Creating a European AI Powerhouse

A Strategic Research Agenda from the European Learning and Intelligent Systems Excellence (ELISE) consortium

May 2021

elise

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 the Grant Agreement no 951847. ELISE began in September 2020 and will run for 3 years. 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

Contents

Executive Summary

ELISE's vision is of a Powerhouse of European AI. Motivated by the ambition to establish European leadership in AI and create a new generation of trustworthy AI systems, ELISE will build a network of the continent's leading AI researchers. Together, this network will pursue pan-European research collaborations that tackle issues of pressing scientific and social concern.

Many of the recent breakthroughs in the field of AI – breakthroughs that have attracted widespread interest from researchers, policymakers and the wider public – have been enabled by advances in machine learning. Machine learning systems are already successfully deployed in a range of applications, from car driver assistance to language translation and in fields from climate science to drug development. Further advances in AI have the potential to transform economies and society, contributing to better healthcare, safer transport, more productive and competitive industry, and more effective public services.

Recognising this potential, recent policy initiatives have placed AI at the heart of European visions for a thriving economy, healthy planet and effective public administration. Investments in research and development are seeking to promote AI adoption across sectors; emerging legislative programmes are setting regulatory frameworks for AI products and services; and AI has been recognised as an important enabler of major policy agendas, such as the Green Deal.

Realising these visions will require AI systems that are technically sophisticated, robust in deployment, and designed in alignment with the rights and standards set out in European law. AI must meet expected standards of security and data privacy; be designed in ways that allow different stakeholder communities to understand its results; adhere to regulatory standards that verify it is trustworthy; be deployed safely and effectively, and be able to operate under conditions of uncertainty; and uphold ethical standards and principles. Designing such systems is at the core of ELISE's work, through research programmes that advance theory and methods in machine learning and AI, and that translate these methods into practice.

ELISE researchers are already leading projects that seek to advance foundational concepts in AI, to develop AI methods in line with social and regulatory needs, and to deploy AI systems in applications that could bring significant social and environmental impacts. Areas of research interest for the consortium include advancements in machine learning theory and core technical functions, such as computer vision, natural language processing and information retrieval; creation of new learning strategies, through new models and methods in areas such as transfer learning; further development of methods for explainability and robustness; and collaborations to design domain-appropriate systems in areas such as healthcare. These current research programmes will increase the power of today's AI methods and promote their deployment in areas that can boost economic growth and societal wellbeing, while at the same time helping to ensure that these new AI tools work well for all in society.

Drawing together these prorgammes, this Strategic Research Agenda sets out ELISE's roadmap for creating AI technologies that are technically advanced, robust in deployment and aligned with social values. It outlines how technological advances can contribute to

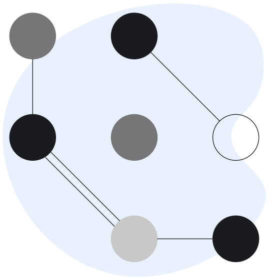
European policy ambitions for AI, and the support needed from technologists and policymakers to maintain European leadership in AI.

Achieving these ambitions for the future of machine learning and AI will require investments in world-class AI research in Europe. By building a network of independent centres of research excellence, Europe can maintain its world class research community, its vibrant research ecosystem, and its leading role in AI development. Each centre will bring its own areas of specialism, allowing countries across Europe to build on the top-class research in their region and to pursue research that reflects the needs of their local innovation ecosystem, while maintaining strong links across borders that foster a wider sense of European AI. ELISE's work will be the basis of such a network, supporting innovations in research, attracting top AI research talent to centres of excellence, and facilitating collaboration across those centres. In support of this aim, and working closely with the European Laboratory for Learning and Intelligent Systems, ELISE will partner with industry to share insights from the cutting edge of AI research and development, and to support wider adoption of trustworthy AI systems. To help build a European AI research community, it will also support early career researcher mobility across Europe's top machine learning research groups, fostering further collaborations.

It is clear that AI will have profound economic and societal impacts on the global scale. ELISE plays an important role in this revolution, bringing scientific and industrial players together to enhance Europe's innovation capacity in AI, creating new market opportunities and strengthening the competitiveness and growth of European industry. As AI research continues to advance, understandings of areas of opportunity and concern in relation to AI will evolve. Reflecting the needs and interests of the AI community, ELISE will continue to update this Strategic Research Agenda throughout its lifetime.

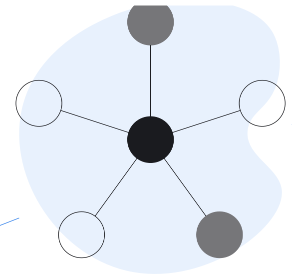
ELISE AI roadmap

ELISE has identified a pathway to achieving trustworthy AI technologies, through advances in technical capabilities; new approaches to deployment; and action to align AI design and implementation with social values. ELISE will:



1 Strengthen technical capabilities

- [1 2] Advance the science of artificial intelligence by better understanding the intelligent behaviour of living systems.
- [1 2] Strengthen the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems, and addressing concerns about privacy and security.
- [1 3] Design new, energy-efficient machine learning algorithms and hardware implementations.
- [1 2] Build bridges between classical AI methods and machine learning to unlock further advances in both.
- [1 2 3] Create machine learning methods that are explainable by design, through advances in surrogate modelling methods, visualisation tools, and approaches to encoding existing knowledge.
- [1 2] Explore the role of causal modelling in bridging the gap between observational and interventional learning and understand the principles underlying interactive learning systems.
- [1 2] Develop new learning strategies to operate in low data-resource environments, advancing research in areas such as one- or few-shot learning; transfer learning; interactive learning; and reinforcement learning.

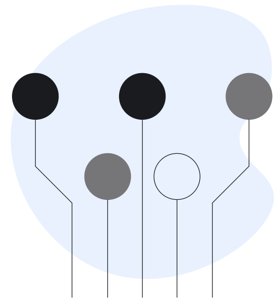


2 Improve performance in deployment

- [1 2 3] Understand the principles for robustness in deployment and develop techniques for calibrating uncertainty to support safe and reliable machine learning systems.
- [1 2] Improve core machine learning functions, for example through general-purpose natural language understanding and generation, and methods for rapid and accurate identification of semantic content from media in deployed environments.
- [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] 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.
- [2 3] Integrate emerging methods for ensuring the robustness of machine learning systems into real-world use cases.
- [1 2] Advance methods for embedding knowledge about the physical world in the design of machine learning systems.

3 Align with social interests

- [2 3] Create AI systems to support delivery of effective public services, for example creating 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.
- [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.
- [1 2 3] Foster collaborations at the interface of machine learning and human-computer interaction to understand how human and algorithmic decision-making interact.
- [2 3] Advance the foundations and application of explainable AI.
- [3] Build collaborations with policymakers, legal experts and social sciences to understand the ethical implications of advances in AI.



[1 2 3] Highlighted flags indicate overlap between pathways

1. Background and purpose of this document

European Learning and Intelligence Systems Excellence (ELISE) is a consortium of artificial intelligence (AI) research hubs. ELISE conducts research and knowledge exchange activities to create a new generation of trustworthy AI systems, which can be deployed reliably in real-world applications to support economic growth and benefit all in society.

The ELISE Strategic Research Agenda lays the foundation for research from the ELISE consortium. It starts by summarising the key themes that will guide ELISE research and that inform the network's activities. It then outlines the ways in which ELISE will collaborate with industry partners and European research initiatives to help share the benefits of its work. In so doing, this document presents a roadmap for the development of European AI through academic research that maintains European R&D excellence.

This Strategic Research Agenda combines the technical expertise of ELISE members with input from key stakeholders working to ensure that AI is used safely and effectively, creating a framework for AI research that is 'made in Europe'. It was produced following consultation with AI researchers from across Europe, including those leading research programmes for the ELLIS initiative (the European Laboratory for Learning and Intelligent Systems)[1] (Annex 1).

ELLIS is an independent, pan-European organisation that promotes the development of machine learning and AI in Europe by supporting high-quality research, attracting global talent to European labs, and disseminating knowledge across academia and industry. It was formed in 2018 to create a network of excellence that would drive research breakthroughs in AI, and today consists of 30 ELLIS Units from across 14 countries. ELISE works in close collaboration with ELLIS and builds on its work by mobilising ELLIS researchers and connecting them with the wider European landscape of businesses, policymakers and research networks.

The ELISE Strategic Research Agenda will be a 'live' document, which develops as ELISE's work progresses, responding to technological innovations and emerging areas of need.

1 For further information on the ELLIS Initiative, see: https://ellis.eu

2. Context: a vision for trustworthy European AI

AI in Europe

THE POTENTIAL OF AI IN EUROPE

AI has the potential to transform economies and society, contributing to better healthcare, safer transport, more productive and competitive industry, and more effective public administration.[2] Machine learning systems are already successfully deployed in some industry sectors, but many further benefits of AI have yet to be realised. If widely adopted across sectors and organisations, for example, productivity gains and new forms of economic activity enabled by AI could increase economic output by up to 14% by 2030.[3]

With the ambition of harnessing this potential for economic growth and social benefit, governments across Europe have set in place strategies to advance the use of AI while safeguarding fundamental rights. Reflecting these needs, the EU's strategy for data and AI seeks to promote four key pillars of activity: technology that works for people; a fair and competitive digital economy; an open, democratic and sustainable digital society; and Europe as a global digital player.[4] Together, these pillars highlight the importance of supporting technical advances, deploying safe and reliable AI systems in areas of social and economic need, and ensuring AI systems are designed to uphold the rights and values that form the foundation of European laws.

PLACING EUROPEAN VALUES AT THE HEART OF AI DEVELOPMENT

Recent policy and research debates across Europe have highlighted a range of areas in which AI systems interact with social values or concerns. There is now a clear demand for AI to be developed and implemented in ways that ensure it respects applicable laws, upholds ethical principles and values, and takes into account the changing demands of both the technical and social environment.[5]

Recognising these needs, excellence and trust are the cornerstones of the EU's AI strategy[6]

2 European Commission (2019) Factsheet: AI for Europe, available at: https://ec.europa.eu/digital-single-market/en/news/factsheet-artificial-intelligence-europe

3 Various estimates of the economic impact of AI have been produced. Some are explored here: European Parliamentary Research Service (2019) Economic impacts of artificial intelligence, available at: www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf

4 Through dialogues with sectors that could make use of AI, the Commission will seek to advance the use of AI in healthcare, rural administrations and public service operators, and by developing a governance framework for high-risk applications of AI, the EU will seek to give citizens and organisations confidence that AI is being used safely and effectively. For further information, see: European Commission (2019) European Digital Strategy, available at: https://ec.europa.eu/digital-single-market/en/content/european-digital-strategy

5 European Commission (2019) Ethics guidelines for trustworthy AI, available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai; Euractiv (2021) Von der Leyen assures MEPs: We'll 'go further' on AI that harms fundamental rights, 30 March 2021, available at: www.euractiv.com/section/digital/news/von-der-leyen-assures-meps-well-go-further-on-ai-that-harms-fundamental-right; and European Commission (2021) Coordinated Plan on AI 2021 Review, available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

6 For further information see: https://ec.europa.eu/digital-single-market/en/artificial-intelligence

Harnessing the potential of AI: AI and the Green Deal

AI could play an important role in helping countries to meet their sustainability goals. It offers opportunities to more accurately monitor land and natural resource use; it can be deployed to analyse and optimise patterns of energy use and supply, reducing energy consumption and carbon-intensity across sectors; and it can be used to streamline or optimise activities across industries – from agriculture to transport – with the aim of reducing negative human impacts on the environment.[7]

AI and machine learning are already delivering products that can be used in the fight against climate change, including tools that can detect climate extremes and that can help researchers investigate the physical processes driving changes to the Earth's climate systems. These existing methods have enabled significant progress in tackling problems that seemed insurmountable only some years ago. For example, the accuracy of weather forecasting has improved significantly and the ability of researchers to continuously monitor the earth system has increased, allowing better understanding of the location of water and carbon resources, better detection of land use changes, and better tracking of desertification. The insights created as a result can be used by policymakers seeking to achieve the sustainable development goals and by industry players seeking to minimise their environmental impact.

Recognising this potential, the EU's Green Deal seeks to deploy AI and digital technologies to build a sustainable economy and help Europe achieve climate neutrality by 2050.[8] One key initiative under this banner – Destination Earth – is developing "a very high precision digital model of the Earth to monitor and simulate natural and human activity".[9] To extract actionable insights from such modelling initiatives, researchers will require machine learning methods that can interrogate datasets from different sources and provide insights for policymaking.

ELISE research on climate and earth sciences will improve the accuracy and effectiveness of machine learning systems for earth and climate science by bringing together existing knowledge about the physical dynamics of earth's systems with machine learning approaches. The resulting algorithms will combine the dynamics of a physical system – its invariances, conservation laws, and spatio-temporal effects – with data from observational studies, physical models of the climate and the outputs from advanced simulations. By blending these types of data, researchers will be able to better predict the future dynamics of the Earth-Climate system, generating insights that can contribute to the fight against climate change.

7 For further examples of these applications, see: www.climatechange.ai

8 European Commission (2019) The European Green Deal, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1596443911913&uri=CELEX:52019DC0640#document2

9 Destination Earth project; see: https://ec.europa.eu/digital-single-market/en/destination-earth-destine and its vision of an 'ecosystem of trust' in which AI systems comply with EU rules and their deployment respects citizens' fundamental rights. The seven key characteristics of trustworthy AI that it seeks to advance are:[10]

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental wellbeing
7. Accountability

This emerging European approach to AI seeks to unlock a new wave of AI development, taking a human-centric perspective that both supports AI uptake across industry and ensures that AI systems serve the interests of citizens.

Technological advances can play a role in helping achieve these outcomes, boosting economic growth, increasing digital inclusion, and advancing new applications that improve individual and public wellbeing. There is an opportunity now for a fresh European AI research agenda that advances the development of these trustworthy AI systems and promotes their adoption across sectors. ELISE will be at the forefront of these advances, driving innovations in machine learning theory, methods, and applications to create AI technologies that can benefit people and society.

POSITIONING EUROPE AT THE FOREFRONT OF AI

If Europe is to realise the benefits of AI for citizens, businesses and society, it needs to retain its position at the forefront of global AI development and ensure AI technologies respect the rights and concerns of citizens.[11] Nations across the world are investing significantly in machine learning and AI, and their use in data-intensive industries such as health, finance, tourism, logistics, retail, services and smart manufacturing. These investments are already bearing fruit, with the US and China overtaking the EU in AI research publication outputs and other metrics of AI leadership, such as venture capital and start up success.[12] Without further action to bolster the foundations of independent AI research across the continent, and to ensure these foundations respect the rights and concerns of citizens, Europe risks losing its role at the forefront of global AI development.[13]

Investments in world-class AI research and science in Europe can help counter this risk, increasing its attractiveness to international research talent and attracting additional private sector resources. By building a network of centres of research excellence – each with their own independent academic research directions, regional specialisation and ecosystem of applications – Europe can maintain its world class research community, its vibrant research ecosystem, and its leading role in AI development. ELISE's work will be the basis of such a network, supporting innovations in research, attracting top AI research talent to centres of excellence, and facilitating collaboration across those centres.

10 European Commision High Level Group on AI (2018) Requirements of trustworthy AI, available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Accountability

11 European Commission (2020) AI White Paper, available at: https://ec.europa.eu/digital-single-market/en/artificial-intelligence

12 Centre for Data Innovation (2019) Who is winning the AI race? Available at: https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states

13 European Commission (2021) Keynote speech by President von der Leyen at the 'Masters of Digital 2021' event, available at: https://ec.europa.eu/commission/presscorner/detail/en/speech_21_419

Harnessing the potential of AI: AI for health

With healthcare costs amounting to at least 10% of GDP across most European countries and AI applications in medicine predicted to add billions of dollars of value to the global economy,[14] healthcare has been identified as a strategically important sector for AI deployment. Reflecting both potential economic benefits of AI in health and the potential benefits to citizens of access to improved diagnostics and treatments, the EU's AI strategy seeks to promote the development and deployment of AI across sectors, while taking regulatory action to manage the risks associated with AI in safety-critical domains.[15]

AI can be deployed across a range of healthcare activities. It can streamline decision-making by helping physicians manage diverse sets of patient data", and can help identify new treatments or therapeutics by helping researchers interrogate large biomedical datasets. In intensive care, for example, AI has been deployed to integrate data streams to detect early warning signals that a patient might be at risk of impending organ failure, supporting rapid interventions that improve patient outcomes.[16] Further applications of AI in healthcare could improve patient wellbeing, helping patients to take preventative action before experiencing serious illnesses.

These applications of AI for health interact with a range of different fundamental rights and stakeholder responsibilities:

- The effectiveness of AI tools must be maintained across different communities, and its application must uphold principles of non-discrimination.
- AI must be capable of providing explanations that allow physicians to understand why it has reached a decision and patients to interrogate why different treatment pathways have been recommended.
- Health data is sensitive by nature, and strict standards of security and privacy must be upheld in the operation of AI systems.
- Regulators must have confidence that AI systems meet required safety or efficacy standards, or have been certified appropriately.

14 European Commission Joint Research Centre (2020) AI Watch: AI in medician and healthcare: applications, availability and societal impact, available at: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-medicine-and-healthcare-applications-availability-and-societal

15 European Commission (2020) AI White Paper, available at: https://ec.europa.eu/digital-single-market/en/artificial-intelligence

16 Hyland et al, (2020) Early prediction of circulatory failure in the intensive care unit using machine learning, Nature Medicine, 26 (364-373), see: https://www.nature.com/articles/s41591-020-0789-4

Realising the potential of AI in healthcare will require researchers and policymakers to set standards around the transparency and security of AI systems, to design technical architectures that enable data access while managing privacy concerns, and to bridge the gap between research and commercialisation of AI systems.[17]

There already exist systems for managing some of these concerns. Security of health data, for example, can be maintained using specialised compute systems that manage data access for those performing research on medical data.[18] Research advances can help make these processes more effective, enabling data access while protecting fundamental rights.

ELISE research in healthcare will help create diagnostic and therapeutic tools that can be deployed safely and effectively in real-world healthcare settings. Close collaborations between machine learning experts, healthcare practitioners and legal experts will develop methods for AI in health that reflect the competing technical, legal and societal demands in this crucial domain. By examining potential failure modes and designing certifications to ensure their reliability, ELISE research will help ensure machine learning systems are robust in deployment.

The ELISE network

ELISE's vision is of a Powerhouse of European AI. Motivated by a shared desire to establish European leadership in the development and deployment of trustworthy AI, ELISE has established a network of excellence that connects leading researchers across Europe. By enabling high-quality research, inspiring a new generation of research leaders and establishing partnerships that use research knowledge to support economic growth and improvements to public services, ELISE will facilitate a new generation of AI research that serves European citizens and wider society. The ELISE project will establish a European network of AI excellence, by:

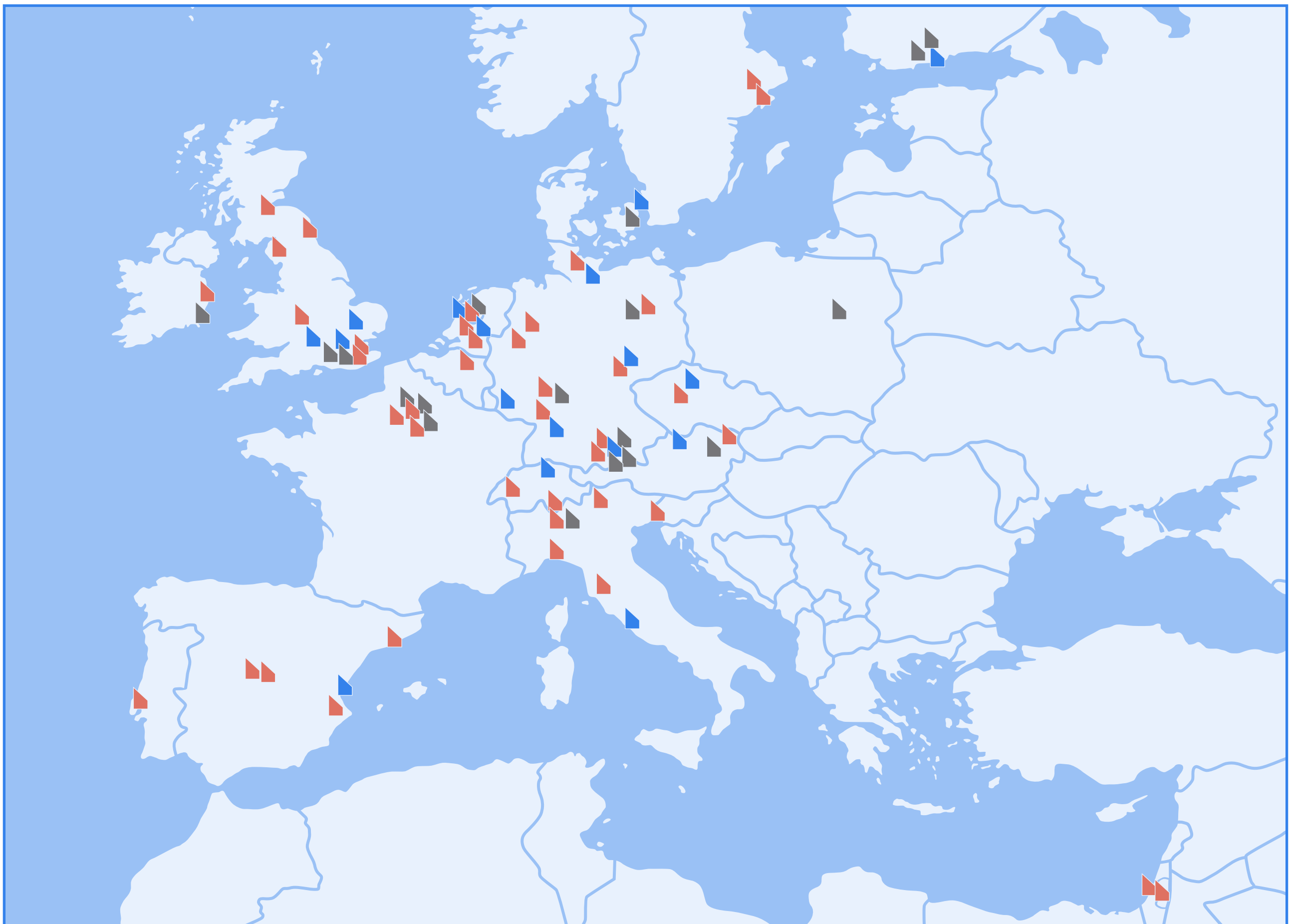
- connecting over 200 researchers and research groups to tackle shared research challenges (see Figure 1);
- supporting cross-European mobility through world-class European PhD and Postdoc programmes;
- enabling industrial innovation and applications through industry engagement activities.

ELISE builds on and works in cooperation with the fast-expanding European Laboratory for Learning and Intelligent Systems (ELLIS)

17 European Commission Joint Research Centre (2021) How can Europe become a global leader in AI in health?, available at: https://knowledge4policy.ec.europa.eu/file/how-can-europe-become-global-leader-ai-health_en

18 One example of such systems is Leonhard Med. For further information, see: https://blogs.ethz.ch/its/2019/02/21/secure-scientific-platform-for-confidential-data and the research programmes that it supports. Both ELISE and ELLIS are set up to attract students and experienced researchers, to sustain high level independent research in academia, and to spread AI knowledge and methods across research, industry and society. ELISE emerges from machine learning, the current core technology of AI, but the network is encompassing all ways of reasoning, and has interests in a variety of types of data from almost all sectors of science and industry.

Figure 1. **ELISE organising nodes**



Participants
- Work package duties
- Manage networking actions between participants and associate partners
- Participate in networking actions
- Participate in the PhD program

Associate partners
- Participate in networking actions
- Participate in the PhD program
- Travel costs covered by the participant organizing the networking action
- Host fellows

Company or other non-academic organization

AI and machine learning

AI is an umbrella term. It refers to a suite of technologies and methods that seek to simulate characteristics associated with intelligence in humans or other living systems. Machine learning is an approach to AI in which models process data, learning from this data to identify patterns or make predictions. Machine learning is the underpinning technology that has driven many of the impressive advances in AI, bringing the field to the wider attention of publics and policymakers in recent years. Progress in machine learning has also enabled further advances in fields such as natural language processing and computer vision, which contribute to the creation of AI systems.

Much of ELISE's work focuses on achieving advances in machine learning, as the method that is most widely deployed and consequently the approach which is generating the most urgent need for new research. However, the necessary advances will only be achieved through collaborations between machine learning and other sub-domains of AI. This report therefore uses both machine learning and AI in describing ELISE's work.

This Strategic Research Agenda is the first output from ELISE. It sets out current ELISE research programmes[19] and the themes that connect these, pointing to the role that technology can play in helping advance wider ambitions for trustworthy AI.[20] It then summarises how ELISE and ELLIS will help secure European leadership in AI by fostering a network of centres of excellence that together act as a beacon to attract talent and investment to support AI in Europe.

19 Many of these programmes are shared with ELLIS; summaries are presented in Tables 1 and 2.

20 This Strategic Research Agenda focuses on the role that technical advances can play in contributing to these ambitions, in line with wider policy objectives. While recognising the importance of a range of different stakeholders and interventions, this scope of this document does not extend to the wider policy, cultural, legal or organisational factors that would also contribute to the development of trustworthy AI systems.

3. ELISE Strategic Research Agenda: a framework for European AI research

Research for a new generation of AI and machine learning systems

Machine learning has transformed technical fields such as computer vision, natural language processing, robotics, information retrieval, recommender systems, and more. Recent studies into the foundations of deep neural networks, data representation and information extraction, learning with limited data, and learning with uncertain and noisy data have helped develop machine learning tools with capabilities that can be deployed to a wide range of challenges. Systems deployed today have already revolutionised various fields of science – from medical image diagnosis to climate science – and are becoming pervasive in commercial applications, including quality control in industrial production and optimisation of logistics and supply.

Despite these advances, most of machine learning's enormous potential is still to be realised. Central to Europe's future success in AI will be its ability to pursue excellent research that advances both foundational concepts in AI and its application to areas of social and scientific need.

The ELISE community brings together Europe's leading researchers in machine learning. This concentration of expertise makes ELISE uniquely placed to drive forward both the science of machine learning and its application across domains. Together, this community will build a new generation of machine learning systems that are technically powerful, effective when deployed in real-world environments, and aligned with society's values and concerns. Reflecting this approach, this Strategic Research Agenda provides a framework for coordinating ELISE research.

The Strategic Research Agenda combines the technical advances coming from ELISE and ELLIS research programmes with five cross-cutting themes that connect these programmes to wider scientific or social issues. These themes cover:

- security and privacy;
- explainable and transparent AI;
- trustworthiness and AI certification;
- AI integration across systems;
- AI ethics and societal impact.

Table 1 illustrates the intersection of these programmes and themes. Together, these lay the foundation for a new generation of AI methods and applications that benefit all in society.
The sections that follow summarise some of the areas of research that are currently being pursued by ELISE researchers, setting out how this research might develop in future.

Table 1. **ELISE/ELLIS research programmes and their connection to ELISE research themes**

Programme	Aim	Security and privacy	Explainabiilty and transparency	Trusworthiness and certification	AI integration	AI ethics and societal impact
Theory, algorithms and computations of modern learning systems	Advance the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems.	■	■	■	■	■
Machine learning for health	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.	■	■	■	■	■
Quantum and physics based machine learning	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.				■	■
Geometric deep learning	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.		■		■	

Programme	Aim	Security and privacy	Explainabiilty and transparency	Trusworthiness and certification	AI integration	AI ethics and societal impact
Machine learning for earth and climate sciences	Create AI tools that can contirbute to humanity's response to the climate crisis, increasing understanding of climate extremes, changes to earth systems and potential areas for intervention.		✓		✓	✓
Natural intelligence	Advance the science of artificial intelligence by better understanding the intelligent behaviour of living systems and how this emerges.		✓			✓
Human-centric machine learning	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.	✓	✓	✓	✓	✓
Robust machine learning	Understanding the principles and develop the techinque for machine learning that reliably performs well.	✓	✓	✓	✓	✓
Machine learning and computer vision	Build bridges between classical algorithms and machine learning to unlock further advances in computer vision.		✓	✓	✓	
Natural language processing	Build systems for general-purpose natural language understanding and generation.		✓		✓	

Programme	Aim	Security and privacy	Explainabiilty and transparency	TruswWorthiness and certification	AI integration	AI ethics and societal impact
Multimedia/ multimodal information	Enable meaningful, accurate, fast, and scaleable identification of semantic content from existing and emerging media, including cross and multi-modal media.		✓		✓	
Robot learning: closing the reality gap	Create robotic systems that can interact intelligently with the world around them.		✓	✓	✓	
Interactive learning and interventional representations	Explore the role of causal modelling in bridging the gap between observational and interventional learning and understand the principles underlying interactive learning systems.	✓	✓		✓	✓
Information retrieval	Improve the ability of information systems to understand human behaviour and produce appropriate answers in response.	✓	✓	✓	✓	✓
Explainability and fairness in data mining	Address concerns about explainability and fairness in data mining, by developing new tools and methods.		✓	✓		✓
Symbolic machine learning	Combine symbolic and data-driven AI methods to develop more effective AI systems.		✓	✓	✓	

Research in action: energy-efficient machine learning

While having great potential as an enabler of sustainability, current machine learning and AI systems are often a source of significant energy demand.

Increased computational power and data availability, as well as algorithmic advances, have led to impressive machine learning applications in areas such as computer vision, pattern recognition, robotics and AI. At the same time, conventional semiconductor technology is reaching its physical limits and the energy consumption of computing is growing. There is a pressing need to design novel computing paradigms to address these challenges.

The Quantum and physics based machine learning (QPhML) programme will apply concepts from quantum physics and statistical physics to develop novel machine learning algorithms, with the ultimate aim to realise novel future, possibly energy efficient, hardware implementations.

Research in action: robot learning

Autonomous robots could support humans to perform a wide variety of daily tasks – from driving, to logistics, manufacturing and support-assistants in the home. To function effectively, these robots must be able to sense and respond to their environments. Training such systems would typically require access to a large amount of data with examples of the tasks or situations that a robot might encounter when in use. However, this data is not usually available to those developing autonomous systems, as a result of their novelty or the high-cost of obtaining such data.

To create autonomous systems that can interact with the world around them, learning strategies are needed that allow a robotic agent to recognise when it is operating outside of known circumstances, and to develop strategies for responding to novel situations. Development of such strategies can be supported by high-quality simulations, which allow the agent to try different responses in an environment without the costs of real-world failures. The robot learning research programme will explore how robotic agents can be designed to perform reliably in real-world situations, through research into new learning strategies, principles for robustiness, and the use of simulations.

ELISE research themes

SECURITY AND PRIVACY

Context

Today's machine learning methods process large amounts of data to identify patterns and make predictions. These interactions between algorithms and data create security vulnerabilities in machine learning systems and contribute to concerns around data privacy, as potentially sensitive data is used and re-used in new ways.

The complexity of AI systems leads to security challenges that must be tackled to make these systems trustworthy. These include cybersecurity risks, such as those from adversarial attacks – security attacks that manipulate data sources with the aim of introducing errors in a system or attempts to contaminate training data to create flaws in models. There is already a large research field focusing on such security risks and defences against them. Research from this community in recent years has provided new defensive approaches, for example through mathematical proofs that show which defence mechanisms can be designed into a system to make it more robust against different forms of security attack.

Another form of security or privacy risk relates to the ability of machine learning systems to effectively manage concerns about data protection. While data use can create benefits for individuals or society, the ability to combine datasets in ways that generate sensitive insights – or the misuse of data for unintended purposes – can lead to new digital harms. Seeking to prevent these harms, the EU has led the world in developing governance frameworks that give individuals rights over how data about them is used, setting a framework for data use while protecting fundamental values.

Data governance practices can help translate these rights into practice, bridging the desire to share data for public benefit – for example, enabling analysis to develop medical therapies for rare diseases – and concerns about protecting privacy and individual rights. Technology can play a role in enacting these governance practices, by integrating privacy and security concerns into the operation of AI systems. Research advances are bringing new opportunities to design machine learning systems in privacy-conscious ways: cryptographic and differential privacy techniques can enable or restrict different forms of analysis, while federated machine learning can create systems in which algorithms access data to learn without the need to first centralise sensitive personal data. Further development of these methods will be necessary, if the field is to respond effectively to policy demands and citizen expectations of AI.

Areas of research interest

Privacy and security by design: advances in technical concepts and methods are creating tools for managing privacy and security concerns. For example, promising approaches to privacy-preserving machine learning include: differential privacy, secure multi-party computation, and privacy-secure federated learning. Further work is required to both progress the technical sophistication of these methods and to ensure these tools work well in practice. This includes research to create demonstrably reliable privacy-preserving systems and translation of these tools and concepts to ensure they are accessible to practitioner communities. In the longer-term, further technological and governance developments could allow individuals to control the use of their personal data while at the same time drawing critically valuable scientific information from aggregate data.

Working with practitioners to manage privacy in practice: in some domains, the data required to develop machine learning systems is sensitive by nature. The resulting data privacy concerns can be managed by both procedural and technical approaches. Collaborations between machine learning and domain experts – including legal specialists, social scientists, affected communities, and relevant industry sectors – are necessary to design governance systems that meet the needs of practitioner communities. In healthcare, for example, data access mechanisms and contracts that govern data use ensure that only appropriately-skilled researchers have access to sensitive information, and that these researchers adhere to relevant standards in data processing. Such data is further protected by restricting access only to those researchers that have been approved to use specialised IT systems. While privacy-preserving approaches that aim to provide access to medical data do exist, given the complexity of this environment, such methods will require further development and user-testing before they can be reliably deployed.

The ELISE approach

ELISE collaborators are working to increase the security of machine learning systems and to create machine learning methods that can help address privacy issues. ELISE research into machine learning methods seeks to increase the robustness of machine learning to adversarial challenges and its collaborations with partner organisations will foster design practices that embed privacy by design in deployed systems.

ELISE research will:

- 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.
- 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.

EXPLAINABLE AND TRANSPARENT AI

Context

The innovations that have supported rapid advances in machine learning research have contributed to the creation of complicated models, algorithms and patterns of data use. In some cases, the complexity of these systems means that their workings can be difficult for even expert users to explain. While not a feature of all machine learning systems, concerns about so-called 'black box' methods are subject to growing attention.

A central tenet of many of the ethics guidelines and principles that have been created by AI research and policy stakeholders is that AI systems should be explainable; that there should be methods by which human users can interrogate how or why a machine learning system has produced a particular output.[21] This demand for explainable AI stems from concerns about human dignity – that machine learning should serve to empower human users and maintain human agency in the face of automation;[22] from the need to create procedures that maintain accountability where decision-making chains may be in-part automated;[23] and from recognition of the important role that explainability plays in creating machine learning systems that work well in practice.[24] Noting the importance of these needs, transparency is one of the EU High Level Group's requirements for trustworthy AI.[25]

This demand for AI systems whose workings can be understood by non-expert users has also spurred a vibrant field of research. The term 'Explainable AI' describes a cluster of AI methods and techniques that aim to enable human users to understand the outputs of an AI-enabled analysis. These tools for explainable AI operate at different levels:

- some seek to identify the key factors contributing to a particular decision or output from a system using machine learning;
- some focus on explaining how a machine learning model is working, with the aim of identifying the factors that influence its performance;
- others look at decision-making pathways involving different automated and non-automated components, seeking to ensure accountability through this pipeline.

These explanations can serve a variety of different purposes. Explanations can help: developers to examine how a system is working, and to debug or improve it; practitioners to understand the strengths and limitations of the system they are working with, and anticipate potential issues in deployment; and those subject to algorithmically-enabled decisions to challenge the outcome of a decision-making process. They can also help users audit a prediction or decision if some irregularities appear; monitor or test for safety standards; and build confidence with a prediction or decision.

The reason for providing an explanation and the type of explanation required are

21 A summary of these principles is available in Table 2 of the Royal Society (2019) Explainable AI policy brief, available at: https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf

22 European Commission (2019) Ethics guidelines for trustworthy AI, available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

23 Doshi-Velez et al. (2017) Accountability of AI under the law: role of explanation, available at: https://arxiv.org/pdf/1711.01134.pdf

24 Bhatt, U. et al.(2020) Explainable AI in deployment, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020, pp. 648–657, https://doi.org/10.1145/3351095.3375624

25 High Level Group on AI (2018) Requirements of trustworthy AI, available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1#Accountability contextual. Questions about what technical approaches to pursue in developing explainable AI can therefore only be resolved within the context of an application, understanding the needs of different users. In climate science, for example, researchers using machine learning to understand earth systems typically look for machine learning models that are encoded with well-accepted physical laws and that demonstrably adhere to these laws, with the aim of ensuring the predictions from such systems are reliable. Researchers may also seek machine learning methods that can point to underlying causal structures – or the reasons behind research outputs – in order to use such systems to advance scientific knowledge on a subject. Design of explainable AI therefore requires close collaboration between machine learning experts and affected user communities to design appropriate forms of explainability and transparency.

Areas of research interest

Advancing explainable AI tools and methods: different technical approaches can support different forms of explainability. Progress across these different methods will be needed to create a suite of tools that meet the needs of different users in different circumstances. For example, research advances can create systems that:

- are inherently interpretable, based on known physical laws;
- are inherently interpretable, based on other structured methods, for example prototype-based systems;[26]
- visualise which inputs help contribute to particular types of output or prediction;
- test how well a model responds to different types of perturbation;
- use interpretable surrogate models to approximate the outputs from 'black box' systems.

Supporting practitioners to implement explainable AI methods that meet stakeholder needs: explainable AI has a huge range of potential applications – from helping citizens challenge algorithmic decision-making processes to helping scientists better understand the phenomena they study. The variety of explainable AI methods that exist today can be deployed in different ways to serve the needs of different stakeholders. However, surveys of explainable AI in deployment suggest that this diversity of approaches is not well-reflected in the type of explainability being designed into real-world AI systems. The explainable AI implemented today largely serves to help AI designers de-bug machine learning models.[27] If explainable AI methods are to meaningfully contribute to efforts to ensure accountability, transparency and fairness in the use of machine learning, further effort is needed to help practitioners to develop and implement forms of explainability that better serve the needs of diverse stakeholders. This in turn requires better understandings of the needs of different stakeholders, support for those developing AI to co-design explainable AI systems with relevant stakeholder communities, and wider accessibility of explainable AI tools and methods.

26 For examples of prototype-based methods, see: Angelov, P. and Soares, E. (2020) Towards Deep Machine Reasoning: a Prototype-based Deep Neural Network with Decision Tree Inference, IEEE International Conference on Systems, Man, and Cybernetics (SMC2020), pp. 2092–2099, DOI: 10.1109/SMC42975.2020.9282812

27 Bhatt, U. et al.(2020) Explainable AI in deployment, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020, pp. 648–657, https://doi.org/10.1145/3351095.3375624

The ELISE approach

ELISE research programmes will create explainable AI tools that serve the needs of diverse user communities. Through close collaboration with partners working in application domains, ELISE will explore how the demand for explainable AI varies across different contexts and what type of support practitioners need to implement explainable AI methods that meet their needs.

ELISE research will:

- 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.

TRUSTWORTHINESS AND AI CERTIFICATION

Context

Development of trustworthy AI – AI that is demonstrably explainable, competent and reliable[28] – is central to the variety of policy programmes, research initiatives, and organisational AI ethics initiatives that have emerged in recent years. If AI systems are to be considered trustworthy, their users will require confidence that these systems will perform the function they are designed to do, safely and effectively.

To meet this demand, organisations across Europe are investigating the role that certification schemes could play in guaranteeing AI performance standards. These schemes include regulations, accreditations and certificates to help demonstrate the performance of technical systems. Emerging policy approaches include risk-based assessment frameworks;[29] kitemarks

28 For further discussion of these principles, see Onora O'Neil's Reith Lectures, which are available from: www.bbc.co.uk/programmes/p00ghvd8

29 See, for example, work by Germany's Data Ethics Commission, available at: www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html or quality-assurance schemes;[30] sector-specific standards or evaluation processes;[31] and voluntary certification schemes.[32]

In addition to these sectoral or national interventions, policy developments within the EU during 2021 are expected to create a risk-based regulatory framework for AI applications. This framework will seek to ensure that machine learning systems in safety-critical domains are appropriately regulated, through the creation of assessments or processes for testing, inspection and certification of algorithms and data sets used in their development.[33]

With a significant portion of Europe's economy being based on building machines – from automotive to industrial robotics and manufacturing – the ability to certify the safety integrity level offered by machine learning systems has potential for great impact. Technological advances and technology-focused certification methods can play a role in supporting such certification schemes. Some techniques exist for certifying and performance-guaranteeing AI in specific settings; such techniques are critical for the use of AI in safety-critical domains. However, further work is required to improve the effectiveness of these local schemes, while comprehensive solutions remain in the early stages of development.

Areas of research interest

Certifying or guaranteeing the performance of AI systems: machine learning already benefits from well-developed theory surrounding the type of problems on which it can perform well and its standard of performance on those challenges. However, the ability to certify machine learning systems – guaranteeing that they will not fail in deployment or that they are able to identify when their predictions may be wrong – is less well-developed at present. Such certifications will be necessary for practitioners to confidently deploy machine learning systems, particularly in domains where failure comes with significant personal or societal costs. In some of these domains, there already exist processes that could be adapted to help certify machine learning interventions. In healthcare, for example, certification of medical devices and decision-support systems has been pursued through international clinical trials. However, few such studies of machine learning systems exist; pursuing this work would require close collaboration between physicians and machine learning experts. In other domains, for example autonomous driving, further work is required to better understand the limitations of such certification schemes in ensuring safety. Developing these approaches to certification requires methodological advances that allow researchers to analyse the robustness of machine-learned systems, identify relevant benchmarking standards, and understand different potential failure modes (as well as how to fail safely, based on an analysis of the consequences).

Verifying and validating machine learning systems: improved mechanisms for verification and validation of machine learning systems would complement certification methods. Machine learning research has already established

30 For example, Denmark's Data Ethics Seal (see: https://oecd.ai/wonk/an-independent-council-and-seal-of-approval-among-denmarks-measures-to-promote-the-ethical-use-of-data) and the European Economic and Social Committee's certification proposals (see: www.eesc.europa.eu/en/news-media/news/eesc-proposes-introducing-eu-certification-trusted-ai-products).

31 Council of Europe (2020) Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary, available at: https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4

32 For further information, see Malta's AI Strategy, available at: https://malta.ai/wp-content/uploads/2019/11/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf

33 European Commission (2019) White Paper on AI – A European approach to excellence and trust, available at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

methods for evaluating performance – test-set or cross-validation methods, for example. However, these existing methods for verification and validation have been largely oriented towards static systems – proving that a method developed 'in the lab' performs to the same standards under expected conditions 'on the shop floor' – or for less complex machine learning models. While valuable in assessing machine learning performance in conditions where the deployment environment is well-characterised or in verifying performance in the face of known system issues (such as likely faults), these approaches are not designed to validate performance in dynamic environments. To work in environments where machine learning systems are expected to encounter unforeseen issues, which might range from security breaches to unexpected human interactions, new approaches to verification and validation will be needed.

Improving the robustness of machine learning in deployment: when deployed, machine learning systems will likely encounter a range of situations that were not represented in the datasets on which they were trained. To be robust in the face of these unexpected situations, methods are needed that allow these systems to self-diagnose when they are faced with circumstances that fall outside the bounds of their training data, to operate effectively in the face of adversarial manipulation, and to recognise where undesirable biases might be affecting their accuracy. Such robustness can be developed through a combination of theoretical advances, algorithmic innovations, and approaches to deployment. For example:

- machine learning theory can help develop principles for robustness or to understand the types of error that a system can produce;
- methods such as meta-learning or AutoML[34] can create machine learning systems that identify appropriate actions in new circumstances;
- research into benchmarks can set standards for performance;
- further developments of tools for causal inference can help researchers better understand likely system behaviour in different circumstances; and
- collaborations with domain experts can help identify likely failure modes.

The ELISE approach

By providing new insights, theories, and methodologies for evaluating machine learning systems, ELISE will help create quality guarantees that can be used in example domains. Recognising the significance of this area of research, ELISE will pursue the technical challenge of robustness as both a technical problem and a cross-cutting design principle.

34 The use of machine learning to create more effective machine learning systems.

ELISE research will:

- advance the technical sophistication of core machine learning methods, including computer vision, Natural Language Processing, and information retrieval.
- improve understanding of the principles and techniques that can make machine learning robust.
- 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.

AI INTEGRATION ACROSS SYSTEMS

Context

Machine learning is today both an active area of academic research and an applied field. Applications across industry sectors offer the opportunity to streamline business processes, to help prevent machinery failures and to tailor services to different consumer groups. The economic and social benefits that follow are potentially significant, from better healthcare to safer transport, a thriving manufacturing sector and more sustainable energy.[35]

Despite these potential benefits, there persists an implementation gap between the capabilities of machine learning methods 'in the lab' and experiences of using machine learning 'in the wild'. Surveys of organisations seeking to use machine learning in their day-to-day operations repeatedly report challenges and delays in deployment – arising, for example, from bias in data, lack of organisational capability, or high costs of use – with a high portion of deployments failing.[36] Further progress will require research that investigates concrete use cases, engaging with a wide range of users and stakeholders to identify barriers to deployment and strategies to overcome them.

From a technical perspective, achieving robust and reliable AI will require design

35 European Commission (2020) Execllence and trust in artificial intelligence, available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence

36 See, for example, Wiggers, K. (2019) IDC: For 1 in 4 companies, half of all AI projects fail, 2019, available at https://venturebeat.com/2019/07/08/idc-for-1-in-4-companies-half-of-all-ai-projects-fail approaches that bring together the data-driven world of machine learning with understandings of the wider physical world and dynamics of real-world systems, to create systems that can respond appropriately to unforeseen or novel circumstances. To achieve this goal, further work is needed: to develop methods to understand and anticipate system performance in a deployed context; to calibrate uncertainties in machine learning systems; to advance the suite of machine learning algorithms that can demonstrably perform safely and reliably in real-world contexts; and to address issues of data availability. These research challenges create both opportunities for foundational research advances and for collaborations that deploy machine learning to tackle real-world challenges.

Areas of research interest

Improving performance in deployment: when deployed in practice, machine learning algorithms usually form part of a larger-scale, algorithmically-enabled decision-making system. Such systems use multiple machine learning sub-components alongside other automated and non-automated processes. Their performance depends on the effectiveness of each machine learning sub-component, but also on the ways in which the sub-units of the system interact, both with each other and their environment. Each sub-unit needs to be designed using an understanding of the challenges of deployment, with the ability to continually monitor performance in deployment for factors such as accuracy, bias, fairness, uncertainty and consistency. It must also be capable of detecting when data changes or the wider system suggests signs of failure, for example by monitoring the prediction quality of the machine learning sub-unit, or through advanced methods for calibrating uncertainty. AI-assisted design and monitoring of these deployed systems can help maintain their safe and accurate operation in real-world environments, but will require new approaches to system design, software, and organisational culture, if they are to be developed and adopted across organisations.

Designing effective simulators and emulators: simulations offer a tool to bridge the gap between data-enabled machine learning and real-world systems, helping decision-makers to explore the consequences of different interventions or to explore the design approaches that help manage different uncertainties. Wider use of such simulations could help increase the reliability of the machine learning that is used in real-world systems. However, many simulators are computationally demanding to run, limiting their accessibility or usability. These demands can be reduced through the use of statistical emulators – a system that uses a statistical model to reconstruct a simulation and assist in uncertainty quantification, inform decisions about how to use simulations, and contribute to the interpretability of AI systems.[37] Achieving this goal requires high quality tools for simulation and emulation – tools that can model real-world interactions involving physical systems, AI systems and human users – based on technical innovations in the mathematical composition of emulator models and sophisticated understandings of the ways in which uncertainty propagates through a machine learning system.

Understanding interactions with human users: AI-enabled systems are often deployed in support of human decision-making, or in environments where they interact with human users. To fulfil these functions safely and effectively, AI systems will require models of human users that will enable them to anticipate different forms of interaction in real-world environments, alongside careful design

37 Lawrence, N. (2019) Meta-modelling and deploying ML software, available at: https://inverseprobability.com/talks/notes/meta-modelling-and-deploying-ml-software.html

of the interfaces between human and machine at those points of interaction. Further advances in the study of human-computer interaction and the inclusion of models of human users in advanced simulations could help create AI systems that better serve human needs.

Combining data-driven and structural insights: Another approach to designing real-world AI seeks to integrate existing knowledge, be it physical laws, biological or physiological knowledge, or social guidelines, into the AI system. The advantage of such structured networks is that existing knowledge does not have to be learned from training data and system designers can provide guarantees that the network cannot operate beyond the laws embedded in its ways of operating. Hybrid systems – that use machine learning alongside, for example, reasoning-based approaches – could also help achieve this goal. For physical laws of geometric invariance, for example, it is already clear how to achieve structured networks, and there are first attempts to integrate higher level knowledge into the network. Further progress using multi-task and representation learning as well as causal inference could play a role in integrating high-level knowledge in large-scale systems.

Increasing data availability: access to high-quality data underpins the success of any machine learning system. However, practitioners creating machine learning systems face a range of data access issues: poor data quality, incorrect modelling assumptions, or re-deployments of models outside the parameters for which they are trained can all contribute to the failure of machine learning systems to perform as anticipated.[38] In some domains, such implementation issues are compounded by concerns around data governance; in healthcare, for example, the ability to create machine learning systems relies on access to potentially sensitive personal data, requiring governance systems that enable data use while protecting individual rights and respecting stakeholder concerns. A range of national and international data policies are seeking to increase access to data across sectors while supporting individuals and companies to maintain control over the ways in which their data is used.[39] In pursuit of this aim, a variety of technical, legal and policy tools can be used, but further work is needed to understand the role that each can play in enabling data access while protecting individual rights.

The ELISE approach

ELISE will develop a set of technical tools and research practices that can integrate AI into real-world applications, creating AI systems that are reliable, maintainable and interpretable. By working closely with industry partners, ELISE researchers will create machine learning systems that can be used to tackle the challenges facing industry today. By advancing research into the safety and reliability of machine learning, ELISE will lay the foundation for further deployment of these systems in the future.

38 Paleyes, A. et al. (2021) Challenges in deploying machine learning: a survey of case studies, available at: https://arxiv.org/pdf/2011.09926.pdf

39 European Commission (2020) European Data Strategy, available at: https://ec.europa.eu/digital-single-market/en/european-strategy-data

ELISE research will:

- 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.

AI ETHICS AND SOCIETAL IMPACT

Context

Machine learning is already part of daily life, mediating a range of digital activities. As machine learning systems influence people's interactions and habits, action is needed to develop human-centric AI, by designing AI tools that uphold fundamental rights and by taking action to ensure that AI serves all in society.

Recent years have already brought examples of harms that mis-use of AI can bring, particularly in the marginalisation of vulnerable groups. Existing social inequalities can be coded into machine learning models as a result of biases embedded in the data used to train these systems or choices made during their design. If action is not taken to counter these biases, the outputs of these models can unfairly discriminate against or exclude certain groups, often those already marginalised in society.

Such action requires efforts to understand the technical and social elements of machine learning, including how data has been gathered, how data is used in a machine learning model, and what happens when the resulting system is deployed as one element in a complex socio-technical environment. Both foundational, theoretical work in machine learning and fresh approaches to co-design of AI systems will be necessary, through research that brings together stakeholders across government, industry, civil society and affected communities.

While taking action to avoid these harms, there is a pressing need to deploy AI to help tackle major social and scientific challenges. Across healthcare, transport, the public sector, and more, machine learning could play a role in creating products and services that improve health, wealth and wellbeing across society. For all these applications, there is a need for continuing dialogue between AI research disciplines, policymakers and citizens to create shared understandings of society's desired future relationships with AI. ELISE research will contribute to these dialogues by bringing technical expertise to policy discussions and by fostering human-centric AI research that prioritises collaborations between experts in machine learning, the law, policy, and the ethical implications of science in society.

Areas of research interest

Advancing foundational research to create human-centric AI: if AI is to effectively serve human needs, AI technologies will require certain core characteristics: the ability to provide explanations, or insights into causality, and account for concerns about fairness; robustness in dealing with such issues when deployed; and alignment with rights around data use and privacy. Advances across these areas can help create human-centric AI, which reflects the needs and values of the individuals, communities and societies it serves and which contributes to effective governance of AI technologies.

Putting ethical AI principles into practice: building on advances in the theoretical foundations of ethical AI, further action is needed to develop mechanisms for the design, control and application of AI that connect social values to technological practices. These mechanisms will differ across contexts, depending on stakeholder need and the values and risks in play. Their design must draw from multidisciplinary collaborations to understand how different stakeholders view concepts such as fairness, how humans and machines interact in practice, and how to evaluate the operation of machine learning systems.

Designing governance frameworks to support trustworthy AI: these technical approaches to minimising the potential harms of AI form one part of a wider AI governance ecosystem. AI governance spans a range of activities – from regulations that direct the implementation of AI systems in domains such as healthcare, to legislation that creates data rights and provides means to enact them, to codes of ethics and professional standards that influence how organisations adopt machine learning technologies. Which combination of governance interventions is required for any specific machine learning system will depend on the context of its deployment. For many applications of machine learning, data governance frameworks play a particularly important role in contributing to trustworthy data use and human-centric AI. ELISE members will be actively engaged in contributing to the development of such governance frameworks.

Developing AI applications in areas of societal interest: to further advance the use of AI for wider public benefit, ELISE's research programmes will investigate application domains in which the use of AI could bring significant social or environment benefit. A research programme on the use of AI for healthcare will develop systems to support medical data analysis, genomic medicine, biomedical imaging, drug discovery and clinical support; and a programme on earth and climate sciences will advance modelling and understanding of the earth system to unlock insights that can help society tackle climate change.

The ELISE approach

ELISE will integrate AI ethics into foundational AI research, supporting close collaboration with law and ethics experts and promoting best practices in responsible AI. Ethical issues will be taken into account in and by all of the research programmes and cross-cutting themes. ELISE will work to advance the development of technological methods that can help protect fundamental rights and of fresh approaches to data governance. ELISE's work seeks to widen access to machine learning, enabling all in society to make use of this technology and ensuring that it is developed in ways that meet the needs of diverse user communities.

Some of today's most powerful machine learning methods are compute- and energy-intensive, raising questions about the energy consumption of AI and how best to manage these resources. In developing these machine learning theories and methods, ELISE researchers are grappling with the parallel challenge of managing increasing demands for compute power from machine learning systems.

ELISE will:

- 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 hardwares.

Implementing the Strategic Research Agenda: research programmes

ELISE and ELLIS research programmes advance the areas of interest set out in the cross-cutting themes, connecting aspirations for the characteristics that AI should demonstrate to the research advances needed to create these characteristics. Together, these research programmes map out a pathway to achieving trustworthy AI – AI that reflects European values and benefits all in society.

Table 2 summarises current areas of ELISE research. Lead researchers for each programme are listed in Annex 1.

Table 2. **ELISE and ELLIS research programmes: aims and context**

Programme	Aim	Context
Theory, algorithms and computations of modern learning systems	Advance the theoretical underpinnings and algorithmic capabilities of machine learning, creating more reliable, efficient and usable machine learning systems.	The creation of contemporary machine learning systems requires significant resources – the time and energy of highly-trained experts; access to data; and compute power. While some machine learning solutions now rival human performance in restricted tasks, these typically use significantly more energy than humans performing the same tasks and rely on careful design and implementation to perform well. Advances in machine learning theory can help researchers better understand which machine learning approaches work well for different challenges, which methods do not, and the reasons behind these variations in performance. This understanding can be used to enable easier development of future machine learning systems – reducing the barriers to creating these systems by requiring less expert input – and to produce advances in machine learning design that improve the performance in these systems. Performance improvements could lead to increased speed, more accurate predictions or estimates of uncertainty, and overall greater robustness. In turn, these can contribute to guarantees of the performance of machine learning algorithms, which will be critical in enabling their deployment in a wide range of applications.
Machine learning for health	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.	The application of AI to healthcare challenges could lead to research breakthroughs and novel therapies that could significantly improve society's health and wellbeing. Data from a variety of sources – including complex biomedical data – can be used to help monitor patient health and design treatments to improve healthcare outcomes. This research programme seeks to build bridges between advanced AI research and applied fields across all areas of biomedicine and public health, with the aim of creating machine learning systems that can support a step-change in healthcare provision.
Quantum and physics-based machine learning	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.	Increased computational power and data availability, as well as algorithmic advances, have led to impressive machine learning applications in many areas such as computer vision, pattern recognition, robotics and AI. At the same time, conventional CMOS technology is reaching its physical limits and the energy consumption of computing is increasing in ways that could create tensions between the desire to develop new machine learning systems and the imperative to reduce overall resource consumption. There is therefore a great need to design novel computing paradigms that face these challenges. Concepts in quantum and statistical physics offer insights into how to develop such energy-efficient hardware solutions and machine learning algorithms.

Programme	Aim	Context
Geometric deep learning	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.	Deep learning is a powerful tool for processing data to detect signals and could be deployed to extract insights from a range of different data types, including those that do not confirm to Euclidean spaces (for example, data on graphs, point clouds, data on spheres, or manifolds). The development of deep learning methods that perform well on non-Euclidean data could unlock further applications in a range of domains, including a variety of scientific disciplines. For example, recent advances in the study of protein-folding through the AlphaFold project could significantly improve understandings of the dynamics of disease and its treatment; similar advances may be possible in materials and drug design, or in any field with graph-structured data.
Machine learning for earth and climate sciences	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.	Tackling climate change will require insights into the ways in which human actions are changing the Earth's atmosphere and the impact of these changes on local conditions. Machine learning can be deployed in earth and climate sciences to better understand the interactions between people and the planet. Recent years have demonstrated progress in the use of machine learning in some key areas – weather forecasting and monitoring land use, for example – as well as the potential for further advances in the use of these techniques. New generation, machine learning-enabled climate models could help advance research into earth and climate sciences, combining data-enabled insights with knowledge of physical laws and environmental interactions from these disciplines. Areas of potential interest range from improving the performance of solar panels to optimising industrial processes to support a shift to a green economy.
Natural intelligence	Advance the science of artificial intelligence by better understanding the intelligent behaviour of living systems and how this emerges.	Different forms of intelligence exist across the natural world. Studying the intelligent behaviour of animals and humans can give insights into how the field of artificial intelligence could develop and the tasks that intelligence machines could help solve. Such research could, for example, identify evolutionarily-tailored inductive biases that can be useful in the development of AI and help understand the modes of failure that can lead to dysfunction in AI-enabled systems. It could further lay foundations for advances in computational psychiatry, with different classifications of disease and opportunities for improving prognosis and treatment selection.

Programme	Aim	Context
Human-centric machine learning	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.	As AI systems pervade different aspects of daily life, there is a pressing need to ensure that such systems are well-aligned with human needs. Current machine learning methods and systems are, however, rigid and not human-oriented. Human-centric machine learning systems would reflect the rights or values that society prioritises – for example, fairness, accountability, privacy – and function well for all in society. Achieving this goal will require foundational and theoretical work in machine learning, as well approaches to AI development that involve a wide range of stakeholders, including governmental, regulatory, legal, industry and civil society players, and collaborations across research domains, including ethics, law, computer science, and human-computer interaction.
Robust machine learning	Understand the principles and develop the techniques for machine learning that reliably performs well.	As machine learning technologies are deployed progressively more across sciences and the real world, there is a pressing need for tools and methods to guarantee that these systems perform well. These would demonstrate, for example, that the system performs in settings different to those encountered during training, that it can continue to perform well under conditions of adversarial manipulation, and that are capable of dealing with unbalanced, messy and heterogeneous data. Such capabilities are particularly important in high-stakes domains, such as healthcare, where theoretical advances in understandings of the stability of machine learning models could increase the confidence of different stakeholders in deploying these models in practice.
Machine learning and computer vision	Build bridges between classical algorithms and machine learning to unlock further advances in computer vision.	The field of computer vision has been revolutionised by machine learning, in particular deep learning. Recent years have brought improvements to the state-of-the-art performance on many problems that have been studied for decades, as a result of the use of artificial neural networks to tackle these challenges. Such approaches draw from insights generated from the study of classical algorithms, including how to design datasets, cost-functions and architectures. Better connections between the study of classical algorithms and modern machine learning could unlock further advances. In turn, this would enable a range of applications – from autonomous vehicles to video search to robotic navigation – and address important challenges around the trustworthiness and explainability of machine learning methods such as deep learning.

Programme	Aim	Context
Natural Language Processing	Build systems for general-purpose natural language understanding and generation.	Natural language processing enables automated systems to process and extract insights from human language, recognising the nuances of its use in different contexts. Advances in natural language processing already support the operation of digital assistants, machine translation services, and information extraction from different documents. Further advances could help democratise access to AI – enabling easier interaction with digital services across groups in society.
Multimedia/ multimodal information	Enable meaningful, accurate, fast, and scalable identification of semantic content from existing and emerging media, including cross and multi-modal media, and develop advanced methods to allow content generation, summarisation, decomposition, and re-composition.	Humans interact and communicate with each other and the environment in a variety of ways. The ability to sense, decipher and understand information and communication is central to these interactions. As these interactions move into the digital realm, new techniques are needed to replicate these abilities, from having information and communications in digital formats that go beyond storing and transmitting over networks to building better understanding of the meanings in these digital communications, including videos, images, text, audio and other media. Such techniques would improve existing services based on these media types and open the possibility of new applications using this content.
Robot learning: closing the reality gap	Create robotic systems that can interact intelligently with the world around them.	Despite recent advances in robot learning, there remain important challenges to address if these systems are to operate effectively in real-world contexts. Key questions include: how should the robot move? How to act? How to interact? How can sensorimotor behaviour be improved by machine learning approaches? This programme will pursue research that takes out of research labs into the real world, for example learning assembly skills.
Interactive Learning and Interventional Representations	Explore the role of causal modelling in bridging the gap between observational and interventional learning and understand the principles underlying interactive learning systems.	Interactive learning and interventional representations enable the development of agents that interact with the environment (including humans) in a robust, adaptive, and dependable manner. This interaction is the key to using machine learning for solving a much larger set of practical problems and deploying systems that can interact effectively in real-world contexts.

Programme	**Aim**	**Context**
Information retrieval	Improve the ability of information systems to understand human behaviour and produce appropriate answers in response.	The recent introduction of deep language models has brought tremendous progress on several information retrieval tasks, such as search, recommendations, and conversational assistance. However, further advances would improve their performance. For example, current systems lack deep understanding of their inputs and the context of this input. This lack of contextual understanding can contribute to spread of incorrect information, in particular that generated by machines instead of humans, and result in answers being provided for incorrect reasons. Future information retrieval systems need to significantly improve their understanding of human information interaction behaviour and act based on this understanding. This requires substantial progress in representation learning of human behaviour; semantic matching methods based on very limited volumes of interaction data; counterfactual learning to rank methods that are able to learn user preferences, in an unbiased manner, from interaction data collected using a variety of information retrieval environments; and mixed-initiative scenarios in which information retrieval systems need to learn to ask questions that are minimally intrusive but maximally informative to the system. Such advances are required across information types – not only texts – and in languages other than English.
Explainability and fairness in data mining	Address concerns about explainability and fairness in data mining, by developing new tools and methods.	The past decades have witnessed significant progress in the development of intelligent methods for analysing large volumes of data and performing challenging tasks with success that matches, or in many cases exceeds, the performance of human experts. However, these methods are largely non-interpretable. In addition, they suffer from data biases and produce models that do not meet expected standards for fairness. The research community has recognised these issues and a lot of recent work has been devoted to design models that are fair, accountable, and transparent. Most of the work, however, focuses on supervised learning. This ELISE programme will address the challenges of explainability and fairness in unsupervised learning.
Symbolic machine learning	Combine symbolic and data-driven AI methods to develop more effective AI systems.	Early waves of AI development were dominated by symbolic approaches, as researchers sought to formalise complete human knowledge in formal representations of logic. Recent years have brought data-enabled AI methods to the fore. Further progress in each subdomain can be achieved by learned the lessons of their respective approaches.

4. Building the European AI ecosystem

In advancing this agenda, ELISE will collaborate closely with ELLIS. ELISE and ELLIS share the aim of developing AI for the benefit of society and the economy. ELISE and ELLIS also share the vision of a future European AI ecosystem in which strong local centres of research excellence – that each bring different strengths – work closely together and present a coordinated European identity, which is maximally attractive to talent and investment.

The foundation of work by ELISE and ELLIS is excellence in research. Excellence in machine learning research has been one of the most important drivers of innovation over the last decade and continuing excellence in this domain is vital for future economic growth. If Europe is to remain at the forefront of this growth in machine learning and AI, investment in world-class science and world-class researchers is vital.

The concentration of knowledge and economic drive in the ELISE and ELLIS networks means they are strongly positioned to lead AI research in Europe and to build European research collaborations. These networks can be leveraged for most value across Europe through the creation of a network of centres of excellence, each with their own regional specialisation and ecosystem of applications. In this network of institutes, each regional hub will be able to take advantage of top-class research in the region, its innovation ecosystem, and specific economic structure, while maintaining strong links with other members of the European network.

In addition to supporting high-quality, independent academic research, common features of all ELISE and ELLIS activities include:

- engagement with industry to translate AI knowledge into practice;
- collaborations to spread AI expertise across research domains and countries;
- activities to develop the next generation of European AI talent.

Engaging industry to translate AI knowledge into practice

Increasing innovation is a priority area for action under the ELISE programme. By working with industrial collaborators on a series of use cases, ELISE research programmes will generate insights into the challenges that arise in real-world AI applications, designing AI methods that can be deployed safely and effectively. Table 3 summarises the use cases that will be developed in the first wave of ELISE research activity, drawing from the interests of current research programmes and expertise of ELISE collaborators.

Table 3. **ELISE 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 machine learning techniques which are able to explain binary classifier decisions.
Robust ML Benchmark and Challenge	DeepMind	Robust Machine Learning	Suite of benchmarks for robustness of machine learning methods and a real world challenge to foster innovations in robust machine learning.
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 Multi Modal 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.
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 optimisation using reinforcement learning.
Experimental Environment for Real World Reinforcement Learning	Zalando	Interactive Learning and Interventional Representations	Optimal business decision making in a given situation using reinforcement learning.

In addition to these current collaborations, ELISE will further support innovation acceleration activities through a programme of SME engagement, which will coordinate open calls for support in developing novel AI applications and host incubators that support SMEs seeking to make use of AI methods. This programme will lead to further use cases through which ELISE researchers can increase the uptake of their research and help spread AI expertise across the European innovation community. This work will be further bolstered by ELISE's dedicated knowledge transfer programme, which together with the industrial PhD and Postdoctoral programmes allows effective dissemination of results and methods.

Collaborating to spread AI expertise across research domains

There is a high demand for machine learning methods across industry sectors and in a range of other research disciplines. To help address a bottleneck in the availability of such expertise, ELISE will establish connections with leading European initiatives that help share insights and expertise from its networks. ELISE has already identified a set of complementary initiatives where it will be developing joint activities in the near-term. These include:

AI4EU: To ensure that ELISE's work is shared widely across the European innovation ecosystem, ELISE will collaborate with AI4EU, implementing five selected use cases as external assets on this platform.

LifeTime: Is a European consortium of >90 research institutions extensively building on single-cell technologies to revolutionise healthcare by mapping, understanding, and targeting human cells during disease. The ELISE Health program will interface with this consortium through its research leads, facilitating joint workshops and other collaborations. The collaboration between ELISE and LifeTime will foster novel computational advances to analyse large-scale high-dimensional data (for example, molecular profiles of cells from large disease cohorts), such as methods for dimensionality reduction, causal inference of regulatory dependencies or the integration of molecular data and medical data from health records to predict clinical endpoints. LifeTime is building a collaboration platform with >70 companies across sectors. This will enable identification of key industry challenges for AI applications to earlier diagnosis and better treatment, trigger new scientific questions and foster take-up of ELISE scientific advances.

Humane AI: A human-centric approach to AI development is core to several ELISE research programmes. Such an approach requires both solving core technical AI challenges and interdisciplinary collaboration. Such collaboration will be enabled through a continuing relationship with the Humane AI initiative facilitated by ELISE partner K4All and coordinated by ELISE research leads.

Robotics: Robotics is one of the key application fields of machine learning and modern AI. Coordination between the ELISE and CENTRIS programmes will identify areas for collaboration, which may include the use of multimodal datasets, control policies for soft robots, data-efficient representations and resource-aware computing, and safety in human-robot collaboration.

Hardware and Edge AI: Advances in infrastructure, understood broadly, have been one of the main enablers of the current machine learning revolution, and maximising the information transfer between developers of the next generations of hardware and algorithms may radically increase the opportunities for the next breakthroughs. Edge AI is a new application domain requiring novel AI solutions. ELISE has already identified complementarity with the AI@Edge research programme and identified persons to coordinate collaboration.

International partnerships: For international collaboration and exchange of best practices, in addition to the extensive collaboration networks of the individual partners, ELLIS has built collaboration partnerships with high-profile organisations outside Europe, notably CIFAR, RIKEN, and Data Science Africa.

Supporting the next generation of AI talent in Europe

The ability to attract and retain top AI talent in Europe is crucial both to the effectiveness of the ELISE research agenda and the success of European AI more broadly. In support of this goal, ELISE and ELLIS pan-European PhD and postdoctoral researcher programmes will seek to attract top AI early career research talent in Europe and prepare the next generation of talent for European research and industry.[40] Already a recognised brand, this programme will function as an entry point for early career researchers seeking to work at the cutting edge of machine learning in Europe. By embedding industrial collaboration in its ways of working, ELISE and ELLIS will encourage early career researchers to undertake internships, short projects and joint workshops with companies and start-ups across Europe. A travel programme across the network will promote mobility between ELISE sites, further enabling the spread of expertise and skills across Europe.

40 For further information, see: www.elise-ai.eu/work/postdoc

5. Advancing AI: the importance of access to compute, skills and data

Excellent research requires access to high-quality resources and talent.[41] Important enablers of the long-term success and sustainability of initiatives such as ELISE include:

Investments in data and compute infrastructure

Modern AI methods require training on large datasets and improving upon the state-of-the-art methods requires extensive experimentation. These in turn create demands for access to major computing capacity. While AI groups across Europe are producing excellent research and publishing at top conferences, computing capacity is becoming a limiting factor in their impact, contributing to a declining share of top research publications in the field coming from Europe. Further investments in high performance computing suitable for machine learning will be necessary to maintain European leadership in this area.

Initiatives to build skills and attract research talent

Central to the success of European AI will be the ability to attract and retain leading research talent, in the face of strong demand for such talent across industry sectors. ELLIS is already supporting early career researchers, through its PhD scheme that attracts excellent students to study in labs across Europe and through enrichment activities – summer schools, workshops and exchanges – that help foster a sense of European community. Beyond these schemes, further efforts are needed at all levels to equip citizens with the skills they need to use AI, build communities of practice across sectors, and create attractive career pipelines for AI researchers working in Europe.

Access to data

The machine learning methods at the core of ELISE's work require access to high-quality data. In seeking to promote a thriving European data economy, the EU's data strategy makes the case for further investments in the development of data access frameworks and governance mechanisms that will enable data use across sectors. Such policy measures will be critical to nurture the development of AI in Europe.

41 For further information on this topic, see https://ellis.eu/news/leading-european-ai-scientists-issue-a-call-to-action

6. Developing the Strategic Research Agenda

Europe already benefits from being home to some of the world's leading AI research groups. ELISE and ELLIS are building from these local clusters to create a European 'critical mass' of AI excellence, through an ecosystem that enables researchers and students in institutions across Europe to benefit from each others' expertise.

ELISE and ELLIS seek to create an environment that fosters fundamental research, ground-breaking ideas, and excellence in machine learning. Together, they create a beacon of European AI excellence that can help maintain European leadership in AI amidst intense global competitionfor AI technological superiority.

ELISE will convene top European talent in AI to tackle the issues of social and technological importance set out in this Strategic Research Agenda, combining academic independence to pursue issues of pressing social and scientific interest with industry engagement to secure impact for this work. Research carried out by ELISE programmes will drive a new generation of AI technologies, and alumni from these programmes will lead a new wave of AI research and adoption. The success of this work will be demonstrated through high-impact research papers and the adoption of ELISE-generated methods and tools by industry.

As AI research continues to advance, understandings of areas of opportunity and concern in relation to AI evolve. Reflecting the needs and interests of the AI community, ELISE will continue to update this Strategic Research Agenda throughout its lifetime.

ANNEX 1

Current ELISE and ELLIS research programmes

ELISE and ELLIS currently support 16 research programmes, on the following topics:

1. Geometric deep learning: graph, group and gauge convolutions: Max Welling (University of Amsterdam, Qualcomm AI Research) and Michael Bronstein (Imperial College London).

2. Robust machine learning: 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).

3. Interactive learning and interventional representations: Nicolò Cesa-Bianchi (Università degli Studi di Milano), Andreas Krause (ETH Zürich), Bernhard Schölkopf (MPI-IS Tübingen).

4. Machine learning and computer vision: Cordelia Schmid (INRIA), Yair Weiss (Hebrew University), Bernt Schiele (MPI Informatics).

5. Robot learning – closing the reality gap: Aude Billard (EPFL), Jan Peters (TU Darmstadt), Tamim Asfour (Karlsruhe Institute of Technology).

6. Human-centric machine learning: Nuria Oliver (DataPop Alliance, Royal Academy of Engineering, University of Alicante), Plamen Angelov (Lancaster University), Adrian Weller (Alan Turing Institute).

7. Theory, algorithms and computations of modern learning systems: Francis Bach (INRIA), Philipp Hennig (University Tübingen/MPI Tübingen), Lorenzo Rosasco (Universita' di Genova and MIT).

8. Quantum and physics-based machine learning: Bert Kappen (Radboud University Nijmegen), Riccardo Zecchina (Bocconi University Milan).

9. Natural intelligence: Matthias Bethge (University of Tübingen), Y-Lan Boureau (Facebook AI Research), Peter Dayan (MPI for Biological Cybernetics).

10. Health: Gunnar Rätsch (ETH Zürich), Mihaela van der Schaar (U Cambridge, ATI), Oliver Stegle (German Cancer Research Center and EMBL).

11. Machine learning in earth and climate sciences: Gustau Camps-Valls (Universitat de València), Markus Reichstein (MPI for Biogeochemistry).

12. Natural language processing: Mirella Lapata (University of Edinburgh).

13. Multimedia / multimodal information: Alan Smeaton (Dublin City University).

14. Information retrieval: Maarten de Rijke (University of Amsterdam).

15. Explainability and fairness in data mining: Aristides Gionis (KTH Stockholm).

16. Symbolic machine learning: Volker Tresp (Siemens).

ANNEX 2

Glossary

Adversarial machine learning
A machine learning method that tests the performance of machine learning models by providing inputs designed to prompt the model to make a mistake.

Computer vision
A research field exploring how computer systems can extract information from and understand the content of images or videos.

Deep neural networks
A machine learning method that uses artificial neural network models to extract end-to-end information about the goal from raw data.

Differential privacy
A method for sharing specific group-level information (for example, medical information indicative for a certain disease) hidden in a dataset of personal records without giving away any individual details (ie from the details it cannot be reconstructed who the one data point was).

Federated machine learning
A method of training machine learning algorithms on data held across a decentralised system or edge devices without requiring access to a single centralised dataset.

Information retrieval
The process by which computers search and find relevant information, usually in the form of text, from large amounts of data that may be largely unstructured.

Machine learning
Machine learning is an approach to AI which learns from (raw) data and the goals associated with them for the purpose of identifying patterns, reproducing goals or predicting the course in new, unseen data.

Natural language processing
A research field exploring how computer systems can process an extract insights from human language.

One-shot learning
The ability of an intelligent agent to learn how to perform a task based on a small number of data points or examples. For instance, an algorithm that is able to learn what a car is after seeing only a single image of a car.

Reinforcement learning
Reinforcement learning creates intelligent agents that learn through a process of interacting with real or virtual environments by maximising the reward from those interactions.

Transfer learning
Transfer learning is the technique of learning to recognise a new category from an existing classifier trained on a related but different set of data.

Acknowledgements

This document was produced with support from Jessica Montgomery (University of Cambridge), Neil Lawrence (University of Cambridge), Samuel Kaski (Aalto University), Arnold Smeulders (University of Amsterdam), Sami Heinäsmäki (Aalto University), and the leaders of ELLIS and ELISE research programmes (see Annex 1).

designbysoapbox.com

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.