



SHERPA
Advocating for Ethical AI

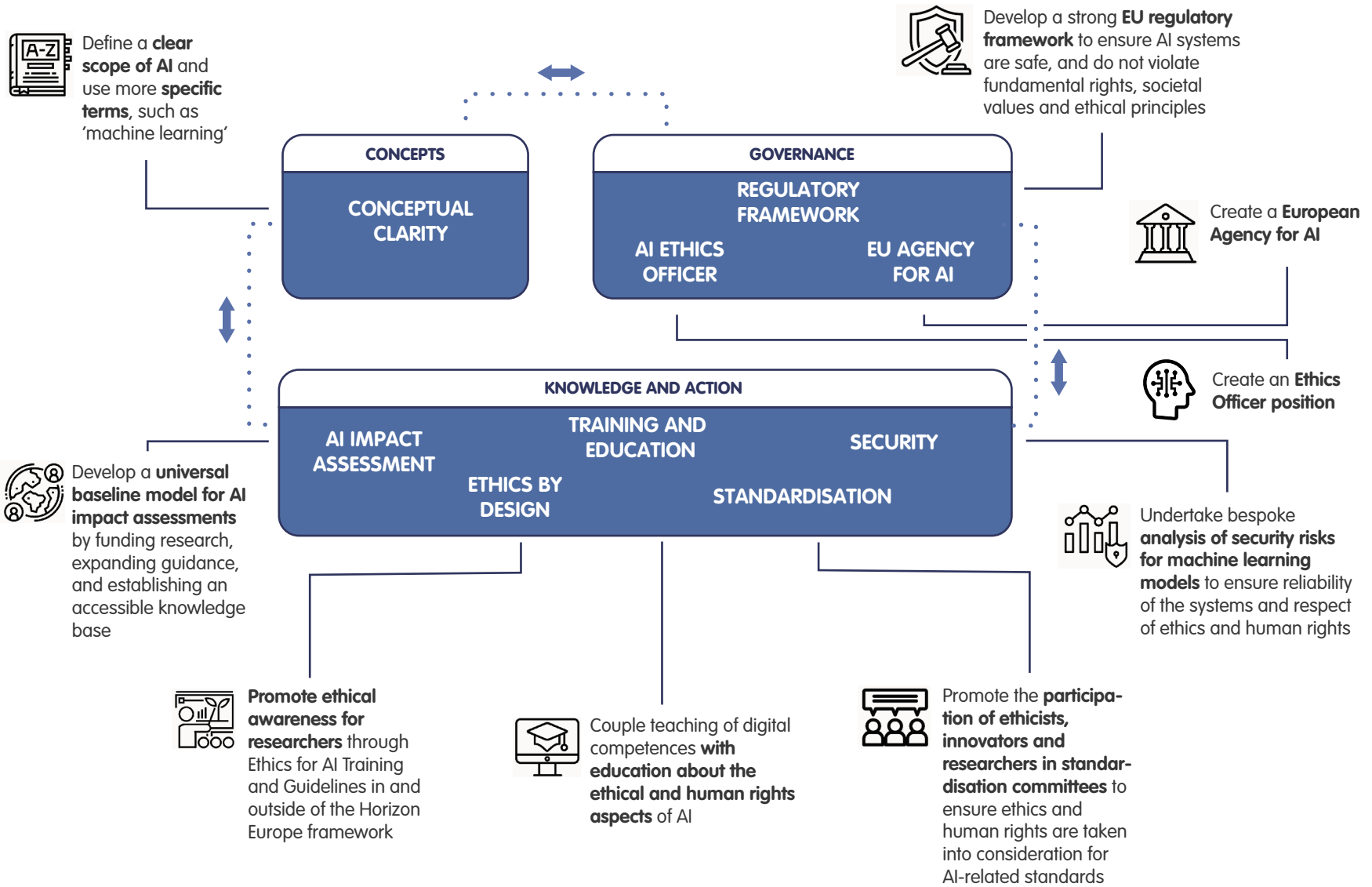
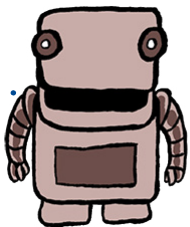
SHERPA RECOMMENDATIONS

HOW CAN WE ENSURE THAT ARTIFICIAL INTELLIGENCE IS ETHICAL AND SUPPORTS HUMAN RIGHTS ?

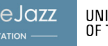
About the SHERPA Project

In collaboration with stakeholders, the SHERPA project investigates, analyses and synthesises our understanding of the ways in which smart information systems (SIS) i.e systems that incorporate AI and big data analytics, impact ethics and human rights issues.

Now is the time to harness the benefits of SIS and simultaneously address the downsides. The SHERPA project is uniquely placed to guide the ongoing debate, focus it and develop actionable recommendations and advocate them to ensure that SIS promote the public good.



Partners



<https://www.project-sherpa.eu/recommendations/> • @project_sherpa

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 786641





AI, Ethics and Human Rights - Designing a Better World

Recommendations from the SHERPA project



USE APPROPRIATE AND CLEAR DEFINITIONS OF ARTIFICIAL INTELLIGENCE (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', 'deep neural network', or 'reinforcement learning'.
- It is important to use appropriate and clear terminology to distinguish between different technologies that fall under the umbrella of AI, because specific technologies bring about unique ethical and human rights implications.

Who should act?

The key stakeholders include bodies developing policies and guidelines for ethical or trustworthy AI, including the European Commission, national governments and standardisation bodies.

Related SHERPA Resources

SHERPA has developed **case studies and scenarios** that inform the categorisation of AI in terms of narrow AI (machine learning), converging socio-technical systems and artificial general intelligence. These definitions can help delimit ethical and human rights issues.

DEVELOP A BASELINE MODEL FOR AI IMPACT ASSESSMENTS



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 risk level for a technology or application
- Issues to be included in AI impact assessments

It's important to develop a rigorous and standardised approach to AI impact assessments in order to define, measure, interpret and address relevant ethics and human rights risks associated with these technologies. This will help researchers and developers to harness the benefits of AI while mitigating potential risks or negative implications.

Who should act?

The key stakeholder is the European Commission, ideally led by a **EU Agency for AI**.

Related SHERPA Resources

SHERPA has developed **case studies and scenarios** that provide accounts of ethical and human rights risk associated with AI in organisations. Identifying and understanding these issues is key to developing a successful baseline model for AI impact assessment.





SHERPA partners are also 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 of ethics by design methods, for example.



PROMOTE ETHICS BY DESIGN FOR RESEARCHERS IN EC-FUNDED PROJECTS

Ethical AI should be encouraged through the promotion of an **Ethics by Design** approach to technology design and development. This means assessing, identifying, and addressing ethical concerns from the conception through implementation phases of technology development.

Ethics by design is important because 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. Adopting an ethics by design approach will help to mitigate the risk of funding projects developing or involving AI which are damaging in terms of ethics and human rights.

Who should act?

The key stakeholders are the Ethics sections in European Commission, in particular the EC's **Directorate-General Research and Innovation** and the **Research Executive Agency**

Related SHERPA Resources

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

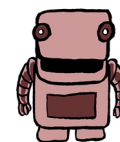
These guidelines for developers and users of AI provide the basis for the further development of Ethics by Design undertaken in the SIENNA project. This work feeds into the EC's guidance for ethics review of Horizon Europe projects. SHERPA also plans to offer **Ethics by Design training** to interested audiences such as NCPs, adopting the SIENNA approach, and 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** and thus will remain available after the project.

CREATE TRAINING AND EDUCATION PATHWAYS THAT INCLUDE ETHICS AND HUMAN RIGHTS IN AI



Current technology-oriented curricula should be updated to ensure that ethical and human rights aspects of AI and related technologies are covered. Curricula should include:

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



Creating training and education pathways is important because the ability to recognise and deal with ethical and human rights aspects of emerging digital technologies (such as AI) depends on better awareness and understanding of the risks and harms.





Who should act?

The key stakeholders are the European Commission, Council of Europe, national governments, in particular national education bodies, and any bodies involved in designing curricula.

Related SHERPA Resources

SHERPA has investigated the importance of education and training in ethics of AI in both the Online Survey and Delphi Study. SHERPA has also developed:

- Accounts of ethical and human rights issues of AI as part of its **case study** research.
- Likely future issues as discussed in the SHERPA **scenarios**.
- An analysis of the **human rights aspects**.
- Two sets of guidelines for operationalization of Ethics by Design for **developers** and for **users** 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**.

In addition, 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.



INCLUDE RESEARCH FINDINGS ON AI ETHICS IN STANDARDISATION

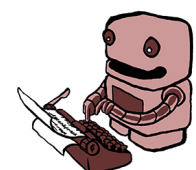
- Ethicists, innovators and researchers should be included in standardisation committees, and be actively engaged in the standardisation process.
- Including AI ethics experts in the standardisation process is important to ensure that the ethical, social, legal and human rights implications of new and emerging technologies inform and help to shape standards.

Who should act?

The key stakeholders are standardisation committees, as well as ethicists, innovators and researchers.

Related SHERPA Resources

SHERPA has produced a set of guidelines for operationalisation of ethics by design for **developers** and for **users** of smart information systems that can inform standardisation. SHERPA has also contributed to the standard on health and wellness apps that includes an annex on ethics; and members of the project are represented on national mirror committees of **ISO/IEC JTC 1/SC 42 Artificial intelligence**. Designers, developers, integrators, and operators of AI systems should undertake bespoke analysis of security threats and risks for AI-powered systems. This includes threat enumeration and analysis, as well as real-time monitoring and analysis of external inputs.



UNDERTAKE SECURITY ANALYSIS FOR MACHINE LEARNING SYSTEMS



Conducting security analyses is important because AI systems may have security vulnerabilities that render these systems subject to cyberattacks. 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?

The key stakeholders are the designers, developers, integrators, and operators of AI systems.

Related SHERPA Resources

SHERPA has produced reports on **Technical Options and Interventions for Responsible Development of Smart Information Systems** as well as on **Security Issues, Dangers and Implications of Smart Information Systems**. These can inform best practices for undertaking a security analysis of AI systems.



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. Such a framework would set a **strong legal standard** at the EU-level that establishes high standards of protection of fundamental rights and societal values.

Developing a mandatory EU regulatory framework for AI is important because existing soft law initiatives do not sufficiently address the potential human rights and safety implications of new and emerging technologies. Additionally, EU Member States are developing and implementing their own regulatory frameworks, creating fragmentation and confusion.

Who should act?

The key stakeholders are the European Commission, the European Parliament and the European Council.

Related SHERPA Resources

SHERPA has produced a **report on regulatory options** and **policy brief for EU policymakers**.

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

ESTABLISH AN INDEPENDENT EUROPEAN UNION AGENCY FOR AI



The EU should establish an independent **European Union Agency** for AI. The Agency should be responsible for activities including: making recommendations; identifying restrictions on AI development, deployment, and use; setting benchmarks for enforcement; cooperating, liaising, and promoting public dialogue, and promoting the EU's AI approach through international cooperation.

Establishing an EU agency for AI is important because 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 key stakeholders are the European Commission, the European Parliament and the European Council.

Related SHERPA Resources

SHERPA developed a **Terms of Reference for a European Agency for AI** and **policy brief for EU policy makers**. These documents, based on the project's research and expert consultations, including interviews, a focus group, and feedback from the SHERPA Stakeholder Board, can be used as a basis for establishing an independent European Union agency for AI.



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.
- Creating the role of AI Ethics officer is important because, at present, there is no single individual or role bearing responsibility for AI ethics in organisations using and developing AI-assisted systems. This could allow for unethical practices to develop and prosper.

Who should act?

The key stakeholders are the European Commission, the European Parliament, and the European Council.

Related SHERPA Resources

SHERPA has produced **reports on ethical and human rights concerns** arising in relation to AI and **10 case studies** 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** and for **users** 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.

