

**TRILATERAL
RESEARCH**
Ethical AI



We build **Ethical AI solutions** to tackle complex societal problems.

www.trilateralresearch.com

UK

One Knightsbridge Green, London,
SW1X 7QA

IRELAND

Marine Point, Belview Port, Waterford,
X91W0XW

***Looking back, moving forward:
AI research ethics***

Forum on Digital Ethics in Research
17 October 2022

Rowena Rodrigues, PhD
Head of Innovation & Research Services

rowena.rodrigues@trilateralresearch.com

@RRodriguesTRI

Trilateral Research

Ethical AI company

based in UK/IE, est. 2004, 120+ personnel

Pioneering AI solutions

for law enforcement, defence, health, climate, and public agencies

Supporting services

including Innovation & Research, Data Protection, Cybersecurity, Sociotech Insights



This presentation will cover

- **SIENNA work**
- **Existing Codes and Guidelines on AI ethics**
- **Briefly, what they cover and should consider**
- **Implications of the EU AI regulatory proposals for researchers.**

IRECS

Project title: improving Research Ethics Expertise and Competences to Ensure Reliability and Trust in Science

Type of action: HORIZON Coordination and Support Actions

Granting authority: European Research Executive Agency

Duration: 1 Oct 2022 - 30 Sept 2025

Project number: 101058587



New tech needs in RECs



Training materials

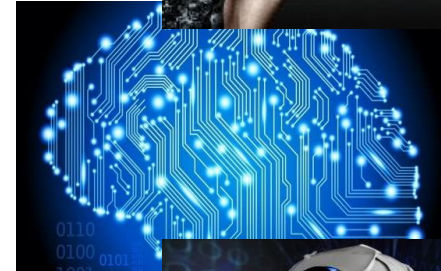


Training programmes



SIENNA

- Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact
- EU-funded research project (Horizon 2020)
- Focussed on three emerging tech fields: **Artificial Intelligence (AI) & Robotics, Human Enhancement, Human Genomics**
- Developed ethical frameworks, policy recommendations and operational tools
- 8 EU and 5 non-EU partners
- Duration: Nov 2017 to April 2021
- Budget: 4.0 M euros
- <https://www.sienna-project.eu>



Surveys of research ethics committee (REC) approaches and AI and robotics codes

International Codes and Guidelines: examples

UNESCO RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE (2021)
EDPS OPINION 3/2018 ON ONLINE MANIPULATION AND PERSONAL DATA (2018)
Mission Statement and Berlin Statement (2009)
Humanitarian UAV Code of Conduct & Guidelines (2014)
Asilomar AI Principles (2017)
REPORT OF COMEST ON ROBOTICS ETHICS (2017)
SHERPA Guidelines for Use and Developments of AI or big data system (2019)
Opportunities 4/2019 Towards a new digital ethics: Data, Dignity and Technology
AI HLEG ETHICS GUIDELINES FOR TRUSTWORTHY AI (2019)
COMMISSION OF THE EUROPEAN COMMUNITIES (2002) Code of Ethics for Professionals in the Information Society
ACM Code of Ethics and Professional Conduct (1992/2018 update)
IEEE Code of Ethics (revised 2017)
The Montreal Declaration for a Responsible Development of Artificial Intelligence (2017)
Top 10 Principles For Ethical Artificial Intelligence (2017)
SOFTWARE ENGINEERING CODE OF ETHICS
INCOSE Code of Ethics

See also: Algorithm Watch's AI Ethics Guidelines Global Inventory

National codes and guidelines: examples

China	France	Germany	Netherlands	Sweden	USA	UK
<ul style="list-style-type: none">• The Ethical Norms for the New Generation Artificial Intelligence, <i>National Governance Committee for the New Generation Artificial Intelligence, 2021</i>	<ul style="list-style-type: none">• Research Ethics in Machine Learning, <i>CERNA 2017</i>	<ul style="list-style-type: none">• Automated and connected driving, <i>Ethics Commission, 2017</i>	<ul style="list-style-type: none">• Human Rights in the Robot Age: Challenges Arising from the Use of Robotics, Artificial Intelligence, and Virtual and Augmented Reality, <i>Rathenau Institute, 2017</i>	<ul style="list-style-type: none">• Robots and surveillance in the care of older - ethical aspects, <i>Swedish National Council on Medical Ethics, 2014</i>	<ul style="list-style-type: none">• The National Artificial Intelligence Research and Development strategic plan, <i>National Science and Technology Council, 2016</i>	<ul style="list-style-type: none">• Ethical Principles for AI in Defence, <i>Ministry of Defence, 2022</i>• Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <i>The Alan Turing Institute, 2019</i>• Statement of Ethical Principles, <i>Engineering Council and the Royal Academy of Engineering in 2005 (revised 2017)</i>



National – application areas covered by Codes

- ❖ Health: telemedicine, robotic surgery, mobile phone apps, medical imaging and diagnosis, care robots.
- ❖ Military/defence: robots in defence and security, autonomous weapon systems and autonomous weaponised drones
- ❖ Intelligent systems and products: intelligent manufacturing
- ❖ Robots that interact with people and groups.
- ❖ Use of algorithms.
- ❖ Use of machine learning: machine image recognition, natural language processing

Ethical issues covered

accountability	autonomy	conflicts of interest	consent	control over technology	de-socialization	deception
deskilling	discrimination	fair trial	fairness	family life	freedom of expression	human contact
human dignity	isolation	justice	objectification	peaceful enjoyment of possessions	privacy	productivity
proportionality	protection of personal data	responsibility	safety	social relations	stability of society	transparency
			well-being			

Considerations for Codes: key messages



Clear objectives: precise and useful.



Means or pointers to interpret or help achieve the objectives



Pragmatic - easy to operationalise



Include monitoring and enforcement mechanisms (including organisational) and review



Ensure continued relevance and sustainability

AI REC-related challenges

- Current focus in RECs on harm to human participants. Need to move extend/adapt this focus.
- RECs need support to develop **expertise in evaluating societal impacts** of AI research
- Lack of **guidance** on ethical risks and **training** on assessment
- Inadequate/lack of RE assessment or access to RECs = risks continue into products and create liabilities
- REC **structures and flexibility**
- Need for **multi-stage/repeat** ethics reviews in AI research
- Need for further **field-specific good practices** for ethics disclosure and review.



EU legislative proposals

Three inter-related legal initiatives for building trustworthy AI:

- ❖ A [European legal framework for AI to address fundamental rights and safety risks](#) specific to the AI systems
- ❖ An [AI liability directive](#) - adapting liability rules to the digital age and AI
- ❖ A [Proposal for a product liability directive](#) (Revision)



Implications for researchers and RECs

- Will affect and/or influence research activities.
- No intent to stifle innovation but boost it responsibly.
- Push for the development of **secure, trustworthy and ethical** artificial intelligence, protect ethical principles
- Use a “specific **ethically-focused approach** during the development, deployment, and/or use of AI-based solutions” – ethics by design (HE Ethics Guidance).
- Need to address the risks of AI systems (specifically, high-risk).
- Issues related to **researcher access** and **use** of high-quality datasets need addressing.
- Ensure **legitimacy of purpose** and **no harm**.
- Research should be in line with **recognised ethical standards** for scientific research.
- AI systems or techniques - **clearly described** and demonstrated for **technical robustness and safety**.
- Carry out enhanced risk assessments, including on **human rights**.



**TRILATERAL
RESEARCH**
Ethical AI



We build Ethical AI solutions
to tackle complex societal problems

trilateralresearch.com

UK _ One Knightsbridge Green, London, SW1X 7QA
IRELAND _ Marine Point, Belview Port, Waterford, X91W0XW