**Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization**
**TAILOR**
**Grant Agreement Number 952215**

# Joint SRA v1 Report

| | |
|---|---|
| **Document type (nature)** | Report |
| **Deliverable No** | D2.7 |
| **Work package number(s)** | WP2 |
| **Date** | Due June 30, 2023 |
| **Responsible Beneficiary** | LiU, ID # 1 |
| **Editors and Authors** | **Editors**: Fredrik Heintz (TAILOR) & Jessica Montgomery (ELISE).<br>**Editorial board:**<br>Paul Lukowicz, HumaneAINet<br>Filareti Tsalakanidou, AI4Media<br>Alin Albu-Shaeffer, euRobin<br>Mario Fritz, ELSA |
| **Publicity level** | Public. Published on both VISION and TAILOR website on July 6, 2023 |
| **Short description** | The EU's six Networks of AI Excellence Centres (NoEs) are providing a Joint Strategic Research Agenda (SRA). The European Union's aspirations for AI, Data and Robotics (ADR) that are "made in Europe" demand an ambitious approach to advancing European AI research and development. The EU's six AI Networks of Excellence (NoEs) – AI4Media, ELISE, ELSA, euROBIN, HUMANE-AI-Net, and TAILOR – are providing a framework for delivering these ambitions, by advancing the frontiers of AI, data and robotics research and its translation to real-world impact in different domains.<br>The joint SRA complements the different Strategic Research Agendas (SRAs) that have already been published by AI4Media, ELISE, HUMANE-AI and TAILOR. |

| History | | | |
|---|---|---|---|
| **Revision** | **Date** | **Modification** | **Author** |
| version 1 | - | - | - |

| Document Review |
|---|
| The Joint SRA underwent a thorough process of reviewing by the participants of the six NoEs, as managed by the editors and the editorial board. |

# AI, data and robotics “made in Europe”:

## Research agendas from the European AI and robotics Networks of Excellence

Supported by:

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

June 2023

## Copyright



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# Table of Contents



These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# Summary

The European Union's aspirations for AI, Data and Robotics (ADR) that are 'made in Europe' demand an ambitious approach to advancing European AI research and development. The EU's six AI Networks of Excellence (NoEs) – AI4Media, ELISE, ELSA, euROBIN, HUMANE-AI, and TAILOR – are providing a framework for delivering these ambitions, by advancing the frontiers of AI, data and robotics research and its translation to real-world impact in different domains.

The NoEs, funded by the European Commission in the AI, Data and Robotics Co-programmed partnership, each bring unique strengths and specialisms to the European ADR landscape.

**AI4Media** provides a forum for researchers and practitioners with a focus on the media industry, responding to pressing concerns about the interaction between AI, the information environment, and wider society. **ELISE** convenes leading researchers in machine learning, pursuing research to accelerate innovation and adoption of these technologies in ways that safely and effectively address real-world challenges. **HUMANE-AI** focuses on the development of AI systems that work alongside human users, leveraging AI capabilities to enhance human activities. **TAILOR** brings together AI researchers with an interest in building the scientific foundations of trustworthy AI, integrating these methods across research and practice. **euRobin** gathers researchers dedicated to increasing the performance of robots with a holistic approach combining increased cognitive capacities, enhanced learning performance, greater levels of interaction and better suitability for users. **ELSA** is spearheading research in foundational safe and secure AI methods, pursuing the development of robustness guarantees and certificates, privacy-preserving and robust collaborative learning, and human control mechanisms for the ethical and secure use of AI.

Complementing the Strategic Research Agendas that have already been published by AI4Media, ELISE, HUMANE-AI and TAILOR, this joint document provides an overview of the areas of research interest pursued across the networks. It highlights shared themes relating to:

1. Building the technical foundations of safe and trustworthy ADR;
2. Integrating AI into deployed or embedded systems, including robots;
3. Enhancing human capabilities with collaborative AI and robotics;
4. Accelerating research and innovation with ADR;
5. Understanding interactions between ADR, social needs and socio-technical systems;
6. Advancing fundamental theories, models, and methods;
7. Ensuring legal compliance of ADR systems;
8. Advancing hardware for safe and energy efficient interaction between ADR technologies, humans, and the environment.

Progress in each of these areas is delivered through NoE-convened research programmes that advance knowledge, understanding, and applications in specific areas, alongside

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

NoE-led efforts to accelerate research, education, and knowledge transfer. Across their portfolios, NoE activities have engaged over 1,000 researchers and 100 industry organisations. Research activities from the networks have brought together researchers from different countries to deliver high-impact research. Educational activities have created accessible resources and provided an entry point for talented early career researchers from across the world to join the European ADR community. Knowledge transfer activities have created industrial partnerships that are integrating AI and deploying AI and robotics in diverse sectors, supporting small and medium-sized businesses to make use of AI and robotics technologies. By embedding these activities in local innovation ecosystems across Europe, programmes supported by the networks have been able to leverage local industrial strengths and opportunities to maximise their impact. This work has seeded an ecosystem of AI and robotics research and development activities that connect the EU's ADR policy ambitions to real-world, on-the-ground benefits for citizens and businesses in local communities across Europe.

The NoEs' achievements so far demonstrate that diversity of ADR research in Europe can be a strength, shaping ADR deployment across multiple sectors to align with policy aspirations for ADR technologies. However, amidst continuing international competition for ADR leadership – in terms of the ability to attract talent, shape research agendas, develop policy frameworks, and build novel AI-driven services and applications – increased investment in this European AI and robotics ecosystem is needed to deliver these aspirations.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# 1 Introduction

Europe aspires to be an international leader in ADR technologies. It aims to be at the forefront of the next generation of ADR, developing these technologies for the benefit of citizens in Europe and beyond, and in accordance with the rights and values enshrined in European law and regulations. Research and innovation are central to these aspirations. Amidst growing international competition for research leadership, research talent, and economic benefit, Europe can leverage its diverse strengths in research to shape the future development of ADR technologies. Recognising this opportunity, the European Union has set out a vision for AI 'made in Europe' that encompasses both core AI technologies and robotics, pursuing international leadership in ADR technology development alongside the deployment of these technologies across sectors for social and economic benefit.

To deliver this vision, a similarly ambitious approach to driving progress in AI, data, and robotics research, innovation and deployment is required. Horizon 2020's ICT48 Networks of Excellence and similar projects, funded by the European Commission in the AI, Data and Robotics Co-programmed partnership, are an engine for such progress. Established to strengthen Europe's ADR research excellence, these networks are pursuing a wide-ranging research agenda, strengthening collaboration and knowledge exchange, and building networks that facilitate research and innovation across the continent. Six ADR-related NoEs are currently in operation, each driving forward research to tackle a collection of important scientific and technological challenges: AI4Media; ELISE; ELSA; euRobin; HUMANE-AI; and TAILOR working alongside the **VISION** consortium as a forum for cooperation. Together, in their first years of operation, these networks have engaged almost 1,000 researchers and developed over 100 industry partnerships and use cases across more than 20 countries, while building educational programmes that are attracting global interest (see Box 1).

### Box 1: Overview of the work of the work of the AI and robotics NoEs

***AI4Media*** **(https://www.ai4media.eu/)**

AI4Media (A European Excellence Centre for Media, Society and Democracy) brings together a critical mass of top AI researchers, media professionals, social scientists and legal experts to create a Network of Excellence and a European Powerhouse in Media AI. AI4Media aims to deliver the next generation of AI technologies for the Media Industry and reimagine AI as a human-centred, trusted and beneficial enabling technology that can be used to offer innovative solutions to major challenges facing the media, the society and democracy.

The AI4Media consortium includes 142 researchers and 63 media and ICT industry professionals from 30 organisations engaging across 17 countries and collaborating in the context of 10 research & innovation WPs and 7 media-related use cases. It supports 60 researchers and 28 media and ICT industry professionals from 20 third-party organisations in 11 countries developing 20 media-related projects funded by

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

the two AI4Media Open Calls. It also engages 70 associate members, including academia, research, industry, and public service organisations and NGOs from more than 20 countries in the EU and beyond. Research in AI4Media has already resulted in more than 260 scientific publications, 50 open software and 15 datasets as well as 7 AI-enhanced demonstrators for the media industry.

***ELISE*** **(https://www.elise-ai.eu)**

European Learning and Intelligence Systems Excellence (ELISE) is a consortium of 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. Its focus is the development of next-generation machine learning technologies. Progress in machine learning has been the driving force behind recent advances in AI; it has unlocked a wave of new AI applications across research and industry, as well as enabling advances in sister fields such as Natural Language Processing and Computer Vision.

ELISE engages a network of world-leading machine learning researchers, bringing together 260 research fellows across 23 sites in 10 countries, with 40 associated partners across 14 countries. Its PhD programme is now the most internationally-competitive in the field. In 2021, its first round attracted over 1300 applications from over 70 countries, and gave over 60 PhD applicants the opportunity to study at top EU AI labs across 13 countries. Interest in subsequent rounds of this programme continues to grow. ELISE mobility programmes, which support research exchanges across Europe, have connected 40 scientists working at 25 sites across 10 countries. In addition to engaging over 30 start-ups through its network of incubators, the ELISE SME engagement programme has attracted over 500 applications, and provided grants and mentorship support to companies across 12 countries. Core to the success of this work has been a structure that connects knowledge institutes and industry in the context of local innovation ecosystems to deliver real-world benefits. To help build these connections, ELISE works in collaboration with ELLIS (European Laboratory for Learning and Intelligent Systems), a cross-European network of leading AI researchers that focuses on fundamental science, technical innovation, and societal impact, with the aim of securing European leadership in AI. This collaboration facilitates bottom-up engagement across Europe, creating a community of fellows committed to advancing the frontiers of machine learning, which is both scale-able – the community having grown rapidly in its first five years – and sustainable for the long-term, generating new research income and programmes.

***ELSA*** **(https://www.elsa-ai.eu)**

European Lighthouse on Secure and Safe AI – ELSA – is a virtual center of excellence on safe and secure AI aiming to address fundamental challenges hampering the deployment of AI technology. The ELSA community of world-class researchers seeks to integrate robust technical approaches with legal and ethical principles supported by meaningful and effective governance architectures. The end goal – AI in service of all European citizens, promoting and advancing core European values. ELSA has 26 founding members (20 scientific and 6 industry partners). The whole network works

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

closely with ELLIS - European Laboratory for Learning and Intelligent Systems - committed to shared standards of excellence, covering all aspects of safe and secure AI.

The research work in ELSA is carried out across three research programmes that focus on technical robustness and safety, privacy preserving techniques and infrastructures, and human agency and oversight. In addition, three grand challenges are designed to address the key obstacles preventing the industry from further AI development across various sectors – health, autonomous driving, robotics, cybersecurity, multimedia, and document intelligence. Grand challenges are implemented with industry partners co-leading six use cases through benchmarking and long-run competitions. Furthermore, the first of the two planned calls for SME funding is currently open. Funding will be awarded to selected SMEs and start-ups offering innovative solutions on applications of ELSA-developed methodology, software or tools in the industry environment, as well as contributions to use case benchmarks. Based on the core research and practical implementation of grand challenges, ELSA will produce a Strategic Research Agenda to support decision making on the topics of safe and secure AI both on the EU level, nationally, and in the industry.

***HUMANE-AI*** **(https://www.humane-ai.eu)**

The Humane-AI network seeks to facilitate a European brand of trustworthy, ethical AI that enhances Human capabilities and empowers citizens and society to effectively deal with the challenges of an interconnected globalized world. It focuses on systems capable of what could be described as “understanding” humans, adapting to complex real-world environments, and appropriately interacting in complex social settings. A key concern is also going from interactions between individual users and individual systems to an in depth understanding of large-scale human-AI socio technical systems.

The HumanE AI Net consortium consists of 53 partners from across Europe uniting a variety of communities within and beyond AI including in Human Computer Interaction (HCI), social sciences and complexity science. The consortium includes a combination of large European industrial champions, top universities and research centres. To foster network building and synergies between partners HumanE AI Net has developed a unique agile micro-project concept where researchers from within and outside the project can flexibly collaborate on short, focused questions aiming and outside the project can flexibly collaborate on short, focused questions aiming to create tangible results for the benefit of the entire European AI community.

***euROBIN*** **(https://www.eurobin-project.eu)**

euROBIN is a Network of Excellence that brings together European expertise on Robotics and AI. Distinguished research labs across Europe are jointly researching AI-based Robotics. Goals include both significant scientific advances on core questions of AI-based robotics as well as strengthening the scientific robotics community in Europe by providing an integrative community platform. As a main scientific and technological challenge hampering the breakthrough of robotics today,

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

we will address Transferability of cognition-enabled robotics methods between systems and among companies. The network is open to the entire robotics community and provides mechanisms of cascade funding.

The euROBIN consortium comprises 31 partners (24 scientific and 7 industry partners) across 14 countries. About 130 researchers are working in 3 application fields (Robotic manufacturing for a circular economy, Personal robots for enhanced quality of life and well-being, Outdoor robots for sustainable communities). The four science areas are Embodied Interaction, Learn, Know, and Human Centric Robotics. A core element is the data and code repository (EuroCore), designed as a central exchange platform. Networking tools of euROBIN are hackathons and cooperative competitions, brain magnet PhD and fellow programs, workshops / summer school and the science-industry dialogue. Cascade funding is offered within five open research calls for involving the larger robotics community in the research and the network.

***TAILOR*** **(https://tailor-network.eu)**

TAILOR aims to build the capacity to provide the scientific foundations for Trustworthy AI in Europe. TAILOR develops a network of research excellence centres, leveraging and combining learning, optimisation, and reasoning. These systems are meant to provide descriptive, predictive, and prescriptive systems integrating data-driven and knowledge-based approaches.

The TAILOR consortium consists of 54 partners, including 10 industry partners, from 20 countries and more than 250 researchers and professionals. The network also has more than 100 network members from 25 countries including both academia and industry. Its research is organised in five research programs and have resulted in more than 250 publications of which more than 150 in A/A* venues. TAILOR has also organised 4 challenges and 5 Theme Development Workshops with strategic thinkers and AI researchers from industry as well as academic institutions to jointly identify strategic AI research areas and challenges in specific sectors. The TAILOR Connectivity Fund provides funding for research visits between TAILOR partners and TAILOR non-partners to encourage and support new joint research.

Reflecting the strength of the European Union, these networks are ‘united in diversity’; their accommodation of diverse research perspectives, priorities, and participants is helping to build a vibrant European ADR research landscape. To help build understanding of this landscape, and of the work of the NoEs, this document presents an overview of the research challenges currently engaging network members and introduces the areas of technology development that are helping to address these challenges. It complements the Strategic Research Agendas that have been produced by some of the networks; those Research Agendas provide detailed roadmaps for progress in AI research and should be considered alongside the overview provided here. Annex 1 provides a short summary of each NoE Strategic Research Agenda, which illustrate their areas of focus and strategic approaches.

This document starts by introducing eight areas of research interest shared across these networks:

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

1. Building the technical foundations of safe and trustworthy ADR
2. Integrating AI into deployed or embedded systems, including robots
3. Enhancing human capabilities with collaborative AI and robots
4. Accelerating research and innovation with ADR
5. Understanding interactions between ADR, social needs and socio-technical systems
6. Advancing fundamental theories, models, and methods
7. Ensuring legal compliance of ADR systems
8. Advancing hardware for safe and energy efficient interaction between ADR technologies, humans, and the environment

It then highlights the research enablers that can help build the EU's ADR ecosystem, and the role of the NoEs' in providing an enabling environment.

Across these areas, technology, policy, and practice are moving at a rapid pace. Rapid shifts in AI capabilities can change both their technical power and their deployability, as demonstrated by current debates about generative AI, with implications for research, policy, and practice. This document therefore presents a snapshot of current work in AI and robotics; it was produced in Spring 2023 and reflects areas of activity at that time. Updates to the Strategic Research Agendas of each NoE will continue to provide up-to-date information about new directions in the field.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# 2 Delivering a European AI agenda

## 2.1 Research challenge: Building the technical foundations of trustworthy ADR

Respect for human dignity and rights, freedom, democracy, equality, and the rule of law are the foundation of European law and policy frameworks. The challenge for those developing and deploying AI technologies has been to translate these foundational values to practical interventions that align technology development with societal interests. To support this translation, the High-Level Expert Group on Artificial Intelligence (AI HLEG)[1] has articulated seven characteristics that AI systems should demonstrate to be considered trustworthy[2]. These are:

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

While technology alone cannot deliver on all these characteristics, advances in the technical capabilities of AI systems can contribute in each of these areas, providing a foundation for trustworthy AI. Box 2 below describes these connections between technological advances and the development of trustworthy AI systems.

### Box 2: Research directions in building the foundations of trustworthy AI and robot-based systems

#### *Human agency and oversight*

Integrated effectively into decision-making systems, AI can enhance human decision-making, extracting insights from diverse data sources to suggest actions that might not otherwise be visible. This integration involves careful consideration of the interplay between human decision-making and AI-enabled/robotics-based systems, and between individual machine learning or AI components within a decision-support system, especially visible during real-time physical interaction, but critical in a range of non-physical interactions between AI and humans. These interfaces – between human and AI, between physical and digital, and between interacting technical systems – give rise to research challenges where progress is needed to create AI systems that enhance human agency, safety and oversight. While the nature of the interface will vary according to the context of decision-making, required capabilities include: the ability to interrogate how and why a recommendation has been made; the nature of the uncertainty connected to that recommendation or output, and how that

[1] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai
[2] https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

might affect confidence in the system's workings; and the impact of the decision on different user groups and the operating environment. These capabilities and their integration into system design are active areas of research.

### *Technical robustness and safety*

Achieving the benefits of AI relies on the ability of ADR systems to function safely and effectively when deployed, especially in safety critical environments. While much progress has been made in testing and improving the performance of ADR technologies against performance benchmarks 'in the lab', further progress will require ensuring technical robustness and safety in real-world environments. ADR systems should be resilient to changes in the deployment environment, capable of interacting safely with environments and actors they have not encountered previously, robust against manipulation by adversarial actors, and able to adhere to expected standards of security and safety. Advances in verification and validation, the ability to certify or guarantee performance of AI systems that interact, and a better understanding of failure modes can all contribute to building these capabilities.

Safety in robotics is an overarching issue to enable robots working closely with people in the same environment (collaborative, assistive robots), in contact with humans (exoskeletons, teleoperated robots, cobots for instance) or even inside human bodies (medical robots). Safety is also crucial to avoid damage to the environment where the machine moves. Safety guarantees require to be made both in software, through comprehension of intention, situation awareness, and in hardware, with the utilisation and control of soft material systems that ensure safety even when in direct contact, for instance in assistive technologies and medical robotics. Ultimately, a holistic approach that takes into account AI, software and hardware aspects is essential to ensure the safety of humans and living beings in an environment where robots adapt, learn and evolve. Being able to validate the evolution of robot performance over time is critical to the deployment of advanced AI in robotics.

### *Privacy and data governance*

Privacy is a fundamental human right, which can be affected by a variety of practices in data collection, management, and use. A strength of AI technologies is the ability to combine multiple, complex data sources and process large amounts of data to identify insights that would not otherwise be available. To deliver this function, AI systems may require access to data about individuals that contains personal or sensitive data; they may also generate such sensitive data by analysing and combining datasets that may individually not appear to contain information that would cause concern. These complex patterns of data use contribute to a wider socio-technical environment in which it is challenging for individuals to understand or exert control over what data about them is used and for what purpose. Technical advances can help alleviate these concerns: progress in data-efficient AI is enabling new methods that can deliver accurate results without access to large datasets; privacy-preserving AI methods are demonstrating the ability to process data without revealing personal data; AI at the edge has the potential to process data locally, and share only the pieces of information needed for a given application instead of raw data; and growing interest in the generation of high-quality synthetic data offers an alternative to accessing personal

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

information in the creation of AI systems. Unlike in many other areas of trustworthy AI, there exists a widely accepted framework, differential privacy, that provides formal privacy guarantees. Formal guarantees are an important component of sustainable privacy solutions as they enable long-term anonymity. However, legal and technical understandings of anonymization may differ and this is an active area of research. In parallel, innovations in data and AI governance are needed to empower individuals and communities to set boundaries on the use of data about them, and a variety of data intermediaries – such as data stores, cooperatives, and trusts – are emerging in response to this need. New approaches are required to ensure data integrity and quality, especially for self-learning systems, in order to avoid malfunction or malicious function of AI systems.

***Transparency***

Connected to the principle of enhancing human agency, the requirement for transparency of ADR systems in operation can take different forms. It can invoke notions of:

- explainability or interpretability (the ability to provide an explanation of how the system works or why a result was provided, in terms meaningful to the users in a domain of application and to AI experts);
- justifiability (the ability to justify why a conclusion or recommendation was made, in the context of what is deemed reasonable for the overarching system);
- auditability (the ability to trace or track how different decisions are made or informed by AI systems, which might include how data is used by the system, and how it interacts with human users);
- accountability for decision-making, especially important post-failure where liability is at stake; or
- non duplicity of robots and decisional agents.

Technical approaches for responding to these needs are being pursued across different stages of the ADR development pipeline, from tracking what data is used and how, to the characteristics of the AI model ADRs used, to the design of human-machine, machine-machine and machine-environment interfaces.

***Diversity, non-discrimination, and fairness***

Examples of AI and robotics failures in deployment have shown how, if not carefully designed to accommodate the needs of different stakeholders, the use of ADR risks harming individuals and groups, with those harms typically accruing to already marginalised communities and groups of users. Consideration of diversity and inclusion is needed at all stages of AI design and development, including:

- The inclusion of communities and stakeholders affected by technology development or use in AI R&D, through consultation and co-design of research questions and technology development.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- Careful stewardship and curation of data resources deployed to develop ADR systems to ensure that the resulting system does not reinforce unfair bias encoded in training data or in assumptions about interaction and cultural context.
- Analysis of the impacts of ADR deployment, to identify unfairness in implementation or the consequences of AI-enabled decision-making.
- Human-centric ADR design and implementation, including deployment practices to ensure that user interfaces are accessible and reflect the needs of different communities.
- Accountability towards fostering inclusion and diversity in teams developing or deploying AI solutions, especially in leadership positions.
- Integration of interdisciplinary knowledge and non-academic perspectives in the education and upskilling of researchers and industrial professionals to foster a nuanced understanding about the potential harms of AI and responsibilities towards addressing them.

### *Societal and environmental wellbeing*

Who benefits from AI and robotics, how, and who bears the risks will be shaped by the purposes for which AI and robots are deployed and who has the power to make decisions. Policymakers and citizens have already expressed diverse ambitions for the use of AI, as shown in commitments relating to the Sustainable Development Goals, EU Innovation Missions, and public dialogues on AI. To achieve these ambitions, ADR technologies must deliver on the capabilities described above in an environment that facilitates their deployment for societal benefit. This necessitates continuous monitoring and research of the effects of AI on individuals and groups (e.g. in terms of health, social relationships or agency) but also research on the impact of AI on society and democracy at large. Additionally, more transparent, participatory, and inclusive AI development practices are necessary to stimulate accountability towards societal well-being.

AI research and development also need to be considered in the context of the Green Deal goals towards climate neutrality. There is an urgent need to address the environmental impact of the widening use of ADR. The raw materials used for constructing digital devices, the energy required to develop large-scale AI systems, and continuing use of such systems all contribute to the environmental footprint of AI. The extent to which the net environmental impact of AI will be positive or negative will depend in part on the extent to which these technologies are deployed for environmentally beneficial outcomes. Changing technical capabilities and practices can also play a role in mitigating their negative impacts, for example through energy-efficient AI methods such as transfer or continual learning. Here, interdisciplinary collaborations can again serve as a foundation necessary to introduce solutions that embed concerns for environmental well-being by design.

Counterbalancing the negative externality impact of AI, beneficial applications of ADR include:

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- Energy and waste: ADR could support a transition to a sustainable circular economy, through applications that optimise energy systems, support green energy management, or enhance recycling.
- Social care: ADR systems could help mitigate the declining carer ration through support for elderly, frail, and disabled people, enabling people to live independently for longer through ADR systems in the home or in clinical environments. This requires deeper development of technologies that enable motion and dexterity in any kind of human centric environment, Based on innovative combinations of robotics and AI.
- Online information: while ADR technologies have been a tool to create misinformation or disinformation – for example in priority policy areas such as vaccination or climate change – they could also provide tools to help mitigate its effects. Rapid progress is needed in capabilities to detect such online fakes.
- Health: AI can improve diagnostic tools, for example helping more accurately diagnose cancer and predict its spread, as well as enabling the delivery of personalised medicine.
- Climate and environmental management: AI is helping to build understanding of the Earth's climate system, building more powerful predictive systems that support local decision-makers to develop environmental management plans in the context of a pressing need for climate adaptation.

***Accountability***

Often linked to the notions of transparency described above, accountability of ADR systems implies close integration with human decision-making processes, to ensure the creation of mechanisms to assign responsibility for the use of AI in decision-making and to hold those responsible to account in the event of a failure of decision-making, especially where decisions have significant personal or social impacts. Connected research areas include:

- Auditability and decomposability of ADR systems: the ability to interrogate how different sub-components, including physical components, in the system work, their contribution to a system output, and how different environments, models, data sources and sensing have influenced that output.
- Explainability and interpretability: the ability to generate accurate, reliable explanations of how and why different outputs have been produced, in line with the needs and interests of different domain users or AI experts. In the case of robotics, this requires environmental and decision tracking that captures all the key information that leads to decisions, including proprioceptive information and data from external sensors and services, including other physical actors.
- System design validation: methods to assess how ADR methods are integrated into wider socio-technical systems for decision-making and physical interaction.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- Human-machine interactions: scrutiny of the interactions between human decision-making processes and ADR systems, and how these are influenced by AI or by the behaviour of robots.
- Law and regulations: work to address questions about assignment of liability and the regulation and certification of high-risk AI applications, alongside documentation to adhere to legal requirements, for example on key decisions throughout the system lifecycle that can be used for auditing and assigning responsibility and liability.

***Trade-offs and interactions***

Trustworthy AI guidelines list a number of positive aims for AI systems, many of which are reproduced above. The different aims may be mutually contradictory and have a negative impact on system utility; for example, the transparency and explainability of an AI system may be in tension with attempts to increase the privacy of such systems. More work is needed to understand these interactions and learning costs to allow making informed decisions on which aims to prioritise to which degree in a given application.

***Human Trustable and Understandable AI***

An extension of the concept of trustworthy and explainable AI is required, from a definition focused on technical aspects to a user-oriented approach that emphasizes systems that act and interact in a way that people and the society feel comfortable trusting and using.

These characteristics of trustworthy AI continue to be a cornerstone for AI development as technologies progress: current policy debates about generative AI, for example, highlight the importance of stewarding the development of AI in line with European values. The imperative to develop trustworthy AI and robot systems is a core area of interest for both technical and policy communities. Reflecting its significance, this research theme interacts with many of the themes described in the sections that follow. For example:

- The characteristics described above often shape how ADR technologies are integrated into deployed systems;
- Whether these characteristics are effectively implemented influences the extent to which ADR systems can enhance human capabilities, while respecting fundamental rights underpinning collaborative interaction;
- Accelerating ADR research, innovation and deployment with AI requires methods and tools that can be deployed reliably to analyse a variety of data types, knowledge areas and environments, while working alongside human users;
- The trustworthiness of ADR systems influences their acceptability within and influence on large-scale socio-technical systems;
- Efforts to address many of the research areas required above contribute to advancing fundamental theories, models, and methods across ADR technologies

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

## 2.2 Research challenge: Integrating AI into deployed or embedded systems, including robots

The power of ADR stems from its widespread applicability. As a general-purpose technology, ADR can unlock innovation across industry sectors and public administration; at work and at home; and in areas of critical societal interest, such as healthcare, media, environmental or economic sustainability, food production, energy supply and transport. Delivering this value will require ADR technologies that can be deployed safely and effectively to tackle real-world challenges.

Despite much progress in the technical capabilities of ADR technologies, deploying ADR into real-world systems remains challenging. A catalogue of examples of AI failures demonstrates the diverse implementation issues that affect the performance of AI systems 'in the wild', with consequences for individuals, communities, and society.[3] These issues arise in part from the complexity of real-world environments, and the complexity of integrating AI into deployed systems as effective decision-support tools and as tools to carry out physical functions.

Deployed ADR systems are typically formed of multiple interacting components, both hardware and software, each specialised for a specific function and contributing to the ability of the system to tackle a more complicated task.[4] When deployed in real-world contexts, these complex systems must operate in an environment that is itself dynamic, as human behaviours, environmental conditions, or interactions with other autonomous agents all influence the system's performance. These interactions can lead to a form of technical debt,[5] or a deterioration of performance, if the complexity of the deployed system and the dynamic nature of its environment mean that overall performance becomes difficult to maintain.

A further challenge in the deployment of ADR in real-world contexts is the integration of AI into decision-making systems as a decision-support tool. By combining data from diverse sources and identifying new patterns or insights, including data captured from agents in the working environment, AI offers the opportunity to enhance human decision-making. It can help make sense of complexity, enabling professionals in leadership roles to explore different scenarios for intervention or develop new strategies. To achieve this goal, careful design of the interfaces between human users and AI systems are needed, based on a nuanced understanding of the needs of human users, the complexity of the environment and sensory data, how human users interact with ADR systems, and what forms of knowledge transfer between user and AI can best inform decision-making, planning, and action.

Another AI deployment challenge concerns the high computing power required to train and deploy AI systems. The most advanced AI services demand costly computing power,

[3] https://partnershiponai.org/aiincidentdatabase/

[4] Autonomous vehicle control systems, for example, may draw from different components for image analysis, environmental analysis, and user interactions.

[5] Cunningham, W. (1992) The WyCash Portfolio Management System, OOPSLA '92 Experience Report, http://c2.com/doc/oopsla92.html

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

to be able to quickly process large volumes of data and perform complex mathematical calculations in intense machine learning workloads. The Cloud and GPU market is currently dominated by large US companies like Amazon, Google, Microsoft, and NVIDIA. A new architecture is needed for the fast-evolving AI field in Europe, for limiting Europe's dependence on large US corporations, and also for preventing AI from becoming an environmental burden. Towards that direction, research in AI at the edge can be considered critical for reducing energy consumption and AI hosting costs.

To deliver ADR systems that perform safely and effectively, and that can be integrated into decision-making processes, further progress is needed to build the capabilities of ADR for deployment, to create the measurement and evaluation mechanisms to scrutinise its performance, and to design effective interaction interfaces between humans and ADR systems In response, research in robustness seeks to build ADR systems that can respond to – or continue functioning effectively despite – unanticipated changes, system failures or unforeseen challenges in the deployed environment. Progress in automated performance monitoring and management, based on techniques such as AutoML/AutoAI and neuro-symbolic and social AI, propose new strategies for adjusting how an AI system functions in response to feedback from its environment, in line with performance requirements set by system designers. To demonstrate the effectiveness of AI systems at specified tasks, research in verification and validation is creating benchmarks for performance and mechanisms to guarantee system functioning are critical and particularly so in human-centric interactions.

Robots can be seen as the ultimate expression of complexity for the deployment of AI. There it is not only a matter of integration of AI technologies rather than achieving an entanglement between software and hardware at the image of the living. The challenge there is to create an intricate structure embedding tightly physical, sensorial, cognitive, decisions capacities, and human factors amongst other.

Areas for progress are described in Box 3.

## Box 3: Research directions to integrate AI into robot and deployed systems

### *Robustness*

When deployed in real-world environments, ADR systems are likely to encounter circumstances not directly represented in the data on which they were trained, or in the interpretation and planning systems they use. Reliability in dynamic, uncertain environments is central to the successful integration of ADR in real-world systems. Building on recent progress in research into the robustness of ADR systems, further advances are needed to develop principles and methods for robust AI, allowing users to understand how to improve robustness in practice; to quantify and verify methods proclaiming robustness; to improve methods in adversarial learning, allowing systems to respond to attacks from malicious users; and in associated areas like explainability and human-machine interaction, which can contribute to improved scrutiny and accountability of how ADR works in practice for human users.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

### *AutoML and AutoAI*

Designing a machine learning model to function effectively on real-world tasks can be time-consuming and expertise intensive. Automating elements of model selection and design can ease some of these demands, lowering barriers to deployment. In support of this goal, AutoML offers tools to automate design tasks such as algorithm selection or model tuning, thus accelerating the development of ADR systems.

Once deployed, a variety of different factors affect how an ADR system operates: changes to environmental conditions, shifts in data, or unexpected interactions with users or other autonomous agents can all contribute to difficulties maintaining performance. Operating at a systems-level, AutoAI offers a mechanism to automate monitoring, adjustment, and maintenance of interconnected machine learning components in an AI system, ensuring that the resulting system operates to expected performance standards. Those standards could be specified by users, and may include characteristics relating to its trustworthiness, such as its performance in relation to fairness, explainability, and safety. The resulting system would be able to identify when it had deviated from desired performance levels and automatically initiate corrective action. Further progress in this area is needed to create system design strategies that can deliver such functionality, supported by advances in simulation, emulation, and end-to-end or multi-objective optimisation, alongside progress in integrating core concepts from systems design, control theory, and software engineering. Proving AI automation is able to provide physical performance guarantees is a critical step towards the acceptability of ML and adaptive systems in human centric and safety critical environments.

### *Safety*

Both technical and governance safeguards can help ensure that AI systems do not cause harm. The field of AI safety encompasses a broad spectrum of research, including examination of the social and physical impact of widespread AI deployment and the technical interventions that can ensure AI systems operate safely and reliably in deployment. Areas of research interest include the robustness of AI systems operating in unfamiliar environments, assurance mechanisms to guarantee performance to desired standards, specification schemes to ensure AI systems align with the intentions of human users (avoiding harms and unintended consequences), and the design of fail-safe mechanisms that minimise harm in the event of system failure.

### *Verification, validation, and certification*

Verification and validation in the context of AI development seeks to ensure that ADR systems deliver the function required by human users to the expected standards. Action to verify or validate ADR systems may be needed across the development pipeline, from checking that data is accurate, to benchmarking performance under different circumstances.

Understandings of the most effective methods for verification and validation are evolving, based on changing understandings of deployed environments and their challenges. The research community is developing new benchmarks to demonstrate

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

system performance in ways that align with the needs of real-world challenges, alongside pursuing theoretical developments to prove or guarantee system functionality. Efforts to develop AI standards are also advancing, developing technical specifications that can be used by regulators to set performance expectations.

### *Traceability*

The ability to track how different inputs influence the output of an AI system is important for those seeking to understand how an AI system works, and how its workings relate to regulatory requirements around trustworthiness, data governance, and the legal rights of different parties. In the context of Large Language Models, for example, there are already legal questions relating to the use of data to train open-source AI systems. One research challenge that follows is how to trace the use of data or information in training a large AI system, for example using end-to-end learning. Another important challenge is how to anonymize personal data which appears in the training set while maintaining the performance of the models.

### *AI at the edge*

AI algorithms that run on devices at the point where application data is collected allow an AI system to process data and make decisions in milliseconds, without latency or downtime. In addition, they could reduce battery needs since devices could save energy by not being required to continuously send data to the cloud over a WiFi connection. More research is needed on AI at the edge, in particular methodologies for compressing large AI models to sizes that permit deployment at the edge without sacrificing model accuracy, along with the development of new hardware that overcomes current edge computing limitations, such as the limited scale of computation capacity on edge devices.

### *Human-centric AI*

To function effectively for human users, ADR systems require interfaces that enable communication, physical interaction and knowledge exchange, including procedural knowledge, between people and machines. These interaction interfaces should support planning, decision-making and accountability, providing relevant information, highlighting uncertainties or risks, and encouraging human users to understand how the system works. Human-centric AI seeks to embed these user needs in ADR design, integrating methods in participatory design, trustworthy AI, and human-machine interaction. An important question is assessing the impact of AI technology on human users determining positive, neutral or negative impact with respect to time scale, population (individual, group, corporate), risk (balance of likelihood of harm and size of negative and impacts measures (qualitative or quantitative, in context or general)). For this, a general broadly usable assessment framework is needed.

### *Entanglement between AI, software and hardware*

Beyond the integration of technologies, entanglement is the ultimate way to ensure conception of robots that will behave in a safe and natural way with humans and living beings in the surrounding environment. This entanglement should reinforce in particular, safety, reactivity, robustness of decisions and make the robot a “social”

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

device. This overarching consideration of taking in to account human factors is a key driver to acceptance of robotics by users.

***Cybersecurity***

AI systems are software systems and susceptible to all the same cybersecurity threats as any other software. As AI is becoming part of the IT infrastructure and landscape, it is important to understand that it is also becoming part of the attack surface. Therefore, AI deployment cannot be disentangled from cybersecurity considerations. Policymakers, users and developers need to understand attacks and defences for AI and how they interact with the overall system. This requires a security analysis in order to derive threat models and risk assessment. A failure in this dimension will create systems with novel vulnerabilities. Equally, the using AI also has the potential to support secure and resilient systems with innovations in cybersecurity itself.

## 2.3 Research challenge: Enhancing human capabilities with collaborative AI and robots

Both techno-dystopian and utopian visions of ADR inform public perceptions of how ADR might contribute to society. There already exists in the public consciousness a concern about displacement of human labour because of ADR-enabled automation, or the potential for people to become over-reliant on technologies in ways that cause harm, for example by removing human interaction in contexts where it is valued. An alternative vision is for the use of ADR to support and complement human activities, through human-machine collaborations that enhance human capabilities. Such collaborations across science, public administration, and industry could leverage data to improve human decision-making, automating tasks that are dangerous for humans, helping address pinch-points in the delivery of public services, or providing support where additional labour is needed.

The core research problem is how to facilitate complex, flexible collaboration in mixed AI-human teams of different sizes and compositions at different levels of complexity and different social and temporal scales. When collaborating with such systems, users should be able to adapt the system to new tasks by high level, human understandable representation of the goal and the boundary conditions and rely on the shared domain models to ensure that the system manages to find a solution that is aligned with the user's true intentions, including values and other personal or collective priorities. To facilitate such collaboration, systems should analyse their own role and capabilities, the roles and capabilities of others, and the ability to ask/accept/negotiate tasks with other actors.

Achieving this positive vision of ADR requires technologies that can work effectively alongside people, helping make sense of complex environments, data and phenomena and inputs that might otherwise seem intractable. Collaborative AI and robotic agents can work in support of human activities, identifying relevant goals, providing information and insights to support decision-making and physical task execution in aid of those goals,

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

and enabling human scrutiny of their work. To support these capabilities, autonomous AI and robotic agents would need to:

- Elicit information from human users about their goals or desired outcomes, using this information to identify relevant information or take relevant action, even in cases where these goals might be unclear or subject to change. The result should be a model of the task at hand and the environment surrounding that task, including human users, that establishes a shared understanding of what the user is trying to achieve and how the system can help, based on explicitly eliciting information and sensing or interpreting behaviours and reactions.
- Exchange knowledge with human users, communicating important information – and the limitations of the system's effectiveness – in ways that are intuitive for users, that are adapted to their level of expertise and that facilitate action and interaction. These shared understandings should enable information exchange that facilitates complex deliberation and decision-making.
- Enable human scrutiny, supporting users to interrogate how and why the system is working, or when its results might not be trustworthy, through interfaces that empower human users and discourage over-reliance or confidence in its results. These interfaces might include factual, counter-factual, and other high-level explanations that include both domain knowledge and user background.

The resulting collaborative system would be based on models of collaboration and knowledge representation that enable interaction between different autonomous agents. They would need to be supported by learning strategies that bridge human and data-driven knowledge in a dynamic environment, combining multiple sources of information and communicating that information to different agents, while also meeting expectations in relation to the trustworthiness of AI systems. Together, these functions can support collaborations between intelligent agents – both human and artificial – while aligning the action of AI-enabled systems with the interests and concerns of human users (Box 4).

## Box 4: Research directions to enhance human capabilities with AI and robots

### *Multi-agent collaborations*

The collaborative systems described above require AI agents that can operate alongside other human and autonomous systems. To implement this type of multi-agent collaboration, research is needed to combine situation awareness, an understanding of user needs and goals, and the development of shared representations of the world in the context of autonomous agents. The resulting collaborative agents would be able to understand the motives and goals of human users, with a form of theory of mind that allows analysis of the agent's intention, how it relates to other agents in the collaboration and how this evolves over time.

### *Models of Human AI collaboration*

Everyday interactions with AI-based technology are increasingly moving away from simple human use of computers as tools to establishing human relationships with autonomous entities that carry out actions on our behalf or that extend our capabilities

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

by supporting our decision-making. Towards a better understanding of such relationships there is a need to develop and evaluate comprehensive models of human-AI collaboration at different relevant points of a collaboration space given by a combination of complexity from (simple reactive use in the sense of a passive tool that directly reacts to explicit inputs through situated implicit interaction to dynamic creative cooperation), social complexity (from a simple one-on-one interaction to complex socio-technical systems where large numbers of networked AI agents interact with complex social structures) and temporal extent of the collaboration (immediate effect or target long term developments).

***Common ground and shared representations***

Effective communication, collaboration, and trust all depends on the stability of the involved partners to relate to a common understanding of the world. This includes understanding of the situations, understanding of the effects of actions or events and the understanding of the manner of attaining objectives. Thus, a key problem that human centric AI needs to address is bridging the gap between the human and machine "understanding", including relating human world models and AI/ML representations built from multimodal input data. In developing these methods, modelling approaches that bridge between data-driven and human models of the world will be necessary. In particular current developments in the area of large language models, especially the drive towards extending them to multimodal embodied variants (e.g., Google PaLM-E).

***Active learning, lifelong learning and dynamic feedback***

The ability to learn and reason in social contexts can help create AI agents that interact intelligently with human users. To adapt to changing goals and needs, AI agents will need active learning strategies that enable them to respond to their environment safely and effectively. Critical to success is the ability to carry out active learning over long periods of time without degrading performance either in the short term or for prior learned tasks, and to integrate new data throughout their life cycle.

***Knowledge representations***

Core to the exchange of insights between human and AI systems is the question of how knowledge – both from human, domain insights and from knowledge and data-driven AI systems – can be represented in AI agents and communicated between those agents and human users, drawing from methods in explainable AI, the development of narrative approaches to explaining AI systems, and human-in-the-loop learning, including in physical environments. In developing these methods, modelling approaches that bridge between knowledge and data-driven models and human models of the world will be necessary especially where these also need to account for the hard physical constraints of real-world applications.

***Understanding intentions, understanding the appropriateness of behaviours according to the operating context***

Adapting to human intent and modelling this within operating contexts – both physical and virtual – is key for sustained and successful interaction. Interaction in the physical

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

world demands that the internal models of the system continue to provide useful input to autonomous decision making in the long term and over both short and long timescales adapt to changes in human intent and the dynamics of the environment. For example, in the development of assistance robots for rehabilitation or for use by elderly people or people with cognitive impairment or mobility difficulties, the understanding of intentions, or of "otherness", is the key to increasing human capacities through technical assistance. In online environments, understanding intentions and appropriateness of behaviour is also important for interactions between humans and AI systems. For example, when using chatbots in any application – from customer service support to virtual tutors in education – or when human users interact with virtual characters in virtual worlds, in games or the metaverse. All this relies on the ability to analyse data and to take into account the performance of people, devices and robots in complex dynamic physical environments.

***Illuminating the AI generative space***

When users engage in creative work with an AI-assisted tool, they would benefit in understanding the limits of the tool's generative capabilities: what it can typically create and the range of artefacts it can provide. The tool's ability to illuminate the design space with multiple different and good solutions can help designers identify new designs or to perfect generated content based on their current priorities. Towards that end, expressive range analysis has been introduced as an (often visual) analysis of the output in terms of styles and variety of artefacts generated by the chosen approach, which can highlight biases of the generator towards specific types of content. Evolutionary search towards quality diversity (QD) addresses this by exploring a search space to find the largest possible set of diverse and high-quality solutions in one run. The most prominent QD algorithms already come with visualisation modules and are naturally well suited for this task, which contributes to the explainability of the AI-assisted tool. Research in combining QD search with current state-of-the-art deep learning generative systems is expected to have a strong impact in the explainability and usefulness of AI-human co-creative endeavours.

***Human AI co-evolution as a learning paradigm***

In developing collaborative AI there is an opportunity to move from the current paradigm of training systems on large amounts of data for narrowly defined specific applications to systems that can evolve new functionality through collaboration with the user on the basis of shared representations and common ground.

## 2.4 Research challenge: Accelerating research and innovation with ADR

Addressing many of today's most pressing societal challenges – from adapting to climate change, to increasing food security, to widening access to high-quality healthcare – requires advances in research and innovation. ADR can help accelerate innovation, becoming an enabler of scientific discovery that unlocks progress in research across disciplines to help tackle these challenges.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

AI has already supported high-profile successes in science. It has helped generate new understandings of the science of protein folding, predicting the three-dimensional structures of proteins from amino acids[6]; provided tools to help identify and characterise the behaviour of fundamental particles[7]; and supported the search for new planets.[8] In addition to these flagship discoveries, AI is also becoming integrated in a variety of research domains, becoming part of an analytical toolbox that is helping researchers make sense of complex datasets across the natural, physical, medical, and social sciences. Automated and robotic based systems are able to carry out lab-based experimentation over long periods of time without human intervention. They are also able to capture samples and perform remote analysis, for example in planetary, oceanographic, medical or geological exploration.

These applications of AI and robots in science have also highlighted a variety of areas where further progress is needed to support their wider application and use. Advanced robots and AI tools for scientific discovery would be able to combine datasets from different sources to create sophisticated representations of complex systems that allow researchers to explore their dynamics. They would allow researchers to interrogate how and why those systems work the way they do, considering both insights from data and what researchers already know about the system – such as the fundamental laws that govern its dynamics. And they would deliver these functions in a way that integrates with human research in the lab, working collaboratively with human users. Delivering these capabilities requires new modelling approaches that can bridge data-driven and domain-led, or mechanistic, modelling approaches, creating ADR agents that enhance scientific investigations.

Building these bridges will require technical advances and new ways of working with AI and robot agents. Advances in simulation and emulation can help researchers build representations of complex systems that they can use to interrogate the forces that shape them and understand how systems work. To understand why particular system dynamics, emerge, new approaches to causal AI are needed, taking researchers from an understanding of how a system works to the causal inference and discovery to show what drives it. In developing these ADR tools, researchers can make use of existing, structured knowledge about how the system works by encoding that domain knowledge, aligning insights from data with known laws and theories.

To help deliver these technical advances, researchers will need to build interdisciplinary collaborations that bring together expertise in ADR with expertise in relevant research domains. Sharing their results will require toolkits that allow others in the community to adopt relevant methods, and to understand the limitations of those methods. In support of these efforts, incentives and recognition for interdisciplinary research are needed from research institutions, alongside forums for community-building. Underpinning all these

[6] https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

[7] https://home.cern/news/news/computing/higgs-boson-machine-learning-challenge

[8] https://exoplanets.nasa.gov/news/1565/discovery-alert-two-new-planets-found-by-ai/

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

efforts is the need for an amenable data environment, in which well-curated data is accessible for use in research.

## Box 5: Research directions to accelerate research and innovation with ADR

### *Simulation and emulation*

Simulations have long been used to support scientific discovery, allowing researchers to explore how a complex system operates and to test the impact of different interventions. AI-enhanced tools can leverage advances in machine learning to accelerate and deepen these capabilities, building more sophisticated simulations than possible with traditional methods. AI tools for simulation or emulation already exist, but further work is needed to refine their capabilities through technical advances in machine learning techniques such as simulation-based inference and surrogate modelling, and by understanding which techniques are best-suited to which research challenges.

### *Causal AI*

In the context of research and innovation, AI users typically seek to not only understand how a system operates, but why particular dynamics emerge. Current AI technologies have achieved great success in identifying patterns and predicting outcomes from data, but moving from correlations to causal inference is necessary to understand why those outcomes are produced. This requires new approaches to causal learning. In addressing this challenge, researchers have leveraged new learning strategies – such as transfer learning, or multi-task learning – to improve how an AI system performs a previously unseen task, as well as model design techniques that integrate domain knowledge.

### *Encoding domain knowledge*

In most areas of study, there already exists a wealth of scientific knowledge that can be used to improve the performance of AI systems. This domain knowledge can be leveraged to improve the performance of AI research tools, increasing analytical power and tailoring the performance of the research tool to the research domain of interest. Methods for integrating domain knowledge into AI systems include: constraining model operations according to known laws (for example, ensuring that model outputs adhere to the laws of physics); building collaborative AI that can be integrated into research processes, based on effective human-autonomous agent collaboration; and the design of user interfaces that allow researchers and AI tools to exchange knowledge and insights.

### *Multimodal learning*

One of the aspirations for AI in research is that by combining different data types from different sources, AI-enabled analysis might identify phenomena or patterns not previously observable by human researchers. To deliver this function, AI systems must be able to work with different data types – text, images, video, audio for example –

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

from different sources. This use of multi-modal data can help build more comprehensive models that allow nuanced exploration of the system of study. Advancing this capability will require techniques for multimodal knowledge representation and ways of learning from multimodal data.

***Explainability***

While so-called 'black-box' AI methods can be helpful in identifying phenomena of interest, in many cases researchers seeking to advance scientific understandings of a system of study require some level of explainability to support their work. The nature and type of explainability required in scientific applications will depend on the user or use case, drawing from a variety of techniques that might variously consider how data influences AI outputs, how AI models work, or how changes in the AI system influence the performance of the model.

***AI applications in research and innovation***

Many areas of science are already benefitting from the use of AI to support research and discovery, but further work is required to increase the uptake of AI tools across disciplines, ensuring all domains – from archaeology to zoology – can adapt and deploy these methods.

***Research Methodology and Infrastructures***

There is a pressing need to define an overarching research methodology that bridges the differences between the individual disciplines needed to address core problems of human centric AI. The primary concerns are evaluation metrics and methods that combine technical performance metrics with user acceptance, usability and social aspects. This goes beyond well-known cost factors that weight different types of errors according to their significance towards subtle, dynamic, situation and user specific assessment that takes into account the impact on the effectiveness of the system within the different types of interaction. A related issue is experimental methodology including ethical aspects of data collection and experimentation. In parallel with the methodology definition, we need to provide tools and infrastructure such as data sets, evaluation scripts, repositories, baseline evaluation sets and benchmarking support. Finally, the research methodology aspect needs to be incorporated in education, in particular at Ph.D. level.

## 2.5 Research challenge: Understanding interactions between ADR, social needs and socio-technical systems

Through its Innovation Missions in cancer prevention, sustainable cities, soil health, ocean protection, and climate adaptation, the EU has already articulated policy goals where effective use of robots, data and AI could help deliver public value[9]. Achieving these aspirations for robotics and AI – and the wider goals for the use of ADR to deliver benefits for society – requires deployment at scale. This scale introduces new challenges

[9] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

for ADR technologies and for their integration into large-scale socio-technical systems. In this context a key challenge is to move from the perspective of a single user accessing a single AI system to complex socio-technical systems where a large number of users and AI systems interact with each other in dynamically evolving patterns. This can lead to unforeseen emergent effects, which can only be understood through a combination of AI model complexity science methods, and social science considerations. Another related concern is balancing the usefulness of the system to individual with consequences that the system functionality can have on a collective level.

The technical challenges associated with building trustworthy systems that function effectively in deployment have been introduced in the preceding sections. Delivering public benefit typically either involves deployment in decision support systems or in the delivery of levels of autonomy and human functional augmentation to complex physical tasks. The use of AI in public services typically involves careful integration in high-stakes decision-support systems, where decision-making can have significant personal or social effects, and in complex systems where there can be unpredictable or unintended consequences from decision-making. To serve public interests, further research is needed to create deployable ADR systems that both deliver on trustworthiness and create genuine societal benefit that can be evaluated[10].

As outlined above, advances in technical capabilities are necessary. How that research and development proceeds also matters. Progress towards societal aspirations for ADR– as implied by initiatives such as the Innovation Missions – will require rallying of the research community behind shared goals and creating multidisciplinary collaborations that centre public interests in system design. These research practices need to:

- Identify and implement methodologies that create effective co-design projects with impactful outcomes that engage closely with affected communities, including publics, policymakers, business, and civil society.
- Convene across disciplines to identify mission-oriented research questions, that couple specialised domain knowledge to the design of systems, and measure and scrutinise the potential impact of the use of ADR in the domain of interest.
- Identify best practices in stewardship to ensure trustworthy access, management, and use with the goal of driving uptake and acceleration of impact.
- Embed principles of 'X[11] by design' and responsible research and innovation in the development of ADR-enabled interventions, to ensure the resulting systems reflect shared values, interests, and concerns.
- Develop and implement mechanisms for monitoring the performance and impact of ADR interventions in complex socio-technical systems, with feedback loops

[10] For example: Delivering societal benefit against mission goals in the physical domain demands integration of ADR technologies to carry out complex tasks such as the localisation and removal of tumours, the removal of pollutants, such as plastics, from the oceans and the use of robotics driven by AI interpreted data to ensure environmental parameters, such as soil and air quality are maximised on farms or in cities.

[11] X by design includes: Privacy, Ethics, Safety, Trust, Security, Dependability, Maintainability, Adaptability etc.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

that enable corrective action if needed to address issues in deployment, acceptability, trustworthiness and impact.

These practical considerations in turn highlight areas where multidisciplinary research involving extended dialogue between the humanities, technical and social sciences into the complex interactions between ADR, public interests, and large-scale socio-technical systems is needed to better understand how to direct ADR towards beneficial social outcomes. The pathways along which ADR progresses can also be shaped through the development of careful stewardship practices that allow citizens to influence patterns of use and interaction; through participatory design practices that connect affected communities to the development of systems through the better understanding of how the principles of responsible research and innovation can be applied to ADR systems and through interdisciplinary studies of the impact of ADR on society, including attempts to predict emerging risks or issues; and the implementation of X-by-design' methods (Box 6).

## Box 6: Research directions to understand the interactions between AI, robots, social needs, and large-scale socio-technical systems

### *Data and robotics stewardship*

Many currently popular AI methods rely on access to data. In areas of policy concern – health, climate, environment, media for example – the EU has identified the need for an amenable data environment as an important enabler of AI applications, developing frameworks or repositories aimed at increasing data access and use. In operationalising these policy ambitions, further work is needed to understand the data standards and management practices that can support the development of trustworthy AI. Such management practices include data stewardship interventions that can align the contours of data use with the interests and concerns of citizens and affected communities, giving publics a voice in decisions about how data relating to people is deployed for AI development. The need for such stewardship creates opportunities for citizen engagement in AI development, and new questions about what data stewardship practices can help deliver trustworthy AI. It is here where the disciplines of dataspaces and trustworthy AI meet each other.

Stewardship also applies to physical systems where deployment have impacts on people and environment. Wide scale deployment of robotic systems carries negative externalities that require engagement with the interests and concerns of users who are remote from the design and implementation process. Without guided and considered stewardship, the imposition of robotics technology can result in low levels of acceptance and increased levels of distrust. Examples include the widespread use of drones or pavement robots for delivery, the use of data derived from autonomous systems in everyday environments and the use of robotics in the care of children, the elderly and the vulnerable. Good stewardship is also important in the introduction of robotics into factories and workplaces.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

### *Participatory design*

Centring human interests in system development requires design practices that identify the needs of different communities, explore trade-offs in functionality in deployment, and develop technologies in line with societal interests and concerns. Participatory design has long been used in product design and in human-computer interaction studies, as a means of helping develop effective interfaces and interactions between human users and products (software and hardware based). Such practices can also help empower individuals and communities, creating space for societal interests to directly shape technology use.

### *Responsible research and innovation*

Responsible research and development practices have long sought to address concerns, trade-offs, and risks arising from technology development in an inclusive and transparent way. Different frameworks or tools for responsible research already exist, and their application in AI research is evolving. Oversight of what constitutes "responsible research and innovation" cannot simply rest with the developer, it is critical that external assessment is made of "responsibility". This is increasingly critical when socio-technical impacts must be considered and where ethical considerations exist. Without these safeguards the potential for unintended harm is increased.

### *Dynamics of socio-technical systems*

Interactions between ADR and people are typically part of complex socio-technical systems, characterised by shifting relationships between people and technology and unpredictable or emergent dynamics. The implications of these complex network effects also need to be integrated into technology development, through consideration of how human and artificial agents interact, how both individual and collective goals can be represented in ADR systems, and what forms of social awareness need to be built into ADR agents (for example, perception of user emotions or attitudes, interpretation of social gestures norms and the tracking of social context). Understanding this broader context is critical in analysing the social impact of ADR technologies, and requires interdisciplinary research that combines expertise in ADR, social sciences, and policy. The general challenge is to characterize how the individual interactions of individuals, both humans and AI systems, with their own local models, as well as the social relationships between individuals, impact the outcome of AI models globally and collectively. Using a combination of machine learning, data mining, and complexity theory, there is an opportunity to understand the networked effects of many distributed AI systems interacting together, some (or all) possibly representing human users, therefore comprising a complex human and technical ecosystem. The different layers of this system are in mutual interaction, producing emergent phenomena which may range from synchronization to collapse.

### *AI impact on society*

By ushering a wave of catalytic innovations in nearly all aspects of business and society (from health to transportation, finance, the fight against climate change, the media industry, politics, etc.), AI has the potential to transform the world and our lives in a positive way. However, with its full potential still evolving or largely unknown, there

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

are increasing questions and concerns regarding the social, economic, and political implications of AI. Understanding the impact of AI will require a multidisciplinary approach involving social scientists, ethicists and technology experts that will, on one hand, highlight its transformative potential and opportunities and, on the other hand, identify concerns and risks from its use. This approach is expected to identify what mitigative measures will be necessary to ensure that AI is developed and used in a responsible manner that will create a safer and better world for all.

***X by design***

Building on emerging technical approaches to implementing the characteristics of trustworthy systems, new design practices are seeking to embed these characteristics in the development of AI systems. Privacy-by-design, legal-protection-by-design, safety-by-design and ethics-by-design are amongst these emerging attempts to integrate the ethical basis for responsible AI into AI research and development. While it is easy to state these objectives, the added complexity that they bring to the design process, the ability to develop tools that support their integration, and the methods for engaging relevant external actors, for example in medical ethics, into the design process remain open questions. In simple terms, it is important that design can proceed to implementation without being stalled by over-complication of the process.

***Balancing individual and collective (social) interests***

Social dilemmas occur when there is a conflict between the individual and public interests. Such problems may appear also in the ecosystem of distributed AI and humans with additional difficulties due to the relative rigidity of the trained AI system on the one hand and the necessity to achieve social benefit and keep the individuals interested on the other hand. Policymakers are grappling with questions about the principles and solutions for individual versus social benefits using AI, what balance is desirable, and how that balance can be achieved.

## 2.6 Research challenge: Advancing fundamental theories, models, and methods

While focused on the research challenges that arise in the context of AI deployment, progress in each of the areas described above will rely on advances in fundamental AI research. Delivering the next generation of AI research will require new AI methods and their underpinning theoretical developments. Achieving progress in these areas can help drive a new wave of AI innovation; if Europe wishes to remain at the forefront of this wave of innovation, it must be able to advance these AI fundamentals. Rapid developments in Large Language Models and other forms of generative AI, for example, have relied on innovations in core underlying technologies. Europe needs both the technical capability to drive such innovations, and the ability to apply those innovations in practice.

Such fundamental research can:

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- Provide a rigorous methodological and theoretical basis for the characteristics of trustworthy AI, including robustness, social interaction, safety, explainability, security, and privacy-enhancing technologies.
- Give a foundation for the cross-cutting capabilities that many deployed systems require to function effectively, including causal AI, the ability to work with multimodal data, active and federated learning strategies, and the integration of learning, reasoning, and optimisation.
- Progress core technical domains to deliver functions such as computer vision, Natural Language Processing, video analysis, reinforcement learning, and quantum computing; and
- Help ensure the reproducibility of AI research and reliability of AI methods in deployment.

Advances in this research agenda can be driven through efforts to build the technical foundations of trustworthy AI; develop new learning strategies that can increase the power and accuracy of AI methods; progress in AI-enabled capabilities, such as computer vision, natural language processing, and quantum computing; and the integration of different learning methods (Box 7).

## Box 7: Research directions to advance fundamental theories, models, and methods

### *Characteristics of trustworthy AI: robustness, safety, explainability, security, privacy*

Each of the characteristics of trustworthy AI introduced earlier have theoretical and methodological components. Advancing the technical basis for these characteristics through further research into foundational theories or methods can contribute to more effective implementation of these features in practice.

### *Foundation models*

Much recent excitement has been generated by progress in Large Language Models that can analyse, synthesise, and generate text by analysing large datasets, producing answers to user queries that are convincingly human. The ease with which users can interact with these models opens the possibility of a wide range of AI applications. At the same time, lowering the barriers to use of such AI tools also creates new risks and potential harms. Further work is needed to understand the opportunities and challenges of Large Language Models, and to identify how these tools can be deployed to maximise their society benefits and minimise the associated risks.

Large Language Models are one type of foundation model. These AI models are typically trained on large datasets and then deployed for a variety of tasks, such as text, image, or video generation. Progress across a range of foundation models could unlock more general-purpose AI tools, which could be trained once and used for many purposes. The question of whether it is possible to combine foundation models using different data types is generating new research directions. Given their wide potential applicability, one possibility might be to consider foundation models as commons, and

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

thus facilitate access to the full pipeline needed to create and use them (data collection, structuring, training, tuning, deployment). A key concern is connecting such models to the physical (and social) reality of the real world by developing methods to align their representations with physical simulations and sensing. This follows the initial ideas presented by systems such as the Google PaLM-E to move from today's language based foundational models to multimodal, embodied models.

Two important research challenges for Large Language Models are how to make them more factful and how to keep them up-to-date as new information becomes available. Currently, Large Language Models are designed to produce the most probable next token given the context. A major achievement would be to make these models generate factful and logically consistent text. This also includes the ability to perform mathematical computations and other forms of formal reasoning. Today, these models are trained on historic data, thus they are ignorant of everything that has happened after the data was collected. Developing methods to keep foundation models updated with the latest information would be another major achievement. Other challenges include making them more trustworthy, explainable and compliant with regulation such as GDPR.

***Learning strategies: active learning, deep learning, reinforcement learning, transfer learning, few-shot learning, federated learning; continual learning; multimodal learning; causal inference***

A variety of new learning strategies are increasing the power of machine learning methods, widening the scope of applications where they may be deployed. Progress in reinforcement learning is creating AI agents that can perceive and interact with their environment in support of a goal. Transfer learning is supporting AI agents to apply insights learned from one domain to another environment. Few-shot learning is increasing the ability of AI systems to learn how to perform a task based on one, or very few examples, enabling learning in data-poor environments. Federated learning allows AI agents to learn a task by accessing data held across multiple, decentralised devices, without relying on aggregating data to a central location. Active learning methods are supporting new types of user interactions, creating collaborative AI agents. Continual learning is well-adapted when artificial agents need to learn from streams of data under time, computation and/or memory constraints. Multimodal learning is allowing AI systems to integrate different data types, to create a richer understanding of their environment. Causal inference methods are improving the ability of AI to identify cause-effect relationships in complex systems. Progress in deep learning continues to increase the power of these methods. Further advances in all these areas can increase the technical ability and deployability of AI agents.

***Computer Vision***

Machine learning has already facilitated a wave of progress in computer vision, which enables computer systems to analyse and interpret the visual world. Building on this progress and advances in learning strategies, there are opportunities to further increase capabilities in object recognition, image reconstruction, and the use of vision in autonomous systems, also leveraging the use of other available modalities like audio or text. The availability of foundation models for visual and multimodal data is

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

likely to have a strong impact on the field in the next period, much as it already has for natural language processing.

***Natural Language Processing***

Advances in Natural Language Processing have already helped reshape how humans communicate with AI systems, creating AI systems that can deliver human-like text in response to user questions. The ability to interact 'naturally' with AI systems could help make AI accessible to a wider community of users, as well as unlock applications across sectors – as already being demonstrated by widespread engagement with recent Large Language Models. Further progress will require techniques to integrate different types of data – such as images and text – to produce answers to user-generated questions. More sophisticated reasoning strategies or forms of knowledge integration can also facilitate progress.

***Quantum Computing and Machine Learning***

The field of quantum machine learning uses ideas from quantum physics to create more powerful machine learning methods running on Quantum Hardware, while also applying these methods to help drive new understandings of quantum information processing and Quantum systems. Progress in this area could create more powerful AI technologies, as well as more energy efficient approaches to implementing AI, based on new understandings of the physics of computing beyond current technology.

***Integration of learning methods***

Different approaches to 'intelligence' inform the research areas described across this document. These include the ability to learn, reason, or optimise, based on data-driven, or structured forms of knowledge. Integration of these methods can create new possibilities for AI, providing a spectrum of modelling approaches that can be flexibly deployed in response to the demands of the application under consideration. Methods are needed to integrate existing domain knowledge into learned models, or enable existing knowledge to enhance learning.

## 2.7 Research challenge: Ensuring regulatory and legal compliance of ADR systems

In response to the policy challenges posed by the development and use of AI in business and society, the EU has already set out a package of regulatory interventions. These focus on embedding EU values in AI governance, regulating AI applications, and establishing standards and certification mechanisms that align technical systems with regulatory requirements. Many of the research areas described in this document indicate the role that AI technologies themselves can play in delivering the goals of these regulatory frameworks.

As these technologies advance, questions arise about the scope of existing legal rights and responsibilities, and how the design, development, and deployment of algorithmic systems can progress in compliance with human rights and fundamental freedoms. The impact of AI on privacy, data protection, copyright, and human rights laws are already

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

areas of active research. Monitoring and critical analysis of the growing and rapidly evolving legal landscape, including the Artificial Intelligence Act, Digital Services Act, Data Governance Act, Data Act, Digital Markets Act, Medical Devices Regulation, AI Liability Directive, and the revision of the Machinery Directive can help identify the implications of this legislative agenda for ADR research and researchers. It can also help recognise shortcomings in legislative initiatives, or legal uncertainties that might affect – or hold back – AI or robotics research, or protection and integration of EU values and fundamental rights in the development of AI methods. In so doing, such research can close the gap between legislative requirements and the developments and deployments of AI. It can offer practical guidance for AI and robotics researchers and AI users and create awareness about the legal challenges triggered by ADR development, based on clearer understandings of the scope of this emerging and evolving regulatory framework.

As understanding of the impact of AI and robotics technologies on society increases – and as technological capabilities continue to progress, as demonstrated by recent progress in generative AI – AI and robotics technologies and AI and robotics regulation will need to co-evolve. Policy-informed research can provide analysis to support this co-evolution, increasing regulatory compliance and long-term effectiveness, clarifying legal requirements, and supporting progress in AI and robotics capabilities that facilitate legal compliance (Box 8).

## Box 8: Research directions to advance legal and regulatory compliance of ADR systems

### *Data stewardship*

Trustworthy data stewardship is the foundation for human-centric AI. In recent years, a variety of technical, legal, and organisational interventions have been proposed to help connect public interests and concerns to the ways in which data about people are used in the development of AI. These include technologies to govern who has access for data in what forms, such as Privacy-Enhancing Technologies or Federated Learning; legislative provisions for data rights and the regulation of organisations using data, such as those included in the General Data Protection Regulation and Data Governance Act; and emerging data intermediaries, including Personal Data Stores and data trusts. Implementation of many of these ideas remains challenging. Barriers to implementation where research and development can play a role in supporting progress include lack of clarity in core areas (such as the scope of data rights), low levels of know-how in how to deploy stewardship-enabling technologies, and operational challenges in building new organisational data intermediaries.

### *Impact of AI on fundamental rights*

Alongside the imperative to consider EU values in the development of AI systems – described earlier under the topic of trustworthy AI – there exists a complex web of fundamental rights affected by the design and use of AI. The complex regulatory landscape is a considerable challenge for AI researchers to comprehend and comply with relevant legal obligations. Continuing progress in AI methods also raises new

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

questions about the interaction between AI and legal rights. Active research questions include:

- The nature of copyright protection in relation to AI-generated outputs and reproductions of open-source code using AI;
- The impact of new AI methods or models on those protections, such as the implications of Large Language Models or foundation models trained on large datasets;
- The ethical and legal limits associated with using publicly available datasets;
- 'Extraction fairness' regarding the large-scale exploitation of training data;
- The personal data and privacy limitations of data scraping practices;
- The impact of algorithmic content moderation on the right to freedom of expression;
- The issues of attribution of responsibility and liability regarding automated content and AI systems between users, data scientists, and developers;
- Transparency requirements for the use of AI systems, including algorithmic decision-making in recommender systems, content moderation, and online advertising;
- Bias and discrimination, including gender-based discrimination, in algorithmic decision making and its impact on the right to equality and non-discrimination;
- The extent and scope of legal obligations for AI researchers.

***AI for regulation***

Advances in AI technologies can serve to enhance legal compliance. For example:

- Technical advances in increasing AI robustness can help ensure systems operate as desired in deployment.
- New approaches to verification and validation can help test the performance of AI systems, and linked to new forms of certification can provide a guarantee for their safety.
- Privacy-enhancing technologies, such as private federated learning, can help embed regulatory requirements in how AI technologies interact with data.
- Trust-by-design principles can integrate policy expectations about the trustworthiness of AI systems into how AI tools are developed.

## 2.8 *Research challenge: Advancing hardware for safe and energy efficient interaction between ADR technologies, humans and the environment*

In the context of robotic applications for AI technologies, there is a specific need relating to hardware that can deliver intelligent, safe, and energy-efficient interactions between robots, humans, and the environment. Designing such hardware systems requires transferable embodiment design and interconnection principles – representations that enable transfer of design and advanced control between different embodiments – coupled with low-level sensori-motor loops and reflexes. This involves multidisciplinary methods from soft robotics, cognitive mechatronics, perceptive systems, embodied intelligence, biologically-inspired reflexes and control.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

Traditional robots were built using rigid parts to achieve high position accuracy useful for many industrial applications. Over the past two decades, Europe led the development of the so-called collaborative and soft robotics fields, enabling the development of robotic applications to work alongside humans. Building on this success, further progress in embodied intelligence could come from the exploitation of smart materials to deliver robot designs where actuation, energy, sensing, reflexes, nervous and control functions are mostly embedded in and distributed throughout the body. This necessitates the exploration of new and smart materials, their production techniques and control/AI algorithms as well as new modelling methodologies to understand their multi-physical nature in order to effectively control their behaviour[12]. Aligning these developments with the sustainability aspirations of the Green Deal will rely on research directions that make the complete system, including storage, sensing, processing, and actuation, energy-efficient so that it will also substantially benefit power autonomy in mobile robotics applications.

Many of today's systems work with a single human or are embedded in a static, hierarchically-controlled system like a manufacturing plant. To increase the resilience and robustness of these systems, robots can be designed for interaction: such robots have increased robustness through redundancy and higher flexibility for a wide variety of tasks, moreover multiple robots can solve problems faster using parallelism. Building on this work, and inspired by human populations that are dynamic, flexible, and cooperative, it is possible to envisage physically interacting multi-robot systems with different capabilities, which exploit the resources on the Cloud for extending their computational capabilities and functionalities towards more resilient and adaptive control strategies.

Another strategy to help build robustness and resilience is to design, test, and train control and AI algorithms and predict robot behaviours for the different challenges. This requires reliable emulators, simulators and digital twins that realistically represent the physical world. This allows real-world challenges to be simulated and control strategies fine-tuned, before testing the developments on the real hardware platforms. Such simulations also support interaction with robots in the cyber-physical space using AR and VR technologies.

## Box 9: Advancing hardware for safe and energy efficient interaction between ADR technologies, humans and the environment

***Next-generation soft robots exploiting smart materials***

[12] For example, the development of sensitive synthetic skin is largely underrepresented in robotics, in contrast to development in audio and video. These capabilities are paramount to allow robotic hands and grippers to not only grip but also manipulate objects as needed in, e.g., the manufacturing (for continuum and flexible objects such as cables) and personal robotics challenges (e.g., the clothes).

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

There is an opportunity to develop the next generation of smart materials with sensing, actuation, energy, healing and processing capabilities to implement embodied AI using advanced processing techniques in new robot components (especially new soft grippers) to be validated in the robotic challenges, for example to manipulate soft objects like textiles in the personal **robot's** challenge. Materials need to be sustainable with self-healing, recyclability and biodegradability properties.

***Next-generation electronics (sensors, processing, communication) for physical interaction***

To better interact with the unknown environment, robotic hardware systems need to be equipped with a new generation of electronics and methods to better sense, communicate, and process multimodal sensor information. AI and control algorithms can then leverage on the generated multisensory data which not only incorporates vision and audio, but also touch and other sensing capabilities beyond capabilities of humans (radar, proximity, 3D information, etc.). This is needed to better sense the dynamically changing environments.

***Next-generation actuation technologies for safe and energy-efficient interaction***

To work safely and energy-efficiently, further advances on actuators, locking mechanisms, springs, gears, power electronics, batteries and embedded control strategies are needed. Such technologies are required to help achieve higher payload to mass ratio and more energy efficient operation.

***Collaborative multi-agent systems***

Multi-agent control systems often have very complex dynamics, which are hard to model a priori. In response, the development of combined centralized/decentralised control schemes can help compensate for weaknesses in each individual approach[13]. Progress will allow the collaborative delivery of goods or collaborative manufacturing and deployment in heterogeneous robot systems. Control strategies for role adaptation in cooperative grasping and manipulation as well as dynamic adaptation of tasks priority can support this progress. These strategies will also need to cope with computation or communication delays without ever compromising on physical safe interaction.

***Physics-enabled digital twins***

A digital twin is a virtual representation that serves to simulate a physical object or process and can be used to rapidly prototype ADR applications. In this context of robotic applications, they can replicate the real-world behaviour of the robot and its environment to a degree that is useful, also for applications where visual rendering is necessary. As such, novel AI methods for such applications can be tested before transferring them to the actual platforms. Digital twins in which the accuracy of collision detection and rigid-body or soft dynamics of the simulator is important.

[13] E.g., issues with centralized computation or modelling, data collection, and actuation. A complete centralized or hierarchical control will suffer from network delays, high dimensionality, uncertainty, jitter, unavailability of agents, etc.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# 3 Pursuing the agenda

## 3.1 Advancing a European AI research agenda

The research agenda described above spans fundamental and applied research, with different AI theories, methods, and applications contributing to advancing the frontiers of ADR technologies and their use in critical areas of need. By tapping into diverse research networks, where each NoE is driving advances in ADR research across Europe, and convening to progress research in ways that accommodate the needs of different research priorities. Links between these research challenges and the Work Packages implemented by each NoE are described in the Annex 2. This table is intended as a snapshot of current programmes, based on a loose mapping of NoE Work Packages already established by existing NoEs to the cross-cutting themes in this document; there are overlaps between themes and research topics, the nuances of which will not be represented in this high-level view. As newly initiated NoEs set up their programmes, and existing NoEs adapt theirs to changing needs, further research activities will be developed under each of these themes.

## 3.2 Supporting a European AI ecosystem

Alongside this ambitious research agenda, activity from the NoEs is helping to seed and grow the European AI ecosystem. Reflecting the needs of the wide range of potential AI applications and user groups, a variety of enabling actions are required to support individuals, communities, organisations, and governments to benefit from AI, and manage the risks associated with its use. The EU's AI policy portfolio has already identified the need to grow Europe's AI skills base; to attract leading research talent and maintain European research leadership; to support businesses to adopt AI technologies; and to provide tools, platforms, and datasets to encourage AI development. In response to these needs, the NoEs are taking action to deliver education and training; boost research collaborations and connectivity; encourage industry adoption of AI; and provide tools and toolkits for AI research. The sections that follow illustrate how NoE activities contribute to each of these areas.

### *Education and training*

- The International AI Doctoral Academy[14] (AIDA) is an initiative of AI4Media, founded and supported by the 5 ICT48 projects (AI4Media, ELISE, HUMANE-AI, TAILOR and VISION CSA), which provides training for early career researchers and professionals seeking to boost their AI skills. AIDA will be a vehicle for providing access to top-quality academic material, in various formats, including academic courses, free access to thematically organised academic material, and lectures on hot AI topics. AIDA currently includes 77 members (58 academic institutions and 19 research or industrial organizations), 121 registered lecturers and 191 registered students while it has already offered 66 educational courses in various formats (short courses, semester courses, seasonal schools) attended by more than 1,700 students/researchers/professionals in total.

[14] https://www.i-aida.org/

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- ELISE's PhD and Postdoc programme hosts a variety of networking and training activities, including bootcamps, summer schools, and workshops, to encourage early career researchers to develop networks with European peers that deliver world-leading AI research[15]. The ELISE/ELLIS PhD programme has attracted over 3000 applications, funding over 150 PhD places at leading AI labs across Europe, engaging over 100 research institutions and businesses.

### *Research collaboration and connectivity*

- The AI4Media Junior Fellows Exchange Program[16] facilitates exchanges of early career researchers that want to improve their skills and knowledge in AI for the media and society by collaborating with top European AI researchers and media companies to conduct innovative research that considers media industry needs. The program has already facilitated more than 60 exchanges of junior and senior researchers from over 40 organisations across Europe, producing important outcomes in the form of publications, open software, and open datasets[17].
- Recognising that innovation stems from cross-fertilisation of ideas, ELISE's research mobility programmes encourage both early career and established researchers to develop collaborations with peers in AI labs across Europe. It gives early career researchers the opportunity to be supervised by leading scientists, as they spend six months in a collaborating lab or working with an industry partner. Faculty-level exchanges encourage research visits that enhance scientific dialogue and collaborations across Europe, in support of ELISE's research goals. ELISE's mobility programmes have helped connect researchers across Europe, through 40 research exchanges across 10 countries.
- ELISE workshops have convened researchers from across Europe to explore diverse topics including the fundamentals of machine learning[18]; the role of AI in climate adaptation, soil health, sustainable cities, ocean restoration, and cancer prevention[19]; the ability and limitations of modern AI[20]; and more. A highly successful online crash course in High Performance Computing has also offered training for AI researchers seeking to access European compute infrastructure.[21]
- A joint project call between ELISE and HUMANE-AI is encouraging research collaborations to advance goals relating to human-centric machine learning and semantic, symbolic, and interpretable machine learning.
- To advance research in particular domains or applications TAILOR is organising challenges and benchmarks. There have been challenges related to brain age prediction, smarter mobility, cross-domain meta deep learning, and learning to run a power network.

---

[15] For example: https://www.elise-ai.eu/events/elise-mobility-program-for-phd-students-and-postdocs-two-participants-share-their-experiences
[16] https://www.ai4media.eu/junior-fellows-program/
[17] https://www.ai4media.eu/wp-content/uploads/2023/04/MobilityTestimonialsBooklet_AI4Media_final.pdf
[18] https://www.elise-ai.eu/events/elise-theory-workshop-on-machine-learning-fundamentals
[19] https://www.elise-ai.eu/events/elise-invites-you-to-attend-4-exciting-workshops-in-january-2023
[20] https://www.elise-ai.eu/events/theory-research-program-holds-workshop-on-abilities-and-limits-of-modern-learning-systems
[21] https://www.elise-ai.eu/events/european-hpc-for-ml-elise-online-crash-course-materials

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- The TAILOR Connectivity Fund reaches out to the many excellent labs and organisations across Europe to encourage research collaborations, with a particular focus on supporting young researchers to gain valuable experience. There are three open calls every year to provide funding for research visits and workshops.
- To encourage collaborations across research and industry, TAILOR, HumanE AI Net, VISION and CLAIRE AISBL convene Theme Development Workshops (TDW). These workshops identify strategic AI research areas and challenges in specific sectors, and encourage activities to address them beyond the event (e.g., joint working groups, research papers, challenges, hackathons, transfer labs).

### *Industry collaboration and use cases*

- During 2022-2024, the AI4Media open calls will fund 20 projects led by industry or academia entities with up to €50,000 equity-free funding per project to develop new research in AI or innovative AI applications for the media[22].
- AI4Media implements 7 use cases through close collaboration between AI researchers and media industry professionals (European media organisations or content related companies). Informed by emerging market opportunities and urgent industry challenges, the use cases cover a variety of topics such as disinformation, news research and production, media moderation, organisation of audiovisual archives, game design, human-machine artistic co-creation, social science research etc. They aim to address significant challenges currently faced by different media industry sectors and to highlight how AI applies throughout the media industry value chain, from research and content creation to production, distribution, consumption/interaction, performance and quality measurement[23].
- ELISE's SME support scheme offers grants and mentoring from leading researchers to small and medium-sized businesses seeking to use AI in their work. The scheme so far has attracted over 800 applications. Its first round provided 16 businesses with tailored support to develop new AI tools[24]. This cohort will shortly double in size, after a second round of this scheme. The resulting projects have included AI-enabled medical devices for skin cancer diagnosis, state-of-the-art automated language translation services, enhanced cyber security systems, enhanced diagnostics and personalised medicine systems for treatment of eye disease and breast cancer, automated management of physical infrastructure in energy networks, and more.
- Close collaboration with industry through ELISE's uses cases allows rapid transfer of knowledge and expertise between academic and industry partners, and translation of research advances to real-world impact[25]. 11 use cases are integrated into ELISE research programmes, tackling challenges including environment perception for autonomous driving, benchmarks for machine

[22] https://www.ai4media.eu/meet-ai4media-open-call-1-winners/; https://www.ai4media.eu/open-call-2-winners/
[23] https://www.ai4media.eu/use-cases/
[24] See: https://www.elise-ai.eu/events/how-have-the-16-champions-from-elises-1st-open-call-been-doing
[25] For example: https://www.elise-ai.eu/events/elise-industry-stories

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

learning robustness, optimisation of warehouse logistics, data-efficient video analysis, and more[26].

### *Tools and toolkits*

- AI4Media has created a set of novel open-source AI tools, open access datasets, and open access publications related to AI for the media. These include:
  - Open access software[27] for machine learning, multimedia content analysis and trustworthy AI;
  - Open datasets[28] for media AI research;
  - Open access roadmap[29], white papers[30] and results in brief[31].
  - In addition, AI4Media has launched a Media AI Observatory[32], aiming to monitor, aggregate, study, and interpret information on topics relevant to Media AI, with the purpose to support a better understanding of AI developments and their impact on society, economy, and people.

[26] For further information, see table 3: https://uploads-ssl.webflow.com/5f55e90f6a7294f66f94d30d/609a39235e25e3d64bb65053_ELISE-strategic-research-agenda-web.pdf
[27] https://www.ai4media.eu/software/
[28] https://www.ai4media.eu/open-datasets/
[29] https://www.ai4media.eu/roadmap-ai-for-media/
[30] https://www.ai4media.eu/whitepapers/
[31] https://www.ai4media.eu/results-in-brief/
[32] https://www.ai4media.eu/observatory/

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# 4 Looking ahead

Progress under each of the areas described above is being delivered through work packages across AI4Media, ELISE, ELSA, euROBIN, HUMANE-AI, and TAILOR. The Strategic Research Agendas from each of these networks sets out their vision for AI research, and plans to translate their ambitions for AI research to technological progress (Annex 1). As the work of these networks develops, updates to these Strategic Research Agendas will provide insights into the frontiers of Europe's AI capabilities.

International competition for AI research talent, for technical leadership, and for policy influence will continue to grow. Current debates about generative AI highlight what is at stake. In this environment, investment in the European AI landscape is crucial, if Europe is to have an influence on the pathways for technological development and the ways in which AI technologies shape society. Europe must be at the forefront of technological innovation – pursuing world-leading, excellent ADR research – while building an ecosystem that can translate that research to application, boosting the diverse strengths of local innovation ecosystems across the continent.

To secure the European strategic autonomy needed in a turbulent world, significant further investments are needed in research, innovation, and infrastructure. The EU has already made major investments to lay the foundations for the European AI ecosystem. However, the context for such investments is an ecosystem where large corporate players are able to devote resources beyond the scale possible in the public sector. To give a concrete example, OpenAI received an investment of 10 billion USD besides access to the vast Microsoft computing infrastructure.

To be globally competitive European researchers need to be able to do large-scale ADR research significantly faster than today. This requires an agile research and innovation infrastructure that can respond to emerging areas of interest or need. Community-building by the existing NoEs provides a forum for catalysing new collaborations; scaling these and translating research to impact can be achieved through flexible funding instruments whose design reflects the fast-moving pace of the field while providing a grounding that gives the field confidence to innovate. A lesson learned from the pandemic, for example, is that researchers with stable long-term faculty funding was the quickest to respond to the urgent society need. Having very specific calls on narrow topics, for example, risks missing the most important research questions.

The exascale computing infrastructure being developed as part of EuroHPC is a good step towards enabling researchers to operate at scale. There is an opportunity to push this work further; considering that the EU has 1000's of researchers the compute per researcher or even research institute is still very low. If these resources are allocated evenly, the risk is that none of researchers have the capability of doing large scale research. To give a sense of scale, one exascale computer is sufficient to train one of the largest language models at the time.

The NoEs first 24 months of operation have already generated collaborations, research insights, and practical case studies that are helping to advance AI R&D across Europe. These signals of success demonstrate how the NoEs are seeding an ecosystem that

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

connects from local capabilities to international priorities, taking strength from the diversity of research interests and opportunities across the continent by connecting researchers and industry in local innovation environments that deliver on-the-ground benefits for citizens, businesses, and society. The breadth of these activities shows the opportunity for Europe to pursue an AI agenda that delivers real-world benefits for citizens and organisations. Sustained and increased investment in these networks is needed to reap the benefits of the collaborations established via the ICT-48 programme in the long term.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# Annex 1 – Summaries of the Strategic Research Agendas of each Network of Excellence

## AI4Media

AI4Media's vision is that of a European Network of Excellence in Artificial Intelligence for the Media, Society and Democracy that will glue together the pieces of the currently fragmented European AI landscape and promote a unique brand of European Media AI. AI4Media has built a network of experts, including both leading researchers in media AI from academia and research as well as top European media companies that use AI to enhance their operations and business opportunities. Together, they address significant technical, legal, ethical and application challenges, aiming to address pressing needs of the media industry and significant societal problems.

The Media are already benefiting from AI advancements and AI-driven applications that can significantly facilitate, enhance, or transform important tasks, including smart assistants, smart recommender systems, content personalisation, automatic content creation, multi-modal content search, multilingual translation, disinformation and manipulated content detection, social media analysis and trend detection, online debate analysis, forecasting and decision support-systems, and many more. Further advances in AI have the potential to transform the media industry and revolutionise how operations run and how content is created, delivered, and consumed while they can also offer trustworthy solutions with a societal impact, aiming to improve political participation, increase social cohesion, equip citizens against disinformation, and encourage healthy debates and social interaction.

To realise this enormous potential of AI will require breakthroughs in several domains such as:

- Machine learning, aiming to address important challenges of current ML techniques, including learning with few data, learning on-the-fly, transfer of knowledge and optimal AI architectures. In addition, research should also focus on distributed AI systems running on heterogeneous devices but also disruptive technologies currently at the laboratory stage such as Quantum-assisted Reinforcement Learning.
- Content-centred AI technology, valuable for the media industry and marketable as end-user services, such as multimedia metadata extraction, summarisation, and clustering, automatic audiovisual content generation and enhancement, linguistic analysis, and media-specific core technologies to improve learning performance.
- Human and society-centred AI technology, to equip citizens and media professionals with a set of tools that can be used to counter the effects of media manipulation and disinformation, enhance the understanding of online debates,

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

support the analysis of perceptions of social media and the effects of online data sharing, improve content recommendation and moderation, and improve local news understanding without being limited by language barriers.

- Trustworthy AI techniques, aiming at providing a framework for the development of the technologies mentioned above that guarantees their suitability with respect to democratic and ethical values. Research should focus on issues of robustness against threats and malicious attacks, explainability of AI decisions, fairness and mitigation of bias of AI models, and techniques for privacy-preserving AI.

These AI advances will be integrated and evaluated in real-world use cases, aiming to address significant challenges currently faced by different media industry sectors and to highlight how AI applies throughout the media industry value chain, from research and content creation to production, distribution, consumption/interaction, performance and quality measurement. The use cases cover a variety of media and societal topics such as disinformation, news research and production, organisation of media archives and content moderation, game design, human-machine artistic co-creation, and social science research.

In parallel to delivering the next generation of AI research at the service of media, AI4Media aims to establish a Media AI Observatory[33] to monitor the legal and technological landscape as well as the impact of media AI on the society, economy and democracy. The Observatory provides an overview of the existing EU policy and legal initiatives and their impact on future AI research for the media industry, analyses ethical, societal, environmental, and economic concerns, and provides easy access to leading experts in this domain.

Implementing our vision of AI as a human-centred, trusted, and beneficial enabling technology in the service of media and society, requires supporting in practice the next generation of AI talent in Europe by offering opportunities for top AI education and skill development while also supporting entrepreneurship and innovative ideas. To this end, AI4Media has established the International AI Doctoral Academy, a joint ICT-48 instrument to support world-level AI education and training for PhD/postdoc AI researchers. In addition, it provides mobility opportunities for young researchers and media professionals. And lastly, it will fund and support SMEs, start-ups and research labs that want to develop innovative applications and research for the Media. These activities will further strengthen the European AI research community.

There is overwhelming agreement that AI will drive the majority of innovation across nearly every industry sector in the next decade. The media industry should be ready to exploit new AI advances but also mitigate possible risks, to enjoy the full potential of this technology and transform the industry. The AI4Media Network of Excellence aims to play an important role in this transformation, by bringing together leading research and industry players in this domain to strengthen the competitiveness and growth of the European media industry and increase Europe's innovation capacity in media AI.

[33] https://www.ai4media.eu/observatory/

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

## ELISE

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 programmes, this Strategic Research Agenda sets out ELISE's roadmap for creating AI technologies that are technically advanced, robust in deployment

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

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.

Progress delivering the research agenda is achieved through its research programmes, which tackle cutting-edge research challenges in:

- Quantum and physics-based machine learning
- Robust machine learning
- Interactive learning and interventional representations
- Machine learning and computer vision
- Natural Language Processing
- Machine learning in Earth and climate sciences
- Symbolic machine learning
- Human-centric machine learning
- Multimodal learning systems
- Theory, algorithms and computations of Modern Learning Systems
- Machine learning for health
- Natural intelligence
- Geometric deep learning

In pushing forward the frontiers of machine learning research, these programmes engage with five cross-cutting themes that connect AI research advances to wider issues of social and economic concern. These themes relate to:

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

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 (ELLIS), 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.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

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.

In its first three years of operation, ELISE has already:

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

Reflecting the needs and interests of the AI community, ELISE will continue to update this Strategic Research Agenda throughout its lifetime[34].

## ELSA

The ELSA (European Lighthouse on Secure and Safe AI) network is focused on promoting secure and safe AI solutions to the trustworthy AI challenges.

ELSA research builds upon methods with a strong theoretical foundation for provable security and safety, such as certifiable robustness and differential privacy. These approaches provide a basis for sustainable trustworthiness that will endure future technical developments and avoid arms races of increasingly sophisticated attacks and defences.

ELSA promotes a cross-disciplinary approach to questions of human agency and oversight, combining technical expertise with ethical and governance expertise. We also advance rigorous solutions to handling the uncertainty of predictions as well as transparency, explainability and interpretability.

[34] See: https://www.elise-ai.eu/sra-refresh/strategic-research-agenda-refresh

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

ELSA is currently in the process of writing its own more detailed SRA. This will be published in late summer 2023 at https://elsa-ai.eu/sra and updated here when possible.

## HUMANE-AI

The research agenda of the HUMANE-AI project describes HUMANE-AI in terms of the evolution with respect to the research questions described in the proposal. We focus on research questions that are directly related to the project's vision of AI that enhances human capabilities and empowers citizens both in individual and collective/social level while observing ethical and fundamental rights concerns "by design". We leave the work on more generic AI research agenda to collaboration with road mapping activities within VISION and our collaboration with the CLAIRE road mapping effort. We also closely collaborate with the EILSE project towards joint research agenda items that combines our AI-human based angle with ELISE more fundamental ML oriented vision (currently a joint call for proposals of microprojects is being defined).

Most significant evolution of the research agenda has taken place at the interface of the individual WPs as result of synergies between the different communities (core AI, HCI, Ubiquitous computing, Social Science, Complexity Science) producing significant, critical insights on what is needed to move towards a vision of truly human centric European AI. These include:

1. A hierarchical framework to provide a taxonomy of research problems for collaborative AI systems. Solutions to problems at any level can build on techniques and solutions developed at lower levels. The framework is proposed as a research roadmap, grouping related challenges into subcategories according to the information that is processed and the nature of the interaction. This facilitates formulation and comparative evaluation of competing techniques.
2. An understanding how the question of a common ground and shared representations relates to different types of interaction and what are key directions that we need to explore to facilitate our vision of human centric AI. In particular, we emphasize the role that the concept of narrative, leveraging recent advances in NLP and self-supervised multimodal representation learning can play across the WPs.
3. An extension of the concept of trustworthy and explainable AI from a definition focused on technical aspects to a user-oriented approach that emphasizes systems that act and interact in a way that people and the society feel comfortable trusting and using.
4. The insight that we need to work on a research methodology and infrastructure that brings together the different cultures of the discipline involved. This includes a definition of evaluation standards and experimental methodologies as well as the creation of data sets and tools.

Within the individual WP research agendas, the adaptations focus on incorporating new developments in the respective fields and synchronising with the overall crosscutting adaptations outlined above.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

We consider this research agenda to be a “living document” that will be continuously updated as new insights and ideas arise from the project work and overall progress in the field.

## TAILOR

The TAILOR Strategic Research and Innovation Roadmap (SRIR) for Trustworthy AI will define the foundations of Trustworthy AI for the years 2022-2030. It aims to boost research on Trustworthy AI by clearly defining the major research challenges. The objectives of the SRIR are:

1. Providing guidelines for strengthening and enlarging the pan-European networks of research excellence centres on the foundations of Trustworthy AI.
2. Defining paths for advancing the scientific foundations for Trustworthy AI and translating them into technical requirements to be adopted by industry.
3. Identifying directions for fostering collaborations between academic, industrial, governmental, and community stakeholders on the foundations for Trustworthy AI.

Detailed priorities, actions, and timing needs further discussion within and outside the TAILOR project and will be detailed in the second version of the SRIR. For the time being, both short- and long-term recommendations are proposed.

### *Measure and assess Trustworthy AI dimensions*

Short term

- Develop methods for measuring and evaluating the trustworthiness of AI systems.

Long term

- Develop tools for continuously auditing and adapting Trustworthy AI systems: monitoring, dynamically identifying issues, and mitigating them.

### *Scientific challenges*

Short term

- Develop human interpretable formalisms to enable synergistic collaboration between humans and machines with regards to the criteria of being explainable, safe, robust, fair, accountable; and develop standards and metrics to quantify the grade to which these criteria are satisfied.
- Develop methods for integrating model-based and data-driven approaches to autonomous acting.
- Develop a broad range of AutoAI benchmarks to facilitate development and critical assessment of AutoAI techniques and systems.
- Expand current AutoAI techniques to better meet the demands of real-world applications, including multiple interacting design objectives (with aspects of trustworthiness), scalability, scope and ease of use.

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

- Develop integrated representations and frameworks for learning, reasoning and optimisation based on probability, logic, neural networks, ontologies, knowledge graphs and constraints.

Long term

- Develop the science, techniques and tools for adjustable autonomy for autonomous AI agents. In particular, equip autonomous agents with the ability to understand when certain decisions that it could take on its own are questionable or unethical, and human supervision should be required.
- Develop a computational theory of mind that considers mental attitudes such as beliefs, knowledge, goals, intentions, capabilities, emotions, and integrates them in a computational effective fashion into autonomous acting.
- Enable the broad, safe, and efficient use of AutoAI techniques across all sectors of industry and society, especially in contexts where limited AI expertise is available.
- Develop a unifying theory and framework of learning, reasoning and optimisation that bridges the gap between the data- and knowledge-driven and the symbolic and sub symbolic approaches in AI.

*Innovation*

Short term

- Develop generic operational models of hybrid approaches allowing their reuse in various domains and propose metrics/benchmarks for validating these models.
- Consider that transparency (incl. explainability) targets different kinds of users: developers, domain experts, regulators, “users” (citizens, patients, etc.).

Long term

- Implement Trust by Design: Enable the design and verification of trusted AI systems according to appropriate legal, social and technical criteria and aspects, focusing in particular on critical and risky applications.

## euROBIN

www.eurobin-project.eu

A Network of Excellence on Artificial Intelligence-driven Robotics seeking to bring together the robotics community and to benefit science, industry, and society while promoting European values. The network is a facilitator of knowledge transfer and exchange between research institutions and industry partners.

*Objectives*

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

The main vision of euROBIN is that of a European ecosystem of robots that share their data and knowledge and are able, based on their diversity, to jointly learn to perform the endless variety of tasks in human environments.

euROBIN's main goals are:

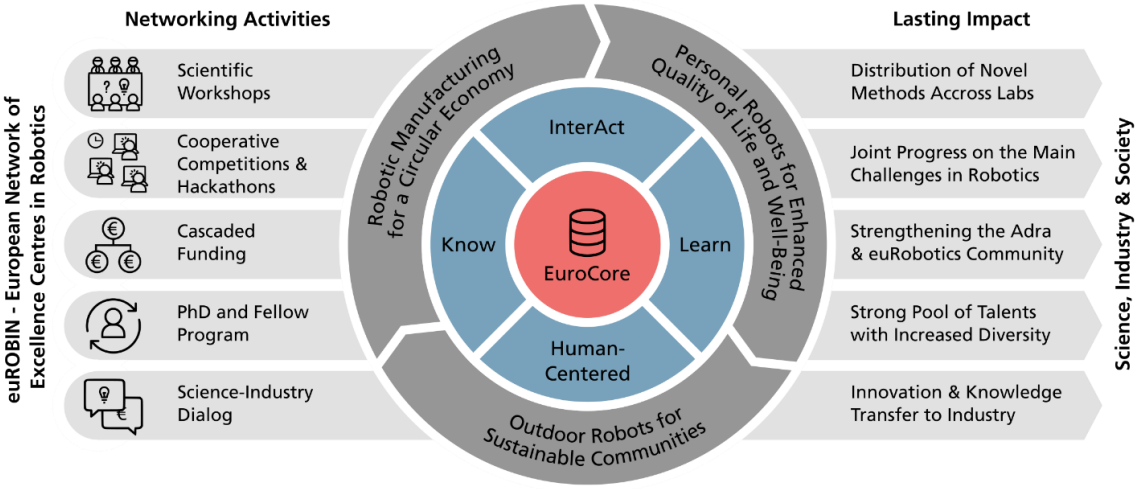
- Addressing the main scientific and technological challenges hampering the breakthrough and large-scale deployment of robotics: euROBIN focuses on making cognition-enabled Robotics solutions more transferable and reusable among scientists and by new industries. This is crucial to better join forces in Europe in this dynamic and very competitive field.
- Providing a stage for cooperation and exchange of scientific knowledge and talents between the most outstanding robotics labs in Europe in the areas of knowledge representation, physical interaction, robotic learning and human-robot interaction
- Generating a nucleus to which the community at large can adhere, enabling ground-breaking new applications in industrial, personal and outdoor robotics in Europe
- The euROBIN network proposes the novel concept of robot cooperative competition focused on the three main application domains defined in the robotics roadmap of Horizon Europe:
    - Robotic manufacturing for a circular economy
    - Personal robots for enhanced quality of life and well-being
    - Outdoor robots for sustainable communities
- The network will strongly interact with and benefit from other collaborative EU-initiatives such as the euRobotics association and the AI DATA Robotics Association (Adra), empowering the strength of AI & Robotics in Europe. It builds on and contributes to the assets on the AI-on-Demand platform.

*Expected impact*

1. Scientific breakthrough: A scientific and technological framework for transferable AI-powered and cognitively-enabled robots is the basis for an ecosystem of heterogeneous, jointly learning intelligent machines.
2. Economic: Increase the share of the European robotics industry at the international robotics market in the areas of interactive industrial manufacturing, personal/healthcare robotics, supply chain robotics.
3. Long-Term economic: In the areas industrial manufacturing, personal robotics, logistics supply chain, emerging mass-market robotics products benefit from the transferability concepts and/or the EuroCore.
4. Societal: Emergence of AI-powered robotics mass markets contribute sustainable solutions to the challenges of international industrial manufacturing competitiveness, demographic change, and cyclic supply chains.

The euROBIN network essentially contributes, through its focus on transferability, interoperability, cooperation, and its instruments including the EuroCore repository, to international leadership of European robotics science and technology.

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).




These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

# Annex 2 – Links between research challenges and the Work Packages implemented by each NoE

| | *AI4Media* | *ELISE* | *ELSA* | *HUMANE-AI* | *euROBIN* | *TAILOR* |
|---|---|---|---|---|---|---|
| ***Building the technical foundations of trustworthy AI*** | • Explainability, Robustness and Privacy in AI (WP4)<br>• Content-centred AI (WP5)<br>• Human- and Society-centred AI (WP6)<br>• European AI Vision, Policy and Common Research Agendas (WP2) | • Machine Learning for Health<br>• Robot Learning: Closing the Reality Gap!<br>• Human-centric Machine Learning<br>• Interactive Learning and Interventional Representations<br>• Machine Learning and Computer Vision<br>• Quantum and Physics-Based | • Focus on fundamental and excellence science in order to deliver rigorous methodology for principled solution in secure and safe AI<br>• 3 grand challenges:<br>○ Robustness guarantees and certification<br>○ Private and robust collaborative learning at scale | • Pillar 1: Human-in-the-loop machine learning, reasoning, and planning<br>• Pillar 3: Human-AI collaboration and interaction<br>• Pillar 4: Societal awareness<br>• Pillar 5: Legal and ethical bases for responsible AI | • Human-centred perspective (WP3)<br>• Forming a Human-Centered Robotics Collegium (WP3)<br>• Human-Aware situation assessment systems (WP3)<br>• Human-Aware synthesis of robotic behaviour (WP3)<br>• Human-Robot Interaction for collaborative | • Trustworthy AI (WP3): Explainable AI systems<br>• Trustworthy AI (WP3): Safety and robustness<br>• Trustworthy AI (WP3): Fairness, equity and justice by design<br>• Trustworthy AI (WP3): Accountability and reproducibility<br>• Trustworthy AI (WP3): Respect for privacy |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Machine Learning<br>• Robust Machine Learning<br>• Semantic, Symbolic and Interpretable Machine Learning<br>• AI certification white paper<br>• AI regulation and trust white paper<br>• Use cases in security, robustness, and AI governance | ○ Human-in-the-loop decision making: Integrated governance to ensure meaningful oversight<br>• Research on core technologies including differential privacy, certification, explainability<br>• Understanding worst case performance in order to mitigate attacks<br>• Understanding attacks and defences for AI | | manipulation tasks (WP1)<br>• Next-generation actuation technologies for safe and energy-efficient interaction (WP1)<br>• Human-robot interaction for coexistence and cooperation (WP2)<br>• Developing open question answering capabilities (WP6)<br>• Ethics Management (WP9) | • Trustworthy AI (WP3): Sustainability<br>• Unifying paradigms (WP4)<br>• Acting (WP5)<br>• Learning and reasoning in social contexts (WP6)<br>• AutoAI (WP7) |
| ***Integrating AI into deployed systems*** | • New Learning Paradigms & | • Machine Learning for Health | • ELSA innovation lab | • Pillar 1: Human-in-the-loop machine | • Three workpackages in euROBIN | • Trustworthy AI (WP3): |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| Distributed AI (WP3)<br>• Explainability, Robustness and Privacy in AI (WP4)<br>• Content-centred AI (WP5)<br>• Human- and Society-centred AI (WP6)<br>• Use cases & demonstrators in media, society and politics (WP8):<br>○ Use Case 1: AI for social media and Against Disinformation<br>○ Use Case 2: AI for News - The Smart News Assistant<br>○ Use Case 3: AI in Vision - High quality | • Robot Learning: Closing the Reality Gap!<br>• Human-centric Machine Learning<br>• Interactive Learning and Interventional Representations<br>• Machine Learning and Computer Vision<br>• Machine Learning for Earth and Climate Sciences<br>• Multimodal Learning Systems<br>• Natural Language Processing | as transfer platform<br>• Open industry call<br>• Cybersecurity aspects of AI<br>• Threat modelling | learning, reasoning, and planning<br>• Pillar 2: Multimodal perception and modelling<br>• Pillar 3: Human-AI collaboration and interaction<br>• Pillar 4: Societal awareness<br>• Pillar 5: Legal and ethical bases for responsible AI | are dedicated to integrating AI in deployed robotic systems:<br>• WP1: Robotic Manufacturing for a Circular Economy: Selection of industrial challenges, Setting up of manufacturing research framework, Research on manufacturing challenges outcomes<br>• WP2: Personal Robots for Enhanced Quality of Life and Well-Being: Scenario definition and "Personal | Explainable AI systems<br>• Trustworthy AI (WP3): Safety and robustness<br>• Trustworthy AI (WP3): Fairness, equity and justice by design<br>• Trustworthy AI (WP3): Accountability and reproducibility<br>• AutoAI (WP7)<br>• Trustworthy AI (WP3): Respect for privacy<br>• Unifying paradigms (WP4)<br>• Acting (WP5)<br>• Learning and reasoning in social contexts |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| ○ Video Production and Content Automation<br>○ Use Case 4: AI for Social Sciences and Humanities<br>○ Use Case 5: AI for Games<br>○ Use Case 6: AI for Human Co-creation<br>○ Use Case 7: AI for Content Organisation and Content Moderation | • Robust Machine Learning<br>• Semantic, Symbolic and Interpretable Machine Learning<br>• Use cases in infrastructure management, industrial digitalisation, cybersecurity, healthcare, and farming. | | | Robotics Kit", Human-robot interaction for coexistence and cooperation, Translational efforts toward robot deployment in retail stores<br>• WP3: Outdoor Robots for Sustainable Communities: Personal delivery, multi-robot cooperative delivery, Delivery in complex scenario with knowledge transfer, multi-user list delivery in unknown scenario, Translational | |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| | | | | efforts toward logistics applications | |
| ***Enhancing human capabilities with collaborative AI*** | • New Learning Paradigms & Distributed AI (WP3)<br>• Explainability, Robustness and Privacy in AI (WP4)<br>• Human- and Society-centred AI (WP6)<br>• Use cases & demonstrators in media, society and politics (WP8) | • Machine Learning for Health<br>• Robot Learning: Closing the Reality Gap!<br>• Human-centric Machine Learning<br>• Natural Intelligence<br>• Robust Machine Learning<br>• Use cases in healthcare and industrial digitalisation. | • Human in the loop decision-making<br>• AI governance and oversight | • Pillar 1: Human-in-the-loop machine learning, reasoning, and planning<br>• Pillar 2: Multimodal perception and modelling<br>• Pillar 3: Human-AI collaboration and interaction<br>• Pillar 4: Societal awareness<br>• Pillar 5: Legal and ethical bases for responsible AI | • HRI for collaborative manipulation tasks (WP1)<br>• Human-robot interaction for coexistence and cooperation (WP2)<br>• Transfer of task knowledge to jointly move forward in personal assistance (WP2)<br>• Translational efforts toward robot deployment in retail stores (WP2)<br>• Next-generation actuation technologies for safe and energy-efficient | • Trustworthy AI (WP3): Explainable AI systems<br>• Trustworthy AI (WP3): Safety and robustness<br>• Trustworthy AI (WP3): Accountability and reproducibility<br>• Unifying paradigms (WP4)<br>• Acting (WP5)<br>• Learning and reasoning in social contexts (WP6) |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | interaction (WP4)<br>• Developing open question answering capabilities (WP6)<br>• Human-Aware situation assessment systems (WP7)<br>• Human-Aware synthesis of robotic behaviour (WP7) | |
| ***Accelerating research and innovation with AI*** | • Explainability, Robustness and Privacy in AI (WP4)<br>• Human- and Society-centred AI (WP6)<br>• Content-centred AI (WP5)<br>• Use cases & demonstrators in media, society and politics (WP8) | • Machine Learning for Health<br>• Human-centric Machine Learning<br>• Interactive Learning and Interventional Representations<br>• Machine Learning for Earth and | • Competitions and benchmarks to accelerate research and target key challenges<br>• Challenges defined jointly by academia and industry | • Pillar 1: Human-in-the-loop machine learning, reasoning, and planning<br>• Pillar 2: Multimodal perception and modelling<br>• Pillar 3: Human-AI collaboration and interaction | • The European Robotics Collaborative Repository (EuroCore) will foster knowledge and technology transfer to small and medium robotics enterprises and large | • Trustworthy AI (WP3): Explainable AI systems<br>• Trustworthy AI (WP3): Safety and robustness<br>• Trustworthy AI (WP3): accountability and reproducibility<br>• Acting (WP5) |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Climate Sciences<br>• Multimodal Learning Systems<br>• Natural Language Processing<br>• Robust Machine Learning<br>• Semantic, Symbolic and Interpretable Machine Learning<br>• Use cases in health and climate sciences | | | companies (WP8)<br>• Cascade Funding Outreach: Up to 38 projects will be funded in open calls for industry and research (WP8)<br>• Science industry dialogue (WP8) | • Learning and reasoning in social contexts (WP6)<br>• Industry, Innovation and Transfer (WP8) |
| ***Understanding the interactions between AI, social needs, and large-scale socio-technical systems*** | • European AI Vision, Policy and Common Research Agendas (WP2)<br>• Explainability, Robustness and Privacy in AI (WP4) | • Machine Learning for Health<br>• Human-centric Machine Learning<br>• Machine Learning for Earth and | • Interdisciplinary approach<br>• Outreach and awareness | • Pillar 1: Human-in-the-loop machine learning, reasoning, and planning<br>• Pillar 3: Human-AI collaboration and interaction | • Translational efforts toward robot deployment in retail stores (WP2)<br>• Human-robot interaction for coexistence | • Trustworthy AI (WP3): Explainable AI systems<br>• Trustworthy AI (WP3): Safety and robustness<br>• Trustworthy AI (WP3): fairness, |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| • Content-centred AI (WP5)<br>• Human- and Society-centred AI (WP6)<br>• Interviewing more than 20 AI Speakers in the AI-Cafe about the interaction between AI and the social needs in the context of the SDG (Societal Development Goals) (WP7)<br>• Use cases & demonstrators in media, society and politics (WP8) | Climate Sciences<br>• Robust Machine Learning<br>• White paper on AI regulation and trust | | • Pillar 4: Societal awareness<br>• Pillar 5: Legal and ethical bases for responsible AI | and cooperation (WP2)<br>• Transfer of task knowledge to jointly move forward in personal assistance (WP2)<br>• HRI for collaborative manipulation tasks (WP1)<br>• Next-generation actuation technologies for safe and energy-efficient interaction (WP4)<br>• Developing open question answering capabilities (WP6)<br>• Human-Aware situation assessment systems (WP7) | equity and justice by design<br>• Trustworthy AI (WP3): accountability and reproducibility<br>• Trustworthy AI (WP3): Respect for privacy<br>• Trustworthy AI (WP3): Sustainability<br>• Acting (WP5)<br>• Learning and reasoning in social contexts (WP6)<br>• Industry, Innovation and Transfer (WP8) |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| ***Advancing fundamental theories, models, and methods*** | • New Learning Paradigms & Distributed AI (WP3)<br>• Explainability, Robustness and Privacy in AI (WP4)<br>• Content-centred AI (WP5)<br>• Human- and Society-centred AI (WP6)<br>• | • Machine Learning for Health<br>• Geometric Deep Learning<br>• Human-centric Machine Learning<br>• Interactive Learning and Interventional Representations<br>• Machine Learning and Computer Vision<br>• Multimodal Learning Systems<br>• Natural Intelligence<br>• Natural Language Processing<br>• Quantum and Physics-Based Machine Learning | • Rigorous methodology is at the core<br>• 3 research programs<br>• Technical robustness and safety<br>• Privacy preserving techniques and infrastructure<br>• Human agency and oversight<br>• Threat modelling<br>• Adversarial approach | • Pillar 1: Human-in-the-loop machine learning, reasoning, and planning<br>• Pillar 2: Multimodal perception and modelling<br>• Pillar 3: Human-AI collaboration and interaction | • Three workpackages in euROBIN are dedicated to integrating AI in deployed robotic systems:<br>•<br>• WP1: Interact: Physical interaction, Collaborative multi-agent systems, Physics-enabled digital twin.<br>• WP5: Learn: Machine learning for robotics: Representation learning: theory, analytic and data-driven models, learning from small datasets, sequential | • Trustworthy AI (WP3)<br>• Trustworthy AI (WP3): Safety and robustness Trustworthy AI (WP3): accountability and reproducibility<br>• Trustworthy AI (WP3): Respect for privacy<br>• Trustworthy AI (WP3): Sustainability<br>• Unifying paradigms (WP4)<br>• Acting (WP5)<br>• Learning and reasoning in social contexts (WP6)<br>• AutoAI (WP7) |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| | • Robust Machine Learning<br>• Semantic, Symbolic and Interpretable Machine Learning | | | decision-making and Reinforcement Learning, Theory of skill transfer and adaptation, Interaction transfer: integrating perception, learning and control,<br>• WP6: Know: Knowledge Representation for robotics, Robot reasoning methods, Memory systems for robot agents, developing open question answering capabilities.<br>• WP6: Human-centred Robotics, | |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Human-Aware situation assessment systems, Human-Aware synthesis of robotic behaviour | |
| ***Ensuring legal compliance of AI systems*** | • European AI Vision, Policy and Common Research Agendas (WP2)<br>• Moderating AI-Cafes on ethical and legal compliance of AI systems (WP7) | • Robot Learning: Closing the Reality Gap!<br>• Human-centric Machine Learning<br>• Robust Machine Learning<br>• Semantic, Symbolic and Interpretable Machine Learning<br>• Use cases in AI governance<br>• White papers on certification and regulation | • Human agency and oversight<br>• Key techniques (e.g., certification, differential privacy) in order to comply with regulations<br>• Risk assessment<br>• Threat modelling | • Pillar 4: Societal awareness<br>• Pillar 5: Legal and ethical bases for responsible AI | • euROBIN will form a “Human-Centered Robotics Collegium” in WP7, which will advancing legal frameworks for human-robot interaction | • Trustworthy AI (WP3): fairness, equity and justice by design<br>• Trustworthy AI (WP3): Respect for privacy |

These projects have received funding from the European Union’s Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | | |
|---|---|---|---|---|---|---|
| ***Advancing hardware for safe and energy efficient interaction between ADR technologies, humans and the environment*** | • New Learning Paradigms and Distributed AI (WP3) | • Quantum and Physics-Based Machine Learning<br>• Robot Learning: Closing the Reality Gap!<br>• Machine learning for Earth and climate sciences<br>• Roadmap for European compute infrastructure | | • Pillar 3: Human-AI collaboration and interaction | • WP1: Robotic Manufacturing for a Circular Economy: Selection of industrial challenges, Setting up of manufacturing research framework, Research on manufacturing challenges outcomes<br>• WP2: Personal Robots for Enhanced Quality of Life and Well-Being: Scenario definition and "Personal Robotics Kit", Human-robot interaction for coexistence and cooperation, Translational | • Acting (WP5)<br>• Learning and reasoning in social contexts (WP6) |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).

| | | | | | |
|---|---|---|---|---|---|
| | | | | efforts toward robot deployment in retail stores<br>• WP3: Outdoor Robots for Sustainable Communities: Personal delivery, multi-robot cooperative delivery, Delivery in complex scenario with knowledge transfer, Multi-user list delivery in unknown scenario, Translational efforts toward logistics applications | |

These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).







Supported by:


These projects have received funding from the European Union's Horizon 2020 research and innovation programme under the following Grant agreements: No 951911 (AI4Media), No. 952070 (VISION), No. 952026 (HumanE-AI Net), No 951847 (ELISE), No. 952215 (TAILOR), No. 101070596 (euROBIN), and No. 101070617 (ELSA).