SHERPA

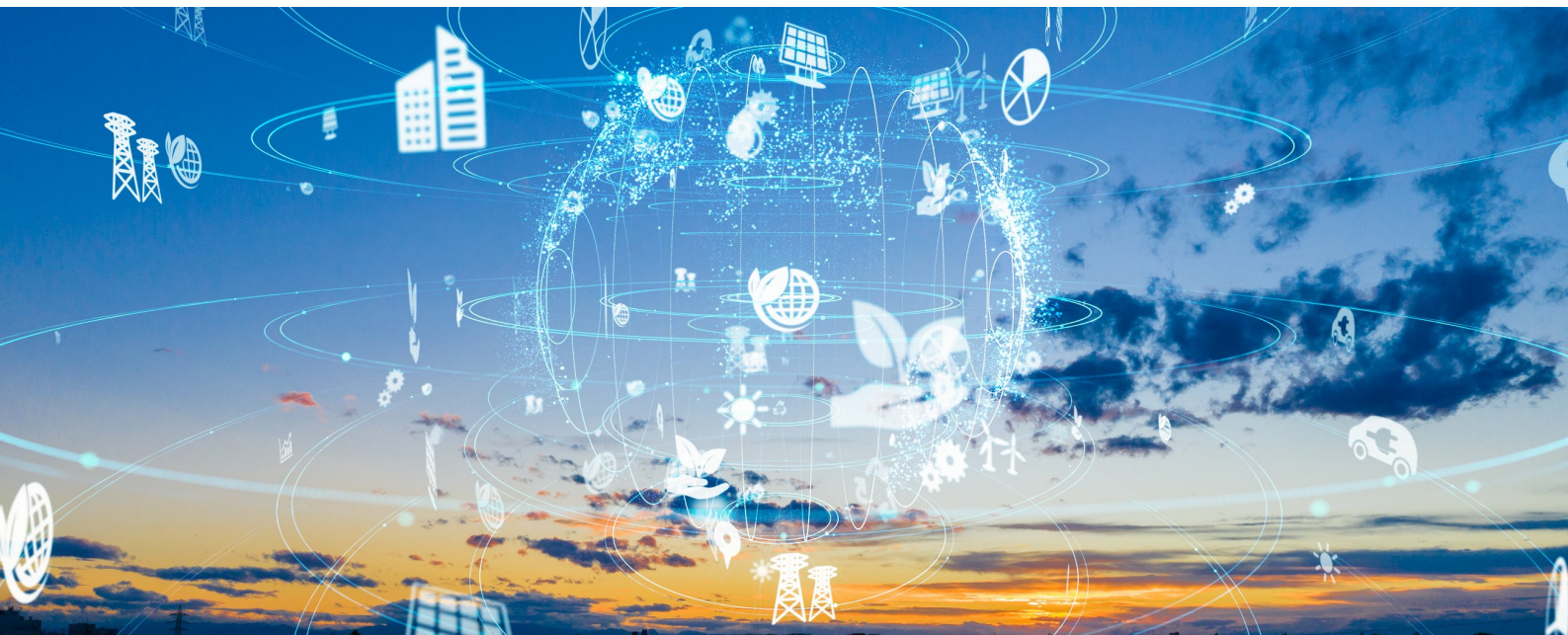## Shaping the ethical dimensions of smart information systems– a European perspective (SHERPA)

Deliverable No. 4.3: SHERPA Final Recommendations

30 April 2021

Document Control

Deliverable	D4.3 SHERPA final recommendations
WP/Task Related	WP4: Evaluation, validation and prioritisation
Delivery Date	April 2021
Dissemination Level	PU
Lead Partner	Trilateral Research (TRI)
Lead Contributors	Nicole Santiago (TRI) Bernd Stahl (De Montfort University)
Other Contributors	All members of the SHERPA Consortium *With contributions from* the SHERPA Stakeholder Board
Reviewers	Rowena Rodrigues (TRI) Doris Schroeder (UCLan Cyprus) External reviewer
Abstract	This report presents the recommendations of the SHERPA project to ensure that the benefits of Smart Information Systems (the combination of artificial intelligence and Big Data) are harnessed, and related ethical issues and human rights concerns are addressed. Understanding AI as a set of overlapping ecosystems of interlinked stakeholders, technologies and processes, the SHERPA project proposes ways in which these AI ecosystems can be shaped to ensure they are beneficial to people, uphold human rights and, more generally, promote human flourishing. These recommendations are based on the ethical and human rights issues identified during the project, including through 10 case studies, 5 scenarios, literature reviews, an online survey with more than 300 respondents, 45 stakeholder interviews, an expert Delphi study, and consultation with the SHERPA Stakeholder Board.
Key Words	artificial intelligence, smart information systems, ethics, human rights, human flourishing

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	Oct 2019	Bernd Stahl; SHERPA Consortium		Internal; to guide discussion
0.2	May 2020	Bernd Stahl; SHERPA Consortium		Internal; to guide discussion
0.3	Sep 2020	SHERPA Consortium		Initial recommendations list; shared with SB

0.4	Dec 2020	SHERPA Consortium		Updated draft recommendations; shared with SB
0.5	Jan 2021	SHERPA Consortium		Finalised language for SHERPA website
0.6	9 April 2021	Nicole Santiago; Bernd Stahl	Rowena Rodrigues; Doris Schroeder; external reviewer	First report draft for external and internal QA
1.0	30 April 2021	Nicole Santiago; Bernd Stahl		EC Submission

Table of Contents

Executive Summary

Artificial Intelligence (AI) refers to a set of powerful technologies that are increasingly prevalent in the lives of individuals around the world. Like all technologies, AI can be used for different purposes and therefore holds the potential to hugely benefit individuals and society. At the same time, AI raises many ethical and social concerns, ranging from concerns about biases and resulting discrimination, to the redistribution of socio-economic and political power and the impact on democracy.

The SHERPA project has developed a set of *Final Recommendations* to ensure that the benefits of AI are harnessed, and related ethical issues and human rights concerns are addressed. These Recommendations are based on the ethical and human rights issues and potential governance measures identified throughout the duration of the project, through: 10 case studies[1]; 5 scenarios[2]; literature reviews on ethical and social impacts[3], existing human rights law[4], regulatory options[5], standardisation activities[6], and technical interventions[7]; an online survey with more than 300 respondents[8]; 45 stakeholder interviews[9]; an expert Delphi study[10]; and consultation with the SHERPA Stakeholder Board (SB). All partners in the SHERPA consortium (the Consortium) contributed to these *Final Recommendations*.

SHERPA's *Final Recommendations* understand AI as a set of overlapping ecosystems of interlinked stakeholders, technologies and processes. The Recommendations propose ways in which these AI ecosystems can be shaped to ensure they are beneficial to people, uphold human rights and, more generally, promote human flourishing. The Recommendations should be read in the context of the European discussion on AI, notably the European Union (EU) regulatory proposals, and the work of the High-Level Expert Group on AI (AI HLEG).[11]

---

[1] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies. SHERPA Project. https://doi.org/10.21253/DMU.7679690

[2] Wright, David (2019): D1.2 SIS Scenarios. SHERPA Project. https://doi.org/10.21253/DMU.8181695

[3] Ryan, Mark; Brey, Philip; Macnish, Kevin; Hatzakis, Tally; King, Owen; Maas, Jonne; et al. (2019): D1.4 Report on Ethical Tensions and Social Impacts. SHERPA Project. https://doi.org/10.21253/DMU.8397134

[4] Andreou, A.; Laulhe-Shaelou, S.; Schroeder, D. (2019): D1.5: Current Human Rights Frameworks. SHERPA Proejct. https://doi.org/10.21253/DMU.8181827

[5] Rodrigues, Rowena; Laulhe Shaelou, Stephanie; Lundgren, Björn (2020): D3.3 Report on regulatory options. SHERPA Project. https://doi.org/10.21253/DMU.11618211

[6] Bijlsma, Marlou; Zijlstra, Thamar (2020): D3.4 From Good Practices to Standard Practices. SHERPA Project. https://doi.org/10.21253/DMU.13553684

[7] Kirichenko, Alexey; Marchal, Samuel (2020): D3.5 Technical Options and Interventions. SHERPA Project. https://doi.org/10.21253/DMU.13187081

[8] Brooks, Laurence; Stahl, Bernd; Jiya, Tilimbe (2020): D2.3 Online survey report. SHERPA Project. https://doi.org/10.21253/DMU.11777478.

[9] Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D2.2 Report of Interview Analysis. SHERPA Project. https://doi.org/10.21253/DMU.13168265

[10] Santiago, Nicole; Stahl, Bernd; Keene, Paul; Hatzakis, Tally; Rodrigues, Rowena; Wright, David (2020): D2.4 Delphi Study Report. SHERPA Project. https://doi.org/10.21253/DMU.13168553

[11] European Commission High-Level Expert Group on AI (AI HLEG), https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

The Consortium was guided by the following criteria that the recommendations should:

- Focus on a clearly identified problem that can be addressed.
- Be clear, implementable and actionable.
- Target a clear audience and be time-bound.
- Link to SHERPA work and build on the Consortium's strengths, expertise and credibility.
- Be capable of promotion by SHERPA through dissemination, exploitation, and advocacy activities.

The nine SHERPA recommendations are divided into three domains: Concepts, Knowledge, and Governance.

In order to prioritise ethics and fundamental rights, **three main conditions** should be fulfilled. One, **concepts of AI need to be clear**, and AI ecosystems need to be clearly delineated from one another. Two, there must be a **sustainable knowledge base of technical, social, ethical and legal aspects**. And three, **governance of AI ecosystems should set the framework** and support individuals and organisations.

Figure 1: SHERPA final recommendations

List of figures

Glossary of terms

Term	Explanation
AI	Artificial intelligence
AI HLEG	EC High-level Expert Group on AI
Consortium	SHERPA Consortium (Aequitas, De Montfort University, EUREC, European Business Summit, F-Secure, Mutual Shoots, NEN, Trilateral Research, University of Twente, UCLan Cyprus, and Pineapple Jazz)
D	Deliverable
DPO	Data protection officer
EC	European Commission
EP	European Parliament
EU	European Union
OECD	Organisation for Economic Co-operation and Development
SB	SHERPA Stakeholder Board
SIS	The combination of Artificial Intelligence and big data analytics
T	Task
WP	Work Package

1. Introduction

This report presents the *Final Recommendations* of the SHERPA project (Shaping the ethical dimensions of smart information systems–a European perspective)[12]. The Recommendations aim to ensure that the benefits of AI are harnessed, and related ethical issues and human rights concerns are addressed. These recommendations are intended to contribute to shaping the AI ecosystem in a way that is conducive to human flourishing. To have practical impact, the SHERPA project has developed the Recommendations to be actionable and realistic so they can be implemented by stakeholders.

Conceptual framework

Artificial Intelligence (AI) refers to a set of powerful technologies that are becoming increasingly prevalent in the lives of individuals around the world. Like all technologies, AI can be used for different purposes and therefore holds the potential to hugely benefit individuals and society. At the same time, AI raises many ethical and social concerns, ranging from concerns about biases and resulting discrimination, to the redistribution of socio-economic and political power and the impact on democracy.

The SHERPA *Final Recommendations* understand AI as a set of overlapping ecosystems of interlinked stakeholders, technologies and processes. The Recommendations propose ways in which these AI ecosystems can be shaped to ensure they are beneficial to people, uphold human rights and, more generally, promote human flourishing. They should be read in the context of the European discussion on AI, notably the EU's regulatory proposals and the work of the High-Level Expert Group on AI (AI HLEG).[13] The Consortium was guided by the following criteria that the recommendations should:

- Focus on a clearly identified problem that can be addressed.
- Be clear, implementable and actionable.
- Target a clear audience and be time-bound.
- Link to SHERPA work and build on the Consortium's strengths, expertise and credibility.
- Be capable of promotion by SHERPA through dissemination, exploitation, and advocacy activities.

SHERPA Final Recommendations

The nine SHERPA Recommendations are divided into three domains: **concepts**, **knowledge**, and **governance**.

Recommendation: Concepts
This recommendation addresses the need for clarity of concepts of AI and clear delineations of AI and digital ecosystems.

> Use appropriate and clear definitions of AI and digital technology

---

[12] SHERPA project, https://www.project-sherpa.eu
[13] European Commission High-Level Expert Group on AI (AI HLEG), https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

Recommendations: Knowledge

This set of recommendations calls for a sustainable knowledge base of technical, social, ethical and legal aspects of AI.

> Develop baseline model for AI impact assessment

> Promote Ethics by Design for researchers in EC-funded projects

> Create training and education pathways that include ethics and human rights in AI

> Include research findings on AI ethics in standardisation

> Undertake security analysis for machine learning systems

Recommendations: Governance

This set of recommendations calls for governance of AI ecosystems to set the framework and support individuals and organisations.

> Develop a regulatory framework and enforcement mechanisms for AI

> Establish an independent European Union Agency for AI

> Establish role of AI (Ethics) Officer in organisations

Structure of the report

Section 2 describes the methodology for developing these recommendations. **Section 3** presents the organising vision and criteria that guided the selection of the recommendations. **Section 4** presents each recommendation with its context and justification for, the stakeholder responsible for taking a leading role, key considerations for implementing the recommendation, and/or related concepts. **Section 5** briefly outlines SHERPA's past and ongoing contributions to putting these recommendations into action. The **Appendices** provide more context for the conceptional framework and vision for the *Final Recommendations*. Appendix 1 provides insights from SHERPA on ethical and human rights issues; Appendix 2 provides insights on possible mitigation strategies.

2. Methodology

The *Final Recommendations* were developed following the methodology described in the SHERPA Description of Action (DoA):

> Based on the outcomes of the evaluation and validation activities in task 4.2, the SHERPA consortium will undertake an exercise of prioritisation of different options using the 4.1 strategy. The insights gained from all activities of the project prior to this point, i.e., the representation of the ethical and human rights issues of SIS in WP1, the broad range of stakeholder engagement of WP2, the proposals for responsible development of SIS in WP3, and the earlier tasks of WP4 will inform the proposals that will be included in the final set of recommendations of the project. The criteria for the evaluation of the project recommendations will be developed as part of this task, but will include: potential to strengthen human rights; reduction of ethical tensions; enhancement of RRI; coverage of stakeholders; impacts on stakeholders; impact on innovation; ease of use; likelihood of adoption; value for money etc. The outcome of the task will be shared with all stakeholders involved in SIS, for example, via the SIS workbook, the project website, press release (see WP5).

2.1 Related SHERPA activities

These *Final Recommendations* are the culmination of Work Package 4 (WP4), *Evaluation, validation and prioritisation*, and are based on all previous work in the SHERPA project in WP1 (*Representation and visualisation of ethical and human rights issues in SIS*), WP2 (*Stakeholder analysis and consultation*), WP3 (*Responsible development of SIS*), and the two previous Tasks in WP4: T4.1 Develop an evaluation and validation strategy[14], and T4.2 Stakeholder evaluation and validation.[15]

These Recommendations are based on the ethical and human rights issues and potential governance measures identified during the project, through 10 case studies[16], 5 scenarios[17], literature reviews on ethical and social impacts[18], existing human rights law[19], regulatory options[20], standardisation activities[21], and technical interventions[22], an online survey with more than 300 respondents[23], 45 stakeholder

---

[14] Antoniou, Josephina (2019): D4.1 Evaluation and Validation Strategy. SHERPA Project. https://doi.org/10.21253/DMU.9205289

[15] Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D4.2 Evaluation Report. SHERPA Project. https://doi.org/10.21253/DMU.12917717

[16] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[17] Wright, David (2019): D1.2 SIS Scenarios.

[18] Ryan, Mark; Brey, Philip; Macnish, Kevin; Hatzakis, Tally; King, Owen; Maas, Jonne; et al. (2019): D1.4 Report on Ethical Tensions and Social Impacts.

[19] Andreou, A.; Laulhe-Shaelou, S.; Schroeder, D. (2019): D1.5: Current Human Rights Frameworks.

[20] Rodrigues, Rowena; Laulhe Shaelou, Stephanie; Lundgren, Björn (2020): D3.3 Report on regulatory options.

[21] Bijlsma, Marlou; Zijlstra, Thamar (2020): D3.4 From Good Practices to Standard Practices.

[22] Kirichenko, Alexey; Marchal, Samuel (2020): D3.5 Technical Options and Interventions.

[23] Brooks, Laurence; Stahl, Bernd; Jiya, Tilimbe (2020): D2.3 Online survey report.

interviews[24], an expert Delphi study[25], and consultation with the SHERPA Stakeholder Board (SB). All partners in the SHERPA Consortium contributed to these Final Recommendations.

2.2 Decision-making process

The SHERPA Consortium began developing these *Final Recommendations* in October 2019. The first step was to internally collect ideas from all Consortium partners. The ideas were discussed during monthly online Consortium meetings from November 2019 through August 2020. All partners were invited to contribute to a shared document, which was continuously updated and revised in response to findings from past and ongoing SHERPA activities and the monthly Consortium calls.

In September 2020, the Consortium met online for a dedicated meeting for prioritisation and selection of recommendations from among the full list of ideas proposed, based on criteria from T4.2 Stakeholder evaluation and validation[26]. The result was an initial draft of the recommendations, which was presented to the SB for feedback during an October 2020 SB meeting (online).

In advance of the October 2020 meeting, EUREC circulated the draft recommendations to all SB members. Individual SB members were asked to prepare a few comments and lead discussion on one recommendation. For each recommendation, the SB discussion was guided by four questions:
1. Is this recommendation clear and understandable?
2. Is this recommendation needed and helpful?
3. Is this recommendation practicable and feasible?
4. Is there a need to modify this recommendation?

The SB was also asked to provide feedback more generally on the recommendations as a whole, along the following lines:
- Are these the most important recommendations?
  - Gaps – is anything vital/indispensable missing?
  - Lack of importance – should a recommendation be removed from the list?
- Is the set of recommendations coherent?
- Is the structure comprehensible?
- Is the classification in three categories reasonable?
- Is the allocation of the recommendations to the categories comprehensible?
- Is there any other comment you would like to make?

In November 2020, the Consortium met online to discuss each recommendation individually, and to finalise the language of the recommendations; each recommendation had a dedicated meeting. The partner responsible for leading the relevant work in the SHERPA project related to the recommendation led the discussion. The revised recommendations were sent to the SB for additional feedback in December 2020. Further work on finalising the language of the recommendations continued in January 2021, at which point the Final Recommendations were published on the SHERPA website.[27]

---

[24] Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D2.2 Report of Interview Analysis.
[25] Santiago, Nicole; Stahl, Bernd; Keene, Paul; Hatzakis, Tally; Rodrigues, Rowena; Wright, David (2020): D2.4 Delphi Study Report.
[26] Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D4.2 Evaluation Report.
[27] SHERPA Final Recommendations, https://www.project-sherpa.eu/recommendations/

2.3   Understanding of AI

There are many definitions of Artificial Intelligence. In discussions of ethics and AI, there are at least three different meanings of the term AI:

1. Machine learning is currently the most prominent "narrow" form of AI – a technology that replicates one specific aspect of natural intelligence, such as speech recognition or pattern classification. Machines can be very good at these individual tasks, but cannot easily apply them to other areas.
2. AI is often used to refer to larger socio-technical systems that incorporate and build on AI, such as machine learning, but that go beyond the immediate AI technology. Examples include autonomous vehicles (self-driving cars), credit rating systems, or commercial recommender systems, i.e., systems used in electronic commerce (e.g., Ebay, Amazon) that recommend items to customers on the basis of prior purchases.
3. General AI (or artificial general intelligence) stands for machines that can perform human-level cognitive functions. No such systems currently exist, but they figure strongly in the general literature and public imagination. They also serve as inspiration for AI research and development.

For example, the Organisation for Economic Co-operation and Development (OECD) *Recommendation of the Council on AI* states that an "AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy."[28] A similar policy-oriented definition from the European Commission's *White Paper on Artificial Intelligence* suggests that "AI is a collection of technologies that combine data, algorithms and computing power."[29] One of the most cited academic definitions, from Li and Du (2007), notes that AI combines: "… a variety of intelligent behaviors and various kinds of mental labor, known as mental activities, … [to] include perception, memory, emotion, judgement, reasoning, proving, identification, understanding, communication, designing, thinking and learning."[30]

For the purpose of these SHERPA Final Recommendations, we understand AI as a set of powerful technologies that increasingly affect most countries, organisations and individuals. Like all technologies, AI can be used for different purposes. AI holds the potential to hugely benefit individuals and society. For example, AI can help to better understand and cure diseases, to revolutionize transport, to optimize business processes or reduce carbon emissions. At the same time, AI presents many ethical and social challenges, ranging from concerns about biases and resulting discrimination, to the redistribution of socio-economic and political power and the impact on democracy.

---

[28] Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, adopted 22 May 2019. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[29] European Commission. (2020a). White Paper on Artificial Intelligence: A European approach to excellence and trust (White Paper COM(2020) 65 final). https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

[30] Li, Deyi; Du, Yi (2007), Artificial Intelligence with Uncertainty, p.1.

3. Vision: Fostering an AI Ecosystem for Human Flourishing

Ethical and human rights aspects of smart information systems, i.e., those digital technologies building on artificial intelligence and big data analytics, are a key issue that needs to be addressed as a matter of priority.

SHERPA's *Final Recommendations* are designed to work within the existing framework of EU and national legislation and regulation, in particular with the European Commission (EC) White Paper[31] and Inception Impact Assessment[32], the European Group on Ethics in Science and New Technologies statement on Artificial Intelligence, Robotics and 'Autonomous' Systems[33], and the AI HLEG guidelines[34], policy and investment recommendations[35], sectoral considerations[36], and assessment list[37]. The SHERPA *Final Recommendations* aim to be non-obvious and avoid replicating existing recommendations.

The overarching idea that provides the framework in which to understand the *Final Recommendations* is that the current AI scene can be interpreted as an interlinking set of ecosystems consisting of numerous actors (including individuals, companies, civil society organisations, public sector organisations, states, international actors). AI technologies are developed, deployed and used in the context of the complex relationship between these different actors. This description of AI using the metaphor of an ecosystem is now widely established, including in the EC's White Paper[38]. It has also been applied to the ethics of AI[39]. The *Final Recommendations* of the SHERPA project push these ideas further, and are based on the

---

[31] European Commission. (2020a). White Paper on Artificial Intelligence: A European approach to excellence and trust.

[32] European Commission. (2020d). Inception impact assessment for the Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence - Ares(2020)3896535. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements

[33] EGE, European Group on Ethics in Science and New Technologies. (2018). Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Publications Office of the European Union. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

[34] HLEG on AI. (2019a). Ethics Guidelines for Trustworthy AI. European Commission - Directorate-General for Communication. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[35] HLEG on AI. (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence. European Commission - Directorate-General for Communication. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence

[36] HLEG on AI. (2020a). Sectorial considerations for Trustworthy AI - taking AI's context specificity into account. European Commission. https://futurium.ec.europa.eu/en/european-ai-alliance/blog/sectorial-considerations-trustworthy-ai-taking-ais-context-specificity-account

[37] HLEG on AI. (2020b): Assessment List for Trustworthy AI (ALTAI). European Commission. https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence

[38] European Commission. (2020a). White Paper on Artificial Intelligence: A European approach to excellence and trust.

[39] Digital Catapult. (2020). Lessons in practical AI ethics: Taking the UK's AI ecosystem from 'what' to 'how.' Digital Catapult. https://assets.ctfassets.net/nubxhjiwc091/xTEqMcYudwQ7GHZWNoBfM/c2a2d55a0ee1694e77634e240eafdfdf/20200430_DC_143_EthicsPaper__1_.pdf

following insight: In order for AI to meet ethical expectations and safeguard human rights, measures should be taken that guide the entire AI ecosystem to foster and support human flourishing.

Priority should be given to recommending measures that are likely to have a desirable effect on the ecosystem as a whole. The proposed *Final Recommendations* achieve this by focusing on three main areas:

1. Conceptual clarity: Interventions into ecosystems requires clarity on which aspects are included or excluded, and the clear and consistent use of terminology

2. Knowledge base: Ecosystem actors require technical, conceptual and procedural knowledge, and the ecosystem needs mechanisms to store and disseminate this knowledge

3. Governance of the ecosystem: Ecosystems' governance needs to be flexible, adaptable and able to learn. Governance structures need to be conducive to this and involve all stakeholders.

On the basis of these principles and insights, the SHERPA Consortium proposes its *Final Recommendations*.

3.1 Shaping AI ecosystems

AI consists of many different types of technology, which are developed, deployed and used by an array of different stakeholders. AI requires many components, including technical, social and organisational ones. It is therefore helpful to think about AI in terms of an ecosystem of interlinked stakeholders, technologies and processes.
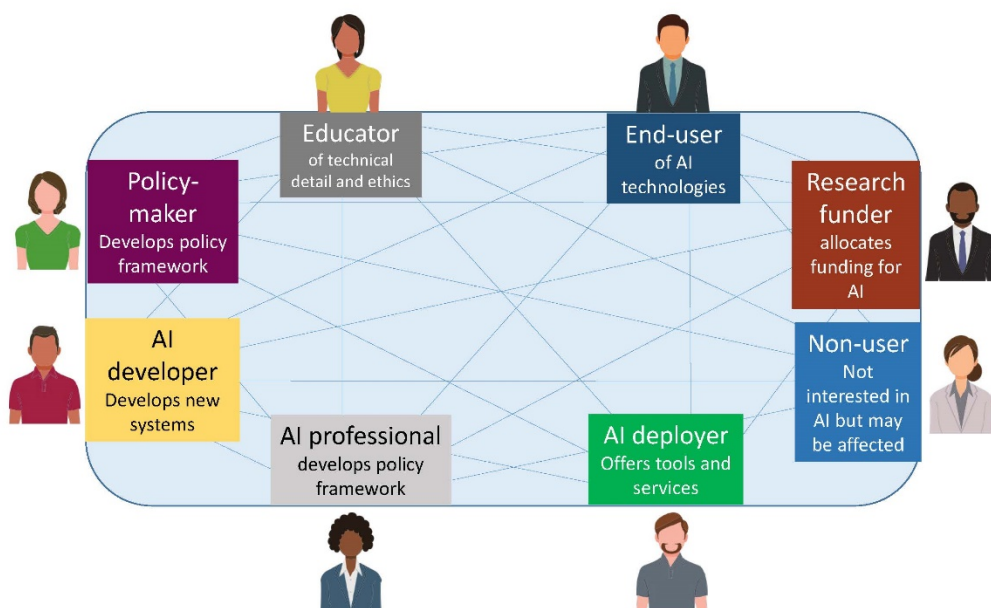
Figure 2: AI ecosystems stakeholders

The *Final Recommendations* are based on the view of AI as a set of interlinking ecosystems. They propose ways in which AI ecosystems can be shaped to ensure they are beneficial to people and communities, uphold human rights and, more generally, promote human flourishing.

Why talk about AI ecosystems?
AI consists of a multitude of technologies that are applied to many different tasks by a variety of stakeholders. Combined with the complexity of ethical issues of AI, this means that there is no simple way

to address the ethics of AI as a whole. Therefore, we need to find ways of thinking about AI that allow for a broader perspective.

One way of thinking about AI is to use the metaphor of ecosystems. This metaphor has been widely accepted, for example, the European Commission talks about an ecosystem of excellence in AI, and an ecosystem of trust[40].

AI Ethics Stakeholders
There are different types and groups of stakeholders involved in ethical issues of AI:

- Policy stakeholders including national, regional and international policymakers and those involved in implementing policies, such as research funders.
- Organisations including developers, deployers and users of AI, and organisations with special roles, such as standardization bodies, educational institutions.
- Individuals such as technical experts and developers, but also users and, importantly, people who don't use AI but may still be affected by it.
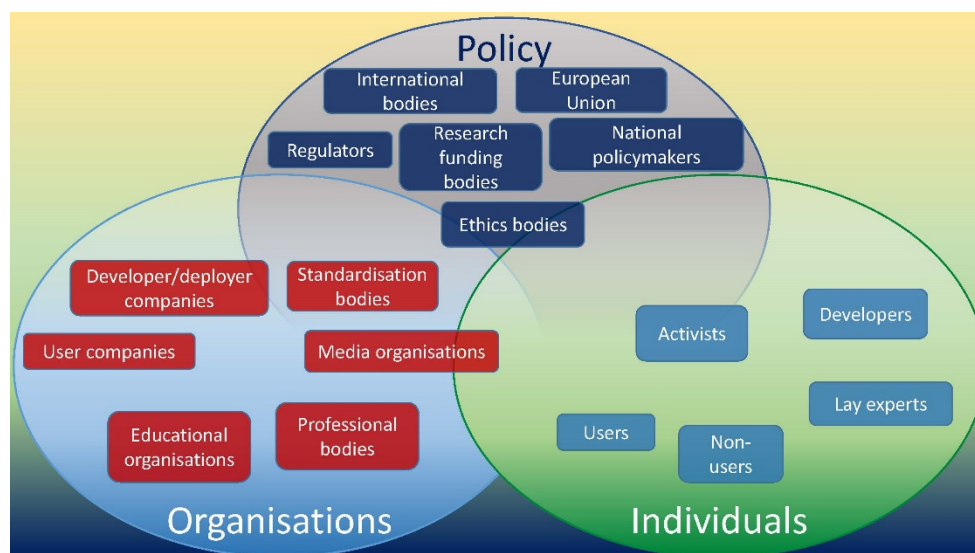
Figure 3: AI ethics stakeholders

Characteristics of innovation ecosystems
Innovation ecosystems mirror natural ecosystems in that their boundaries are often difficult to define. Members of ecosystems co-evolve; they compete but they also collaborate and learn from each other. There are mutual interdependencies between members, but the relationships are typically dynamic.

---

[40] European Commission. (2020a). White Paper on Artificial Intelligence: A European approach to excellence and trust.
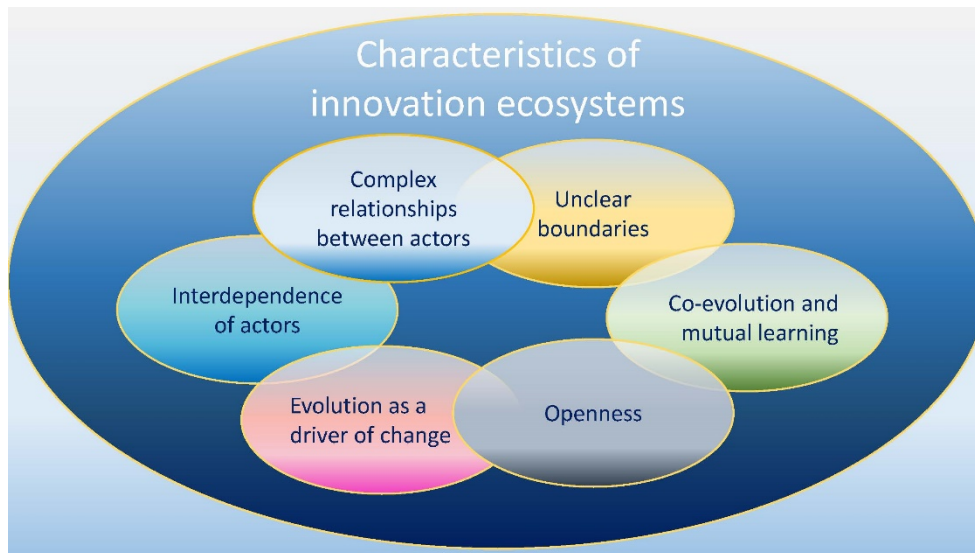
Figure 4: Characteristics of innovation ecosystems

3.2   Human flourishing

Most human beings want to live fulfilled lives that allow them to reach their potential, successfully meet challenges and, as far as possible, determine their destiny. Briefly, they want to flourish. These ideas have a long history in many ethical theories, principles and values and remain current in the 21st century. They are expressed in human rights frameworks, such as the *Universal Declaration of Human Rights*[41] or the *European Charter of Fundamental Rights*[42]. Flourishing is realised in the individual, but typically requires supportive social environments.

Artificial Intelligence (AI), like all technologies, can be used for different purposes. From an ethical and human rights point of view there are three main purposes of AI use: efficiency and optimisation, social control, and human flourishing. These three purposes are not contradictory or mutually exclusive, but point to different driving forces for development and utilisation of these technologies. Framing AI ethics in terms of human flourishing is consistent with numerous national and international ethics guidelines and principles, including those developed by EU institutions.

Principles to Promote Flourishing
In order to intervene in AI ecosystems to promote human flourishing, the SHERPA *Final Recommendations* apply three principles:
- the ecosystem in question needs to be clearly delineated, e.g., in terms of geography, technology, but also conceptually, i.e., with regards to a shared understanding of human flourishing;
- a successful AI ecosystem will need to develop and maintain a knowledge base covering technical but also conceptual and procedural knowledge; and
- governance intervention needs to be adaptive, flexible and geared towards learning.

---

[41] *Universal Declaration of Human Rights*, adopted by General Assembly Resolution 217 A(III) of 10 December 1948.

[42] European Union, *Charter of Fundamental Rights of the European Union*, 2000/C 364/01, entry into force 7 December 2000.

Requirements for interventions into AI ecosystems

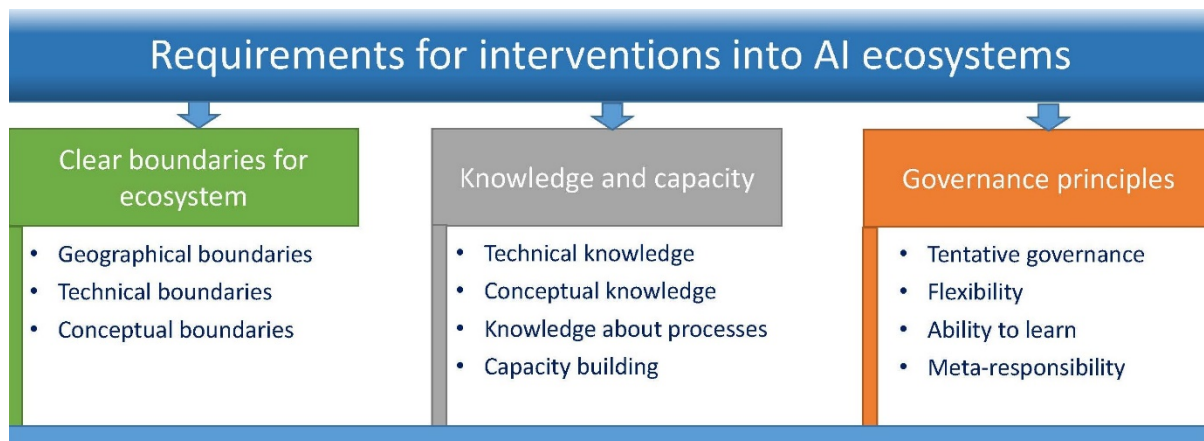| Clear boundaries for ecosystem	Knowledge and capacity	Governance principles
• Geographical boundaries • Technical boundaries • Conceptual boundaries	• Technical knowledge • Conceptual knowledge • Knowledge about processes • Capacity building	• Tentative governance • Flexibility • Ability to learn • Meta-responsibility

Figure 5: Requirements for interventions into AI ecosystems

3.3 Ethics as integral to excellence in AI

The SHERPA project supports the vision of a world where excellent AI and smart information systems promote human flourishing by integrating ethical concerns by design and safeguarding human rights. Achieving this vision will benefit individuals and consumers and strengthen trust in AI systems.

SHERPA identified a significant number of ethical and human rights issues (See Appendix A for a overview of those issues). We use the term 'ethical issues' to denote the issues that were perceived to be problematic by our respondents. These explicitly include issues that are already covered by human rights and other legislation (e.g., privacy, discrimination), but also cover issues that are less clearly specified (e.g., transparency, loss of human contact), or are closely related to technical aspects (e.g., security).

The ethical and human rights issues are different in nature and scope depending on the category of AI. Some issues may be subject to simple and straightforward resolutions, others will require political interventions, while some may be impossible to resolve and require continuous reflection. To illustrate this, we present examples from three different categories of AI: machine learning, socio-technical systems, and artificial general intelligence, with some of the associated ethical and human rights issues identified in the SHERPA project

Machine Learning: Some ethical issues are directly related to AI in the narrow sense, most prominently to machine learning, which is currently often implemented through neural networks. This type of AI is characterised by opacity, unpredictability and, typically, the need for large data sets for training and validation. Ethical issues linked to this type of AI include:
- bias
- discrimination
- security breaches
- data protection issues

Socio-Technical Systems: This understanding of AI points to ethical issues arising from living in a digital world. These socio-technical systems appear to act autonomously, structuring the way humans can act, and have significant social impact. They lead to ethical issues such as:
- unequal access to power and resources
- unfair distribution of the costs and benefits of technology

18

- impact on warfare, and the killing of humans by machines

Artificial General Intelligence: Currently no AI exists that can be described as artificial general intelligence, i.e., that has human cognitive capabilities. However, these systems figure prominently in the literature and in people's imagination. Such systems would potentially raise ethical issues such as:
- hostility towards humanity by super-intelligent machines
- changing human perceptions based on close interaction with machines (e.g., neural implants)

Any recommendation that aims to address the ethics of AI needs to be aware of the breadth of issues, and ensure that the recommendations are clearly delineated and the relevant concepts are well-defined.

4. SHERPA Final Recommendations

These *Final Recommendations* aim to ensure that ethical and human rights issues of AI are recognised and addressed. They are based on the view of AI as a set of overlapping ecosystems. In order to steer these ecosystems, the recommendations are grouped around three main conditions that should be fulfilled:

- Concepts need to be clear and the ecosystems need to be clearly delineated.
- There must be a sustainable knowledge base of technical, but also social, ethical and legal aspects.
- Governance of AI ecosystems should set the framework and support individuals and organisations.

Thus, nine SHERPA recommendations are divided into three domains: concepts, knowledge, and governance. Each recommendation is presented with (1) its context and justification (**What is the problem?**); (2) the stakeholder responsible for taking a leading role (**Who should act?**); and, where relevant, (3) **Key considerations** for implementing the recommendation; and/or (4) **Related concepts**. Section 5 briefly outlines SHERPA's past and ongoing contributions to putting these recommendations into action.

4.1 Concepts: Delineate AI ecosystems

4.1.1 Recommendation: Use appropriate and clear definitions of AI and digital technology

The scope of AI needs to be clearly defined in each use context with regards to relevant issues. **Where appropriate, 'AI' should be replaced with more specific terms**, such as 'machine learning'.

SHERPA research highlights that the concept of AI is a multi-faceted term, as are the concepts of ethics and human rights. Conceptual clarification will aid the development of ecosystems of AI for human flourishing. Some issues, such as changes to employment, or political and economic power redistribution, are only peripherally linked to AI. When addressing them, it may be more appropriate to use inclusive terms such as 'emerging digital technology'. Other issues (e.g., autonomy of machines) are already well understood and categorised (e.g., the taxonomy of the levels of autonomy in vehicles), which may help develop similar categorisations in other application domains.

What is the problem?
The term 'AI' is scientifically contested. Ethical and human rights issues are often not directly linked to AI in a narrow sense. Specific technologies (e.g., machine learning, deep neural networks; reinforcement learning) can have properties leading to concerns that are not relevant to other technologies. For example, machine learning using deep neural networks requires large training datasets which can raise issues of data protection and security and may also perpetuate biases that are contained in the datasets. This would not apply to AI techniques that do not rely on the analysis of large data sets, such as symbolic AI or systems that use purely technical data.

Who should act?
All bodies developing policies and guidelines for ethical or trustworthy AI, including the European Commission, national governments, standardisation bodies.

Key Considerations
Use a concept of AI that points to the features of the technology that are ethically relevant, such as opacity (can hide bias) or automation (replaces jobs). Characteristics or examples may be more helpful than definitions.

The concepts used influence the scope of the technology in question but also the responses (e.g., scope of a risk or impact assessment).

Whichever concept of technology is used, the ethical and human rights implications depend heavily on the application area. Machine learning, for example, may have very different consequences in healthcare as compared to in gaming.

Related Concepts
The focus of this recommendation is on the definition of technology. If the focus is on ethical issues of AI, then the concept of ethics needs to be clearly understood. Human flourishing is a useful term to highlight ethical ideas, but there are many other ethical positions worthy of consideration.

Human rights are outlined in various legal documents including the *Universal Declaration of Human Rights*[43], *Charter of Fundamental Rights of the European Union*[44], *European Convention of Human Rights*[45], OECD's *Framework for the Classification of AI Systems*[46].

---

[43] *Universal Declaration of Human Rights*.
[44] European Union, *Charter of Fundamental Rights of the European Union*.
[45] Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14*, 4 November 1950, ETS 5, entry into force 3 September 1953.
[46] Perset, Karine; Murdick, Dewey; Clark, Jack; Grobelnik, Marko, A first look at the OECD's Framework for the Classification of AI Systems, designed to give policymakers clarity, OECD, 24 November 2020.
https://oecd.ai/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers

4.2 Knowledge: Establish and maintain a knowledge base

4.2.1 Recommendation: Develop baseline model for AI impact assessment

A baseline model for AI impact assessments Is needed to introduce a standardised approach to impact assessments. This model must provide clear guidance on:

- Processes for conducting and evaluating AI impact assessments
- Measures and metrics for AI impacts
- Determination of the risk level for a technology or application
- Issues to be included in AI impact assessments

This is a composite recommendation, requiring further work on different subject areas and processes. The recommendation is therefore broken down into three sub-recommendations:

Sub-Recommendation: **Fund research on appropriate measures for broader ethical, social and other impacts, and to design relevant impact assessment for AI technologies.**

In many cases it is not clear whether a new technology has net positive or negative consequences on society. One key problem is that there is no agreed way to measure and compare different consequences. Therefore, there is a need to define processes and metrics for AI Impact Assessment.

Who should act?: European Commission and national funding organisations, when preparing AI-related calls.

Sub-Recommendation: **Include provisions on how to assess and measure risks in guidance for AI impact assessment and fund research to develop rigorous risk classification for AI.**

There is no agreement on how risk in AI is assessed or measured, therefore there is a need to include a determination of risk levels in impact assessment.

Who should act?: European Commission.

Sub-Recommendation: **Establish a publicly accessible and current knowledge base for ethical issues of AI to inform AI impact assessment.**

There are numerous ethical and human rights issues arising from AI. An impact assessment needs to be informed by current research on these issues. Therefore, there is a need for guidance on assessment of specific issues.

Who should act?: European Union through the proposed EU Agency for AI, and researchers on AI and related topics, including AI centres of excellence in research.

What is the problem?
Despite much academic research, standardisation activities and policy-oriented work (notably including the HLEG (2020) Assessment List),[47] there is no universally accepted baseline of what an AI impact assessment should entail.

There is broad agreement that a risk-based approach to AI is appropriate. For such an approach to work, there must be guidance on how to define, measure, interpret and address relevant risks.

Who should act?
European Commission, ideally led by a EU Agency for AI[48].

Key Considerations
AI impact assessment can build on and incorporate numerous existing impact assessments, including:
- Data protection impact assessment
- Algorithmic impact assessment
- Human rights impact assessment
- Socio-economic impact assessment
- Environmental impact assessment
- Ethical impact assessment
- Responsible innovation assessment

Initial developments of AI impact assessments exist, for example in the IEEE 7010-2020 standard on *Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*[49]*,* or the Dutch Platform voor de Informatie Samenleving (ECP) AI Impact Assessment**.**[50]

AI impact assessment should not be an *a priori* activity, but should be an ongoing process that is regularly reviewed and updated.

AI impact assessment could be implemented as part of due diligence processes.

4.2.2 Recommendation: Promote Ethics by Design for researchers in EC-funded projects

The development of ethical AI should be encouraged through the **promotion of Ethics by Design through the creation of Ethics for AI Training** and Guidelines for Horizon Europe Researchers, and organisations outside the Horizon Europe framework. In preparation for calls involving AI-assisted systems in Horizon Europe:
- Condense the body of knowledge on ethics, human rights, and AI, including mitigation measures, into training for researchers.

---

[47] HLEG on AI (2020b): Assessment List for Trustworthy AI (ALTAI). European Commission.
[48] Rodrigues, Rowena; Santiago, Nicole; Macnish, Kevin; Antoniou, Josephina; Wright, David (2020): D3.6 Feasibility of a new regulator and proposal for a European Agency for AI. SHERPA Project. https://doi.org/10.21253/DMU.13168295
[49] IEEE, *IEEE 7010-2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*. https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html
[50] Platform voor de Informatie Samenleving (ECP), *Artificial Intelligence Impact Assessment*. https://ecp.nl/publicatie/artificial-intelligence-impact-assessment-english-version/

- Implement the best means of delivering this training to Horizon Europe-related projects and organisations within the EU more broadly.
- Include a requirement for evidence of ethical use of AI in the societal impact section of relevant Horizon Europe applications
- Include a requirement for ethical use of AI in project research and development in the Ethics Self-Assessment section of Horizon Europe applications.

What is the problem?
There is a current lack of ethical awareness by researchers working on AI-assisted systems in EU framework programmes, and a lack of guidance provided to those researchers about specific issues arising from AI. The result is a risk of funding projects developing or involving AI which are damaging in terms of ethics and human rights.

Who should act?
Ethics sections in European Commission, in particular the EC's Directorate-General Research and Innovation and the Research Executive Agency.

Key Considerations
Ethical impact should be a starting point, not an add-on, to all Horizon Europe-funded research.

Some research uses AI-assisted systems and tools without being specifically AI research. All researchers using AI-assisted systems and tools, as well as those developing AI, need to be trained in the relevance of the ethical and human rights implications of AI.

Guidelines on Ethics by Design should be developed that clearly specify who should do what and how it should be done.

This recommendation needs to be implemented in accordance with the EU digital education action plan (2021-2027)[51]**.**

Training should cover regulatory frameworks and AI impact assessments.

Once available, the agreed method of Ethics by Design should be made available to organisations beyond the Horizon Europe programme.

4.2.3 Recommendation: Create training and education pathways that include ethics and human rights in AI

Teaching digital competence needs to be coupled with education about its ethical and human rights aspects. **Current technology-oriented curricula should be updated to ensure that ethical and human rights aspects of AI and related technologies are covered**. Specific curricula should include:

- Formal education
    - Primary education
    - Secondary education
    - Relevant undergraduate and postgraduate courses
- Non-formal education

---

[51] European Commission. (2021). *EU digital education action plan (2021-2027)*.
https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en

What is the problem?

The ability to recognise and deal with ethical and human rights aspects of emerging digital technologies (such as AI) depends on awareness and understanding, which is currently lacking throughout the educational system.

Who should act?

European Commission, Council of Europe, national governments, in particular national education bodies, any bodies involved in designing curricula.

Key Considerations

Key considerations in the creation and/or implementation of education about ethics and digital technologies are that it complements the following initiatives:

- Digital Competence Framework for Educators (DigCompEdu)[52];
- Policy and investment recommendations of the HLEG on AI[53]; and
- Digital Education Action Plan 2021-2027[54].

The relevance of education pathways is shown by SHERPA work with regards to:
- Training of AI ethics officers
- Overview of the education pathways by the EU Agency for AI[55].

4.2.4 Recommendation: Include research finding on AI ethics in standardisation.

It is important that ethicists, innovators and researchers participate in standardisation. Therefore, **standardisation committees should invite ethicists, innovators and researchers**. Ethicists, innovators and researchers should engage with standardisation committees. Their research can be a valuable addition to standards. Contributing to standardisation is a recognized way to ensure that research has a lasting societal impact.

What is the problem?

Standards describe requirements and characteristics for products, methods and services. When ethics and human rights are not taken into consideration for AI-related standards, there is a risk of serious negative implications for society, and a reduction of trust in AI. However, it is challenging to reach worldwide agreement on what ethical AI is, as stakeholders around the world have a different idea of ethical

Who should act?
- Standardisation committees
- Ethicists, innovators and researchers

---

[52] European Commission, Digital Competence Framework for Educators (DigCompEdu). https://ec.europa.eu/jrc/en/digcompedu.

[53] European Commission High-Level Expert Group on AI (AI HLEG), https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[54] European Commission. (2021). *EU digital education action plan (2021-2027)*.

[55] Rodrigues, Rowena; Santiago, Nicole; Macnish, Kevin; Antoniou, Josephina; Wright, David (2020): D3.6 Feasibility of a new regulator and proposal for a European Agency for AI.

Key Considerations
There are numerous ethics guidelines on AI that standardisation committees could draw on, such as those of the AI HLEG[56].

4.2.5 Recommendation: Undertake security analysis for machine learning systems

It is necessary to undertake bespoke analysis of security threats and risks for AI-powered systems. This analysis should include:

- Careful and comprehensive threat enumeration and analysis as a crucial pre-requisite for designing effective AI system protection methods (different systems and attacks against them often require different protection approaches).
- Careful analysis of assumptions about training data (for instance, what can be realistically assumed about the training data distribution across digital service customers).
- Real-time monitoring and analysis of external inputs as attacks against AI systems can cause substantial harm (in many scenarios, it is hard to ensure operational resilience only by design and implementation time efforts).
- Consideration of a managed monitoring service in cases when the precision or confidence of an automated real-time monitoring system in identifying malicious inputs is not sufficiently high.

What is the problem?
Most safeguards for ethics and human rights in AI rely on the integrity and reliability of technical systems. However, these AI systems may have security vulnerabilities and be subject to novel attacks. For example, digital services based on Machine Learning models can be targeted by model poisoning or model inversion attacks. Technical security of AI-powered systems is therefore a necessary condition for their robustness and reliability and an enabler of ethical and human rights safeguards.

Who should act?
Designers, developers, integrators, and operators of AI systems.

4.3 Governance: Institute appropriate governance mechanisms

4.2.1 Recommendation: Develop a regulatory framework and enforcement mechanisms for AI

The **EU should develop a mandatory regulatory framework to ensure that AI systems are safe, and do not violate fundamental rights and ethical principles**. The European Commission, the European Parliament and the European Council should set a strong legal standard at the EU-level that establishes a baseline and encourages high standards of protection of fundamental rights and societal values. This framework should include *ex-ante* (before) and *ex-post* (after) enforcement mechanisms, and should have a high degree of detail and specificity so that developers and users understand their legal obligations, and EU and national authorities can monitor compliance.

---

[56] European Commission High-Level Expert Group on AI (AI HLEG), https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

The regulatory framework should include provisions for:

- Establishment of a European Agency for AI**;**
- Centralised safeguards and mechanisms to identify and monitor risks and abuses ('risk alarms') of particular AI applications in specific sectors and use cases, particularly with respect to vulnerable populations (e.g., AI impact assessment);
- Voluntary labelling scheme/certification requirements;
- Mandatory compliance requirements for all or certain types of AI with high societal impacts and their enforcement;
- Complaint and redress-by-design mechanisms;
- Responsible development, implementation and use (e.g., through privacy by design, data protection by design and default, Ethics by Design);
- Prohibitions or limits (red lines) on certain uses and applications (e.g., ban/moratorium on the use of lethal autonomous weapons systems (LAWS), mass urban biometric surveillance);
- Addressing power imbalances.

What is the problem?

The protection of human rights and safety is a societal priority. Existing soft law initiatives do not address legal issues of liability and other ensuing harms in AI, even though they may be useful in some contexts to complement legislation. Additionally, EU Member States are developing and implementing their own regulatory frameworks, contributing to further fragmentation and confusion.

Who should act?

The European Commission, the European Parliament and the European Council.

Key Considerations

An EU regulatory framework must:

- Balance between precise language for legal clarity and the need for adaptability to accommodate technological developments and changing societal expectations.
- Be strict enough to achieve desired outcomes, but without making the cost of compliance too high for small- and medium enterprises (which could risk further entrenching of power asymmetries with Big Tech).
- Be finalised quickly, but decisions must be informed by broad stakeholder consultation.
- Have safeguarding fundamental rights as the overarching aim.

4.2.2 Recommendation: Establish an independent European Union Agency for AI

The EU should **establish an independent European Union Agency for AI**. The Agency should:

- Make recommendations addressed to the European Parliament, the European Council, or the Commission for legislative amendments;
- Identify potential red lines or restrictions for AI development, deployment and use that violates human rights and/or has significant negative societal impacts;
- Develop and promulgate general guidance on legal concepts and regulatory issues of AI;
- Set benchmarks for enforcement;
- Support and advise EU-level institutions, bodies and agencies and national competent authorities in Member States to fulfil their ethical and human rights obligations and to protect the rule of law;
- Maintain an AI risk alert system;

- Assist in coordinating the mandates and actions of the national competent authorities of Member States;
- Develop harmonised and objective criteria for risk assessment and/or conformity assessment;
- Monitor and/or coordinate the evaluation of the operation of conformity assessment and/or certification schemes;
- Cooperate, liaise, exchange information, promote public dialogue, best practices and training activities;
- Ensure complementarity and synergy between its activities and other Community programmes and initiatives;
- Promote the adoption of regulatory sandboxes; and
- Promote the European Union's AI approach through international cooperation.

What is the problem?
The current regulatory landscape for AI in the EU is fragmented, and concerns have been raised regarding cooperation, coordination and consistent application of EU law.

Who should act?
The European Commission, the European Parliament and the European Council should address this in the development of the proposed legislative framework for AI (expected 2021).

Key Considerations
In creating and/or implementing the European Union Agency for AI, key considerations are:
- Making the Agency operational as soon as possible, even if on a provisional or pilot basis.
- Strength of the underpinning legislative framework (establishing the Agency and its mandate, and setting clear boundaries and scope)
- Ability to complement and support (not duplicate) work of existing regulatory bodies
- Genuine independence and impartiality (e.g., guaranteed funding)
- Ability to adapt to reflect technological developments, changing societal needs and expectations
- A structure that incorporates the right competencies and expertise, including multi-stakeholder representation from diverse backgrounds.

4.2.3 Recommendation: Establish role of AI (Ethics) Officer in organisations

To promote ethical awareness, **the role of AI Ethics Officers should be established to oversee AI ethics within organisations using and developing AI-assisted systems**. To be as effective as possible:
- AI developers and user organisations should have an internal position that combines scientific and technical understanding with ethical and human rights awareness to advise on the development and potential use of AI systems.
- Similar to the Data Protection Officer (DPO), this position should consist of a single, identified point of responsibility for overseeing compliance with and developing a culture of preserving AI ethical and human rights.
- The role must have necessary independence to speak out to prevent or mitigate risks of harm (akin to DPO and in-house counsel).
- The role could be *required* in high-risk cases, and only encouraged in others.
- The role may be aligned and combined with existing positions, e.g., DPO, Chief Ethics Officer, Corporate Social Responsibility Officer, or Business and Human Rights Officer.
- Qualification standards and education pathways for this position should be clarified.

There is no single role bearing responsibility for AI ethics concerns in all organisations using and developing AI-assisted systems. This could allow for unethical practices to develop and prosper.

Who should act?
The European Commission, the European Parliament, the European Council.

Key Considerations
In creating and/or implementing the role, key considerations are:
- Criteria for success:
  - Definition of the role in law
  - Presence of role-holders in organisations.
- Definitions, mandate, and parameters of the role need to be very clear to avoid creation of a powerless/"lame duck" position.
- Clear ethical guidelines are needed to direct and support role-bearers.

5. SHERPA Contributions to the AI Ethics Ecosystem

The SHERPA project, itself part of an ecosystem consisting of other EU-funded projects (e.g., SIENNA,[57] PANELFIT[58]), and numerous research groups and projects, continues to work on activities that can support and promote the SHERPA *Final Recommendations*. This section highlights past and ongoing contributions of the SHERPA project to creating the AI ethics ecosystem.

Recommendation: **Use appropriate and clear definitions of AI and digital technology.**
SHERPA's work on case studies[59] and scenarios[60] has informed the categorisation of AI in terms of narrow AI (machine learning), converging socio-technical systems and artificial general intelligence. This definition can help delimit ethical and human rights issues.

See also:
Ryan, M., Antoniou, J., Brooks, L., Jiya, T., Macnish, K., & Stahl, B. (2020). The Ethical Balance of Using Smart Information Systems for Promoting the United Nations' Sustainable Development Goals. *Sustainability*, *12*(12), 4826. https://doi.org/10.3390/su12124826

Recommendation: **Develop baseline model for AI impact assessment.**
SHERPA has provided accounts of ethical and human rights issues of AI in organisations as part of its **case study research**.[61] Likely future issues are discussed in the SHERPA **scenarios**.[62]

---

[57] SIENNA project (*Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact*), https://www.sienna-project.eu

[58] PANELFIT project (Participatory Approachers to a New Ethical and Legal Framework for ICT), https://www.panelfit.eu

[59] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[60] Wright, David (2019): D1.2 SIS Scenarios.

[61] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[62] Wright, David (2019): D1.2 SIS Scenarios.

SHERPA partners are currently undertaking a **structured review of the current landscape**, including critical analysis of the state of the art, and using this to develop a **proposal for an AI Impact Assessment baseline model**, outlining necessary requirements of an AI impact assessment for Horizon Europe researchers. Such AI impact assessments are likely to play an important role in the AI ecosystems and form part, for example, of ethics by design methods. There are several proposals for AI impact assessment, but there is currently a lack of a comprehensive overview.

Recommendation: Promote Ethics by Design for researchers in EC-funded projects

SHERPA has produced a **report of ethical concerns**[63] arising in relation to AI, and **10 case studies**[64] demonstrating how ethical issues arising from AI impact on different sectors of the economy.

SHERPA has also created sets of **guidelines for operationalisation of ethics by design for developers**[65] and for **users**[66] of smart information, further developed by the SIENNA project. These can be used as a foundation from which to develop guidelines for Horizon Europe research applicants and organisations.

The SHERPA guidelines for developers and guidelines for users of AI have provided the basis for the further development of Ethics by Design undertaken in the SIENNA project. This work feeds into the EC's guidance for ethics review of Horizon Europe projects. The further development of the Ethics by Design methodology will help researchers. It will also have to be made available to interested parties who intend to bid for Horizon Europe funding. SHERPA therefore plans to offer **Ethics by Design training** to interested audiences such as NCPs, adopting the SIENNA approach.

SHERPA furthermore intends to develop an Ethics by Design training programme that will be made available as a commercially available resource through the non-profit spin-out company ORBIT[67] and thus will remain available after the project.

Recommendation: Create training and education pathways that include ethics and human rights in AI

SHERPA has provided:

- Accounts of ethical and human rights issues of AI as part of its **case study** research[68].
- Likely future issues as discussed in the SHERPA **scenarios[69]**.
- An analysis of the **human rights aspects**[70].

[63] Ryan, Mark; Brey, Philip; Macnish, Kevin; Hatzakis, Tally; King, Owen; Maas, Jonne; et al. (2019): D1.4 Report on Ethical Tensions and Social Impacts.

[64] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[65] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design Approach. SHERPA Project. https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf

[66] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Use of AI and Big Data Systems. SHERPA Project. https://www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf

[67] ORBIT, www.orbit-rri.org

[68] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[69] Wright, David (2019): D1.2 SIS Scenarios.

[70] Andreou, A.; Laulhe-Shaelou, S.; Schroeder, D. (2019): D1.5: Current Human Rights Frameworks.

- Two sets of guidelines for operationalization of Ethics by Design for **developers**[71] and for **users**[72] of smart information, further developed by the SIENNA project.
- **Suggestions** from Stakeholders on the possible role of Education**,** as discussed in the analysis of the Stakeholders' focus groups and **interviews**[73].

SHERPA partners are currently contributing to the development of AI education through:
- **Publication of an open access Springer book** with novel case studies that highlight ethical and human rights issues and solutions in line with the SHERPA recommendations;
- **Development of teaching-oriented case studies** building on real-life case studies; and
- **Development of training** based on SHERPA-related outcomes, notably ethics by design, for relevant audiences, e.g., national contact points, researchers, designers, and developers.

Recommendation: **Include research finding on AI ethics in standardisation.**
SHERPA has also created sets of **guidelines for operationalization of ethics by design** for **developers**[74] and for **users**[75] of smart information, further developed by the SIENNA project.

SHERPA has contributed to the **standard on health and wellness apps** that includes an annex on ethics.

SHERPA members are represented on national mirror committees of **ISO/IEC JTC 1/SC 42 Artificial intelligence**.

Recommendation: **Undertake security analysis for machine learning systems.**
SHERPA has produced a deliverable **report on technical options and interventions** for Responsible Development of Smart Information Systems[76] and a deliverable **report on security issues, dangers and implications** of Smart Information Systems.[77]

Recommendation: **Develop a regulatory framework and enforcement mechanisms for AI.**
SHERPA has produced a deliverable **report on regulatory options**[78] and **policy brief for EU policymakers**[79].

SHERPA has also provided feedback to European policy-makers through **public written consultations**, including on the European Parliament Committee on Legal Affairs (JURI) report on a framework of ethical

---

[71] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design Approach. f

[72] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Use of AI and Big Data Systems.

[73] Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D2.2 Report of Interview Analysis.

[74] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design Approach.

[75] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Use of AI and Big Data Systems.

[76] Kirichenko, Alexey; Marchal, Samuel (2020): D3.5 Technical Options and Interventions.

[77] Patel, Andrew; Hatzakis, Tally; Macnish, Kevin; Ryan, Mark; Kirichenko, Alexey (2019): D1.3 Security Issues, Dangers and Implications of Smart Information Systems. SHERPA Project. https://doi.org/10.21253/DMU.7951292

[78] Rodrigues, Rowena; Laulhe Shaelou, Stephanie; Lundgren, Björn (2020): D3.3 Report on regulatory options.

[79] Trilateral Research, *Policy Briefing: Moving forward on regulating AI and big data in Europe*, 2020, SHERPA project, https://www.project-sherpa.eu/moving-forward-on-regulating-ai-and-big-data-in-europe-2/

aspects of artificial intelligence, robotics and related technologies[80], and on the European Commission White Paper on AI[81] and Inception Impact Assessment on the proposal for a legal act for AI[82].

Recommendation: Establish an independent European Union Agency for AI.
SHERPA developed a **Terms of Reference for a European Agency for AI**[83] and **policy brief for EU policy makers**[84].

Recommendation: Establish role of AI (Ethics) Officer in organisations.
SHERPA has produced **reports on ethical**[85] **and human rights concerns**[86] arising in relation to AI and 10 **case studies**[87] demonstrating how ethical issues arising from AI impact on different sectors of the economy.

SHERPA has also created sets of **guidelines for operationalisation of ethics by design** for **developers**[88] and for **users**[89] of smart information, further developed by the SIENNA project. These can be used as a foundation from which to develop guidelines to support the role of AI Ethics Officer.

---

[80] SHERPA Project, *SIENNA and SHERPA provide feedback on JURI report on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 22 May 2020, available for download at: https://www.project-sherpa.eu/commentary-on-the-european-parliament-committee-on-legal-affairs-draft-report-with-recommendations-to-the-commission-on-a-framework-of-ethical-aspects-of-artificial-intelligence-robotics-and-related/.

[81] SHERPA Project, *Supplemental Feedback to the European Commission on its White Paper on Artificial Intelligence*, 14 June 2020, https://www.project-sherpa.eu/wp-content/uploads/2020/07/sherpa-supplemental-document-for-ec-consultation-on-white-paper-for-ai-and-questionnaire-response.pdf

[82] SHERPA Project, *Feedback from the SHERPA Project to the European Commission on an Inception Impact Assessment on the proposal for a legal act*, 10 September 2020, available for download: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F550985

[83] Rodrigues, Rowena; Santiago, Nicole; Macnish, Kevin; Antoniou, Josephina; Wright, David (2020): D3.6 Feasibility of a new regulator and proposal for a European Agency for AI.

[84] Trilateral Research, *Policy Briefing: A European Agency for AI – Terms of Reference*, 2020, SHERPA project, https://www.project-sherpa.eu/a-european-agency-for-ai-terms-of-reference/

[85] Ryan, Mark; Brey, Philip; Macnish, Kevin; Hatzakis, Tally; King, Owen; Maas, Jonne; et al. (2019): D1.4 Report on Ethical Tensions and Social Impacts.

[86] Andreou, A.; Laulhe-Shaelou, S.; Schroeder, D. (2019): D1.5: Current Human Rights Frameworks.

[87] Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies.

[88] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design Approach.

[89] Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Use of AI and Big Data Systems.

6. Conclusion

Artificial Intelligence (AI) refers to a set of powerful technologies that are increasingly prevalent in the lives of individuals around the world. Like all technologies, while AI can be used for different purposes and therefore holds the potential to benefit individuals and society, it also raises many ethical and social concerns.

The use of the ecosystem metaphor to describe AI has the advantage of allowing a simultaneous view of various different technologies, individuals, organisations and jurisdictions. In order to find ways to move forward and help policy- and decision-makers to prioritise relevant actions, it is helpful to conceptualise AI ethics in terms of innovation ecosystems. Moreover, ecosystems are characterised by complexity and non-linear relationships. AI ecosystems contain many different technologies and stakeholders. They are dynamic and changing, which means that neither membership nor technical nor ethical positions remain static. Speaking of ecosystems means that concepts need to be clear (e.g., what is meant by ethics, human flourishing) and boundaries of systems need to be defined and set in the context of external environments.

Drawing on insights gained in the SHERPA project through development of case studies and scenarios, literature reviews, an online survey, stakeholder interviews, an expert Delphi study, and consultation with the SHERPA Stakeholder Board, we have developed these *Final Recommendations* to help ensure that the benefits of AI are harnessed, and related ethical issues and human rights concerns are addressed. To have a practical impact, the SHERPA project has developed these recommendations to be actionable and realistic so they can be implemented by stakeholders.

In order to ensure AI ecosystems prioritise ethics and fundamental rights, the SHERPA recommendations are grouped around three main conditions that should be fulfilled. One, interventions need to clearly delineate the ecosystem in terms of geographical and jurisdictional borders, but also in terms of technologies covered. Two, the successful steering of AI ecosystems in desired directions requires the creation of a knowledge base that its members can draw on for technical but also procedural and conceptual knowledge as well as processes for disseminating this knowledge and building capacity. And three, there is the need to define suitable measures of ecosystem governance that provide incentives, are flexible and allow the ecosystem as a whole to learn.

In the final months of the SHERPA project, we will continue to work on activities that can support and promote the SHERPA *Final Recommendations.* The development of the recommendations has highlighted several areas that would benefit from further work and to which SHERPA partners can continue to contribute. These activities are especially pressing due to the upcoming expected EU regulation on AI and the start of the Horizon Europe research framework programme where ethical and human rights issues of AI will play a prominent role. SHERPA intends to utilise remaining resources from the project to fund these activities and support the creation of AI ecosystems that are beneficial to people, uphold human rights and, more generally, promote human flourishing.

Appendix 1: SHERPA Insights: Ethics and Human Rights of AI

The following sections highlight some of the findings and insights arising from SHERPA work. They provide some of the background that justifies the recommendations listed above. Full detail of the work undertaken in the project is available from the website.

Concepts of AI

The following figure represents three different views of AI that have different characteristics that raise different types of ethical concerns:
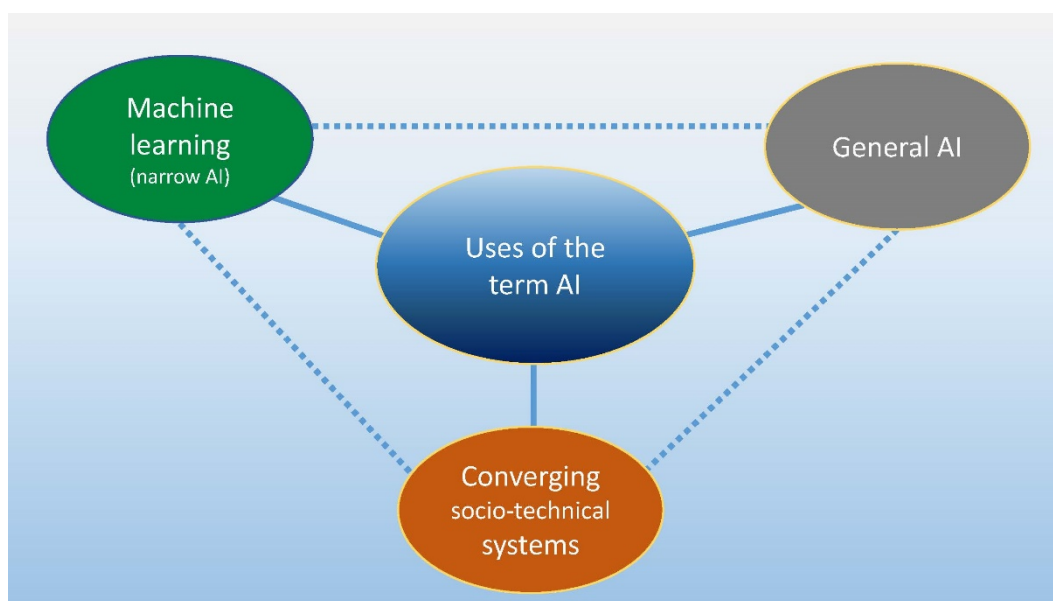

Figure 1: Views on the use of the term 'AI'

Machine learning as the primary example of narrow AI that has stoked the recent AI successes is based on statistical techniques, which require large quantities of data and significant computer resources. General AI which mimics human cognitive abilities does not currently exist, but is of interest because it raises numerous fundamental questions. The use of the term 'AI' to denote converging socio-technical systems points to the apparent autonomy of technology, its potential to manipulate human actions and the technology's social impact.

Purpose of AI

Artificial Intelligence (AI), like all technologies, can be used for different purposes. From an ethical and human rights point of view one can distinguish three main purposes of AI use: efficiency and optimisation, social control and human flourishing.
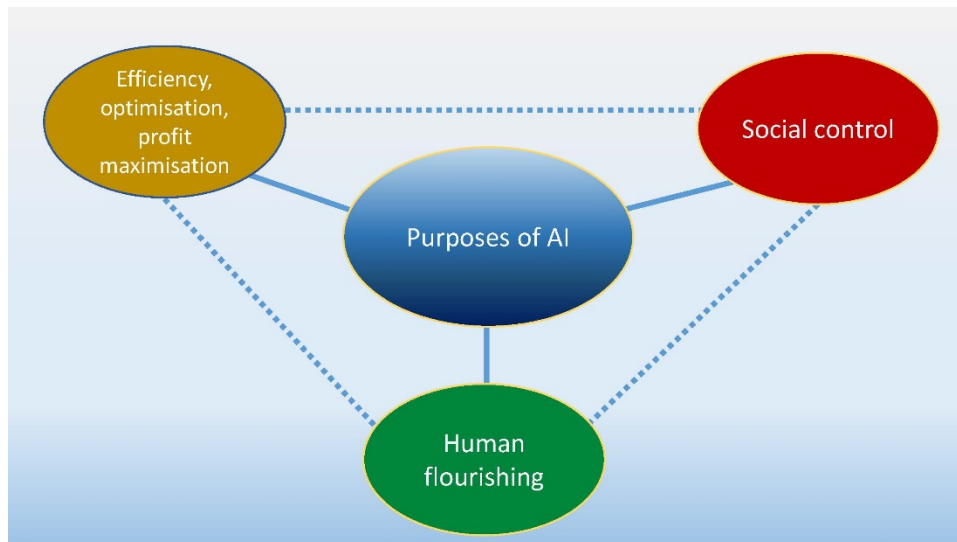
Figure 2: Purposes of the use of AI

These three different purposes are not contradictory or mutually exclusive, but point to very different driving forces for development and utilisation of these technologies. The framing of AI ethics in terms of human flourishing is consistent with numerous national and international ethics guidelines and principles, including those published by the EU's High Level Expert Group (2019)

Ethical issues

Using 10 case studies, 5 scenarios, reviews of various bodies of literature, an online survey with more than 300 respondents, 45 stakeholder interviews, and an expert Delphi study, SHERPA identified a significant number of ethical issues, as shown in the following figure.

We use the term ethical issues to denote issues that were perceived to be problematic by our respondents. These explicitly include issues that are already covered by human rights and other legislation (e.g. privacy, discrimination) but also cover issues that are less clearly specified (e.g. transparency, loss of human contact) or closely related to technical aspects (e.g. security)

Figure 3: Ethical Issues of AI as identified by SHERPA

A closer analysis shows that these ethical issues can be categorised into three types of issues, related to the views of AI introduced above: specific issues arising from machine learning, general questions about living in a digital world, and metaphysical questions.
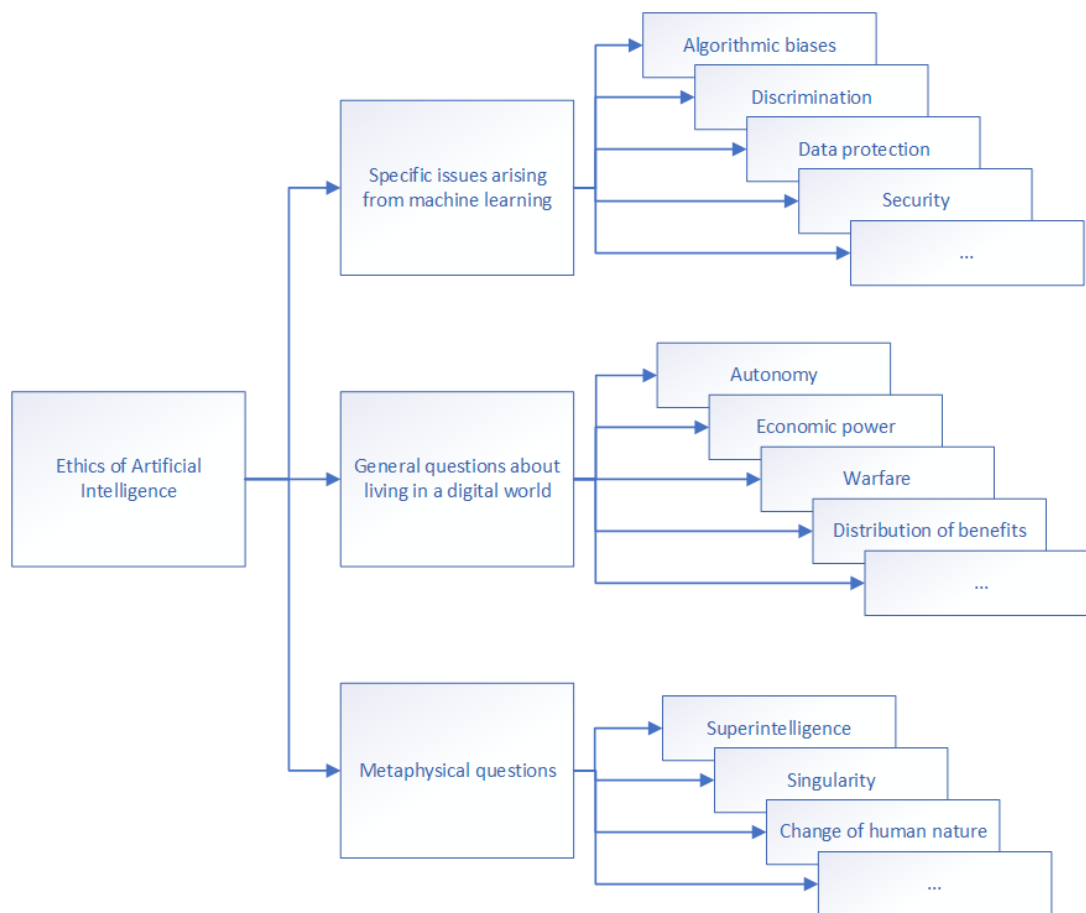
Figure 4: Categorisation of Ethical Issues of AI

Empirical research in SHERPA has demonstrated that this categorisation helps to represent the ethics of AI. This categorisation is consistent with different concepts of AI and gives an initial pointer to ways of addressing them.

Appendix 2: Mitigation of Ethical Issues of AI

In order to address the ethical issues, it is useful to first think about stakeholders that are critical to address and engage with in mitigating such issues.
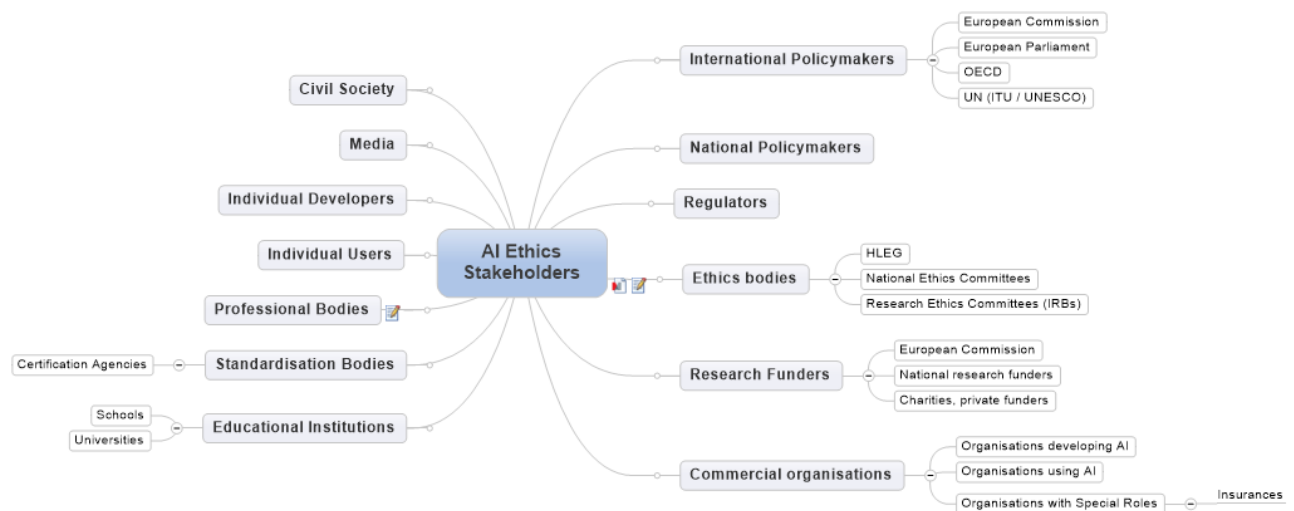

Figure 5: Stakeholders in AI Ethics

Categories of Mitigation Options

There is a large array of possible mitigation options that are currently being discussed, which address particular issues, aspects, technologies or applications. The following figure provides an overview of these split into options on the policy level, options for companies and commercial organisations, and activities to be undertaken by individuals.
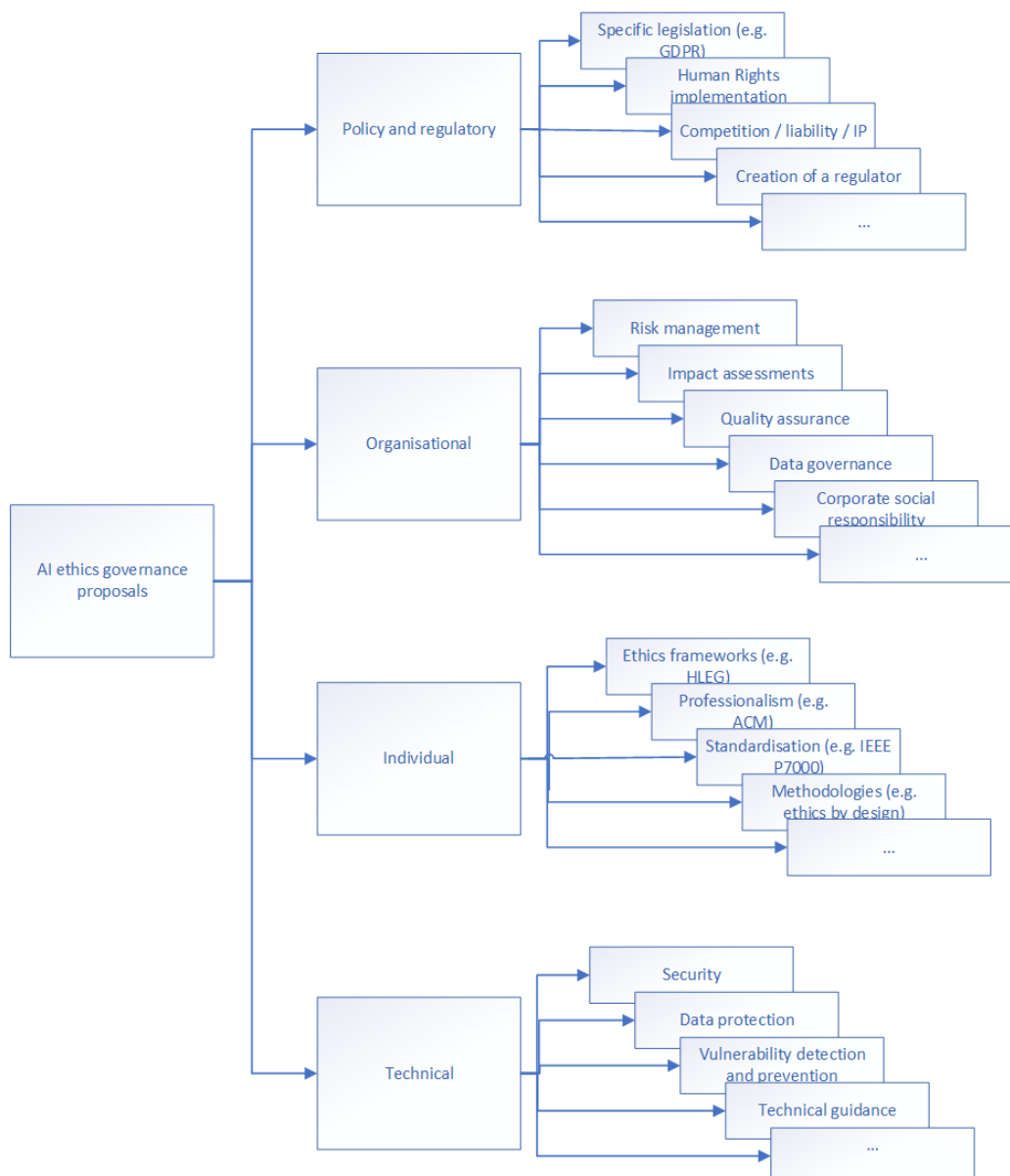
Figure 6: AI Ethics Mitigation Options by Type of Stakeholder

The options covered in figure 6 are not meant to be comprehensive, but give an idea of what type of action is open to various types of stakeholders.

AI Ethics as the Alignment of the AI Ecosystem

The earlier overviews of types of AI, ethical issues, stakeholders involved and possible mitigation options adds to an understanding of the complexity of the subject matter. At this point it is clear that there is no simple way of dealing with the ethics of AI that can cover the various levels and categories. In order to find ways to move forward and help policy- and decision-makers to prioritise relevant actions, it is helpful to conceptualise AI ethics in terms of innovation ecosystems.
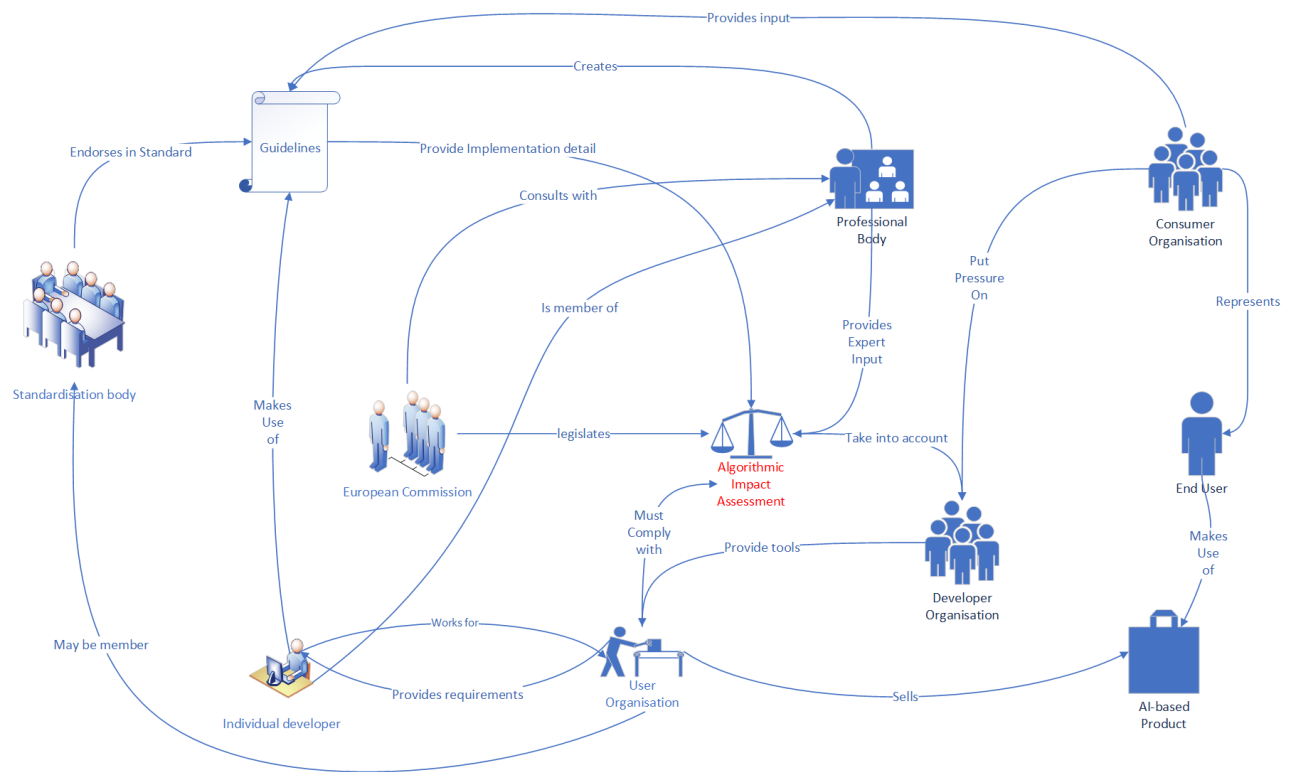
Figure 7: Systems-inspired view of stakeholders and activities to implement algorithmic impact assessment

Figure 7 is an attempt to capture the systems-theoretical nature of one particular mitigation strategy, namely the creation and enforcement of an algorithmic impact assessment (inspired by data protection impact assessments (DPIAs) mandated by the GDPR). It shows sample stakeholders, actions, activities and interdependencies that would be involved in this process. It does not claim completeness and similar diagrams could be created for all other mitigation strategies. The main purpose of the figure is to demonstrate that interpreting AI ethics in terms of (innovation) ecosystems can help bring together and synthesise the many disparate activities in AI Ethics.

Challenges of Ethical Governance of AI Ecosystems

The use of the ecosystem metaphor to describe AI has the advantage of allowing a simultaneous view of various different technologies, individuals, organisations and jurisdictions. However, it raises a number of additional questions. These start with the implications for ethics arising from using a metaphor from a biological system, which require clear justifications of the use of normative statements.

Moreover, ecosystems are characterised by complexity and non-linear relationships. AI ecosystems contain many different technologies and stakeholders. They are dynamic and changing which means that neither membership nor technical or ethical positions remain static. Speaking of ecosystems means that concepts need to be clear (e.g. what is meant by ethics / human flourishing) and boundaries of systems need to be clearly defined and set in the context of external environments.
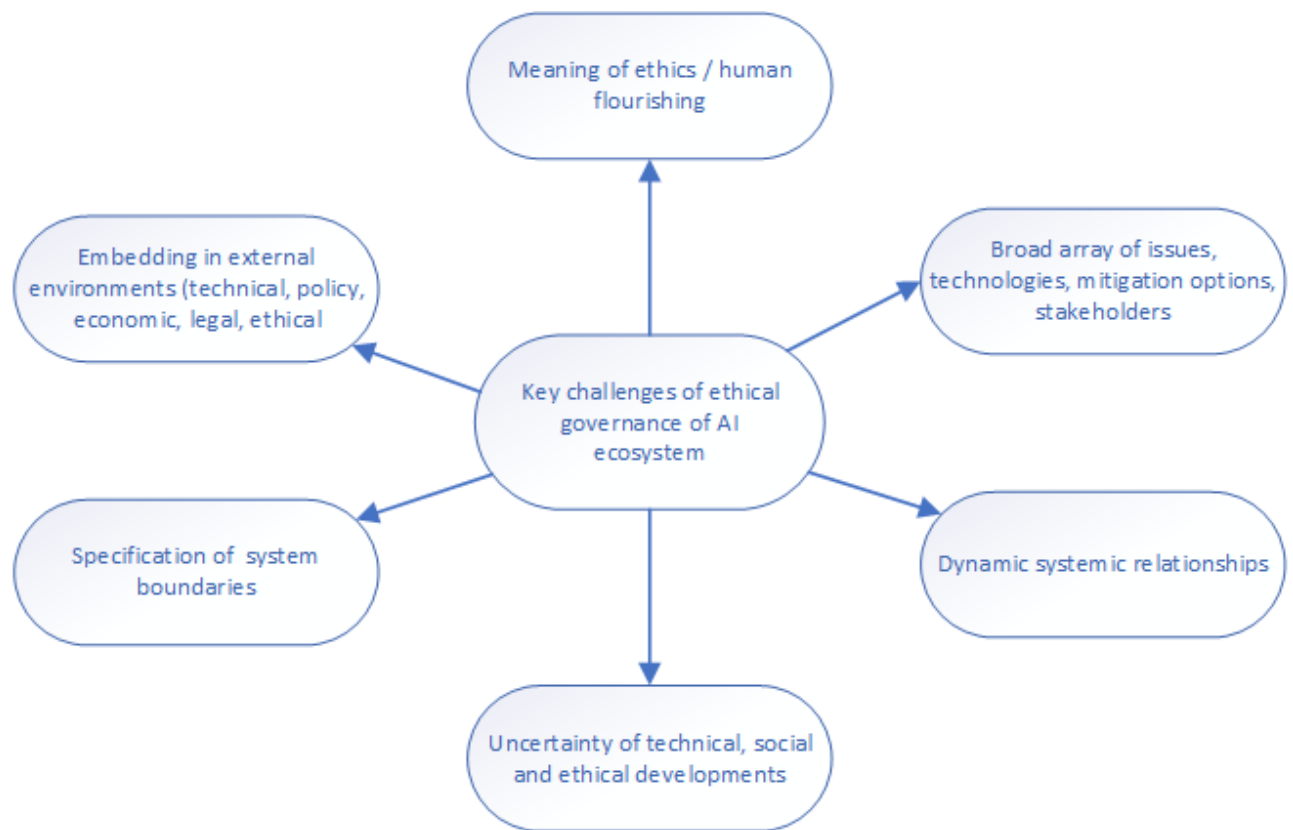
Figure 8: Key challenges of ethical governance of AI ecosystems

Figure 8 provides an overview of the key challenges that the use of the ecosystem metaphor raises for ethics of AI.

These challenges can be translated into requirements that interventions into AI ecosystems need to fulfil, if they are to have their desired consequence of shaping systems to promote human flourishing. The interventions need to clearly delineate the ecosystem in terms of geographical and jurisdictional borders, but also in terms of technologies covered. Secondly, the successful steering of AI ecosystems in desired directions requires the creation of a knowledge base that its members can draw on for technical, but also procedural and conceptual knowledge, as well as processes for disseminating this knowledge and building capacity. Finally, there is the need to define suitable measures of ecosystem governance that provide incentives, are flexible and allow the ecosystem as a whole to learn. These requirements, shown in figure 9, have guided the development of the SHERPA *Final Recommendations*.
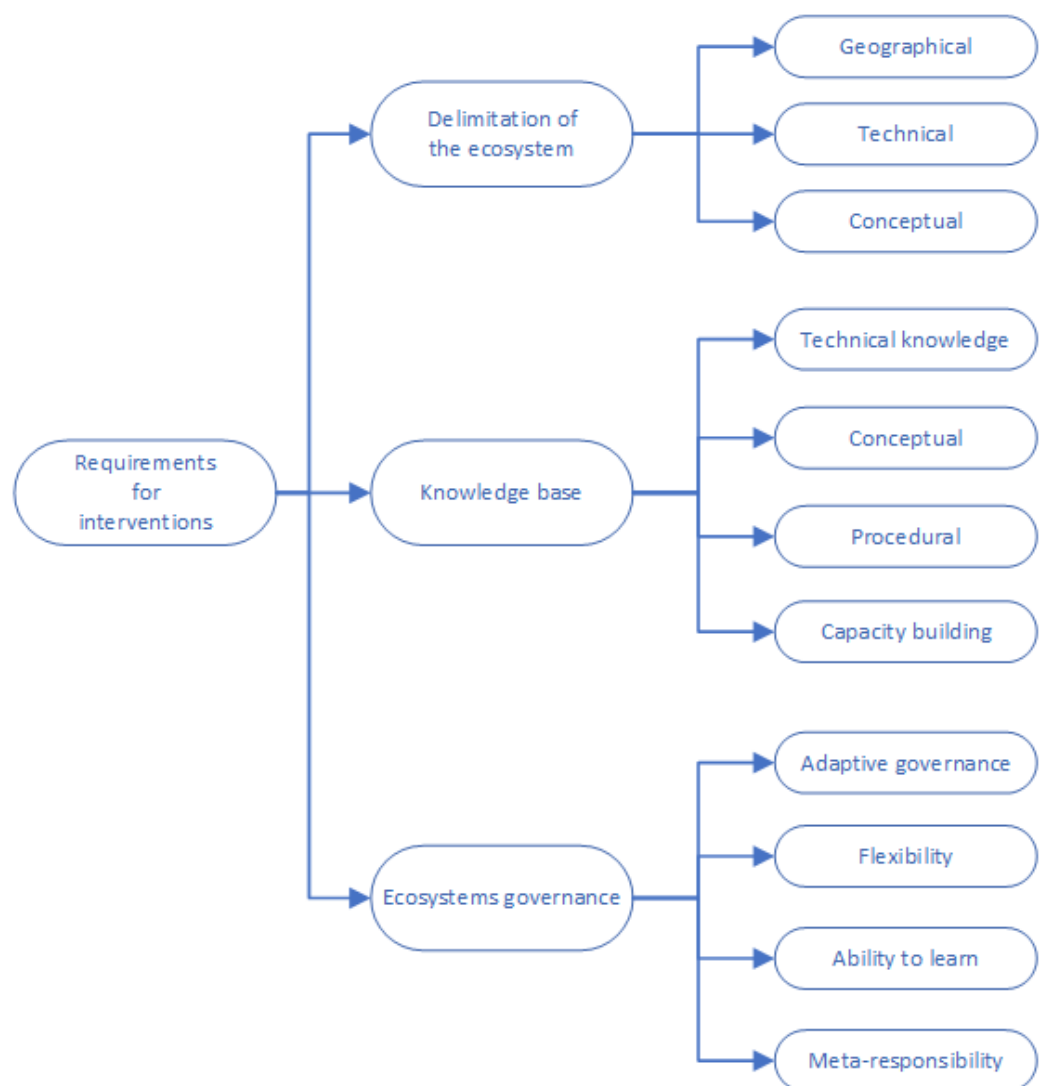
Figure 9: Requirements for interventions into AI ecosystems

The SHERPA *Final Recommendations* are based on the logic outlined in this appendix. They arise from the work done within the project and seek to promote those interventions that are most likely to shape the European AI ecosystem in a way that is conducive to human flourishing.

References

AIEI Group. (2020). From Principles to Practice—An Interdisciplinary framework to operationalise AI ethics (p. 56). VDE / Bertelsmann Stiftung. https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf

Andreou, A.; Laulhe-Shaelou, S.; Schroeder, D. (2019): D1.5: Current Human Rights Frameworks. SHERPA Proejct. https://doi.org/10.21253/DMU.8181827

Bijlsma, Marlou; Zijlstra, Thamar (2020): D3.4 From Good Practices to Standard Practices. SHERPA Project. https://doi.org/10.21253/DMU.13553684

Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design Approach. SHERPA Project. https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf

Brey, Philip; Lundgren, Björn; Macnish, Kevin; Ryan, Mark (2019): Guidelines for the Ethical Use of AI and Big Data Systems. SHERPA Project. https://www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf

Brooks, Laurence; Stahl, Bernd; Jiya, Tilimbe (2020): D2.3 Online survey report. SHERPA Project. https://doi.org/10.21253/DMU.11777478.

Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14*, 4 November 1950, ETS 5, entry into force 3 September 1953.

Digital Catapult. (2020). Lessons in practical AI ethics: Taking the UK's AI ecosystem from 'what' to 'how.' Digital Catapult. https://assets.ctfassets.net/nubxhjiwc091/xTEqMcYudwQ7GHZWNoBfM/c2a2d55a0ee1694e77634e240eafdfdf/20200430_DC_143_EthicsPaper__1_.pdf

European Commission. (2020a). White Paper on Artificial Intelligence: A European approach to excellence and trust (White Paper COM(2020) 65 final). https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission. (2020b). Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (COM(2020) 64 final). European Commission. https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf

European Commission. (2020c). "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Shaping Europe's Digital Future," No. COM/2020/67 final, Brussels: European Commission, February 19. (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:67:FIN)

European Commission. (2020d). Inception impact assessment for the Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence -

Ares(2020)3896535. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements

European Commission. (2021). *EU digital education action plan (2021-2027)*. https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en

European Commission, Digital Competence Framework for Educators (DigCompEdu). https://ec.europa.eu/jrc/en/digcompedu.

EGE, European Group on Ethics in Science and New Technologies (2018). Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Publications Office of the European Union. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

European Parliament. (2020). DRAFT REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)). European Parliament, Committee on Legal Affairs. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/PR/2020/05-12/1203395EN.pdf

European Union, *Charter of Fundamental Rights of the European Union*, 2000/C 364/01, entry into force 7 December 2000.

HLEG on AI (2020a). Sectorial considerations for Trustworthy AI - taking AI's context specificity into account. European Commission.https://futurium.ec.europa.eu/en/european-ai-alliance/blog/sectorial-considerations-trustworthy-ai-taking-ais-context-specificity-account

HLEG on AI (2020b): Assessment List for Trustworthy AI (ALTAI). European Commission. https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence

HLEG on AI (2019a). Ethics Guidelines for Trustworthy AI. European Commission - Directorate-General for Communication. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

HLEG on AI (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence. European Commission - Directorate-General for Communication. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence

IEEE, *IEEE 7010-2020 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*. https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html

Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D2.2 Report of Interview Analysis. SHERPA Project. https://doi.org/10.21253/DMU.13168265

Iordanou, Kalypso; Christodoulou, Eleni; Antoniou, Josephina (2020): D4.2 Evaluation Report. SHERPA Project. https://doi.org/10.21253/DMU.12917717

Kirichenko, Alexey; Marchal, Samuel (2020): D3.5 Technical Options and Interventions. SHERPA Project. https://doi.org/10.21253/DMU.13187081

Li, Deyi; Du, Yi (2007), Artificial Intelligence with Uncertainty.

Macnish, Kevin; Ryan, Mark; Gregory, Anya; Jiya, Tilimbe; Antoniou, Josephina; Hatzakis, Tally; et al. (2019): D1.1 Case studies. SHERPA Project. https://doi.org/10.21253/DMU.7679690

Macnish, K.; Ryan, M. (2019). D3.2 Guidelines for the development and the use of SIS. SHERPA. https://doi.org/10.21253/DMU.11316833.v1

Miller, C., & Ohrvik-Stott, J. (2018). Regulating for Responsible Technology—Capacity, evidence and redress: A new system for a fairer future. Doteveryone. https://www.doteveryone.org.uk/wp-content/uploads/2018/10/Doteveryone-Regulating-for-Responsible-Tech-Report.pdf

Organisation for Economic Co-operation and Development, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, adopted 22 May 2019. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Patel, Andrew; Hatzakis, Tally; Macnish, Kevin; Ryan, Mark; Kirichenko, Alexey (2019): D1.3 Security Issues, Dangers and Implications of Smart Information Systems. SHERPA Project. https://doi.org/10.21253/DMU.7951292

Perset, Karine; Murdick, Dewey; Clark, Jack; Grobelnik, Marko, A first look at the OECD's Framework for the Classification of AI Systems, designed to give policymakers clarity, OECD, 24 November 2020. https://oecd.ai/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers

Platform voor de Informatie Samenleving (ECP), *Artificial Intelligence Impact Assessment*. https://ecp.nl/publicatie/artificial-intelligence-impact-assessment-english-version/

Rodrigues, Rowena; Laulhe Shaelou, Stephanie; Lundgren, Björn (2020): D3.3 Report on regulatory options. SHERPA Project. https://doi.org/10.21253/DMU.11618211

Rodrigues, Rowena; Santiago, Nicole; Macnish, Kevin; Antoniou, Josephina; Wright, David (2020): D3.6 Feasibility of a new regulator and proposal for a European Agency for AI. SHERPA Project. https://doi.org/10.21253/DMU.13168295

Ryan, Mark; Brey, Philip; Macnish, Kevin; Hatzakis, Tally; King, Owen; Maas, Jonne; et al. (2019): D1.4 Report on Ethical Tensions and Social Impacts. SHERPA Project. https://doi.org/10.21253/DMU.8397134

Santiago, Nicole; Stahl, Bernd; Keene, Paul; Hatzakis, Tally; Rodrigues, Rowena; Wright, David (2020): D2.4 Delphi Study Report. SHERPA Project. https://doi.org/10.21253/DMU.13168553

SHERPA Final Recommendations, https://www.project-sherpa.eu/recommendations/

SHERPA Project, *Feedback from the SHERPA Project to the European Commission on an Inception Impact Assessment on the proposal for a legal act*, 10 September 2020, available for download: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F550985

SHERPA Project, *SIENNA and SHERPA provide feedback on JURI report on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 22 May 2020, available for download at: https://www.project-sherpa.eu/commentary-on-the-european-parliament-committee-on-legal-affairs-

draft-report-with-recommendations-to-the-commission-on-a-framework-of-ethical-aspects-of-artificial-intelligence-robotics-and-related/.

SHERPA Project, *Supplemental Feedback to the European Commission on its White Paper on Artificial Intelligence*, 14 June 2020, https://www.project-sherpa.eu/wp-content/uploads/2020/07/sherpa-supplemental-document-for-ec-consultation-on-white-paper-for-ai-and-questionnaire-response.pdf

Trilateral Research, *Policy Briefing: A European Agency for AI – Terms of Reference*, 2020, SHERPA project, https://www.project-sherpa.eu/a-european-agency-for-ai-terms-of-reference/

Trilateral Research, *Policy Briefing: Moving forward on regulating AI and big data in Europe*, 2020, SHERPA project, https://www.project-sherpa.eu/moving-forward-on-regulating-ai-and-big-data-in-europe-2/

Universal Declaration of Human Rights, adopted by General Assembly Resolution 217 A(III) of 10 December 1948.

Wright, David (2019): D1.2 SIS Scenarios. SHERPA Project. https://doi.org/10.21253/DMU.8181695