybersecurity where only automated intelligent responses are rapid enough to counter automated security threats.

Integration of AI into the public sphere continues to bring disruptions, demonstrating the importance of securing the policy foundations for trustworthy AI and data governance.

Combined with the growth of online platforms, shifting patterns of information consumption, and new uses of personal data, AI has contributed to changes in how people interact online. The consequences of the enhanced personal tailoring – or targeting – that follows have already been visible in debates about ethical data use, and its consequences for public and political debate. Such debate is shaped by a variety of social and digital divides, which create power asymmetries in the digital landscape. These asymmetries affect whose interests are reflected in the development of AI, to whom the benefits and risks of deployment accrue, and what levers are available to direct technology development towards wider societal benefit. This trend highlights the importance of policy foundations that ensure trustworthy data governance and responsible AI research and innovation.

Despite progress in data efficient AI and privacy-enhancing technologies, the accumulation of personal data continues.

While progress in data-efficient AI offers a route to developing AI tools without relying on access to large volumes of training data, an increasing amount of data continues to be generated and captured from everyday activities. Removing names or obvious personal details is no longer an effective strategy for privacy preservation, in the context of AI algorithms that can draw connections between personal details available online or embedded in data.[99] A variety of policy questions follow, from how to disrupt the market dominance of those companies with pre-existing access to data resources, to how to ensure the right to privacy of individuals amidst changing patterns of data use.

Market forces continue to favour monopolisation, repeating a historical trend for general-purpose technologies; disrupting these is a long-term policy project.

Monopolisation or centralisation has been a long-standing concern in AI policy. Access to data confers a first-mover advantage on a small number of companies; the resources required to train large AI models give those same companies a competitive advantage in a field increasingly interested in foundation models; and access to knowledge and skills is unevenly distributed. History suggests that disrupting the monopolies surrounding general-purpose technologies requires policy intervention, and that such intervention typically takes decades to deliver results. Investments in European AI to grow the AI research community and connect that community to local innovation ecosystems – sharing the benefits of AI across the continent – can help disrupt this influence. Regulatory interventions to disrupt the power asymmetries that shape AI development also play a role.

The research pursued across the ELISE network is at the forefront of these trends. ELISE's 14 Research Programmes continue to advance the capabilities of AI technologies – across theory, methods, and application – in turn, creating new questions and research challenges. Progress can be seen in the shifting contours of research and practice across the five cross-cutting themes set out in the first ELISE Strategic Research Agenda.

Progress in ELISE's cross-cutting themes

TRUSTWORTHINESS AND CERTIFICATION

The development of trustworthy AI that can be safely and effectively deployed to deliver real-world benefits continues to be a cornerstone of AI research and AI policy. Delivering the EU's seven characteristics of trustworthy AI – human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability – requires theoretical, methodological, and practical advances in AI technologies. Current policy debates highlight the importance of bolstering these technical characteristics with standards, certifications, and guarantees that ensure trustworthiness is not only a principle for development, but is also delivered in practice. Both the technical capabilities to underpin trustworthy AI and the certifications that can prove whether an AI system is trustworthy have been areas of focus for ELISE research (see Box 6).

A variety of methods to deliver on each of the seven characteristics of trustworthy AI now exist, including methods for working

99 Kosinski, M., Stillwell, D., Graepel, T. (2013) Private traits and attributes are predictable from digital records of human behavior, *PNAS*, 110 (15) 5802–5805, https://doi.org/10.1073/pnas.1218772110

with real-world data, explainability, robustness, privacy preservation, and bias-reduction. AI systems can perform well where they are trained on large datasets, where there is a defined task, and where there is a clear understanding of how to validate performance at that task. There remains, however, a gap between these controlled conditions and the real-world environments in which many AI systems are developed and deployed. The size of the AI models underpinning Large Language Models and their equivalents poses a further challenge to implementation of methods for trustworthiness by design. Bridging this gap requires progress in AI robustness, and new understandings of verification and validation. This in turn requires close collaboration between technologists, practitioners, and policymakers to identify appropriate benchmarks, certifications, and validation procedures that connect technical capabilities to effective risk evaluation in deployment.

Table 1. **Areas of research interest in trustworthiness and certification**

Theme	Areas of Research Interest	Approach
Trustworthiness and certification	▪ Certifying or guaranteeing the performance of AI systems; ▪ Verifying and validating machine learning; ▪ Improving the robustness of machine learning in deployment.	▪ Advance the technical sophistication of core machine learning methods, including deep learning, computer vision, natural language understanding and generation, and semantic, symbolic, and interpretable machine learning. ▪ Improve understanding of the principles and techniques that can make machine learning robust, from theory to their application in practice. ▪ Create robotic systems that can interact intelligently with the world around them by combining robot learning approaches with machine learning methods, such as reinforcement learning. ▪ Explore the role of causal modelling in bridging the gap between observational and interventional learning and understand the principles underlying interactive learning systems. ▪ Design models that respond appropriately to situations that were not well-represented in their training data by accurately identifying instances of 'domain shift' and advance the use of transfer learning or AutoML techniques to address such scenarios.

Box 6. Research highlights: Certification of AI systems

The ability to assess the risks associated with AI systems will be central to the implementation of the EU's AI Act. Understanding how to certify AI, and the mechanisms that can deliver trustworthy certification, is essential.

The EU's AI Act proposes a suite of legislative interventions that aim to ensure citizens can have confidence in AI-enabled products and services, by minimising the risks associated with AI deployment.[100] An important pillar of this approach is the development of certification mechanisms to demonstrate that an AI system functions as expected. Certifications provide guidelines for those developing AI to help ensure the resulting products are reliable; they provide information for consumers about product performance; and they can support regulatory assurance processes that guarantee only safe and effective products are brought to market. While AI development does have established best practices, to date it has lacked clear standards and guidelines to form the basis of certification mechanisms. Responding to this policy need, researchers at TÜV Austria Group and Johannes Kepler University Linz, supported by the ELISE network, have created a framework for a new AI certification mechanism.[101]

Certification is a process through which an independent body signifies that a product or service has been tested against objective performance standards, and has been found to meet those standards. The development of AI certification mechanisms is challenging for a variety of reasons. Creating and adhering to certification processes requires:[102]

- a theoretical understanding of core technical concepts;
- standardised quality assessment processes and ways of clarifying requirements around correct usage of AI technologies;
- the ability to account for domain gaps between training and 'real-world' data;
- the ability to accommodate rapid developments in the capabilities of AI technologies;
- mechanisms to translate ethical considerations into technical review;

100 See overview of the digital policy agenda, at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

101 See ELISE (2022) First commercial Certificate for AI systems now on the market, available at: www.elise-ai.eu/events/first-commercial-certificate-for-ai-systems-now-on-the-market and Winter, P.M., Eder, S., Weissenböck, J., Schwald, C., Doms, T., Vogt, T., Hochreiter, S. and Nessler, B. (2021) Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications. Arxiv, https://doi.org/10.48550/arXiv.2103.16910

102 Winter, P.M., Eder, S., Weissenböck, J., Schwald, C., Doms, T., Vogt, T., Hochreiter, S. and Nessler, B. (2021) [n101]

- access to qualified personnel with the skills and knowledge to evaluate AI applications;
- the ability to evaluate robustness in deployment, including to adversarial attacks.

Responding to these challenges, the certification framework proposed in *Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications* sets out a process by which an independent third party could test the quality and safe use of AI applications.

In this certification process, the organisation being certified receives a requirements catalogue, which sets out what capabilities or guarantees an AI application should demonstrate. This requirements catalogue can be used by AI developers to analyse the performance of their system, identify gaps or deficiencies, and take action to address any safety or security concerns. It can be used to help define the scope of certification sought by developers and application owners. This documentation is submitted to an auditor, and used as a basis for interviews with the developer team and inspections of their work. The resulting audit report reviews the performance of the system, making use of an audit catalogue that specifies the different system characteristics that contribute to its safety and efficacy in deployment.

The audit catalogue proposed by the TÜV Austria Group and Johannes Kepler University Linz team is structured into three main themes:

- security in software development, which considers the core software capabilities needed to ensure a system is safe and secure in deployment;
- functional requirements, which reviews topics relating to model development, including data collection, model selection, and other methodological considerations; and
- ethics and data protection, which considers issues associated with the use of personal data and wider societal interests such as fairness and privacy.

Underpinning these themes are detailed specifications relating to over 200 system requirements.

The first iteration of this framework, published in 2021, focuses on the certification of supervised machine learning methods in low-risk applications. Building on its success, the audit catalogue is being extended and refined, with the anticipation

that this work will become an important contributor to European efforts to create a risk-proportionate regulatory system for AI applications.

This snapshot summarises the findings of work by Philip Matthias Winter, Sebastian Eder, Johannes Weissenböck, Christoph Schwald, Thomas Doms, Tom Vogt, Sepp Hochreiter and Bernhard Nessler.

Box 7. Use case: Automating power grid inspection with AI and drones to prevent wildfires

Machine learning and computer vision methods are being deployed by industry to help manage critical infrastructure. AI-enabled monitoring and maintenance systems rely on robust AI methods that can be deployed in real-world environments.

Sparked when a powerline came into contact with a tree, the second-largest wildfire in California's history scorched almost one million acres and destroyed a small town in 2021.[103] As extreme weather events become more frequent in nearly 75% of the world, including Europe,[104] the regular inspection of power lines will become increasingly important to prevent devastating wildfires.

Currently, some five million kilometres of electricity gridlines across the European Union are inspected by manned helicopters, creating pollution and risks for humans, as well as a hefty bill for power companies of up to €150/km. To enable more frequent, safe, and cost-effective inspections, FuVeX aims to replace helicopters with specially-adapted long-range drones. The start-up's automated solution harnesses AI to enable drones to collect the data energy companies need to ensure gridlines are in good order.

Inspector gadget

"The main issue inspecting power lines is capturing high-resolution data, as the camera has to point to the power line towers" explains Carlos Matilla, FuVeX CEO.

103 CNN (2022) California's second-largest wildfire was sparked when power lines came in contact with a tree, Cal Fire says, available at: https://edition.cnn.com/2022/01/05/us/dixie-fire-power-lines-cause-pge/index.html

104 Madge, G. (2023) Met Office Droughts developing more rapidly says global study, available at: www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2023/droughts-developing-more-rapidly-says-global-study

With ELISE funding, the company has developed a prototype drone with cameras and a computer on-board that can automatically point a gyro-stabilised camera at a power line to capture vital visual data needed for inspections, such as clear images of insulators and joints.

The system can do this by combining computer vision and machine learning. It uses neural networks to first automatically detect and locate power line towers in real-time using a low-resolution camera connected to the computer. Using this information, an algorithm calculates the position of the tower, which enables the gimbal with the high-resolution, colour camera on it to point in the right direction and take photographs.

FuVeX first validated the performance of its machine learning in a lab by simulating a drone flight. The team projected power line images in front of its drone to analyse its reaction, and to ensure the system could point its main camera at lines. More tests followed including flights beside real powerlines and now the company is honing its technology with a powerful partner.

A powerful alliance

“The main challenge to the application of these technologies is the vast amounts of data needed to train the AI models,” Matilla says. Luckily, FuVeX has access to Naturgy’s data and power lines. It is working with Spain’s third largest energy company to test its technology and is the first company to receive authorisation to perform beyond visual line of sight (long–range) flights over power lines in the country.[105]

FuVeX inspected 1,000km of power lines with Naturgy last year and aims to survey the rest this year, totalling 12,000 kilometres. Matilla says Naturgy is already spending a third less on surveying its power lines thanks to FuVeX’s innovation. “In the next two years, we are working towards achieving savings of 50 to 75%. And that’s just the beginning” he adds. The start-up believes that when fully automated, its system could save companies 94% compared with using manned helicopters.

While the company is focused on working with its main customer, the team knows that building a larger data set will be challenging because there are various types of powerlines operated by different companies. Training the system to operate alongside these different types of infrastructure is an important part of ensuring its effectiveness in deployment. However, it is confident it will be able to collect the data needed and hone its drone further while testing its reliability. “Robustness is a huge challenge we’re facing right now,” says Matilla. “The technology has to be completely reliable.” This requires testing the system on different types of infrastructure, as well as embedding safety mechanisms to signal when the system is at risk

105 FuVex (2020) Naturgy,Vodafone and the start-up FuVeX employ the help of cellular-connected drones for power line inspections, available at: www.fuvex.es/en/naturgyvodafone-and-the-start-up-fuvex-employ-the-help-of-cellular-connected-drones-for-power-line-inspections

of failure. “Modern society relies on being able to flip a switch to power its homes and businesses. But this simple action requires millions of kilometres of powerlines to function well, so FuVeX’s system has to guarantee it can always capture accurate data to ensure power lines are well maintained,” Matilla explains.

In the future, FuVeX aims to expand within Europe. The company could grow fast, with utility companies eager to transition to autonomous power line inspections, which could become the norm in as little as five years, according to Matilla.

A hot topic

Climate change could also accelerate adoption, as fears of larger and even more frequent wildfires abound. “Having a reliable power network that doesn’t generate wildfires is extremely important,” he says.

Matilla believes regular inspections are a key prevention strategy. He estimates that power grids inspected by manned helicopters currently only take place every one to three years. Consequently, utility companies are spending time and money mending damage that could have been prevented and there is a risk of fires starting. “Our drone technology could keep flying in continuously over power lines ensuring their stability for the same or much less cost,” he says. The company is even working on a new type of aircraft that could make surveying power lines as cheap as €10/km.

“The business case to make inspections completely autonomous is there,” says Matilla. There is also a moral case if autonomous monitoring of power lines could save habitats and lives that could otherwise be lost to wildfires. “Crewed helicopters are not enough to inspect powerlines to make sure they do not cause fires. We want to change the paradigm of inspections so that power lines, which are critical for society’s transition to green electricity, are reliable and safe.” This makes autonomous monitoring one hot topic.

SECURITY AND PRIVACY

AI has been both a catalyst for fresh concerns about security and privacy and a source of new tools to tackle these concerns. ELISE’s 2021 Strategic Research Agenda highlighted the importance of further progress in the development of principled methods for achieving privacy and security by design, and of integrating user needs and societal expectations around the safety and reliability of AI systems into the research and practice that underpins their deployment.

Research progress has yielded a collection of methods for enhancing the security of AI systems and ensuring they respect fundamental rights in relation to privacy. These include privacy-enhancing technologies, such as differential privacy and federated learning, and advanced methods for attack detection and mitigation, such as adversarial learning.

With new threat vectors emerging and amidst continuing concerns about data privacy – for example, as users submit new data types to Large

Language Models – the challenge for the next wave of AI progress is to make these state-of-the-art methods workable in practice and to establish best practices in data stewardship. This requires further work to test that the assumptions on which theoretical developments are based are valid in practice; to resolve design trade-offs in design that affect overall system performance, such as the ability to deploy advanced methods in differential privacy and cryptography in systems that are computationally efficient; and to better understand what type of security threat is likely to emerge from advances in AI. Across these areas, continued dialogue is needed to understand society's expectations in relation to data use, privacy, and security.

Table 2. **Areas of research interest in security and privacy**

Theme	Areas of Research Interest	Approach
Security and privacy	▪ Privacy and security by design; ▪ Working with practitioners to translate state-of-the-art technology to practice; ▪ Better understand future threats.	▪ Advance the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems. ▪ Design novel machine learning methods, including methods for differential privacy and adversarial machine learning, to help manage concerns about security and privacy by design, and resolve trade-offs in their implementation. ▪ Develop principled methods that demonstrate machine learning systems are robust in deployment (where distributions may shift), and that are robust to adverse circumstances and/or adversarial manipulations, making use of software verification, machine learning verification methods, and causal modelling to help secure these advances. ▪ Work with practitioner communities – for example in healthcare – to help develop machine learning systems that manage concerns about security and privacy in practice.

Box 8. Use case: Using AI to automate cyber security

New AI applications are an essential tool to combat novel, AI-enabled security threats.

The view that a major cyberattack poses a threat to financial stability is accepted. It's not a question of if such an attack will arise, but when, according to the IMF.[106] Meanwhile the global cost of cybercrime was estimated at some $8.4 trillion[107] last year and is predicted to increase.

Keeping track of threats is not simple. "When you work in cybersecurity, you are bombarded with threat intelligence; information about risks and attacks in a semi-structured text format," explains Sven Niedner, Founder and CEO of Synamic Technologies.

On the attack

The goal of Synamic's cyber security project, supported by ELISE, is the automation of cyber security processes. This requires reliable information that computers can read. To deliver this material, their SCR.AI system turns full text incident and threat reports into machine-readable data sets, by using – and extending the capabilities of – a cyber security knowledge graph the company had already developed.

The automation process begins with a web crawler, which finds sources of threat intelligence such as cyber incident reports. When a report is found, it is analysed using transformer models and Natural Language Processing (NLP) to extract useful knowledge by focusing on specific phrases containing malicious code. A multi-label classifier then identifies how that attack model is designed to wreak havoc on victim's computers, for example detecting the technique and attack framework.

"Cyber-attacks are a complex chain of events and prevention is only possible with full picture detection," Sven Niedner, founder of Synamic Technologies says.[108] To make this possible, a knowledge graph is used to enable the threat information to be organised in the proper sequence, forming an 'attack graph', which Niedner likens to a fingerprint that describes how attackers will reach their goals. The edges and nodes of a graph are datapoints following a clearly defined ontology. "You basically get a machine-readable, automatically plausible description of a cyber-attack in question," he explains. To ensure attack graphs can be shared and integrated

106 IMF (2021) The Global Cyber Threat, available at: www.imf.org/external/pubs/ft/fandd/2021/03/global-cyber-threat-to-financial-systems-maurer.htm; The Federal Reserve (2022) Feds notes: Implications of Cyber Risk for Financial Stability, available at: www.federalreserve.gov/econres/notes/feds-notes/implications-of-cyber-risk-for-financial-stability-20220512.html

107 Petrosyan, A. (2022) Estimated cost of cybercrime globally 2016-2027, available at: www.statista.com/statistics/1280009/cost-cybercrime-worldwide

108 SCR.AI Synamic Technologies, available at: www.youtube.com/watch?v=Yi7I0Ej8_H0

with other tools, data is expressed using industry standards, such as MITRE ATT&CK and STIX2.

Troubleshooting with ELISE
Of course, developing such a complex tool is challenging. "You reach a point where you need additional experts to discuss problems and people with expertise in both cyber security and AI are hard to find," Niedner says. ELISE gave Synamic Technologies' data scientist access to a fellow expert who could exchange ideas about the use of transformer models and discuss data quality, specifically the point at which the company should stop training its AI model. This is because if you over-train a model it no longer works as well, so finding the sweet spot where data quality is the best it can be is difficult. "It was very helpful to have this assistance and inspiration," Niedner says.

Ensuring SCR.AI can detect the very latest threats is another challenge. It requires that the company continuously tests its tool's ability on a training data set to make sure its AI is robust and threats are being detected. "The bad guys always come up with new strategies, so we have to think of a way to retrain and keep our models updated," Niedner explains.

Explainability and deployability
To translate these technical processes to actionable insights, Synamic Technologies designed a simple user interface, with the front end featuring a visualization of the attack graph. This allows users to gain an intuitive understanding of complex processes underpinning the different attack steps. By its visual nature, the attack graph offers a degree of explainability. The overall design of the interface is intended to reflect user needs, because most users are cyber security professionals: "They are all programmers themselves and they prefer the tool to come with a more programmer like interface," Niedner explains.

The SCR.AI tool has various applications in the cyber security industry. "It can be used to create an early warning system for upcoming threats and also to automate the processing of threat intelligence," he says. It is currently being used by a customer in the financial services industry to simplify the process of analysing intelligence to prevent cyber security attacks.

Building a secure future
The start-up is working on a proof of concept to help it grow in the financial sector and is in discussions with partners in a bid to expand into the Internet of Things (IoT) domain. This is because as devices get smarter, the components used in them are vulnerable to cyber-attacks. Connected, critical infrastructure is a prime target for criminals and so moving into the IoT area is important for Synamic Technologies.

Niedner considers that the applications for AI in cyber security are only at their beginning, with "huge potential" for deep learning models to improve security; the ultimate aim in AI-powered cyber security is automated defence. This is a logical progression from his company's attack graph. While SCR.AI can automatically detect attacks by recognising anomalies and patterns to warn cyber security professionals of an attack, he says "the next step is for a system to take preventative measures autonomously, without human intervention".

While this could prove invaluable to financial services companies and operators of critical infrastructure, it requires people to have confidence that the AI system will act safely and effectively in their best interests. Furthermore, while 'good' AI could one day automatically protect businesses from the malicious threats, it could also be harnessed by criminals to inflict more harm by staging attacks. This means the cat-and-mouse nature of cyber security will endure.

EXPLAINABILITY, ACCOUNTABILITY, AND DECISION-MAKING

Concerns about AI as a 'black box' in decision-making – with technical and non-technical communities alike unable to understand its workings – have received widespread attention. In the research and policy debates that follow, a collection of ideas about the features needed from AI to ensure its trustworthiness collide. The term explainable AI is used variably to signal:

- whether the use of data or the workings of a model are transparent or interpretable by either system designers or users;
- whether it is possible to explain how or why a model or AI system has produced a particular output;
- whether an output – or an AI system itself – is reasonable or justifiable as part of a decision-making process;
- whether AI is being used in ways that facilitate or diminish accountability for the results of a decision-making process.

Attempts to respond to these questions through innovations in technology and governance have resulted in explainability becoming a popular research topic. Many methods for explainable AI have been developed, in particular methods for increasing model interpretability and for creating counterfactual explanations. These innovations are valuable. However, they become more difficult to deploy in complex systems, and they do not necessarily align with user expectations around explainability, or how humans might approach the task of data interpretation. In response, further technical developments are needed to ground these methods to theory – both in terms of modelling and human understanding – and to connect these methods to user needs. Aligned progress in fields such as causal AI could also help drive progress, by allowing users to interrogate the relationships or causal factors that result in a particular output or decision. The result should be explainable AI methods that function well at scale; that are human-centric, based on an understanding of what sort of explanation is needed for whom in what contexts; and that deliver interpretability-by-design, by aligning user perspectives with technical functionality.

Table 3. **Areas of research interest in explainability, transparency, and decision-making**

Theme	Areas of Research Interest	Approach
Explainability, transparency, and decision-making	▪ Advancing explainable AI tools and methods; ▪ Grounding new methods to theory and practice; ▪ Supporting practitioners to implement explainable AI methods that meet stakeholder needs.	▪ Develop inherently (or 'by design') explainable machine learning methods, including deep learning, and approaches that increase the explainability of machine learning systems, through advances in surrogate modelling methods, visualisation tools, and approaches to encoding existing knowledge. ▪ Combine symbolic and data-driven AI methods to develop AI systems that are inherently explainable. ▪ Foster collaborations at the interface of machine learning and human-computer interaction to understand how human and algorithmic decision-making interact. ▪ Engage with policymakers and legal specialists to explore how machine learning system design can ensure that AI use aligns with the rule of law. ▪ Engage with practitioners, users, and other affected communities to translate new methods to beneficial real-world practice.

Box 9. Use case: knowledge graphs for industrial applications of machine learning

AI is a vital enabler of the digitalisation of industrial processes; many industrial applications require AI systems that can be analysed or scrutinised by human users. Progress in explainable AI is supporting the development of safe and effective AI methods for these real-world challenges.

AI could be the trigger for a wave of innovation across industry, but explainability is one barrier to deep learning techniques revolutionising industrial practices. To be successfully deployed in industrial contexts, some element of interpretability or explainability of AI systems is often necessary. Volker Tresp, Ludwig Maximilian University of Munich professor and Siemens Distinguished Research Scientist, notes that while interpretability is often desired, "it is difficult to get explainability in a meaningful way" to meet people's differing needs and expectations.

However, he says that knowledge graphs "are by nature interpretable," because they present contextual information that directly relates to human-understandable concepts or ways of representing information. In this way, they help make AI more

transparent. A way of presenting relationships between interconnected objects, the graphs extract data from different sources and then identify interrelationships within the data. Machine learning can be employed to infer additional relationships and extract the meaning behind those relationships. The result is a tool that combines data analysis with contextual information, enabling knowledge graphs to be applied across industries and unlock new business benefits in the process.

Wider use of knowledge graphs is part of Siemens' digitalisation strategy, which seeks to drive progress towards intelligent engineering and manufacturing, and they are of personal interest to Dr Tresp, who leads a research team working on machine learning approaches that operate at the human abstraction level, where the world is described by entities, concepts, and their mutual relationships. "We cover the machine learning aspects of knowledge graphs and connect them to text and visual data," he says.

Siemens AG is interested in industrial applications for knowledge graphs, which range from turbine predictive maintenance to smart expert recommendation tools. Research activities focus on integrating the solutions across the large and complex businesses of Siemens, which operates in the fields of automation, electrification, and digitisation.

Building knowledge

One of Siemens' success stories in deploying knowledge graphs is its Totally Integrated Automation (TIA) selection tool, which allows project planners to rapidly set up new digitisation projects by performing hardware selection, planning and configuration tasks within one tool, without a manual or any detailed portfolio knowledge. The result is a streamlined, automated process "for error-free configuration and ordering". The company says it is one of the most prominent examples of tools for supporting engineers who are configuring an industrial automation solution, which can easily consist of hundreds of components with thousands of configurable parameters.

Planning and engineering an automated process or service is a challenging task that requires time, experience and a lot of specific knowledge. The tool uses knowledge graph technology to support engineers in selecting the right components to solve their problems. "It's a bit like a recommendation system, but it uses more technical background about the entities that can be bought," Dr Tresp explains.

Knowledge graphs are helpful for users as they can integrate and evaluate in fractions of a second more knowledge than a human brain can process, enabling rapid data analysis, but are designed to assist us in making better decisions. For this purpose, the tool uses data built from an extensive repository of anonymised automation solutions, and a product knowledge graph that incorporates information about product types, variants and technical features. This enables it to compare components on the technical level and recommends the best-fitting component to an engineer. Siemens AG says the tool aids knowledge transfer between experienced and less experienced engineers while its 60,000 active users save time when completing the configuration process.

The team behind the tool have extended the idea of a knowledge graph further by allowing it to accommodate changes happening over time. Its 'temporal knowledge graph' leveraged more contextual information hidden in the data, including temporal dependencies between the actions performed by the user, which improved the quality of recommendations.

Robustness and explainability
Recommendation tools do not have to be as robust as plant management tools, for example, but users still want their tools to be reliable and effective. The TIA selection tool is part of Siemens' end-to-end approach of Totally Integrated Automation that is designed to deliver maximum consistency and transparency and the company says the tool "makes error-free configuration possible".

Dr Tresp believes language models such as ChatGPT could be useful in adding a level of explainability to tools used by businesses. For example, language models could explain how AI arrived at a decision or recommendation in a way that is understandable to users.

However, there are risks with relying on AI-generated explanations. ChatGPT has already been shown to give plausible, but incorrect, information. Dr Tresp says there is a serious effort underway to make it more robust and accurate, he says, as well as find useful new applications and "we are not in a completely different situation" with ChatGPT and other language models. "One challenge is how to get the methods we're developing in our programme to be effective in the context of Large Language Models and foundation models because there are new challenges and novel use cases," Dr Tresp says.

Looking into the future
Siemens AG believes smart recommendations are going to be an essential feature of all future engineering tools, helping users to navigate the ever-more-complex engineering landscape. Dr Tresp and Siemens are passionate about the importance and value of world-class research, including knowledge graphs, but they also appreciate that potentially profitable applications for new AI technologies are crucial. The business case for AI can be difficult, but needs to be established, Dr Tresp says. As well as testing and integrating new AI technologies, his team is also focused on using them to develop new applications. "It's difficult to have business models with AI," he explains. "If you're solving one problem for one customer, that's great, but a company like Siemens needs to have a multiplication factor so that the same solution can serve several customers."

AI INTEGRATION

Recognising that the difficulties of integrating innovative technologies within existing systems can cause promising AI tools to fail in practice, the field of MLOps focuses on the action required to deploy and maintain machine learning models in production. Effective deployment requires careful consideration of technical, organisational, and regulatory aspects of AI integration, and best practices are emerging across each of these areas. As technologies like Large Language Models offer to increase the accessibility of AI tools to organisations across sectors, the need for effective integration of AI into existing systems – taking into account concerns about safety, explainability, security, and other aspects of trustworthy AI – will become more pressing.

Research in robustness has created methods and approaches that seek to increase the safety and reliability of AI in deployment, creating methods that are better able to respond to real-world challenges. These challenges often fall outside the scope of the data on which a model was trained, and require the ability to adapt to different domains, detect and respond to dataset shift, or transfer learning between domains. A collection of approaches to increasing robustness have been developed, which now need translating into practice. A continuing challenge is how to scale these methods across complex AI systems, comprised of interacting components that mix automated and human elements, taking into account human needs and maintaining overall system functioning. AutoAI offers a route to enhancing the safe and effective deployment of such AI systems through AI-enabled performance monitoring and management.

Alongside these technical considerations, understandings of the organisational and regulatory elements of AI integration are also evolving. Organisations deploying AI need to consider the skills needed by those working alongside AI systems, the experiences of users affected by those systems, and the regulatory requirements associated with AI in their sector. With the field developing fast, dynamic ways of responding to issues in deployment are needed, allowing organisations to trial new technologies without causing harm. Proposals for controlled testing environments such as regulatory sandboxes have already been created as part of current European policy developments and offer a way of stewarding new AI tools into deployment.

Table 4. **Areas of research interest in AI integration**

Theme	Areas of Research Interest	Approach
AI integration	▪ Improving performance in deployment, through more robust AI methods and AI-enhanced monitoring; ▪ Designing effective simulators and emulators; ▪ Understanding interactions with human users; ▪ Connecting research and practice; ▪ Combining data-driven and structural insights; ▪ Increasing data availability through technical and governance interventions.	▪ Design simulators and emulators that can help explore the consequences of different interventions or model designs, and that can extract insights from the analysis of complex systems, such as those found in Earth sciences. ▪ Develop new learning strategies to operate in low data-resource environments, advancing research in areas such as one- or few-shot learning (the ability to learn from a small number of data points or examples); transfer learning (using knowledge learned from one task as the basis for performing another); interactive learning (designing agents that learn through their interactions with their environment); reinforcement learning; and the study of the intelligence of living systems (for example, of the role social reasoning plays in influencing decision-making). ▪ Integrate emerging methods for ensuring the robustness of machine learning systems into real-world use cases. ▪ Advance methods for embedding knowledge about the physical world in the design of machine learning systems. ▪ Develop strategies for testing methods in practice or sandboxing new approaches.

Box 10. Use case: Saidot – Helping start-ups master AI governance

Delivering trustworthy AI requires responsible research and deployment practices, which include both technical and organisational interventions. Building understanding of the risks and ethical concerns associated with AI – and the organisational capability to respond to those concerns – is vital.

As businesses harness the power of artificial intelligence, the race is on to effectively regulate it. The EU's AI Act is building a legal framework to regulate high-risk use cases, with the aim of ensuring that AI is developed and deployed safely. Translating these legal frameworks to implementation will require clear obligations for

developers. However, regulations may seem intimidating and time-consuming for companies, particularly start-ups and small and medium-sized enterprises (SMEs).[109]

To address this challenge, Saidot helps businesses adopt AI governance and transparency best practices for trust and compliance, by helping them apply a systematic AI governance process across any AI tools they are building or use.

A tool for start-ups

The company has developed an AI governance model that is built to consider the ethical challenges from the perspective of AI SMEs. "The spectrum of risks is really broad and context-dependent," says Meeri Haataja, CEO and Co-Founder of Saidot. They include privacy concerns and built-in biases, which could be particularly impactful if connected to healthcare, for example. "We need to recognise that the impacts of AI technologies go beyond their intended purposes," she says.

Saidot's B2B software as a service (SaaS) platform is designed to simplify AI governance workflows and empower companies to create and operate safe, equal and transparent AI systems. An important part of this is helping product teams themselves to govern their systems by legal requirements while providing quick access to expertise whenever needed. "A lot of AI teams in previously non-regulated industries are facing strict new requirements. We need to simplify compliance to make this work" Haataja says.

The platform includes a guided AI ethics self-assessment. "It guides product teams through all ethical and legal requirements that they need to comply with when developing or using AI products in a given context, and helps involve stakeholders in the process," Haataja explains.

Typically, users register on the platform when they begin developing an AI product or plan to purchase a third-party product that uses AI to check how the technology and their use of it can comply with standards and regulations. The tool can also be used to share this information with external stakeholders. "Sharing transparent information on how their technology works and what risks are involved is a key part of responsible AI for AI SMEs" says Haataja. "It's already appearing in ESG reports and will only grow as regulations come into force."

Companies must also keep their records up to date throughout the product lifecycle as their system and use of it evolves. Saidot wants to automate this by monitoring systems and their risk environment, and nudging teams when updates or reviews are needed. Saidot does not audit data entered into its platform but helps get their systems audit ready.

109 See European Commission (2022) Regulatory framework proposal on artificial intelligence, available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Building the business
Saidot joined the ELISE programme to help European SMEs cope with AI-related ethical questions. Through Open Calls, Saidot's ethics self-assessment process has guided more than 100 AI products already. "ELISE has been an incredible opportunity to work with a lot of different kinds of AI product companies from different sectors, helping us dive into the specific problems and risks they face," Haataja says.

"Working with teams developing promising AI tools for mental health and breast cancer detection, for example, we recognise the importance of supporting SMEs on their responsible AI journey to ensure that the full potential of AI for healthcare is realised," she adds.

The company's main platform for systematic AI governance, transparency and communication is used by some 450 customers including the City of Amsterdam and the Scottish Government, along with hundreds of AI-focused SMEs.[110]

Intelligent growth
The need to heed AI regulations is only set to increase as more governments, organisations and traditional businesses embed AI into their processes. Saidot is currently focusing on helping companies and governments transition to meet the EU's GDPR, Digital Services Act and AI Act and expects demand for its platform to continue growing as more regulations are introduced worldwide, and more companies adopt third-party AI products.

Getting ready for AI regulations will concern every company that uses AI, particularly the ones in high-risk domains that have additional regulatory obligations such as financial services, healthcare, education or recruiting, Haataja warns. "Everyone needs to know what kind of AI technology they are purchasing or using and ensure they can buy and use it responsibly. In the ongoing boom of generative AI applications, we've all learned that a lot of responsibility lies in the hands of a user."

EMBEDDING AI ETHICS

AI is a disruptive technology, bringing with it both opportunities to enhance societal wellbeing and potential harms. As interest in the field of AI ethics continues to increase, progress in the technical capabilities of AI systems has created AI tools that can help reduce the risks of those harms emerging. The field of human-centric AI has driven research advances across a variety of areas of societal concern, including algorithmic fairness, explainability and transparency, privacy preservation, and the detection of AI-generated 'deepfakes'.

The wider challenge associated with AI ethics is how to connect technology development

110 See www.saidot.ai

to societal interests, leveraging AI's capabilities to help tackle areas of critical need while embedding human interests and concerns in its development. With technology moving at pace, new applications emerging across sectors, and understanding of the risks and benefits of technology use in flux, continued dialogue between research and practice is needed to influence technology development towards beneficial outcomes.

Table 5. **Areas of research interest to embed AI ethics**

Theme	Areas of Research Interest	Approach
AI ethics	▪ Advancing foundational research to create human-centric AI; ▪ Putting ethical AI principles into practice; ▪ Designing governance frameworks for trustworthy AI; ▪ Developing AI applications in areas of societal interest.	▪ Pursue research collaborations that create AI-enabled solutions to challenges in areas of social need, including healthcare and climate policy. ▪ Create human-centric AI methods and tools, which can be deployed in alignment with fundamental rights or social expectations around privacy, transparency, safety, and fairness. ▪ Advance the foundations and application of explainable AI methods. ▪ Build collaborations with policymakers, legal experts and social sciences to understand the ethical implications of advances in AI. ▪ Bring together methods from quantum computing and machine learning to design more energy-efficient AI methods and hardware.

Box 11. Use case: DeepMammo – Using AI to screen breast cancer

Policymakers, publics, and practitioners have great aspirations for the potential of AI to improve healthcare systems and patient outcomes. Delivering on these aspirations requires trustworthy AI that can be integrated into clinical practice, leveraging methods for the analysis of multimodal data sources and explainable AI to create effective decision-support systems.

Breast cancer is the most common type of cancer among women, with one in eight diagnosed during their lifetime.[111] When the disease is detected early, for example through routine mammograms, treatment can be highly effective,

111 NHS (2023) Overview of breast cancer, available at: www.nhs.uk/conditions/breast-cancer/#:~:text=About%201%20in%208%20women,changes%20examined%20by%20a%20GP

achieving survival probabilities of 90% or higher.[112] However, mass screening increases radiologists' workloads, generating more images to analyse. Despite widespread digitization of images and clinical records, collecting data to build time-saving tools is not easy: the quality of images from screenings varies, and there are complex privacy rules to navigate.

Intelligent diagnoses

AIGEA Medical is applying AI to digital imaging workflows in radiology with the aim of helping radiologists diagnose breast cancer more quickly by speeding up repetitive tasks.[113] "If we can use AI to screen out negative cases and bring suspicious or positive cases to the attention of radiologists, it could reduce the time it takes to get a diagnosis from weeks to hours," says Carlo Aliprandi, cofounder and CEO of the Italian MedTech start-up.

The company says its AI cloud-based software, called DeepMammo, increases the accuracy of diagnoses because it works from multiple sources of data. Its patented 'Multimodal Learning AI' based on Deep Neural Networks is applied to a dataset of exams, combining images from screening with text from clinical reports. "We're not just using images to train DeepMammo, but data from clinical records, such as age and tissue density," Aliprandi explains.

At first, the start-up focused on processing medical images, using AI algorithms to analyse scans and generate a classification of positive or negative for breast cancer by exploiting state-of-the-art Convolutional Networks (CNNs). Connecting these capabilities to clinical practice required additional functions, specifically to take into account other clinical inputs and to present outputs in a way that supported clinicians in their work. Its work on the ELISE programme allowed AIGEA Medical to work on explainability, using Natural Language Generation (NLG) to describe important features of the images and incorporate multimodal data, increasing the power and flexibility of DeepMammo.

The company intends for its technology to plug into clinical systems already used by radiologists to manage their workflow. The idea is to use AI to screen a patient's scans and records before or at the point when a radiologist is due to make a reading of the examination.

ELISE helped AIGEA Medical implement a core step in its roadmap, enriching DeepMammo with a novel AI for automatic generation of medical reports from images, using NLG technology. "With the support of ELISE and the company's mentor, Professor Eneko Agirre of the Computer Science Faculty of the University of the Basque Country UPV/EHU, we could move along with this step. It was very

112 WHO (2021) Breast cancer fact sheet, available at: www.who.int/news-room/fact-sheets/detail/breast-cancer

113 Algea Medical (2022) Meet DeepMammo, the AI that saves women's lives, available at: www.aigeamedical.com

useful for us," Aliprandi says. The novel AI can provide the radiologist with a positive or negative classification for the presence of the signs of cancer, but also a useful report that includes data to inform clinical decision-making, such as relevant details in notes or images descriptions that they can use. DeepMammo now explains how it arrives at a diagnosis and provides transparency to the user. "That's what we mean by explainability – explaining a diagnosis to radiologists in terms that are meaningful to them," Aliprandi says.

Visualising the future

"We are at the very beginning of the age of AI applied to digital imaging diagnostics," Aliprandi predicts. "Work has to be done in order to fully face the ethical and regulatory challenges of AI in healthcare, particularly privacy, data fairness and trustworthiness. Addressing these issues is integral for a solution like DeepMammo, which aims to leverage technology in the interest of democracy and common well-being." He hopes his company's technology and other AI in the wider community will be applied to diagnose different diseases and be used for everything from mammograms to MRI scans. "There are technical challenges, including adapting AI techniques for different modalities and diseases, as well as collecting the necessary amount of data." Aliprandi believes that eventually, AI will be a big support for radiologists and other clinicians, allowing them to leave mundane work to algorithms and focus on tasks that really add value, such as diagnosing complex cancer types or delivering better patient care.

A responsive research agenda

Progress in AI has historically come in waves. Over the last ten years, this has yielded AI technologies that are capable of delivering impressive performance when trained on tightly defined tasks. A new wave of technical advances, based on a trend towards large models, is extending these capabilities, creating AI systems that are broader purpose. These advances can deliver plausible results across a wider range of tasks than previously, and innovations in the coming years could extend their capabilities further (see Box 12).

Box 12. Connecting the research agenda to recent progress in Large Language Models and generative AI

Rapidly-advancing capabilities in Large Language Models (LLMs) have caught the attention of researchers, policymakers, and publics, reinvigorating debate about the impact of AI on society. These are complex systems, based on advanced AI methods, and whose impact on society is shaped by how, where, and by whom they are deployed. Ensuring they deliver for all in society will require action across ELISE's research themes:

Trustworthiness and certification
LLMs generate highly-plausible content. However, this content is often incorrect: it can be fictitious, biased, or out-of-date, in ways that might not be immediately clear to the user. There is a pressing need to identify how each of the EU's characteristics for trustworthy AI can be embedded in further developments in this field. To enable their wider use, new mechanisms for certification may also be needed, creating a proving ground in which users can test the performance of models – and their adherence to regulatory requirements – before implementing them in practice.

Security and privacy
Progress in LLMs has generated a range of security and privacy concerns. It is often not clear what data has been used, or how, in developing the model. Queries submitted to LLMs are visible to the organisation providing the model. To avoid such information being misused or made public, emerging user guidance suggests not submitting such information in search queries; further work is needed to understand how data used by foundation models aligns with current regulations. They also influence the cybersecurity environment: injection attacks can undermine the integrity of the system in operation, and LLMs can be used to generate new malware. Wider use of these models also creates new risks for the information environment in which they operate. Their ability to generate misinformation exacerbates issues associated with AI-generated fake content; the result of wider use could be convincing phishing campaigns, for example.[114] In response to these new AI-enabled threats, AI-enabled tools will be needed to enable rapid response to new security concerns.

Explainability, accountability, and decision-making
Concerns have already been raised about the transparency of LLMs and foundation models. These concerns relate to the ability to scrutinise what data the systems have been trained on, how the models work, or what their performance limitations might be when deployed. Technical methods to enhance their interpretability along some axes are in development, but the size of these models and the size of the datasets on which they are trained make the development of such methods highly challenging.

114 National Cyber Security Centre (2023) ChatGPT and large language models: what's the risk? Available at: www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk

AI integration
Integration of LLMs into existing systems requires a mix of technical and operational interventions. Practices for scrutinising the strengths and limitations of these models are in flux. Users need to be able to understand the limits of these systems – how they should or should not be used in practice – and be able to work effectively alongside them in deployment. Addressing these issues will require technical strategies for AI integration, for example through new interfaces, alongside efforts to build organisational capability and skills.

Embedding AI ethics
The development of large foundation models highlights the power asymmetries that shape AI development. Progress in recent years has been driven by a small number of companies with large-scale resources. While their use could bring a range of social and economic benefits, these new capabilities require careful stewardship to ensure that they do not also result in harm, for example through misuse of data, misuse in deployment, disruptions to the online information environment, or intensive use of natural resources in development. Human-centric AI methods – based in responsible research and innovation practices – can help direct the development of these models towards more socially beneficial outcomes. Action to demystify and de-hype advances in AI is also needed, to enable publics and policymakers to understand the limits, risks, and potential benefits of further developments in this domain.

Tackling these issues requires advances in core underlying methods to increase the power of AI techniques and to embed human-centric perspectives in their development. Progress under each of the cross-cutting themes described in this document can help steward the development of LLMs and other foundation models towards more socially beneficial outcomes, by combining progress in AI methods with action to ensure these systems are more robust and aligned with the needs of society. Alongside these technical advances, Europe's wider innovation ecosystem will also influence the extent to which the benefits of these powerful systems are realised in practice. Capabilities in core technologies need to be connected to an environment where organisations have the skills or know-how to adopt new AI tools, and where start-ups and scale-ups can translate this know-how to new products and services. Creating this innovation ecosystem is at the core of ELISE's work.

The challenge that faces the field is how to translate this performance into real-world benefits for individuals, organisations, and society. Rapid progress in trustworthiness, security and privacy, explainability, AI integration, and AI ethics has provided a constellation of theoretical, methodological, and operational tools that can help deliver safe and effective AI systems. Continued support to drive further advances in these areas – advancing the technical capabilities or AI, overcoming current limitations, and understanding what works in practice – can help translate this progress into economic and social benefits.

1

AI technologies that are technically advanced

Areas for action arising from the Innovation Missions

Methods and tools to analyse real-world, multimodal data.

Strengthened core machine learning capabilities, through methodological and theoretical advances, such as techniques to bridge between data-driven and domain knowledge.

Tools to interrogate the workings of complex systems through advances in simulation, emulation, and causality.

1 2 — Advance the science of artificial intelligence by better understanding the intelligent behaviour of living systems and how this emerges.

1 2 3 — Strengthen the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems.

1 3 — Design new, energy-efficient machine learning algorithms and hardware implementations, drawing from concepts in quantum physics and statistical physics to develop more powerful machine learning systems.

1 2 — Build bridges between classical AI methods and machine learning to advance further progress in computer vision.

1 2 3 — Explore the role of causal modelling as a bridge between observational and interventional learning, identifying the principles for interactive learning systems.

1 2 3 — Push forward the foundational of multimodal learning systems and expand their application.

1 — Improve the performance of deep learning systems.

2

AI technologies that are robust in deployment

Areas for action arising from the Innovation Missions

AI that is robust under dynamic or uncertain conditions.

Human-centric tools that are effective as decision-support.

Methods to enhance explainability in decision-making.

1 2 3 — Understand principles for robustness in deployment and develop techniques for machine learning that reliably performs well and can be integrated into real-world systems.

1 2 — Build systems for general-purpose natural language understanding and generation.

1 2 — Improve core machine learning functions, for example through enhanced methods for deep learning, computer vision, natural language understanding and generation, and semantic, symbolic, and interpretable machine learning.

1 2 — Create robotic systems that can interact intelligently with the world around them by combining robot learning approaches with machine learning methods, such as reinforcement learning; and information systems that can better understand human behaviour.

1 2 3 — Advance methods for embedding knowledge about the physical world in the design of machine learning systems, for example through simulators and emulators that can help explore the consequences of different interventions or model designs, and that can extract insights from the analysis of complex systems.

3

AI technologies that align with societal interests

Areas for action arising from the Innovation Missions

Techniques for trustworthy AI.

Deployed AI that is integrated into areas of critical need.

AI research and development that engages stakeholder perspectives.

2 3 — Create AI systems to support the delivery of effective public services, for example AI systems for healthcare that can monitor patient health, using complex datasets to develop decision-support systems and to foster breakthrough applications in healthcare and biomedicine. Build collaborations with policymakers, legal experts and social scientists to understand the ethical implications of advances in AI.

1 3 — Develop AI tools that can contribute to humanity's response to the climate crisis, increasing understanding of climate extremes, changes to earth systems and potential areas for intervention.

2 3 — Design novel machine learning algorithms that are better aligned with human needs and societal interests – for example taking into account concerns around fairness, privacy, accountability, transparency and autonomy – and that advance the technical frontiers and application of AI, for example in explainability.

1 2 3 — Highlighted flags indicate overlap between pathways

5. Looking ahead

Europe can lead a new wave of progress in AI technologies, creating next-generation AI that is made in Europe and that delivers real-world benefits for communities and businesses. To deliver this European capability, Europe needs to be at the forefront of advancing AI technologies, deploying them in service of European priorities, and connecting their development to the rights and values set out in European law. In pursuing these goals, Europe needs to grow its AI R&D ecosystem, creating a pipeline from development to deployment through collaborations that embed knowledge creation in local innovation ecosystems.

The transformative potential of AI stems from its pervasiveness. AI has been posited as the next General-Purpose Technology, with the ability to disrupt economies and societies, bringing both benefits and risks. The history of General-Purpose Technologies shows that their effects are felt over decades, through waves of technological progress and shifting patterns of adoption. Over the coming decades, the challenge for policymakers is to adapt to these shifting patterns: to steward a technology that is changing, pervasive across sectors, and enmeshed with societal interests and concerns. To position itself for rapid response, Europe can act to strengthen its AI foundations, by investing in research that advances core underlying technologies while advancing the trustworthiness of AI in deployment, and by building a sustainable infrastructure for AI research.

Recent advances in Large Language Models signal the transformational change that rapid progress in AI technologies could bring. Today's applications of these foundation models scratch the surface of AI's potential, and the analysis of the use of AI for the EU's Innovation Missions presented in this report showcases what AI could help achieve across priority policy areas. Progress in foundation models also demonstrates the importance of strengthening the fundamental infrastructure underpinning Europe's AI ecosystem. This progress has been driven by innovations in core underlying technologies, and investment to secure European leadership in these technologies is vital. Ensuring such progress delivers benefits for people and society requires human-centric AI methods and applications, through research, policy, and practice that connects technology development to the rights set out in European law. Translating these technologies to wider economic and social benefit relies on an ecosystem that facilitates AI adoption, through wider access to AI skills and a start-up and business environment that can deploy AI safely and effectively. Support to continue to build this infrastructure, and embed these functions in local innovation ecosystems, is necessary to position Europe to capitalise on recent AI advances and lead technology development in the future.

As a Network of Excellence, ELISE has shown how the EU's investments in AI can advance a strategically important research agenda, attract leading research talent to Europe, and translate research into innovation. The result is both enhanced collaboration across Europe and local connections that ensure communities and businesses benefit from AI progress. Sustained investment can ensure this ecosystem continues to grow, delivering AI made in Europe, for Europe and the wider world.

ANNEX 1

ELISE Research Programmes and Use Cases

ELISE currently convenes the following research programmes:

Program	Program aim	Directors
Machine Learning for Health	To create AI systems that can be used to monitor patient health, using complex datasets to inform decision-support systems and to foster breakthrough applications in healthcare and biomedicine.	▪ Gunnar Rätsch (ETH Zürich) ▪ Mihaela van der Schaar (University of Cambridge, ATI) ▪ Oliver Stegle (German Cancer Research Center and EMBL)
Robot Learning: Closing the Reality Gap!	To create robotic systems that can interact intelligently with the world around them.	▪ Aude Billard (EPFL) ▪ Jan Peters (TU Darmstadt) ▪ Tamim Asfour (Karlsruhe Institute of Technology)
Geometric Deep Learning	To improve the performance of deep learning algorithms in non-Euclidean spaces, and in so doing identify new applications, efficient implementations and symmetries in data that can be used to advance the use of deep learning methods.	▪ Max Welling (University of Amsterdam, Qualcomm AI Research) ▪ Michael Bronstein (University of Oxford)
Human-centric Machine Learning	To develop novel machine learning algorithms that are better aligned with human needs and societal interests, for example taking into account concerns around fairness, privacy, accountability, transparency and autonomy.	▪ Nuria Oliver (DataPop Alliance, Royal Academy of Engineering, University of Alicante) ▪ Plamen Angelov (Lancaster University) ▪ Adrian Weller (Alan Turing Institute)
Interactive Learning and Interventional Representations	To explore the role of causal modelling in bridging the gap between observational and interventional learning and understand the principles underlying interactive learning systems.	▪ Nicolò Cesa-Bianchi (Università degli Studi di Milano) ▪ Andreas Krause (ETH Zürich) ▪ Bernhard Schölkopf (MPI-IS Tübingen)
Machine Learning and Computer Vision	To build bridges between classical algorithms and machine learning to unlock further advances in computer vision.	▪ Cordelia Schmid (INRIA) ▪ Yair Weiss (Hebrew University) ▪ Bernt Schiele (MPI Informatics)

Program	Program aim	Directors
Machine Learning for Earth and Climate Sciences	To create AI tools that can contribute to humanity's response to the climate crisis, increasing understanding of climate extremes, changes to earth systems and potential areas for intervention.	▪ Gustau Camps-Valls (Universitat de València) ▪ Markus Reichstein (MPI for Biogeochemistry)
Multimodal Learning Systems	To push the boundaries of the foundational aspects of this field, to build bridges between researchers and practitioners currently active in multiple unimodal communities, as well as expanding and exploring the applications of multimodal learning systems.	▪ Cees Snoek (University of Amsterdam) ▪ Nicu Sebe (University of Trento)
Natural Intelligence	To advance the science of artificial intelligence by better understanding the intelligent behaviour of living systems and how this emerges.	▪ Matthias Bethge (University of Tübingen) ▪ Y-Lan Boureau (Facebook AI Research) ▪ Peter Dayan (MPI for Biological Cybernetics)
Natural Language Processing	To build systems for general-purpose natural language understanding and generation.	▪ Ivan Titov (University of Edinburgh) ▪ Andre F.T. Martins (Instituto Superior Técnico, Lisboa) ▪ Iryna Gurevych (Technical University Darmstadt)
Quantum and Physics-Based Machine Learning	To design new, energy-efficient machine learning algorithms and hardware implementations, drawing from concepts in quantum physics and statistical physics to develop more powerful machine learning systems.	▪ Bert Kappen (Radboud University Nijmegen) ▪ Riccardo Zecchina (Bocconi University Milan)
Robust Machine Learning	To understand the principles and develop the techniques for machine learning that reliably performs well.	▪ Yee Whye Teh (University of Oxford and DeepMind) ▪ Chris Holmes (University of Oxford and Alan Turing Institute) ▪ Samuel Kaski (Aalto University, Finnish Center for Artificial Intelligence and University of Manchester)
Semantic, Symbolic and Interpretable Machine Learning	To become a cumulation point of like-minded researchers and we expect fruitful interactions with closely related programs covering, e.g., NLP, vision, and geometric deep learning.	▪ Volker Tresp (Siemens) ▪ Kristian Kersting (TU Darmstadt) ▪ Paolo Frasconi (University of Florence)

Program	Program aim	Directors
Theory, Algorithms and Computations of Modern Learning Systems	To advance the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems.	▪ Francis Bach (INRIA) ▪ Philipp Hennig (University Tübingen/MPI Tübingen) ▪ Lorenzo Rosasco (Universita' di Genova and MIT)

Enhancing each of these work programmes, ELISE facilitates a suite of industrial collaborations around the following use cases:

Use case	Partner	ELISE research programme	Research area
Environment Perception for Autonomous Driving	Audi	Machine Learning and Computer Vision; Robust Machine Learning	Environment perception for autonomous driving
AI Explainability for Optical Inspection in Manufacturing	Bosch	Explainability and Fairness in Data Mining	Novel ML-techniques which are able to explain binary classifier decisions
Robust ML Benchmark and Challenge	DeepMind	Robust Machine Learning	Suite of benchmarks for the robustness of ML methods and a real-world challenge to foster innovations in robust ML
Generative Adversarial Networks for Real-time Rendering	EnliteAI	Machine Learning and Computer Vision; Multimedia/ Multimodal Information	Real-time generation of high-quality audio-visual content
Robust and Certifiable Multimodal Learning for Safe Human-Robot Interaction	Inxpect	Robot Learning	Multimodal sensing and understanding
Data-Efficient Activity Recognition in Video	Kepler Vision	Machine Learning and Computer Vision; Health	Data-Efficient Activity Recognition in Video
Audio Representations in Hearing Health Care	Oticon	Human-centric Machine Learning; Multimedia/ Multimodal Information; Health	Learning from hearing aid usage
Algorithmic Validation of Smart City AI System Behaviour	Saidot	Human-Centric Machine Learning; Explainability and Fairness in Data Mining	AI validation for interpretability, transparency and accountability

Use case	Partner	ELISE research programme	Research area
Knowledge Scene Graphs for Industrial Applications	Siemens	Machine Learning and Computer Vision; Geometric Deep Learning: Graph, Group and Gauge Convolutions	Describe relational structures in images
Material Flow Optimization	TGW	Interactive Learning and Interventional Representations	Warehouse logistics optimization using RL
Experimental Environment for Real World Reinforcement Learning	Zalando	Interactive Learning and Interventional Representations	Optimal business decision-making in a given situation using RL

ANNEX 2

Participants in Innovation Mission workshops

Thank you to the researchers, policymakers, and practitioners that contributed to ELISE's workshops on AI and the EU Innovation Missions in Winter 2022 and Spring 2023.[115]

- Aiden Hunter, British Antarctic Survey
- Alexis Perrier, Consultant
- Alfonso Acosta Gonçalves, European Commission
- Alice Antonello, Università degli Studi di Trieste
- Amélie Beucher, Aarhus University
- André Meyer-Vitali, German Research Centre for AI (DFKI)
- Andrew Neal, Rothamsted Research
- Angelo Solimini, European Commission
- Anna Pazola, UN Environment Programme World Conservation Monitoring Centre
- Armi Tiihonen, Aalto University
- Calogero Schillaci, European Commission
- Calum Handforth, University of Cambridge
- Christian Mollière, OroraTech
- Christoph Bock, University of Vienna
- Christos Diou, University of Athens
- Claire Weluda, British Antarctic Survey
- Claudio Michaelis, Eberhard Karls University Tübingen
- Conrad Phillipp, Max Planck Institute for Biogeochemistry
- David Wagg, University of Sheffield
- Devis Tuia, EPFL
- Diana Robinson, University of Cambridge
- Dimitris Zisis, e.Nios
- Dorina Thanou, EPFL
- Eiko Yoneki, University of Cambridge
- Emily Mitchell, University of Cambridge
- Emma Boland, British Antarctic Survey
- Enea Parimbelli, University of Pavia
- Fabienne Jacq, European Commission
- Faiza Samreen, UK Centre for Ecology and Hydrology
- Felix Creutzig, Mercator Research Institute on Global Commons and Climate Change
- Filipe Gomez Valdivia, British Antartic Survey
- Firat Güder, Imperial College London
- Francisco Doblas-reyes, Barcelona Supercomputing Centre
- Georg Dorffner, University of Vienna
- Giovanni Stracquadanio, University of Edinburgh
- Gonzalo Garcia-de-Arboleya, European Commission
- Guido Sanguinetti, University of Edinburgh; International School for Advanced Studies
- Gustau Camps-Valls, University of Valencia
- Gwilym Rowlands, University of Oxford
- Ines Machado, King's College London
- James Herbert-Read, University of Cambridge
- Jaume Jafono, British Antarctic Survey
- Jennifer Freer, British Antarctic Survey
- Jennifer Gabrys, University of Cambridge
- Jennifer Schooling, University of Cambridge
- Jessica van der Hoek, Ellogon
- Jose-Luis Fernandez-Villacanas, European Commission
- Karen Bakker, University of British Columbia
- Kasia Tokarska, Risilience
- Kathrin Grahmann, Leibniz Centre for Agricultural Landscape Research
- Katie Green, University of Cambridge
- Kenji Takeda, Microsoft Research Cambridge
- Kerry Howell, University of Cambridge
- Kerstin Rau, Eberhard Karls University Tübingen

115 This list comprises individuals content to be named in this report.

- Kevin Fleming, NLA International Ltd
- Kevin Luck, Aalto University; Finnish Center for Artificial Intelligence
- Kieran Bjergstrom, Omnikara
- Kim Tallaksen Halvorsen, Institute of Marine Research
- Konrad Schindler, ETH Zurich
- Koray Bingöl, TU Delft
- Lexuan Ye, University of Illinois
- Lis Wollesen de jong, Aarhus University
- Louisa van Zeeland, Alan Turing Institute
- Louise Foley, University of Cambridge
- Lucas Corinzia, ETH, Zurich
- Lucas Carvalho Gomes, Aarhus University
- Lucy Stephenson, British Antarctic Survey
- Luis Sánchez Álvarez, European Commission
- Magnus Rattray, University of Manchester
- Maria Beger, University of Leeds
- Markus Reichstein, Max Planck Institute for Biogeochemistry
- Martin Ulbrich, European Commission
- Matteo Turchetta, ETH Zurich
- Mengyi Gong, University of Lancaster
- Michael Thorne, British Antarctic Survey
- Michael Kenn
- Miguel Alvarez Rodriguez, European Commission
- Miguel Mahecha, University of Leipzig
- Mihaela van der Schaar, University of Cambridge
- Milena Vuckovich, VR VIS
- Nafiseh Kakhani, Eberhard Karls University Tübingen
- Nándor Csikós, MTMT
- Nektarios Kalampalikis, University of Patras
- Niki Kiriakidou, University of Athens
- Nikita Janakarajan, ETH Zurich
- Nuno Mota
- Omar Younis, ETH Zurich
- Orlando Timmerman, University of Bristol
- Peter Smith, University of Aberdeen
- Peter Weber, Aarhus University
- Philipp Rode, London School of Economics
- Pietro Liò, University of Cambridge
- Rachel Cavanagh, British Antarctic Survey
- Rachel Furner, Cambridge University
- Ralf Gitzel, ABB Corporate Research
- Raphael Padilha, Urban Environment Department of Helsinki
- Rebekah Oomen, University of Oslo
- Roisin O'Riordan, Department for Environment, Food and Rural Affairs, UK Government
- Ronan Fablet, IMT Atlantique
- Ronja Rönnback, Tilburg University
- Sancho Salcedo-Sanz, University of Alcala
- Scott Hosking, British Antarctic Survey
- Sebastian Gutierrez, Aarhus University
- Shane Beato, Petreanu Lab, Champalimaud Centre for the Unknown
- Simon Gardner, Natural Environment Research Council
- Simon Thomas, University of Cambridge
- Sophie Meyjes, Imperial College London
- Soumya Banerjee, University of Cambridge
- Sunho Kim
- Suraj Verma, Teesside University
- Swati Sharma, Microsoft
- Syed Haider, Institute of Cancer Research
- Victoria Janes-Bassett, University of Liverpool
- Viviana Bastidas, University of Cambridge
- Yuguang Wang, University of New South Wales
- Zoi Konstantinou, European Commission

ANNEX 3

Acknowledgements

In producing this refresh of ELISE's Strategic Research Agenda, we are grateful for contributions made by:

- Leaders of ELISE research programmes who contributed to surveys and workshops;
- Members of the research and policy communities who contributed to workshops on AI in the Innovation Missions;
- The Oxford Insights team that supported those workshops through a literature review and stakeholder interviews on AI and the Innovation Missions, and the individuals who agreed to be interviewed in this process;
- The Collective Next team that facilitated workshops;
- Reviewers from the ELISE and ELLIS networks who helped develop its content.

European Learning and Intelligent Systems Excellence (ELISE) is a European Network of AI Excellence Centres funded by the European Commission under the Horizon 2020 Framework Programme under Grant Agreement no 951847. ELISE began in September 2020. The information contained in this document reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein. For more information on ELISE, please see www.elise-ai.eu.

This document was produced with support from: Sami Heinäsmäki (Aalto University); Samuel Kaski (Aalto University and University of Manchester); Neil Lawrence (University of Cambridge); Jessica Montgomery (University of Cambridge); Christopher Murray (Aalto University); Sue Sebborn (University of Cambridge); and Arnold Smeulders (University of Amsterdam).

elise

ELISE works in close collaboration with the ELLIS Society (European Laboratory for Learning and Intelligent Systems).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951847.