

AI4EU Deliverable D5.1

Ethical Observatory description of functions, oversight powers, specific agenda and interactions with other groups

WP	5	Promoting European ethical, legal, cultural and socio-economic values for AI
Task	5.1	The AI4EU ethical AI observatory

Dissemination level¹	PU	Due delivery date	30/06/2019
Nature²	R	Actual delivery date	DD/MM/YYYY

Lead beneficiary	UVE
Contributing beneficiaries	XXX, XXX

Document Version	Date	Author	Comments ³
V0	11.06.2019	Teresa Scantamburlo	
V1	13.06.2019	Luc Steels	
V2	14.06.2019	Ulises Cortés	
V3	14.06.2019	Teresa Scantamburlo	
V4	31.07.2019	Teresa Scantamburlo	
V5	07.08.2019	Ulises Cortés	
V6	19.08.2019	Teresa Scantamburlo and Luc Steels	
V7	20.08.2019	Atia Cortés and Ulises Cortés	

¹ Dissemination level: **PU** = Public, **PP** = Restricted to other programme participants (including the JU), **RE** = Restricted to a group specified by the consortium (including the JU), **CO** = Confidential, only for members of the consortium (including the JU)

² Nature of the deliverable: **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

³ Creation, modification, final version for evaluation, revised version following evaluation, final

Deliverable abstract

Deliverable D5.1 reports the activities conducted to create the AI4EU Observatory on Society and AI (OSAI). This effort is part of WP5 and contributes to the promotion of European Values for Ethical, Legal, Socio-Economic and Cultural issues in AI (ELSEC-AI). In particular, this document outlines the functions and the agenda of AI4EU OSAI, including interactions with other groups. Besides, the document presents a demonstrator of the Observatory Website that will support the distribution and discussion of ELSEC-AI. This document will serve as a blueprint for the embedding of the Observatory within the AI4EU platform and its connected activities.

Deliverable Review

Reviewer #1:			Reviewer #2:		
Answer	Comments	Type *	Answer	Comments	Type *
1. Is the deliverable in accordance with					
(i) the Description of the Action?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a	<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a
(ii) the international State of the Art?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a	<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a
2. Is the quality of the deliverable in a status					
(i) that allows it to be sent to European Commission?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a	<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a
(ii) that needs improvement of the writing by the originator of the deliverable?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a	<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a
(iii) that needs further work by the Partners responsible for the deliverable?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a	<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> M <input type="checkbox"/> m <input type="checkbox"/> a

* Type of comments: M = Major comment; m = minor comment; a = advice

ACRONYMS LIST

AI4EU	A European AI On Demand Platform and Ecosystem
AI HLEG	High-Level Experts Group on Artificial Intelligence
ECTL	European Centre for Living Technology
ELSEC	Ethical, Legal, Socio-Economic and Cultural
ELSEC-AI	Ethical, Legal, Socio-Economic and Cultural issues of Artificial Intelligence
IPAI	International Panel on Artificial Intelligence
OSAI	Observatory on Society and Artificial Intelligence
WP	Work Package
WG	Working Group

Table of Content

1. Introduction.....	5
2. The context of OSAI.....	7
2.1. The landscape in Europe.....	8
2.2. Reports.....	9
2.3. National strategies.....	13
2.4. Networks and groups.....	18
2.5. Centres.....	21
2.6. Projects and educational resources.....	24
3. OSAI framework.....	25
3.1. Functions and roles.....	26
3.2. Template.....	28
3.3. Publication frequency.....	30
3.4. Methodology and networking.....	31
3.5. Working Groups.....	31
4. Ongoing work and future plan.....	33

1. Introduction

This document describes the outcomes of the first six months of work involved in creating the AI4EU Observatory on Society and Artificial Intelligence⁴ (OSAI). Specifically, it describes the functions, the agenda and the networking activities of OSAI.

An observatory is a special place where people can observe and study astronomical or terrestrial objects thanks to the specialized equipment and conditions it can offer. Similarly, OSAI will offer a selected set of tools that help people to understand better and study the impact of AI technologies across the European Union.

The main goal of the AI4EU Observatory is to support discussion and facilitate the distribution of information about the Ethical, Legal, Socio-Economic and Cultural issues of AI (ELSEC-AI) within Europe. Its creation is an essential task of WP5 as it contributes to the fulfilment of WP5's overall objective, "*to promote European Values for Ethical, Legal, Socio-Economic and Cultural Issues in AI*" (see the Strategic Objective-SO4 in the Description of the Action). Specifically, the AI4EU Observatory helps WP5 implement the following objectives:

- To stimulate reflection, discussion and due consideration of ELSEC-AI issues within the project through a series of working groups,
- To provide resources to educate the general EU public more accurately about AI and ELSEC issues.

So far the Observatory has been set up and animated by the European Centre for Living Technology (ECLT), an international research centre devoted to the study of complex systems, based in Ca' Foscari University, Venice (Italy)⁵. This task was done in coordination with WP5 and other AI4EU members.

As the primary function of the Observatory is to promote a broad understanding and discussion of ELSEC-AI, OSAI requires a strong web presence and therefore WP5 has focused in a first instant on collecting ideas of what such a web presence could provide and building a demonstrator in the form of a fully operational website (www.unive.it/osai) that follows the logic of an online magazine where various types of information on ELSEC-AI are hosted. The idea is to support, on the one hand, the dialogue among experts and professionals dealing with ELSEC-AI, and, on the other

⁴ A note on terminology: the name of the Observatory makes an explicit reference to the term *society* as the latter, being the locus of interactions between humans and AI, well encompasses all ELSEC issues

⁵ More details about ECLT are available on the website of the centre: <https://www.unive.it/pag/23664/>

hand, the education of lay people living within Europe. In a second stage, as soon as the AI4EU platform is fully functioning, this web presence is to be embedded directly in the AI4EU platform, and the web presence will be operated by the work packages responsible for running the platform (WP2 and WP3) and for community formation (WP4). WP5 will, of course, continue to provide input. However, the focus of WP5 (mainly Task 5.2.) will be directed more to the creation and animation of the working groups described in the proposal.

OSAI evolves in a complex scenario: the field of AI is gaining momentum, and many public and private agencies have begun to consider the opportunities and the risks that lie behind this exciting trend. The OSAI seeks to carve out a well-defined identity and role neither in contrast nor competition with other existing European initiatives (e.g. AI HLEG⁶). It aims to increase connections among these related projects and make accessible a broad range of high-quality articles to the European public at large. OSAI is less concerned with short-term news but more with in-depth background and critical reflection on AI.

The implementation of the OSAI web presence demonstrator follows an incremental methodology. Indeed, the first version of the demonstrator rests on a basic template, and a flexible schedule, with a view to revise and refine the initial design choices. At present, OSAI's website (www.unive.it/osai) is the main channel to solicit discussion and distribute information about the preliminary design ideas. It has been developed in collaboration with the IT service of Ca 'Foscari University, which made its Content Management Systems available. The OSAI demonstrator was linked into the internal communication channel (eXo) on 27 June 2019.

We stress again that this website is a temporary step to make the activities carried out within WP5 visible and reviewable, but developers in WP2/3/4 will embed OSAI into the AI4EU platform (see more details in section 4) as soon as the platform becomes operational. The website is a *demonstrator* of OSAI functions and approach and will be replaced and enhanced as the integration of OSAI into the AI4EU platform/website is entirely operated.

The UVE team has coordinated the creation of an ethical advisory board for the whole AI4EU consortium, which will give a regular assessment of how the project is doing concerning the Ethical issues. Input from this advisory board has been solicited for the initial design of the OSAI web presence. The final list includes the following people:

- Luiza Spiru (Univ of Medicine and Pharmacy Spitalul, University of Bucarest, Romania)

⁶ More details about the AI HLEG are available on the website of the European Commission: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

- Itziar de Lecuona (University of Barcelona, Spain)
- Patrick van der Smagt (AI Research, Volkswagen)
- Mikael Anneroth (Eriksson)
- Viola Schiaffonati (Polytechnic University of Milan, Italy)

Marcello Pelillo (UVE) is the coordinator of this board.

The present document is organised as follows. Section 2) introduces the context in which OSAI is being developed. This includes the presentation of some relevant initiatives that animate the debate on ELSEC-AI, including examples of analogous experiences, and a discussion of how OSAI will relate to them. Section 3) outlines the framework of OSAI, that is, the approach that inspires its main activities, its main functions, the adopted methodology and a brief description of the web presence demonstrator. Also, this section outlines how OSAI will be connected to the implementation of ELSEC-AI working groups (see task 5.2 in AI4EU Annex 1). Section 4) includes a plan for OSAI's future work and interactions with other WPs.

2. The context of OSAI

The explosion of AI technologies with connected benefits and risks are at the centre of a global discussion. Recent AI-related scandals, such as Cambridge Analytica, Uber's fatal crash, YouTube's recommendations of divisive and misleading content, and persistent overstatements and failures in the application of AI to health^{8,9} have solicited a profound reflection on AI's impact among experts but also in the public sphere. Governments have been generating reports and official documents surveying the state of the art of AI applications and suggesting possible directions. For example, in 2014, the White House issued a report after a 90-day study on big data technologies¹¹. Many others followed both in the US and other countries.

The European Union committed to a clear vision, the so-called "*human-centric AI*"¹², and have made significant progress in this regard. As well as official documents, in 2016, the EU approved the General Data Protection Regulation that has become the reference framework for data protection and privacy of all European citizens. More recently, the EU Commission has established the High-Level Expert Group on Artificial Intelligence (HLEG), an independent group of experts

⁸ DeepMind's Latest A.I. Health Breakthrough Has Some Problems.

<https://onezero.medium.com/deepminds-latest-a-i-health-breakthrough-has-some-problems-5cd14e2c77ef>

⁹ How IBM Watson Overpromised and Underdelivered on AI Health Care

<https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson>

¹¹ Executive Office of The President, *Big Data: Seizing opportunities, preserving values*, 2014,

https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

¹² COM(2018)237, <https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF>

tasked with the drafting of two documents: the AI Ethics Guidelines and the Policy and Investment Recommendations. The investigation of the impact of AI is not relegated to governments alone. There is a wide range of research centres and think tanks that carry out independent investigations and develop strategies on AI. Examples include AlgorithmicWatch¹³ in Germany, DataEthics¹⁴ in Denmark, and the Institute for Ethical AI & Machine Learning¹⁵ in the United Kingdom.

2.1 The landscape in Europe

The creation of OSAI takes place in this complex and dynamic context where an imprecise number of AI-related events populate the European calendar¹⁶. This vast multitude may be considered an obstacle or a barrier to the development of the AI4EU Observatory. However, this challenge has been transformed into a positive opportunity, that is, to increase connections among European initiatives and enlighten the European public on ELSEC-AI, in particular in countries that receive less attention (e.g. because of language issues). Indeed, while the abundance of centres and projects dealing with AI and its social and ethical impact is a sign of cultural awareness and a source of knowledge, all these positive undertakings run the risk of isolation and self-referentiality. Therefore, OSAI should try to bridge this gap and promote cooperation and mutual knowledge. In addition, it will focus on areas that extend beyond the ethical and legal aspects, including also socio-economic and cultural elements (e.g. how AI is perceived among European citizens, how the arts are presenting or using AI).

At present, examples of initiatives that present analogies with OSAI are:

- *AI Hub Europe*¹⁷: a website that aims to “*aggregate news and background information on all aspects of AI and machine learning in Europe or relevant to Europeans.*”
- *DigitalWatch*¹⁸: an observatory that collects and spread information about digital policy and Internet Governance.
- *Impact AI*¹⁹: a think tank that has created an Observatory to study the evolution and the perception of AI in French society

¹³ <https://algorithmwatch.org/de/about/>

¹⁴ <https://dataethics.eu/>

¹⁵ <https://ethical.institute/>

¹⁶ Examples include international scientific conferences (e.g. FAT ML conferences, <https://www.fatml.org/>, ECAI 2020 whose general theme will be “Paving the way towards Human-Centric AI” <http://ecai2020.eu/>) but also industrial and cultural events. For example, also AI conferences for businesses comprise exhibitions or tracks on responsible and explainable AI (e.g. <https://london.theaisummit.com/>). Other organisations, like Binary District (<https://binarydistrict.com/about>), run workshops to foster conversations among researchers, entrepreneurs and regulators (see for instance the workshop on Ethical AI (<https://binarydistrict.com/events/conference/development/ethical-ai-is-it-possible>))

¹⁷ <https://ai-europe.eu/>

¹⁸ <https://dig.watch/>

- *Swiss Cognitive. The global AI Hub*²⁰. An observatory with regular news items about AI in Europe.

OSAI differ from these initiatives in several respects. In the first place, OSAI will focus not only on articles and news, but also on people. Indeed, one of the motivating ideas behind OSAI is the creation of a *community of people* who can contribute to the discussion of ELSEC-AI. Such a community can combine various types of individuals such as AI experts (e.g. AI researchers and practitioners), specialists in any ELSEC-related field (ethicists, sociologists, lawyers, policy makers...) and lay people. In the second place, OSAI will approach ELSEC-AI in the context of Europe so as to foster the dialogue among European countries.

To lay the groundwork, and to understand which groups might have the potential to become OSAI partners, an essential task was a survey of existing activities in relation to ELSEC-AI. This task will be iterated throughout the project, as the list of these activities is likely to grow. A partial list of relevant European initiatives is presented in the subsections below. The list has been organised according to the following classification: reports, national strategies, networks and groups, centres, projects and educational resources. These have been collected with two main criteria in mind: they have grown up in Europe²¹ and address ELSEC-AI. This survey serves as a starting point for the resources that will be made available on the OSAI website.

2.2 Reports

The selection of report was limited to those documents that had been released recently (the oldest ones date back to 2017) and had a specific reference to AI and machine learning²².

Table 1

Country	Organisation	Title	Year	Brief description
UK	House of Commons	<i>Algorithms in decision-making</i> ²³	2018	The report identifies the themes and challenges for the 'Centre for Data Ethics & Innovation', an advisory body launched in 2018 by the Government. Key issues include data sharing, bias, discrimination, transparency and accountability.
UK	Royal Society	<i>Machine learning:</i>	2017	The report outlines the significant opportunities

¹⁹ <http://www.impact-ai.fr/?lang=en> see for example the infographics on the perception of AI in France: <http://www.impact-ai.fr/2018/12/03/observatoire-sondage-ia/>

²⁰ <https://swisscognitive.ch/>

²¹ Note that, although most of items matured in a European context, certain networks extend to the world thereby including both European and non-European countries.

²² For example, documents that are more focused on big data were not included.

²³ <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>

		<i>the power and promise of computers that learn by example</i> ²⁴		and challenges introduced by modern machine learning techniques. The report makes a number of recommendations for the Government such as: to promote open data standards; to improve education and training in machine learning methods at all education levels; to ensure that immigration and industrial strategy policies align with the needs of the UK AI development sector, and to facilitate public dialogues on the opportunities and challenges of machine learning.
UK	Ipsos MORI & Royal Society	<i>Public Views of Machine Learning</i> ²⁵	2017	The report provides evidence about public perceptions around the potential benefits and risks of machine learning, based on 978 face-to-face interviews conducted in 2016 and public dialogues. Types of perceived risks: human replacement, depersonalisation, restriction and harms; types of perceived benefits: the time saving and better choice.
UK	Royal Society	<i>Machine learning: the power and promise of computers that learn by example</i> ²⁶	2017	The report documents the Royal Society's machine learning project, which aims at increasing awareness of this technology, suggesting its potential benefits and challenges. Also, it identifies areas of public concern that would need farther investigation: interpretability, robustness, privacy, fairness, inference of causality, human-machine interaction, and security.
EU	Informatics for Europe & ACM Europe Policy Committee	<i>When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making</i> ²⁷	2018	The report reviews the major implications of Automated Decision Making (ADM) with a particular emphasis on technical, ethical, legal, economic, societal and educational aspects. Some of the recommendations put forward are: providing standards to assure that ADM systems are fair; Ensuring that ethics remain at the forefront of ADM development and deployment (e.g. European agency for oversight); promoting value-sensitive design; clarifying legal responsibilities for ADM's use and impacts; in-depth consideration of economic consequences of ADM; increasing public funding for ADM-related research (prioritising research in explainable ADM); expanding public awareness of ADM systems.
EU	European Commission & EurAI	<i>The European Artificial Intelligence Landscape</i> ²⁸	2018	The report collects the result of a workshop held in Brussels that reviewed the current state of AI in Europe. As well as considering some bottlenecks (e.g. bureaucracy and fragmentation), it makes some proposals such as: the establishment of a European research centre for AI modelled on institutions such as CERN; Investment and creation of a pan-

²⁴ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

²⁵ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf>

²⁶ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

²⁷ <https://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf>

²⁸ http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51262

				European data infrastructure; designing mechanisms to re-skill and up-skill the broader population in the use of AI tools.
World	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	<i>Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition</i> ²⁹	2019	The report is the result of a vast collaborative effort and guides stakeholders involved in the design and development of autonomous intelligent systems. Based on eight general ethical principles (human rights, well-being, data agency, effectiveness, transparency, accountability, awareness of misuse, competence), it provides concrete recommendations to address and mitigate ethical issues (well-being metrics, embedding norms into autonomous intelligent systems, value-based design methods)
World	Access Now	<i>Mapping Regulatory Proposals for AI in Europe</i> ³⁰	2018	This report surveys the major regulatory initiatives in AI in the EU and among member states. The analysis is based on published strategy papers and states' consultations with experts. A key contribution in the document is the comparison among regulatory strategies with respect to ten relevant principles: transparency, accountability, the right to privacy, freedom of conscience and expression, the right to equality and non-discrimination, due process, the right to data protection and user control, collective rights, economic rights and the future of work, the laws of war.
EU	High Level Expert Group on AI (HLEG-AI)	<i>Ethics Guidelines for Trustworthy AI</i> ³¹	2019	The report introduces a framework for trustworthy AI based on fundamental rights (respect for human dignity; freedom of the individual; respect for democracy, justice and the rule of law; equality, non-discrimination and solidarity, citizens' rights) and four ethical principles (respect for human autonomy, prevention of harm, fairness and explicability). These principles are then translated into seven key requirements for AI systems: Human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; accountability. These guidelines are complemented by an assessment list that offers guidance for practical implementation.
France	Commission for Information Technology and Liberties (CNIL)	<i>How Can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence</i> ³²	2017	The document presents the result of a public debate organised by the French Data Protection Authority, which has involved 60 partners. It identifies six ethical issues: a threat to human autonomy and free will; discrimination and exclusion; algorithmic profiling; retention of personal data; quality, quantity and relevance of training data;

²⁹ <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf>

³⁰ https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf

³¹ https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

³² https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

				hybridisation between humans and machines. Also, the report suggests six practical policy recommendations such as: fostering the education of all players involved in the <i>algorithmic chain</i> ; making algorithmic systems understandable and setting up a national platform for auditing algorithms.
Belgium	Flemish Academy of Science	<i>Artificiële intelligentie: Naar een vierde industriële revolutie?</i> ³³	2017	This document results from the activity of a working group set up by the Class of Natural Sciences and the Royal Flemish Academy for Sciences and Arts to study the impact of AI. The main purpose of this document is to inform the public as objectively as possible and to propose a series of conclusions and recommendations to concerned parties in order to deal with AI and ensure that our community can adequately benefit from the vast opportunities, as well as get an insight into the risks and what to do about them.
The Netherlands	Nederland ICT	<i>Ethische Code Artificial Intelligence</i> ³⁴	2019	This document is a code of ethics developed by the Ethics Think Tank within the Netherlands ICT, a group of Dutch companies operating in the ICT sector. The code of conduct is in accordance with EU ethics guidelines and will be reviewed annually. Each member company of Nederland ICT commits to eight guidelines such as: awareness of AI technical possibilities and limitations; providing insight into the data that is used by AI application; clarifying when a user is dealing with an AI system and the responsibility of each part; and ensuring that the behaviour of application is actively motored.
Sweden	Vinnova	<i>Artificial intelligence in Swedish business and society: Analysis of development and potential</i> ³⁵	2018	This report is delivered by Sweden's Innovation Agency (Vinnova), which was commissioned by the Swedish government. The report maps the opportunities connected to the use of AI in Swedish industry and in Swedish business and the public sector generally. It analyses the development of AI in Sweden with intending to highlight strengths and weaknesses.
World	Future of Humanity Inst, Oxford Uni, Centre for the Study of Existential Risks, Uni Cambridge, Centre for a new American	<i>The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation</i> ³⁶	2018	The report is written by 26 experts from different institutions and builds upon a workshop held in Oxford in 2017. It surveys the threats of AI in three security domains (digital security, physical security and political security). It also makes high-level recommendations and sets up priorities, such as learning from and with cybersecurity community and promoting a culture of responsibility.

³³ https://www.kvab.be/sites/default/rest/blobs/1489/nw_artificieleintelligentie.pdf

³⁴ <https://www.nederlandict.nl/wp-content/uploads/2019/04/NEDERLAND-ICT-ETHICAL-CODE-FOR-ARTIFICIAL-INTELLIGENCE.pdf>

³⁵ https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf

³⁶ <https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/MaliciousUseofAI.pdf?ver=1553030594217>

	Security, Electronic Frontier Foundation, OpenAI			
Germany	Algorithmic Watch & Bertelsmann Stiftung	<i>Automating Society Taking Stock of Automated Decision-Making in the EU</i> ³⁷	2019	This is an explorative study of the use of automated decision-making in Europe. For example, it considers applications for job profiling (Finland), allocating treatment for patients (Italy) and identifying vulnerable children (Denmark). It builds upon a network of experts (academics, journalists, lawyers...) that contribute to report on national situations. This network is expected to grow in the coming years so as to include countries not yet covered.

2.3 National strategies

The effort of developing a national strategy follows the recommendation put forward by the EU Commission in the document “Coordinated plan on AI”³⁸. Some European countries are in the process of delivering their strategies³⁹ and others plan to adjust the one they have already released. Note that the list of strategies can also be extended to private companies (for example, Ericsson aims to develop a responsible approach to AI⁴⁰).

Table 2

Country	Organisation	Title	Year	Brief description
Europe	EU Commission	<i>Artificial Intelligence for Europe</i> ⁴¹	2018	The document proposes an approach that places people at the centre of the development of AI (human-centric AI) and encourages the use of AI to help solve social challenges such as climate change crime-fighting.
Europe	EU Commission	<i>Coordinated Plan on AI</i> ⁴²	2018	The plan builds upon a declaration of cooperation signed by all EU Member States and Norway in 2018. It outlines a series of objectives and actions agreed by Member States, Norway and Switzerland to increase investments and foster talent and trust. For example, the plan proposes to strengthen private-public partnership; to adapt learning and

³⁷ https://algorithmwatch.org/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

³⁸ https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017 (see p. 5 of the document)

³⁹ For example, Poland released a document that will be useful to devise its national strategy: <https://www.digitalpoland.org/assets/publications/mapa-polskiego-ai/map-of-the-polish-ai-2019-edition-i-report.pdf>

⁴⁰ <https://www.ericsson.com/en/blog/2019/2/responsible-ai--human-right>

⁴¹ <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

⁴² https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56017

				training programmes; to create European data spaces, and to develop ethics guidelines. The document provides also a framework for national AI strategies.
Belgium	AI 4 Belgium & Ministers for Digital Agenda	<i>AI 4 Belgium</i> ⁴³	2019	This document presents the initial work of the AI 4 Belgium Coalition, a group comprising academics, tech-firms, corporations and public institution. It identifies five priorities for the country, such as: set up a new learning deal, develop a responsible data strategy and support adoption of AI in the private sector.
France	<i>AI 4 Humanity</i>	<i>For a Meaningful Artificial Intelligence. Towards a French and European Strategy</i> ⁴⁴	2018	The report is the result of the “Villani mission” led by the mathematician and Member of Parliament Cédric Villani. It is built upon a previous report detailing the state of AI in France ⁴⁵ . The objective of the report is “to provide a meaning to AI revolution”. Key indications include building a data-focused economic policy; promoting agile and enabling research; AI for a more <i>green</i> economy and ethical AI (creation of an AI ethics committee).
Germany	Federal Ministry for Economic Affairs and Energy, the Federal Ministry of Education and Research, and the Federal Ministry of Labour and Social Affairs.	Strategie Künstliche Intelligenz der Bundesregierung ⁴⁶	2018	The strategy focuses on research, transfer, public dialogue, impact assessment, skills and data availability. It is based on a previous report outlining the key points for a Federal Government Strategy on Artificial Intelligence (available in English ⁴⁷). Example of identified priorities include to transfer research findings and AI methods to business; to invest in the labour force and skillsets; to attract skilled workers; to make data available and usable; to revise the regulatory framework, and to engage in dialogue with society.
Italy	The Agency for Digital Italy – AI Task Force	<i>Artificial Intelligence at the service of citizens</i> ⁴⁸	2018	This is a white paper that investigates the impact of AI in Public Administration (PA) to improve services to citizens and business. It identifies some critical challenges (ethical, technological, legal, etc), and puts forward a set of recommendations such as: promoting a national platform dedicated to the development of AI solutions; providing a plan to encourage PA investments in AI; establishing a Trans-disciplinary Centre

⁴³ https://www.ai4belgium.be/wp-content/uploads/2019/04/report_en.pdf

⁴⁴ https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

⁴⁵ https://www.economie.gouv.fr/files/files/PDF/2017/Rapport_synthese_France_IA_.pdf

⁴⁶ https://www.bmwi.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.pdf?__blob=publicationFile&v=8

⁴⁷ https://www.bmwi.de/Redaktion/EN/Downloads/E/key-points-for-federal-government-strategy-on-artificial-intelligence.pdf?__blob=publicationFile&v=5

⁴⁸ <https://ia.italia.it/assets/whitepaper.pdf>

				on AI; define guidelines and processes based on a security-by-design approach.
Denmark	Ministry of Finance and Ministry of Industry, Business and Financial Affairs	<i>National Strategies for Artificial Intelligence</i> ⁴⁹	2019	The document sets four objectives: to have a common ethical basis for AI; to support research and development of AI; to help businesses invest in (responsible) AI; to use AI in the public sector. It recommends specific initiatives in four focus areas: a responsible foundation for artificial intelligence; more and better data; strong competences and new knowledge; increased investment.
The Netherlands	Innovation Centre for Artificial Intelligence	<i>AI Voor Nederland</i> ⁵⁰	2018	The report contains a wide range of possible actions for governments and companies, such as: working on a world-class of national AI research centre; making better use of already available data; attract top AI talent and students.
Spain	Ministry of Science, Innovation and Universities	<i>Spanish RDI strategy in Artificial Intelligence</i> ⁵¹	2019	The strategy is the result of a group of AI experts and establishes six priorities: to achieve an organisational structure that allows the development and the assessment of AI; identifying strategic areas where to apply AI (e.g. public administration and education); transfer of knowledge; promoting cross-disciplinary and vocational training in AI. It also includes seven recommendations that seek to align normative, structural and organisational adaptations to the achievements made in Artificial Intelligence in the different public policies.
Sweden	Ministry of Enterprise and Innovation	<i>Nationell inriktning för artificiell intelligens</i> ⁵²	2019	This document complements the Government's digital strategy and identifies four key conditions to realise the potential of AI in Sweden: education and training; research; innovation and use; framework and infrastructure (e.g. standards and rules)
Finland	Ministry of Economic Affairs and Employment	<i>Finland's Age of Artificial Intelligence</i> ⁵³	2017	A network of experts has produced the document.. It outlines eight actions proposed by a Steering Group of the Artificial Intelligence Programme. Examples of proposed actions include to use data in all sectors; to speed up and to simplify AI adoption; ensure top-level expertise and attract top experts; and to establish a new cooperation model. It also defines some actions to make the work progress: extending the network of experts (in specific focus areas) and

⁴⁹ https://eng.em.dk/media/13081/305755-gb-version_4k.pdf

⁵⁰ <https://icai.ai/wp-content/uploads/2018/11/Rapport-AI-voor-Nederland-AINED.pdf>

⁵¹ http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_EN.PDF

⁵² <https://www.government.se/491fa7/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf>

⁵³ http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

				engaging citizens.
Luxemburg	Digital Luxembourg Innovative Initiative	<i>Artificial Intelligence: a strategic vision for Luxembourg</i> ⁵⁴	2019	The report describes the Government's commitment to a human-centric AI and put forward some proposals to implement that approach. Among those proposals, there are: becoming a living laboratory for AI; consolidating large European data lakes; promoting ethics and privacy regulation; boosting investments and strategic partnerships; and investing in AI projects that provide better public services
Estonia	Ministry of Economic Affairs and Communications	<i>Eesti tehisintellekti kasutuselevõtu eksperdirühma aruanne / Kratt Report</i> ⁵⁵	2019	The document is written by a team of AI experts and outlines strategies to accelerate the application of AI in the private and public sectors. Suggested actions include open-source base components; guidelines on how to manage AI projects; and data governance workshops and data auditing. These strategies are aligned with the wider e-State and information society goals. The document concludes that there is no need for changes in the foundation of the legal system and there is no need for a unified AI law
Portugal	Portugal INCoDe.2030 Initiative	<i>AI Portugal 2030. Portuguese National Initiative on DIGI. An innovation and growth strategy to foster Artificial Intelligence in Portugal in the European context</i> ⁵⁶	2019	The document outlines a number of national priorities to implement by 2030. For example, these include promoting a better society; fostering AI skills and <i>digital minds</i> for all; promoting new jobs and developing an economy of AI services; fostering Portugal as a living lab for experimentation of new developments. Also, it lists several specific areas where Portugal will make a specific effort to lead in Europe (e.g. Natural Language Processing; real-time decision making with AI; AI and sustainable energy systems...)
Lithuania	Ekonomikos Ir Inovacijų Ministerija / Kurk Lietuvai	<i>Lithuanian Artificial Intelligence Strategy: A Vision of the Future</i> ⁵⁷	2018	This document results from a series of meeting between a group of AI experts and the Ministry of Economy. As well as identifying the strength and the weakness of AI in Lithuania, it presents several recommendations organised in the form of mechanisms (e.g. actions, programs and initiatives). For example, the document recommends to establish AI ethics committee; to develop rules and guidelines for the development and use of AI; to support research to minimize bias in AI systems; encourage

⁵⁴ https://digital-luxembourg.public.lu/sites/default/files/2019-05/AI_EN_1.pdf

⁵⁵ https://www.riigikantselei.ee/sites/default/files/riigikantselei/strategiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf

⁵⁶ https://www.incode2030.gov.pt/sites/default/files/incode_aiportugal2030_june19.pdf

⁵⁷ <http://kurklit.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf>

				universities to develop courses on the ethical implications of technology; to create a dedicated artificial intelligence research-funding program.
Malta	Malta.AI	<i>Malta Towards an AI Strategy. High-level Policy Document for Public Consultation</i> ⁵⁸	2019	This document is a result of a task force comprising public and private sector representatives. The document describes Malta's aspiration to become the ultimate AI launchpad ("a place in which local and foreign companies, and entrepreneurs, can develop, prototype, test and scale AI"). The strategy identifies three main pillars (investment, start-ups and innovation; public sector adoption; and private sector adoption) and three strategic enablers: education and workforce; legal and ethical framework; ecosystem infrastructure.
Austria	Austrian Council for Robotics and Artificial Intelligence	<i>Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten</i> ⁵⁹	2018	This is a white paper delivered by an advisory body consisting of AI and robotics experts. It discusses current and future opportunities and challenges arising from the use of Robots and Autonomous Systems (RAS) and AI on a technological, economic, social and legal level. The document includes, among others, opinions and recommendations such as: creating conditions ensuring the exploitation of RAS and AI; establishing a legal framework that complies with EU legal requirement; developing measures to detect and mitigate possible dangers that could arise from RAS and AI
UK	Department for Business, Energy & Industrial Strategy Department for Digital, Culture, Media & Sport	<i>AI Sector Deal</i> ⁶⁰	2018	This document is the first commitment from the government and industry to promote the adoption of AI in the UK. It builds upon Government's Industrial and Digital Strategies and organises the plan of actions in five areas: ideas, people, infrastructure, the business environment and places. Examples of planned actions are: to invest up to £20m in the application of AI in the services sector; to develop a global Turing Fellowship programme to both attract and retain the best research talent in AI from around the world to the UK; to invest £406m in skills, with a focus on maths, digital, and technical education; to improve the digital infrastructure and provide legal certainty over the sharing and use of data in accordance with the UK's strengthened Data Protection Bill.
UK	House of Lords	<i>AI in the UK:</i>	2018	The report presents the

⁵⁸ https://malta.ai/wp-content/uploads/2019/04/Draft_Policy_document_-_online_version.pdf

⁵⁹ https://www.acrai.at/wp-content/uploads/2019/04/ACRAI_whitebook_online_2018.pdf

⁶⁰ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal_4_.pdf

		<i>ready, willing and able?</i> ⁶¹		recommendations of a committee mandated by the House of Lord to consider the economic, ethical and social implications of AI. Examples of recommended include to establish voluntary mechanisms for informing the public when AI is being used for sensitive decisions; to incentivise new approaches to the auditing of datasets used in AI; to promote an a thorough education in AI-related subjects; a global summit in London by the end of 2019 to develop a common framework for the ethical development of AI. It also proposes five ethical principles for an AI code such as: AI should be developed for the common good; AI should operate n principles of intelligibility and fairness; AI should not be used to diminish the data right.
Nordic-Baltic region	Government of Sweden & Nordic Council of Ministers	<i>AI in the Nordic-Baltic region</i> ⁶²	2018	This is a joint declaration signed by Nordic-Baltic ministers to promote the use of AI that better serves people. It proposes several measures such as: improving the opportunities for skills development; strengthening access to data; developing ethical and transparent guidelines and standards; working for international standards to ensure interoperability, privacy, security, trust, usability and mobility.

2.4 Networks and groups

Many networks and groups in the list were appointed by national Governments to develop AI strategies and programmes. The list does not include the initiatives that refer to more general objectives such as the creation of national digital strategies (e.g. see “Digital Luxembourg” in table 2 or the CNIL in table 1). The description of the networks and groups draws on the information available on the dedicated websites (see the footnotes in the table)

Table 3

Country	Name	Brief description
World	<i>AI Partnership</i> ⁶³	The Partnership on AI (PAI) is a multi-stakeholder organisation that brings together academics, researchers, civil society organisations, companies building and utilizing AI technology, and other groups

⁶¹ <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>

⁶² https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nm_r_deklaration-slutlig-webb.pdf

⁶³ <https://www.partnershiponai.org/>

		working to understand better AI's impacts. The Partnership was established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.
World	<i>The AI Governance Forum</i> ⁶⁴	The AI Governance Forum is a multi-stakeholders' platform, open to all interested parties and dedicated to building Human-Trust in AI for the benefit of all. Stakeholders can be from the public or private sector, the scientific community and civil society. Topics of interest of the AI Governance Forum are safety and security, data handling and bias, regulations, accountability, transparency, privacy, and impact on the workforce. It contributes to building an open artificial intelligence for the benefit of all. The AI Governance Forum is supported by the Federal Office of Communications OFCOM of the Swiss Confederation.
Europe	<i>European AI Alliance</i> ⁶⁵	The European AI Alliance is a forum engaged in a broad and open discussion of all aspects of AI development and its impacts. The European AI Alliance will form a broad multi-stakeholder platform, which will complement and support the work of the AI High-Level Expert Group (AI HLEG) in particular in preparing draft AI ethics guidelines, and ensuring the competitiveness of the European Region in the burgeoning field of Artificial Intelligence.
Europe	<i>High Level Expert Group on AI (AI HLEG)</i> ⁶⁶	The AI HLEG is a group of 52 experts comprising representatives from the Academia, civil society, as well as industry appointed by the EU Commission to support the implementation of the European Strategy on Artificial Intelligence. This includes the elaboration of recommendations on future-related policy development and ethical, legal and societal issues related to AI, including socio-economic challenges.
Europe	<i>CLAIRE</i> ⁶⁷	CLAIRE is an initiative by the European AI community that seeks to strengthen European excellence in AI research and innovation. To achieve this, CLAIRE proposes the establishment of a pan-European Confederation of Laboratories for Artificial Intelligence Research in Europe that achieves <i>brand recognition</i> similar to CERN.
Italy	<i>AI Task Force (within AGID)</i> ⁶⁸	The AI task force was set up to discuss the new possibilities offered by AI, in general, in our daily life and, more specifically, in the construction of a new relationship between State and citizens.
UK	<i>Select Committee on AI</i> ⁶⁹	The Select Committee on AI was appointed in 2017 to consider the economic, ethical and social implications of advances in AI, and to make recommendations.
France and	<i>International Panel on</i>	France and Canada established the IPAI in 2018 with

⁶⁴ <https://ai-gf.com/>

⁶⁵ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁶⁶ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁶⁷ <https://claire-ai.org/>

⁶⁸ <https://ia.italia.it/en/>

⁶⁹ <https://www.parliament.uk/ai-committee>

Canada	<i>Artificial Intelligence (IPAI)</i> ⁷⁰	that aim to create a global point of reference for AI research. The IPAI will support and guide the responsible adoption of artificial intelligence, centred on the respect for human rights, inclusion, diversity, innovation and economic growth. Its mission will also be to facilitate international collaboration between the scientific community, industry, civil society, related international organisations and governments.
Austria	<i>Austrian Council for Robotics and Artificial Intelligence (ACRAI)</i> ⁷¹	The ACRAI is an advisory body set up by the Federal Ministry of Transport, Innovation and Technology (BMVIT). It consists of experts on robotics and AI from research, teaching and business and identifies and discusses current and future opportunities, risks and challenges resulting from the use of robots and autonomous systems (RAS) as well as AI.
Malta	<i>Malta.AI</i> ⁷²	Malta.AI is a task force on AI appointed by the Government that is made up of entrepreneurs, academics and experts in the field of AI. It aims to devise a holistic approach on the sector, involving Academia, start-ups and companies, to find ways how to create a sustainable local engine for growth, look into the unknown risks of AI without hindering innovation and economic development, and create a new sector for investment on our shores.
The Netherlands	<i>Innovation Centre for Artificial Intelligence</i> ⁷³	The ICAI is an open national network of academic, industrial and governmental partners that is based at Amsterdam Science Park. ICAI will create innovative AI applications, distribute AI-related knowledge for companies and organisations, train corporate employees through dedicated courses and maintain a connection with other world-level science centres. It will also facilitate commercialization by enhancing start-up and spin-outs.
Belgium	<i>AI 4 Belgium Coalition</i> ⁷⁴	The Coalition is a multidisciplinary team supported by the Government to help Belgian people and organisations to capture the opportunities of AI while facilitating the ongoing transition responsibly. Its main objectives are: policy support on Ethics, Regulation, Skills and Competences; Provide Belgian AI cartography; Co-animate Belgian AI community; Collect EU funding and connect EU ecosystems; Propose concrete action for training in AI; Contribute to the uptake of AI technologies by the industry; Make new products and services based on AI technologies emerge.
Switzerland	<i>The Mindfire Foundation</i> ⁷⁵	The Mindfire Foundation is a Swiss non-profit organisation that tackles human-level artificial intelligence to advance cutting-edge research for the benefit of humankind. In collaboration with its academic and private partners, the Mindfire Foundation implements the Global

⁷⁰ <https://www.gouvernement.fr/en/artificial-intelligence-canada-and-france-work-with-international-community-to-support-the>

⁷¹ <https://www.acrai.at/>

⁷² <https://malta.ai/>

⁷³ <https://icai.ai/>

⁷⁴ <https://www.ai4belgium.be/>

⁷⁵ <https://mindfire.global/>

		Robotics and AI Lab (GRAIL). GRAIL is a global research initiative that unites the best researchers to collaborate in both physical and virtual spaces and labs. GRAIL's mission control has been established at the University of Zurich.
Switzerland	<i>Data ethics group (Swiss Alliance for Data-Intensive Services)</i> ⁷⁶	This is a working group of the Swiss Alliance for Data-Intensive Services, a network for innovative companies, academic institutes and individuals with a focus on data-driven. The goal of the data ethics group is to develop and foster competence regarding ethical issues in data usage, discussing specific industry cases, organizing events with and developing ethical standards for companies and organisations.

2.5 Centres

The selection was based on a wider range of topics. The list in table 4 includes centres that deal with ELSEC-AI but have also broader interests (e.g. Internet and society and responsible ICT research and innovation). The list will be updated in the next months as there are many announcements of new centres that will work on ELSEC-AI in countries not yet covered. For example, it has been recently announced that Slovenia will establish an international AI research centre with UNESCO backing⁷⁷.

Table 4

Country	Name	Brief description
Germany	<i>AlgorithmWatch</i> ⁷⁸	AlgorithmWatch is a non-profit research and advocacy organisation to evaluate and shed light on algorithmic decision-making processes that have a social relevance, meaning they are used either to predict or prescribe human action or to make decisions automatically.
Germany	<i>Tactical Technology Collective</i> ⁷⁹	Tactical Tech is an international NGO that engages with citizens and civil-society organisations to explore and mitigate the impacts of technology on society. Examples of projects include Data Detox Kit; Data and Politics; Data and activism
Italy	<i>Nexa Centre for Internet & Society</i> ⁸⁰	The Nexa Center for Internet & Society is born from the activities of an initially informal interdisciplinary group – with expertise in technology, law and economics – that grew up in Torino from 2003. Since 2016, the

⁷⁶ <https://data-service-alliance.ch/expertise/expert-groups>

⁷⁷ The centre, as well as conducting research, will also run public consultations with the wider public on the impact of AI. <https://sciencebusiness.net/news/slovenia-establish-international-ai-research-centre-unesco-backing>

⁷⁸ <https://algorithmwatch.org/>

⁷⁹ <https://tacticaltech.org/#/>

⁸⁰ <https://nexa.polito.it/>

		Nexa Centre has been involved in two European projects: DECODE and EDU-HACK. In 2018 Nexa Center has collaborated to the writing of the White Paper, released by the Italian AI task force
Denmark	<i>DataEthics</i> ⁸¹	DataEthics is a politically independent ThinkDoTank based in Denmark with a European (and global) outreach. The purpose of DataEthics is to ensure primacy of the human being in a world of data, based on a European legal and value-based framework. It does so by focusing on collecting, creating and communicating knowledge about data ethics in close interaction with international institutions, organisations and Academia.
Denmark	<i>DATALAB – Center for Digital Social Research</i> ⁸²	Centre for Digital Social Research is an interdisciplinary research centre at the School of Communication and Culture (Aarhus University). The centre is based on the vision that technology and data systems should maintain a focus on people and society, supporting the principles of democracy, human rights and ethics. Examples of projects include: “when the smartphone becomes a running partner” and “Trust & Algorithms – Developing Machine Learning Algorithms for Danish Healthcare”
Belgium	<i>Privacy Solon</i> ⁸³	Privacy Salon is a non-profit organisation founded in 2014. Privacy Salon co-organises the annual Computers, Privacy and Data Protection (CPDP) conference, as well as numerous public side events focusing on legal and societal issues posed by current and future technologies, including an annual art exhibition. It has organised events on “micro-targeting and tactical fiction” and “algorithms and society”
The Netherlands	<i>UNICRI Centre for Artificial Intelligence and Robotics</i> ⁸⁴	Announced in 2016, the Centre aims to enhance understanding of the risk-benefit duality of Artificial Intelligence and Robotics through improved coordination, knowledge collection and dissemination, awareness-raising and outreach activities.
UK	<i>Digital Ethics Lab</i> ⁸⁵	The Digital Ethics Lab is a centre based in Oxford University. It aims to identify the benefits and enhance the positive opportunities of digital innovation as a force for good and avoid or mitigate its risks and shortcomings. Ongoing projects include Ethical Auditing for Accountable Automated Decision-Making; AI4People and Ethical and Social Implications of AI

⁸¹ <https://dataethics.eu/>

⁸² <http://datalab.au.dk/>

⁸³ <http://www.privacysalon.org/>

⁸⁴ http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics

⁸⁵ <https://digitaleticslab.oii.ox.ac.uk/>

UK	<i>The Institute for Ethical AI & Machine Learning</i> ⁸⁶	The Institute for Ethical AI & Machine Learning is a UK-based research centre that carries out research into responsible machine learning systems. It is formed by cross functional teams of volunteer including ML engineers, data scientist, industry experts, policy-makers and professors in STEM, Humanities and Social Sciences
UK	<i>The Institute for Ethical AI in Education (IEAIED)</i> ⁸⁷	The institute's responses to the Ethics Guidelines for Trustworthy AI published by the AI HLEG. The IEAIED will work to develop frameworks and mechanisms to help ensure that the use of AI across education is designed and deployed ethically.
UK	<i>Leverhulme Centre for the Future of Intelligence (CFI)</i> ⁸⁸	The CFI explores the opportunities and challenges of this potentially epoch-making technology, short-term as well as long-term. The CFI are based at the University of Cambridge, with partners at the Oxford Martin School at the University of Oxford, at Imperial College London, and at the University of California, Berkeley. Examples of research projects currently run are "AI: Futures and Responsibility"; "AI: Trust and Society" and "Kinds of Intelligence"
UK	<i>Centre for Data Ethics and Innovation (CDEI)</i> ⁸⁹	Announced in 2017, the CDEI is an independent advisory body set up and tasked by the UK Government to investigate and advise on how we maximise the benefits of these technologies. The CDEI gathers people from across sectors and society to shape practical recommendations for the government, as well as advice for regulators, and industry. Examples of current projects are online targeting; Algorithmic bias and AI Barometer.
UK	<i>Centre for Computing and Social Responsibility (CCSR)</i> ⁹⁰	The CCSR was established in 1996 at the De Montfort University (Leicester). Its mission is to undertake research and provide teaching, consultancy and advice to individuals, communities, organisations and governments at local, national and international levels on the actual and potential impacts of computing and related technologies on society and its citizens. Examples of current projects are the Observatory for Responsible Research and Innovation in ICT and Responsible Ethical Learning with Robotics
UK	<i>Data Justice Lab</i> ⁹¹	The Data Justice Lab examines the intricate

⁸⁶ <https://ethical.institute/index.html>

⁸⁷ <http://instituteforethicalaiineducation.org/>

⁸⁸ <http://lcfi.ac.uk/>

⁸⁹ <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>

⁹⁰ <https://www.dmu.ac.uk/research/research-faculties-and-institutes/technology/centre-for-computing-and-social-responsibility/ccsr-home.aspx>

⁹¹ <https://datajusticelab.org/>

		relationship between datafication and social justice, highlighting the politics and impacts of data-driven processes and big data. The lab is hosted at Cardiff University's School of Journalism, Media and Culture. Current projects include: DataJustice; Towards Democratic Auditing; Data Harm Record; and Big Data from the South
Sweden	<i>AI Sustainability Centre</i> ⁹²	The Centre was established in 2018 to create a multidisciplinary hub to address the scaling of AI in broader ethical and societal contexts. Founding partners from industry include Atomico, Bonnier, Cirio, Microsoft, and Telia Company. The initiative has attracted some of the top minds in Academia focused on AI, coming from KTH, Karolinska Institutet, Lund, Umeå, and Linköping universities, as well as public agencies like Skatteverket, the Swedish tax authority.
Luxembourg	<i>The Interdisciplinary Centre for Security, Reliability and Trust (SnT)</i> ⁹³	SnT conducts internationally research in information and communication technology, ICT, with high relevance creating socio-economic impact. In addition to long-term, high-risk research, SnT engages in demand driven collaborative projects with industry and the public sector. SnT's strategic research priorities are: Autonomous Vehicles; Cybersecurity; FinTech; Internet of Things; Secure and Compliant Data Management; Space Systems and Resources
Switzerland	AI Transparency Institute	The AI Transparency Institute is a non-profit organisation, dedicated to AI Governance and Human Trust in AI. It addresses key challenges like digital ethics, AI safety, explainability, fairness, transparency and privacy related issues. It contributes to build an open artificial intelligence for the benefit of all ⁹⁴

2.6 Projects and educational resources

EU funded projects that address ELSEC AI are:

- HUMANE AI⁹⁵ (open): a FET Flagship Project that focuses on human-centred AI. It aims to provide the principles for the design and the deployment of AI systems that enhance human capabilities and empower both individuals and society.

⁹² <https://www.aisustainability.org/>

⁹³ <https://wwwfr.uni.lu/snt>

⁹⁴ Further information is available in an online poster:

https://www.researchgate.net/publication/331773647_Presentation_of_the_AI_Transparency_Institute_at_the_Applied_Machine_Learning_Days

⁹⁵ <https://www.humane-ai.eu/>

- Sherpa⁹⁶ (open): a project that aims to analyse and synthesise our understanding of the ways in which smart information systems impact ethics and human right issues.
- CoHuBiCoL⁹⁷ (open): an ERC-funded research project that investigates how the prominence of counting and computation transforms many of the assumptions, operations and outcomes of the law.
- DATAJUSTICE⁹⁸ (open): an ERC-funded research project on understanding datafication in relation to social justice
- ThinkBig⁹⁹ (closed): an ERC-funded research project that studied the opportunities and the risks of the paradigm shift of data-driven AI.

A partial list of educational material and courses on ELSEC-AI include:

- Executive Masterclass “Digital Ethics by Design” (Delft University)¹⁰⁰
- MOOC on “Responsible Data Scientists”¹⁰¹
- Postgraduate course on “Humans and Intelligent Machines”, University of Bath (UK)¹⁰²
- Postgraduate course on “Professional Practice of Artificial Intelligence”, Universitat Politècnica de Catalunya (ESP)¹⁰³
- Course on Ethics of Artificial Intelligence, University of Edinburg (UK)¹⁰⁴
- Course on Ethics of Artificial Intelligence, Linköping University (Sweden)¹⁰⁵

3. OSAI Framework

OSAI will act in accordance with three verbs, i.e. Observe – Reflect – Report, which synthesizes OSAI’s primary approach. In particular, OSAI will:

- *Observe* facts and events occurring within Europe by monitoring newspapers, online bulletins, newsletters, notifications from the AI4EU partners, scientific literature, etc.
- *Reflect* on particular events or issues thanks to the contribution of ELSEC-AI experts and, in particular, thanks to the activities of the working groups (see task 5.2 in AI4EU annexe).
- *Report* to the general public by using a simple (but not simplistic) language in a way to support mutual understanding among experts and educate laypeople.

⁹⁶ <https://www.project-sherpa.eu/>

⁹⁷ <https://www.cohubicol.com/>

⁹⁸ <https://datajusticeproject.net/>

⁹⁹ <https://thinkbig.enm.bris.ac.uk/>

¹⁰⁰ <http://designforvalues.tudelft.nl/digital-ethics-by-design-executive-masterclass/>

¹⁰¹ <http://www.sobigdata.eu/ethics>

¹⁰² <http://www.bath.ac.uk/catalogues/2018-2019/cm/CM50272.html>, see also the lectures composed by Dr Joanna Bryson: <http://www.cs.bath.ac.uk/~jjb/here/CM50272/bryson-him.html>

¹⁰³ <https://www.fib.upc.edu/en/studies/masters/master-artificial-intelligence/curriculum/syllabus/PPAI-MAI>

¹⁰⁴ <https://marksprevak.com/teaching/ethics-of-ai-ug/>

¹⁰⁵ <https://liu.se/en/education/course/721g28>

3.1 Functions and roles

OSAI will perform two main functions:

1. *Distribution of knowledge on ELSEC-AI.* This function will be implemented through regular publications of articles (the types of contributions are described in section 3.2.) and news about ELSEC-AI. Note that this function closely relates to a characterising feature of AI4EU platform (i.e. providing access to knowledge and news about AI¹⁰⁶). But, while the whole platform will primarily serve AI developers and innovators, OSAI will target people who are specifically interested in ELSEC-AI. Articles¹⁰⁷ will be published in a dedicated section of AI4EU platform. At present articles can be provided spontaneously or under request¹⁰⁸, but the process and the policies for contributing to OSAI will be revised in coordination with WP3 and WP4 in a way to achieve a coherent approach with the overall management of AI4EU platform and ecosystem. Articles will be published in compliance with the intellectual property rights of the legitimate owners.

While there may be different types of contribution, articles could address two targets of readers: experts in any field implied by ELSEC-AI discussion (e.g. computer scientists, ethicists, sociologists, artists, philosophers...), and non-experts, (people who have partial or no knowledge about ELSEC-AI but are nonetheless impacted by them). Note that, in the latter, there might be people who have expertise in one field, e.g. computer science, but with limited or no knowledge in others, e.g. philosophy. For this reason, OSAI seeks to facilitate interdisciplinary dialogue by supporting conversation among people with different backgrounds – in particular, this could be carried out by encouraging the exchange of feedbacks, articles or interviews with opposite viewpoints, and multi-author discussion articles. As for the timing, the publication of early articles is expected by the end of June. More details on publication frequency are presented in section 3.3.

2. *Creating a community of people around ELSEC-AI.* This function is connected to the former, but its implementation will take longer than six months. Indeed, the OSAI

¹⁰⁶ In particular, the news section on AI4EU platform will deliver information about AI resources and tools (see, for example, WP4 tasks). However, this is not in conflict with OSAI function as the target audience of the overall platform includes AI developers and, more in general, people interested in gaining technical knowledge or skills. On the other hand, OSAI focuses on ELSEC issues and will refer to technical notions as long as they support the discussion of ELSEC-AI. For this reason, OSAI will interact with other WPs (in particular, WP2 and WP4) to complement the knowledge and achieve full integration of services.

¹⁰⁷ Articles are not meant to be a scientific work, although an author can contribute an adapted summary of a paper he/she has already published or a work in progress (like ones in arXiv or ResearchGate). It will be authors' responsibility to verify copyright policies of journals/books where they have already published or plan to publish.

¹⁰⁸ A first call for contributions has been sent to all AI4EU partners at the beginning of June 2019

community will rest on social ties that mature over time. In addition, the growth of social connections is also affected by the success of OSAI's first function, so that the higher the quality of published articles, the higher the possibility to attract new readers and contributors. In reality, between the two functions, there is a mutual influence as a community of active and motivated people can positively influence the quality of OSAI's articles. OSAI aims to build a broad community, composed of people that act in different capacities. In particular, as well as editors, the community can involve regular readers, contributors and people who make specific requests (e.g. the explanation of a technical concept or a recent AI achievement). In addition, we plan to engage a group of people with a more permanent role: that of taking care of an OSAI mission and help editors in spotting events and report facts (the so-called "observers"). OSAI will foster participation and engagement in several ways. It will leverage AI4EU communication channels (mailing list and contacts) and events (joint WPs workshop and conference call), support the activities of the Working Groups (WGs, see the subsection 3.5), interact with other centres and initiatives working on ELSEC-AI and make use of social media. OSAI community builds upon AI4EU consortium. So, participation will be solicited across AI4EU partners, the users of AI4EU platform and will be monitored in coordination with WP4 (see task T4.8 "Communication and Dissemination Plan" in the Annex, part A). Possible indicators are web analytics and the number of people contributing articles and attending WGs.

At present, the work on OSAI rests on two leading roles:

- *Editors* are the people in charge of the publication of articles. Specifically, their role implies the following tasks: spotting events or facts that are relevant for ELSEC-AI, proposing topics for discussion, soliciting contributions and feedback, reviewing and publishing articles. Currently, the team of editors consists of UVE members Luc Steels and Teresa Scantamburlo.
- *Contributors* are the fuel of OSAI's engine. They may either propose contributions or write articles under a specific editors' request and contribute on a regular or occasional basis. OSAI will work to involve contributors with different viewpoints so as to increase diversity and pluralism within the community. In the beginning, most of the contributors will be directly invited by editors and selected from the AI4EU consortium. As the Observatory evolves, the network of contributors is expected to extend beyond the AI4EU community and involve various stakeholders, policymakers, people from specific scientific communities (e.g. EurAI, the community of Machine Ethics, etc), companies in the business of AI, and non-profit organisations, etc. Extra value in contributing to OSAI is the possibility to get in contact with people from different backgrounds and establish an interdisciplinary dialogue.

3.2 Template

The Observatory issues four publications per year (one for each season) and, at present, includes four types of contributions:

- *Focus*: an event that editors want to emphasise and discuss. It may regard the launch of a new critical AI application, the introduction of a new law or regulation, a controversial fact, and so forth. For the OSAI demonstrator, editors chose to focus on the HLEG-AI's ethics guidelines and the related assessment activities.
- *Debates*: a working paper that presents one or more original ideas that are open to feedback from the readers; a real case study discussing potential ELSEC-AI issues; a domain-specific concept explained by relevant experts or blurred notions explained from different disciplinary perspectives. For the first issue of OSAI, the working paper is contributed by Vincent Muller, professor at the University of Eindhoven, and a partner in AI4EU project (WP5).
- *Reviews & Reports*: a critical appraisal of recent books, papers or experiences (e.g. conferences, meetings or projects) that relate to AI and its impact on humans. Reports may include the results presented at specific conferences (e.g. "Microwork platforms", held in Paris, June 2019¹⁰⁹) or summary of activities on ELSEC-AI (Dagstuhl seminar 19171¹¹⁰)

The timing and the types of publications will be revised at the end of the first year so as to make appropriate adjustments (for example, articles could be published at shorter intervals). This revision will be conducted in close collaboration with AI4EU partners dealing with the design of platform requirements and management (WP2, WP3, WP4) who will certainly by the end of the year have taken over the operation of the OSAI web presence as soon as the platform is ready to be used.

In addition, the Observatory curates two other sections that can be updated at any time:

- *Resources*: all those educational and informational materials available to people in Europe to gain knowledge about AI and related ELSEC issues (e.g. courses, MOOC, videos, organisations, reports, national strategies, etc). These items will be integrated into the

¹⁰⁹ <http://endl.network/events/conferences/indl-1-conference/>

¹¹⁰ <https://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=19171>

educational resources that will be provided or referenced by WP4 through the AI4EU platform.

- *News*: this is a space where we post brief news about ELSEC-AI issues. They may include the announcement of summer schools, activities, conferences, call for papers, open positions, etc. This news section will be integrated within the regular WP4 news section on the AI4EU platform.

Before the AI4EU platform becomes fully operational, all items will be published on OSAI's temporary website (www.unive.it/osai). The demonstrator version of the website has been created by using the Content Management System of Ca' Foscari University (i.e. Typo3¹¹¹), but this is to be replaced by the AI4EU content management system supporting the AI4EU Platform. Some screenshots of the preliminary version are provided in the images 1 and 2.

Image 1

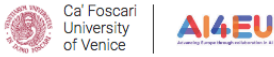


The Observatory on Society and Artificial Intelligence (OSAI) is part of the European project AI4EU and its main aim is to act as a clearinghouse for information and research on the Ethical, Legal, Socio-Economic and Cultural (ELSEC) issues in the development and deployment of AI technologies across the Europe.



¹¹¹ <https://typo3.org/>

Image 2



Home About Debates News



Debates

"Ethics of AI and Robotic" by Vincent Muller

We propose a paper by [Vincent C. Müller](#), professor of Philosophy and Ethics at the Technical University of Eindhoven. The draft has been prepared for Edward N. Zalta (ed.), Stanford Encyclopedia of Philosophy (Palo Alto: CSLI, Stanford University)

Abstract

Artificial intelligence (AI) and robotics are technologies that seem to be of major importance for the development of humanity in the near future. They have raised fundamental questions about what we should do with these systems, what the systems themselves should do, and what risks they have in the long term. They also challenge the human view of humanity as the intelligent and dominant species on Earth. The main division of the article is into issues that arise with AI systems as objects, i.e. tools used by humans (2), vs. AI systems as autonomous subjects, i.e. when ethics is for the AI systems themselves (3). The problem of a future 'singularity' or 'superintelligence' concerns ethical use of AI and the ethics for AI systems – thus the separate final section (4). [Preprint of the [full paper](#) available for comments]

[Call for comments](#)

3.3 Publication frequency

The proposed plan is to publish contents based on two schemes: 1) *at any occurrence* and 2) *at each trimester*. In the first mode, there are contents like news and resources that will be updated at any time the editors collect or receive notification from other partners. While, in the second mode, there are sections that require more extended processing and include: Focus, Debates, and Reviews & Reports. The plan for the publication of articles throughout the project is provided in the tables below:

2019	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						X			X			X

2020	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
			X			X			X			X

2021	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
			X			X			X			X

This plan will be revised to be conform with the publication plans of WP4 and the editorial procedures to be practiced on the upcoming AI4EU platform.

3.4 Methodology and networking

To collect OSAI's articles editors will follow an iterative algorithm exemplified in the table number 5

OUTPUT	Editorial plan with the list of authors
Step 1	Editors meet and create a proposal for the coming issue with names of possible contributors
Step 2	Editors contact authors and coordinate with them
Step 3	Editors adjust the initial plan based on the interaction with authors
Step 4	Repeat steps 1 – 2 – 3 until the list of authors is completed for the deadline

The reviewing process will be constrained by a schedule of deadlines that should be flexible in order to encourage people to contribute. There will also be a flexible approach to the length of contributions, and specific issues will be considered on a case-by-case basis. Contributions will also be encouraged through open calls that can be disseminated through the mailing list of AI4EU partners and AI4EU platform.

Most of OSAI's outcome will result from an intense activity of networking and collaborations. The primary source of collaboration is the network of AI4EU partners. On the one hand, editors will ask AI4EU partners to contribute articles on aspects of their work that may regard ELSEC-AI issues. For example, OSAI plans to invite all AI4EU pilots to contribute case studies that can be discussed from an ELSEC-AI perspective. An opportunity to begin this conversation has been given by the AI4EU joint WP5/6/7/9 workshop held in Paris on May 2019. On the other hand, OSAI can contribute feedback and analysis to people working the AI4EU pilots and the five research areas addressed by WP7, i.e. Explainable AI, Verifiable AI, Collaborative AI, Integrative AI and Physical AI.

Besides, OSAI will work closely with other WP5 members to help disseminate their work, such as the assessment of AI HLEG ethics guidelines and the development of the Value Design Methodology.

3.5 Working Groups

Part of the activities of the Observatory is to stimulate in-depth debates about all ELSEC aspects. Until now, ethical issues have mostly been the focus, but legal, socio-economic and cultural issues are just as important. Hence, the approach (as already specified in the project proposal) is to form working groups with an initial set of core members coming from the AI4EU consortium but extended by experts in the specific themes of each working group. After various initial

consultations, the list of working groups has been revised concerning the list provided in the proposal and will cover the following topics:

- *WG1: Ethical issues in AI* (ethical deliberation, autonomy, agency, responsibility, common/public goods, etc.)
- *WG2: Legal issues in AI* (privacy, bias, discrimination, labour, rights, autonomy, etc.)
- *WG3: Socio-Economic issues in AI* (equality, justice, trust, labour market, trading,
- *WG4: Cultural issues in AI* (representations of AI within society particularly in artistic media, history of AI, developing AI in a multi-cultural society, AI education, interdisciplinary dialogue, etc.)

Each WG will work autonomously on the topics that best reflect the issue at stake (some examples are listed between the parenthesis in the above list). Since their topics are closely intertwined, the WGs are expected to exchange ideas and plan common activities. Examples of possible activities include interviews, user studies, case studies, citizen consultations, brainstorming, summer schools, workshops, seminars, brainstorming meetings, and public events.

Some criteria for selecting members are

- Competence and a track record in AI.
- Competence and prior work on the theme of the working group.
- Geographical diversity.
- Cultural/disciplinary diversity (people from different fields e different sectors).
- Gender and age balance.

The expected output of the working groups

- A workshop
- Regular contributions for the Observatory
- Final report on the activities of the working group
- Final publishable report on the activities of the working group
- Input to other WPs in AI4EU, in particular WP7 and WP9.

At this moment, the working groups are still in the stage of being established, as most emphasis has been on creating the OSAI web presence demonstrator, but first Workshop activities and inputs to the Observatory are expected toward the end of 2019.

4. Ongoing work and future plan

The Observatory activity has the potential to be expanded in various ways and some concrete plans for future are exposed below.

In the short run (i.e. by the end of 2019), the plan is to coordinate with WP2, WP3 and WP4 to achieve the complete integration of the current OSAI web materials within the AI4EU platform/website and to carry over the daily operation and maintenance to WP4. This interaction will be crucial as those WPs deal with the design/implementation of the AI4EU platform (WP2); the management and enrichment of the AI4EU platform (WP3); and the ecosystem creation and development (WP4). In particular, WP5 plans:

- to support WP2 and WP3 in designing the structure of OSAI within AI4EU platform and integrate OSAI activities within AI4EU website
- to help WP4 and WP2 integrate news and resources about ELSEC-AI (i.e. sections “resources” and “news” described in section 3.2) into AI4EU website and platform. In particular, this will help to align with connected tasks in WP4, i.e. “Education and training activities” (T4.6) and “Communication and Dissemination Plan” (T4.8).
- to coordinate with WP4 to identify a group of editors who can take care of ELSEC-AI within the editorial board of the AI4EU platform. Some members of WP5 can initially form this specialised group of editors. This group will help the editorial board of AI4EU platform contribute articles and animate the OSAI section within AI4EU platform.
- to promote OSAI and ELSEC-AI among European organisations, such as scientific communities, universities, non-profit associations, governmental agencies, companies, etc.

As a mid-term plan, OSAI should try to interact with leading conferences and publishers to propose special tracks or issues on ELSEC-AI. Indeed, this would increase OSAI attractiveness among researchers and help WGs to communicate their research findings.

In the longer run, the OSAI will try to create a stable network of experts from different European countries (the so-called *observers*) to spot and report facts about ELSEC-AI within Europe. This would allow for the better identification and understanding of local phenomena that are not reported in the English language, or that are not covered in the public debate.