Grant Agreement N°825619

AI4EU Deliverable D5.2

Ethical Observatory Report

WP	5	Ethical Observatory Report
Task	5.1	The AI4EU Ethical AI Observatory

Dissemination level[1]	PU		**Due delivery date**	30/11/2021
Nature[2]	R		**Actual delivery date**	30/11/2021

Lead beneficiary	**BSC**
Contributing beneficiaries	UVE

Document Version	Date	Author	Comments[3]
v0	27/10/2021	Atia Cortés and Teresa Scantamburlo	First draft
v1	4/11/2021	Atia Cortés and Teresa Scantamburlo	revised version based on first reviewer's comments
v2	23/11/2021	Atia Cortés	Revised version based on second reviewer's comments
v2.1	29/11/2021	Atia Cortés and Teresa Scantamburlo	Final version

[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

This deliverable presents the activity of the Observatory on Society and Artificial Intelligence (OSAI), whose main contributions are the creation of content for the platform regarding the Ethical, Legal, Socio-Economics and cultural aspects of AI (ELSEC-AI) and the coordination of the working groups. This document builds upon the information provided in the deliverable D5.1 The AI4EU Ethical AI Observatory (M6) and D5.3 ELSEC for EU (M30). The report provides a full description of the role of the Observatory within the AI4EU ecosystem through internal collaborations with different work packages of the project. Moreover, the Observatory team has built a network of ELSEC-AI experts composed by members of the AI4EU consortium and external contributors. The different activities coordinated with this group of stakeholders and the results obtained up to today are also included in this work.

Table of Contents

1. Introduction

The Observatory on Society and Artificial Intelligence (OSAI or simply ”Observatory” hereafter) was set up in 2019 within the H2020 EU funded project AI4EU, whose objective is to build the first European Artificial Intelligence on-demand web platform and ecosystem. The OSAI is an example of a vast array of initiatives animating the *ethical turn* of Artificial Intelligence (AI), outlined in table 1. Although at its infancy, it gives us the opportunity to explore how this and similar activities can contribute to stretch the assessment of AI and turn progress towards ethical principles.

As we said, the creation of the Observatory takes place in a complex and dynamic context where an imprecise number of AI-related events populate the European calendar. Table 1 includes a collection of European centers that are specifically dedicated to the research around AI and its impact on Society. These were selected from a larger set based on a search of simple keywords on Google engine (such as “AI”, “ethics”, and “society”). To keep our focus on the European landscape we limited our search to organizations based in Europe or involving European countries.

A common aim of these organizations is to promote designs and developments of technologies that put upfront concepts such as social responsibility, trust or fairness. Some are dedicated to the creation of guides, others to define evaluation methods, but all have in common the will to create spaces for multidisciplinary dialogue.

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, the OSAI aims 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).

The Observatory differs from these initiatives in several respects. In the first place, the OSAI focuses not only on articles and news, but also on people. Indeed, one of the motivating ideas behind the Observatory 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 subjects such as AI experts (*e.g.* AI researchers and practitioners), specialists in any ELSEC-related field (ethicists, sociologists, lawyers, policy makers, artists, etc.) and lay people. In the second place, the OSAI will approach ELSEC-AI in the context of Europe so as to foster the dialogue among European countries.

Table 1: European Centers and Networks for AI and Ethics

Name	Country	Type	Objective
HumanE AI	Europe	H2020 EU Project	To create the foundations for AI systems that empower people and society, with special focus on **Collaborative Humane Computer Interaction** based on a convergence of Human-Computer Interaction with Machine-Learning.
AI Watch	European Commission	Public Institution	An initiative to monitor the development, uptake and impact of AI for Europe
AI4People	Europe	Multi-stakeholder Forum	To **bring together all actors interested in shaping the social impact of new applications of AI**, including the European Parliament, civil society organizations, industry and the media. They published the AI4People's Ethical

			Framework which inspired the European Guidelines for Trustworthy AI
OECD.AI	Inter-governmental	International Organisation	The AI Policy Observatory by the Organization for Economic Co-operation and Development (OECD) combines resources from across its partners and all stakeholder groups to **facilitate dialogue between stakeholders** while providing **multidisciplinary, evidence-based policy analysis** in the areas where AI has the most impact.
Knowledge Centre Data & Society	Belgium	Research Centre	Funded by the Flemish Department on Economy, Science and Innovation, it enables **socially responsible, ethical and legally appropriate implementations of AI in Flanders**.
Karel Čapek Center for Values in Science and Technology	Czech Republic	Research Centre	The center focuses on **ethical and legal issues** connected with the evolution **of contemporary science and technology**, especially in the areas of **Biomedicine**, **Artificial Intelligence** and **Robotics**. In addition to the professional research and international collaboration it is tasked with seeking practical applications to ethical and legal issues in the field of biomedicine, artificial intelligence and robotics.
DataEthics	Denmark	ThinkDoTank	To ensure primacy of the human being in a world of data, based on a European legal and value-based framework. It has a core **focus on AI as the evolution of complex data processing** extended in human decision-making within politics, economics, identity and culture.
DATALAB - Center for Digital Social Research	Denmark	Research Centre	Conducts research in many different aspects of behavioral data within several areas. A special focus is brought to the **social effects of automated data processing** as well as to the social adaptation of automated data systems.
ImpactAI	France	Non-profit Association	Think &Do Tank for Ethics and Responsible AI aiming to **promote the development of trusted AI**, support innovative projects and publish annual reports.
Global AI Ethics Institute	France	Think Tank	A think tank that aims to raise awareness on the **importance of cultures** in the ethical appraisal of AI systems.
Algorithm Watch	Germany	Non-profit Organisation	Based on research and advocacy to **evaluate algorithmic decision-making processes**, raise ethical conflicts and explain its features to the general audience.
AI & Society Lab	Germany	Research Laboratory	Interface and translator between academia on one side and industry and civil society on the other, it functions as **experimental space for new formats to advance knowledge generation and knowledge transfer to AI**.
Institute for Ethics in AI	Germany	Research Centre	To generate global, egalitarian and interdisciplinary **guidelines** for the ethical development and implementation of AI and to integrate ethical **and societal priorities into the development of fundamentally integrative AI** technologies.
German Institute for Standardisation (DIN)	Germany	Public-Private Partnership	An independent platform that is defining an **AI Standartisation Roadmap** that aligns with the German AI strategy. It is composed by a group of

			high-ranking representatives from industry, politics, science and civil society.
AI Sustainability Centre	Sweden	Consultancy	Creation of **AI Sustainability Framework** for identifying, measuring and governing the ethical implications of AI and assisting organizations from a legal, technical and societal perspective.
AI Transparency Institute	Switzerland	Non-profit association	Dedicated to **AI governance and human trust in AI**, they address key challenges in digital ethics, AI safety, transparency, fairness and privacy.
Digital Ethics Lab	UK	Research Centre	Tackles the **ethical challenges of digital innovation** from a multidisciplinary perspective, with the aim to identify benefits and positive opportunities while avoiding risks and shortcomings.
Institute for Ethical AI & ML	UK	Research Centre	Highly-technical, practical and cross-functional research across **8 Machine Learning Principles** and **Explainable AI Framework**
Institute for Ethical AI in Education	UK	Research Centre	As a response to the Trustworthy AI Guidelines, this institute works to **develop frameworks and mechanisms** to help ensure that the use of AI across education is designed and deployed ethically.
Leverhulme Centre for the Future of Intelligence	UK	Research Centre	To build an interdisciplinary community of researchers with strong links to technologists and the policy world to study the **impact of AI in society with a focus on trust, fairness, accountability and democracy**.
Centre for Data Ethics and Innovation	UK	Public Institution	Part of the Department for Digital, Culture, Media & Sport, they connect policymakers, industry, civil society, and the public to **develop the right governance regime for data-driven technologies**.
Future of Humanity Institute - Oxford University	UK	Research Centre	A multidisciplinary research institute at the University of Oxford gathering scholars from mathematics, philosophy and social sciences to bear on big-picture **questions about humanity and its prospects**. Currently the centre includes the following research groups: Macrostrategy - **Governance of Artificial Intelligence** - **AI Safety** - Biosecurity - Digital Minds

2. The Observatory on Society and AI (OSAI)

The Observatory on Society and AI was created in 2019 by the ECLT - University of Venice and the Barcelona Supercomputing Center, with the aim to become the main channel of the WP5 and the AI4EU platform to support discussion and to facilitate the distribution of information about the Ethical, Legal, Socio-Economic and Cultural issues of AI (ELSEC-AI) within Europe. The Coordination of the Observatory is led by Atia Cortés (BSC) and Teresa Scantamburlo (UVE) with the support of Francesca Foffano (UVE), Cristian Barrué (UPC), Ulises Cortés (UPC) and Luc Steels (UVE). Specifically, the following objectives were identified:

- To stimulate reflection, discussion and due consideration of ELSEC-AI issues within the project through a series of working groups. OSAI is attracting a network of experts in different domains of ELSEC-AI that will contribute to bridge the knowledge gap existing today within AI practitioners and users.

- To provide resources to educate the general EU public more accurately about AI and ELSEC-AI issues by generating weekly content in the form of articles, reports, cultural announcements with the objective to promote discussion and awareness on these topics.

The Observatory 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 its own identity and role neither in contrast nor competition with other existing European initiatives (*e.g.* High-Level Expert Group on AI). It aims to increase connections among these related projects and make accessible a broad range of articles to the European public at large. In particular, the OSAI's approach can be described by three verbs:

1. *Observe* facts and events occurring within Europe by monitoring newspapers, online bulletins, scientific literature, etc.
2. *Reflect* on particular events or issues through to the contribution of ELSEC-AI experts and, in particular, thanks to the activities of the working groups
3. *Report* to the general public by using a simple (but not simplistic) language in a way to support mutual understanding among experts and educate lay people.

The work of the Observatory during this project has relied under these three pillars, and is reflected in the outcomes described in the deliverable: the selection and production of content published in the OSAI and the coordination of working groups.

2.1 Integration of the OSAI in the AI4EU platform

The Observatory[4] was born in 2019 as a web demonstrator hosted by University of Venice while the AI4EU platform was under development. Details of the design and purpose of the initial version were largely described in deliverable D5.1 “Ethical Observatory description of functions, oversight powers, specific agenda and interactions with other groups“ (M6). In November 2019, the Observatory team started a discussion with different partners from the AI4EU Consortium (WP1, WP2 and WP4, *i.e.* the operational coordination and ecosystem teams) to organize the process of integration of the Observatory within the AI4EU platform. First, the OSAI team was involved in the design of requirements and user experience of the Observatory and the AI4EU platform in general. The final integration of the Observatory in the AI4EU platform was achieved by May 2020, migrating all the content created until the date. The OSAI team was also involved in the editorial and promotional boards, planning contents, strategic campaigns and improvements of the platform with periodic meetings every two weeks.

Thus, the OSAI team has been an active asset of the platform involved in the process of design and adaptation to the different new versions of the platform since its creation to its latest version after the 2021 migration. Nonetheless, the Observatory has also suffered from changes in the developmental process of the platform (migration, beta version, etc), which have affected our activity (in particular, to the frequency of publication of contents) in certain periods of time.

For each new phase of development of the AI4EU platform, the OSAI team provided feedback on how to improve the structure of the Observatory / Ethics section to 1) augment its visibility, by improving user experience within the web platform and 2) integrate the rest of the work of WP5, being the Observatory one of its outcomes. We have had several interactions with WP2, WP4 and the development teams in charge of each new phase, although our requirements have not always

[4] See the AI4EU platform: https://www.ai4europe.eu/ethics/osai

been included or taken into account. In its latest version, the Observatory is located in the main page of the Ethics section of the AI4EU platform, where the repository of articles, reports, centers and networks is placed. This section includes links to other activities coordinated by the OSAI team as shown in Figure 1*, i.e.* the working groups and workshops, as well as a static page dedicated to introduce the purpose and team members of the observatory.

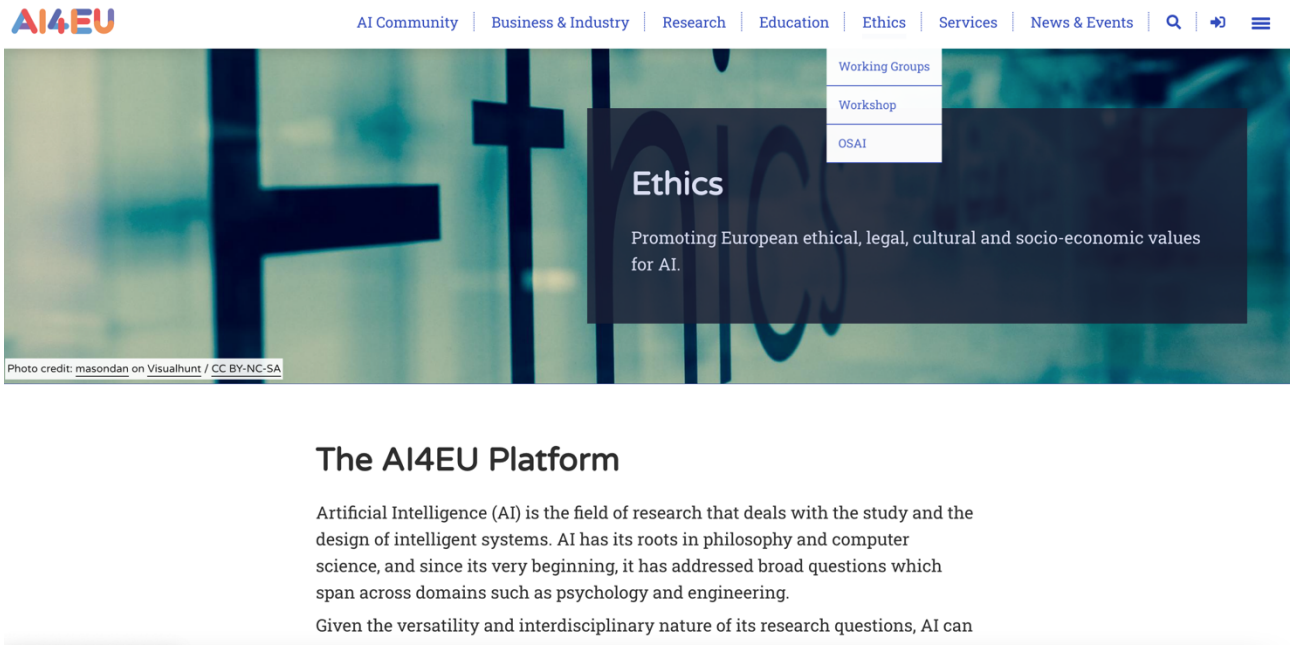

Figure 1: The AI4EU platform

2.2 Editorial line

The Observatory team has been working on an editorial plan to populate the platform with content related to ELSEC-AI aspects. Our initial objective was to create and upload content every week, either created by the OSAI team or by external contributors, but the frequency of publication has decreased drastically since early 2021 due to the migration to the new AI4EU platform. Indeed, during the period of summer 2021, the access to the platform was limited to fewer people of the development / technical team.

The selection and production of contents has followed different criteria, as for example the relation to ELSEC-AI, relevance with respect to occurring events (*e.g.* COVID-19), expression of research collaboration or dissemination activities. The quality of contents has been monitored by the Observatory team (Atia Cortés, Teresa Scantamburlo, Francesca Foffano, Ulises Cortés and Luc Steels) which may at times consult external experts. One of the objectives of the AI4EU platform was to attract a network of experts to foster interactions within the different communication channels, such as discussion forums or the Observatory. A particular editorial line regarded companies and their view on the ethical and social impact of AI. The OSAI team invited 7 companies to answer a few questions about AI and responsible innovation (examples of questions were: “What impact should we expect from AI innovation? Is there any particular example of a positive impact of AI coming from your company or your field that you would like to share? What does Trustworthy AI mean in your view and how does it translate in your everyday research and business?”). Unfortunately, reactions were limited and only three companies contributed (ClearboxAI, Huawei and MediaMonks).

Since the integration of the platform (November 2019), the Observatory have published a total of 119 pieces of content, including:

- **Reports**: brief summaries of documents issued by governments, companies or independent organizations that share investigations, strategies, frameworks or plans concerning the implications of AI.

- **Articles**: they can include working papers presenting original ideas, ideas that are open to feedback, short articles on recent activities dealing with Ethical, Legal, Social, Economic and Cultural issues of AI (*e.g* new laws, facts, events, etc.), case studies, research surveys, interviews with companies, domain-specific concepts explained by relevant experts or blurred notions explained from different disciplinary perspectives.

- **Centers:** a list of public or private organizations working on Ethical, Legal, Social, Economic and Cultural issues of AI (ELSEC-AI) based in Europe.

- **Networks**: a list of international partnerships or forums addressing ELSEC-AI topics involving European countries.

Following are some figures reflecting the activity of the Observatory section during the last 21 months (February 2020 to October 2021).

Table 2: Number of contents by type of contribution

CONTENTS	
Articles	56
Reports	26
Centers	18
Networks	19

Table 3: Relative number contents classified by topic

CATEGORIES	
AI & Society	48
AI & Ethics	46
AI & Law	13
AI & Economics	4
AI & Culture	7
AI & Gender	1

3. The OSAI in the AI4EU Ecosystem

The OSAI has created several connections within and outside the AI4EU project with a view to contribute to the development of an European AI ecosystem. Interactions within the project have allowed the OSAI to join efforts on goals of common interest, such as the development of human-centred AI and the promotion of AI education. While, external collaborations have contributed to the opening of the AI4EU platform to new potential users and the exchange with other relevant European initiatives (such as Tailor, Humane-AI Net, AI4Media and ETAPAS projects).

3.1 Interactions with internal partners

The Observatory has played an active role in the creation and maintenance of the AI4EU ecosystem, offering our service to the whole Consortium in different tasks. This has allowed us to be involved in the design process of the platform, the definition of the editorial and promotional lines and the promotion of trustworthiness and responsible practices of AI.

Within WP4, the OSAI has maintained periodic meetings with the Editorial and Promotional teams to identify relevant topics to populate the platform, promotional strategies or suggestions of improvement of the user experience. In addition, two members of the OSAI team (Atia Cortés and Ulises Cortés) are members of the Gender AI committee, which had monthly meetings during 2020 to define the lines of action (which finally were reduced due to the COVID-19 situation). The main outcome of this committee is the WAIROES campaign in social media channels. Finally, the Observatory has closely collaborated with the AI Education team as part of the work done in one of the working groups (see following subsection).

The Observatory has also collaborated with WP6 and WP8, having a direct interaction with use case providers of AI applications. The objective of these actions was twofold: to introduce AI stakeholders to the notion of ELSEC-AI and trustworthiness and to train them in the implementation of responsible AI practices. Between June and September 2020, the OSAI team had several interactions with the AI4Media pilot from WP6 to identify ELSEC requirements related to the use case and provide a tailored training and final report (see Annex 6). In the case of WP8, we used the self-assessment questionnaire provided in D5.3[5] by University of Umeä to evaluate the 41 projects selected from the open call process[6]. In addition, in September 2021 we provided a training session of the Trustworthy AI requirements and tools to implement them to the mentors of these projects.

The OSAI has also interacted with WP7, where we provided input regarding the European Guidelines for Trustworthy AI to help WP7 identify connections with their areas of interest: Explainable AI, Collaborative AI, Integrative AI, Verifiable AI, Physical AI.

3.2 Interaction with external partners

The creation and the coordination of working groups on ELSEC AI offered remarkable opportunities to open the AI4EU platform and ecosystem to external partners (for a list of working group participants with their short bios see the appendix). Working Groups (WGs) gathered 8 people from the AI4EU project and 15 external partners. For three of them the work period has been 9-month long and gave the opportunity to create meaningful ties among participants.

Working groups' goal was to create a space for reflection on ELSEC AI at the European level and leverage the contributions of experts from different fields and sectors. Participants attended on a voluntary and individual basis to share their personal reflections and, thus, set up a conversation that can serve the research community and the society. Participation has not involved any contracts or registration fees with individuals or their institutions.

Working groups have been coordinated by the OSAI team members: Teresa Scantamburlo (UVE), Francesca Foffano (UVE), Cristian Barrué (UPC), Atia Cortés (BSC), Ulises Cortés (UPC) and Luc

[5] V. Dignum, J.C. Nieves, A. Theodorou, A. Aler Tubella (2021), "An Abbreviated Assessment List to Support the Responsible Development and Use of AI", Responsible Artificial Intelligence Group, Department of Computing Sciences, Umeå University. Technical report 2021/03.
https://webapps.cs.umu.se/uminf/index.cgi?year=2021&number=3

[6] See the AI4EU challenges: https://ai4eu-challenges.fundingbox.com/

Steels (UVE). A more detailed descriptions of Working Groups objective, methodology and activities are provided in the deliverable D5.4 ("ELSEC for EU"[7]) and a brief summary of the performed tasks is offered in the following subsections:

Social Awareness group

The aim of the Social Awareness group was to understand and increase people's awareness of AI capabilities and ELSEC issues. To achieve this goal, the group contributed to the creation of a survey consulting the European population on AI applications and its impact on society. The group collaborated with a company (Marketing Problem Solving[8]) to perform interviews and access a representative sample of the target population. More information about the topics of the questionnaire and the data collected is included in section 4 ("Consultations"). Currently, the group is working on a scientific publication to share the results of the consultation with the research community.

Piloting Trustworthy AI group

The group on Piloting Trustworthy AI has focused on two main tasks. The first one regarded the identification and discussion of ethical and legal concerns in two real-world AI applications. In particular, the group considered a loan application of a Dutch bank (De Volksbank) in the light of the Assessment List for Trustworthy AI (ALTAI[9]) and a research project for the development of an autonomous shuttle (Politecnico di Milano) based on the European Commission's Recommendations on the Ethics of Autonomous and Connected Vehicles[10]. The second task involved the creation of a survey on Trustworthy AI targeting a wide range of experts working on technical and non-technical aspects of AI (more details on this survey are provided in section 4 "Consultation).

Education & AI (AI ethics education)

The purpose of the AI Ethics Education group was to collect information about courses addressing ELSEC in AI and/or Computer Science degree programmes within Europe. Some of the tasks performed include the identification of search criteria, the selection of geographic areas of interest and the data collection. In September 2021, the group collected about 100 courses and covered 19 European countries. The collection includes courses or subject addressing one or more Trustworthy AI requirements in STEM curricula (*i.e.* privacy and data protection, human autonomy and oversight, transparency, accountability, social and environmental impact, diversity and non-discrimination,

---

[7] The deliverable is available online: Foffano, F., Scantamburlo, T. Cortés, A., (2021), "ELSEC for EU" Deliverable_AI4EU_D5.3 https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/_REV__1.Deliverable_AI4EU_D5.3_M30_v4_FINAL.pdf

[8] See the website of the company: https://www.mpsresearch.it/en/

[9] High-Level Expert Group on Artificial Intelligence (AI HLEG), Assessment List for Trustworthy AI, 2020 https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[10] Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. 2020. Publication Office of the European Union: Luxembourg. https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en/format-PDF/source-search

safety and accuracy) and is available in a shared google folder[11]. Note that the results achieved by the group were shared with the Work Package 4 and used to fill the education repository (the integration of the material is ongoing) - the results of this collaboration is detailed in the subsequent sections.

Culture & AI group

The Culture and AI group consisted of a workshop activity held at a joint event organized by three Horizon 2020 projects: AI4EU, MUHAI, ODYCCEUS. The meeting took place on the 24 May 2021 during the workshop "AI & Archives"[12]. The workshop has focused on cultural applications of AI and the associated ethical and legal issues (such as privacy and copyright). This working group built on the idea that AI is playing an important role in the cultural sector which involves the production, dissemination and archiving of cultural goods. This sector is not only important for the cohesion of society, it is also an important economic activity in the EU and one in which Europe has a track record of excellence. But, as for other sectors, the application of AI methods and techniques raises various ethical, legal, social and economic issues, some of them unique to the production or dissemination of cultural artefacts, for example, the issue of fairness to authors in distributing their artistic work or the reuse of social media data to create new work.

3.3 Results of the OSAI contribution to the AI4EU ecosystem

The several contributions introduced in this section have resulted in some tangible outcomes, especially regarding the working groups activities. One of the biggest achievements is the organisation of two workshops, one online in November 2020 and one hybrid in September 2021, that have helped to launch our working groups and show some results of the research carried out during the last year (see Section 5 for further detail). In addition, we have built an education catalogue that has contributed to the educational repository published in the platform. We expect that this repository will grow in the coming months as it becomes more visible within the AI academic community. We have also launched two consultations addressed to citizens and AI experts, with the aim to understand the attitude towards AI and the perception of concepts such as trustworthiness or governance. Section 4 provides a complete description of the questionnaires as well as the presentation of some results.

The OSAI team supported working groups in the production of brief reports summarizing the activity they have completed up to now:

- "A Survey on AI and Ethics. Key factors in building AI trust and awareness across European citizens."[13] (WG on Social Awareness)
- "Safety, Privacy, Fairness, Interpretability and Responsibility of Autonomous Driving"[14] - Case study on autonomous driving (WG on Piloting Trustworthy AI)

[11] https://docs.google.com/spreadsheets/d/148csqzBHtizg8QO5_t6r0bMwWZdXhGaT_QXZHp4yKrI/edit#gid=528097995

[12] The poster of the event is available online: https://muhai.org/images/events/DEF10_brochure_workshop.pdf

[13] https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/Citizen_Consultation.pdf

[14] https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/Driving_case-study.pdf

- "Applying the ALTAI framework to a credit scoring algorithm for mortgages - A case study with the Dutch Volksbank"[15] - Case study on fintech (WG on Piloting Trustworthy AI)
- "A questionnaire to consult European experts on Trustworthy AI"[16] (WG on Piloting Trustworthy AI)
- "Compiling AI Ethics courses across Higher Education in the EU"[17] (WG on AI ethics education)

In addition, we have agreed to open a call for papers in a topic collection called "The culture of Trustworthy AI: public debate, education and practical learning" with the Ethics and Information Technology Journal (springer). To the date, this process is still under evaluation of the editorial board of the journal.

4. Consultations

The OSAI team has coordinated two important activities in the context of the AI4EU Working Groups: a citizen and an expert consultation. The citizen survey was launched on the 7th June and completed on the 14th June, while the expert questionnaire was made available online on the 7th October and the collection of answers will be stopped by the end of the month. Both surveys are meant to strengthen one of the main objectives of the OSAI, that is to bridge the knowledge gap existing today between AI practitioners and AI users. In the subsequent sections we describe the purposes of these activities, the topics of the questionnaires and some of the results achieved (the questionnaires are provided in full in the appendix)

4.1 Citizen consultation

The survey is based on a review of previous studies and consultations focused on public opinion about AI. This includes surveys such as "Public view of Machine Learning" by the Royal Society[18], the European Consultation on AI by the Atomium European Institute (ECAI)[19], the Moral Machine platform by Scalable Cooperation and the MIT Media Lab[20] and the "Trust in Artificial Intelligence: Australian Insights" by KPMG and the University of Queensland[21], among others. Some of the questionnaires were publicly available, such as the one by the Royal Society, while others were not fully accessible as the proposed consultation by ECAI. Our review highlights how a survey on AI and its social impact is still missing at a European level and suggests that our contribution to this topic might be relevant to fill the existing gap.

The questionnaire has been translated into 8 languages and proposed to 4000 people living in 8 European countries (Italy, Spain, The Netherlands, Poland, Romania, Sweden, France, Germany). For each country a representative sample of 500 people was selected based on gender, age

[15] https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/Fintech_case-study.pdf
[16] https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/Expert_Consultation.pdf
[17] https://www.unive.it/pag/fileadmin/user_upload/progetti_ricerca/osai/img/grafica/AI_Ethics_Education.pdf
[18] Ipsos MORI. (2017). Public views of Machine Learning. Report on behalf of the Royal Society. April 2017 https://royalsociety.org/-/media/policy/projects/machine-learning/digital-natives-16-10-2017.pdf
[19] Atomium European Institute for science, media and democracy. The European Consultation on AI (ECAI). Last consultation: 6 August 2021. https://www.eismd.eu/ecai/
[20] MIT. Moral Machines, 2016-2020. https://www.media.mit.edu/projects/moral-machine/overview/
[21] Lockey, S., Gillespie, N., & Curtis, C. (2020). Trust in Artificial Intelligence: Australian Insights. The University of Queensland and KPMG Australia. doi.org/10.14264/b32f129

(spanning from 18 to 75) and geographic areas. The questionnaire was translated in 8 different languages

The questionnaire is composed of 16 questions investigating three key factors:

1. *AI Awareness:* 7 questions which consider people's self-reported knowledge of AI and the perceived impact on their daily life. In addition, this part investigates the general awareness of European Commission's initiatives such as GDPR, the Ethics Guidelines for Trustworthy AI and the recent Proposal for an AI Regulation.
2. *AI Attitude*: 5 questions which focus on citizens' approach toward AI in general and in some specific sectors and scenarios (e.g. job application and energy consumption).
3. *Trust & AI*: 4 questions regarding citizens' ethical priorities and trust in entities which could ensure a beneficial use of AI.

The collection of the data was done in June 2021 in collaboration with Marketing Problem Solving (MPS) based on a Computer-Assisted Web Interview methodology (CAWI). We presented partial results of our study at the AI4EU workshop "The Culture of Trustworthy AI. Public Debate, Education and Practical Learning" held on September 2-3, 2021 in Venice. A sketchy outline of our results is presented in the following paragraphs.

AI awareness

Half of the population don't feel competent on the topic (about 50%) and just a small percentage (20%) of the population believe to have a good education on AI. The perceived competence on AI seems to be associated with respondents' age and digital expertise. For what concerns the initiatives undertaken by the European Commission, the results suggest that the General Data Protection Regulation (GDPR) is the most well-known initiative (66%), especially in countries such as Romania, Poland and Sweden where the percentage exceeds 75%. Other official communications such as the Ethics Guidelines and the Proposal for a Regulation on AI are less known (around 30%).
Regarding their daily interactions, about 30% of our interviewees report to be aware of interacting with products based on AI. With respect to the employment of AI in different sectors, the respondents are cognizant about many domains where AI is applied, especially in the military and manufacturing sectors (the ratio of cognizant citizens over unaware citizens is above 6), but they are less aware when it comes to human resources and agriculture (the ratio is below 3).

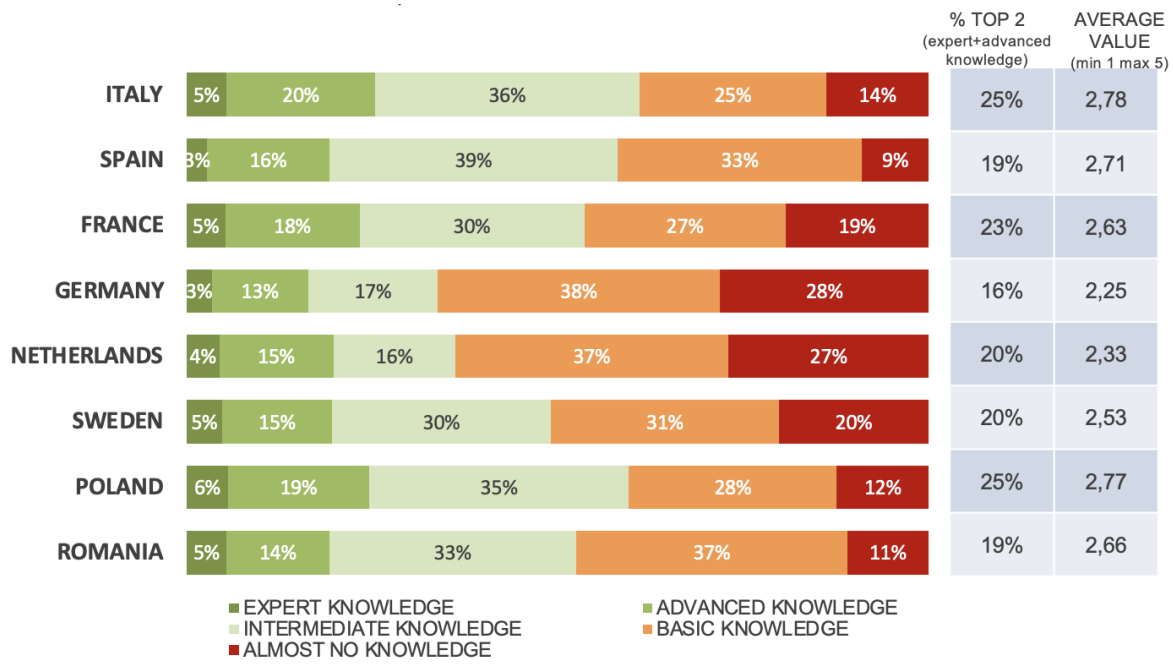

Figure 2: Self-assessed competency on AI - Question: "When it comes to Artificial Intelligence (AI) and its impact on society, I feel my competency on the subject would be: expert knowledge / intermediate / Almost no knowledge / advanced / basic knowledge"

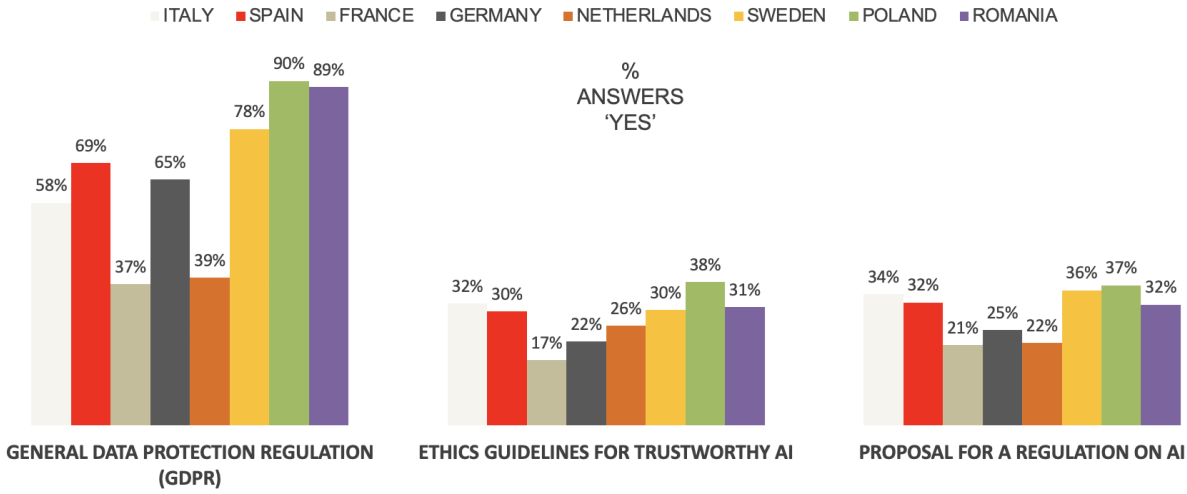

Figure 3: Knowledge of European initiatives regarding AI - Question: "Have you ever heard about the following European initiatives regarding AI? Yes/No"

AI Attitude

In general, European citizens seem largely in favour of the use of AI (60%) in comparison to a lower percentage of the population disapproving its use (10%), and a resulting ratio of approving-to-opposing citizens is close to 6. Approval varies by sector, surpassing 7 for the manufacturing and environmental sectors, and dipping below 4 for human resources, military and transportation. To better understand the actual approval between different AI systems, during the interview the participants were invited to give their opinion also on two case studies. The first regards an AI system

used for recruitment process purposes (scenario 1), while the second presents a smart meter used to improve the home's energy consumption (scenario 2). The results demonstrate that participants are more comfortable using a smart meter (59%) in comparison with a recruiting process based on AI (45%). Differences are found also among the countries involved. In general, the Netherlands, Germany and France are the countries more sceptical in the usage of AI in both scenarios, while Romania is quite comfortable with its adoption.

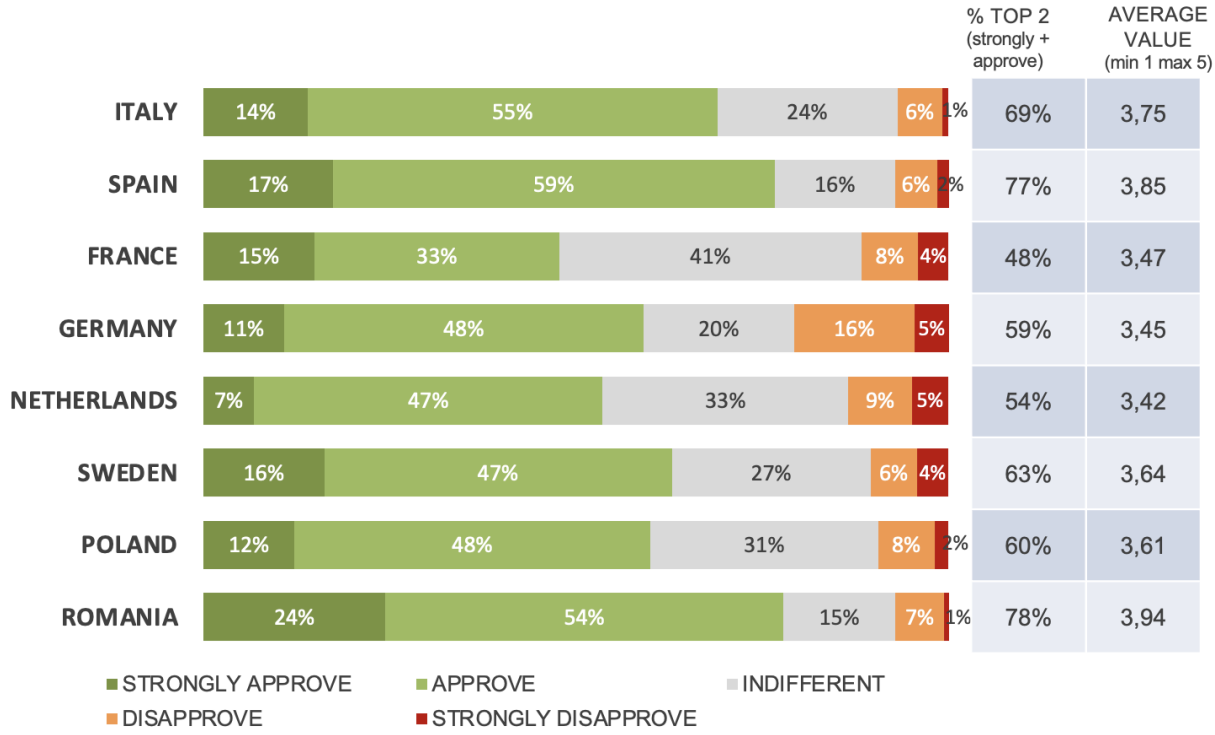

Figure 4: General attitude towards AI - Question: "How would you describe your attitude towards Artificial Intelligence (AI) and its applications? Strongly approve / approve / indifferent / disapprove / strongly disapprove"

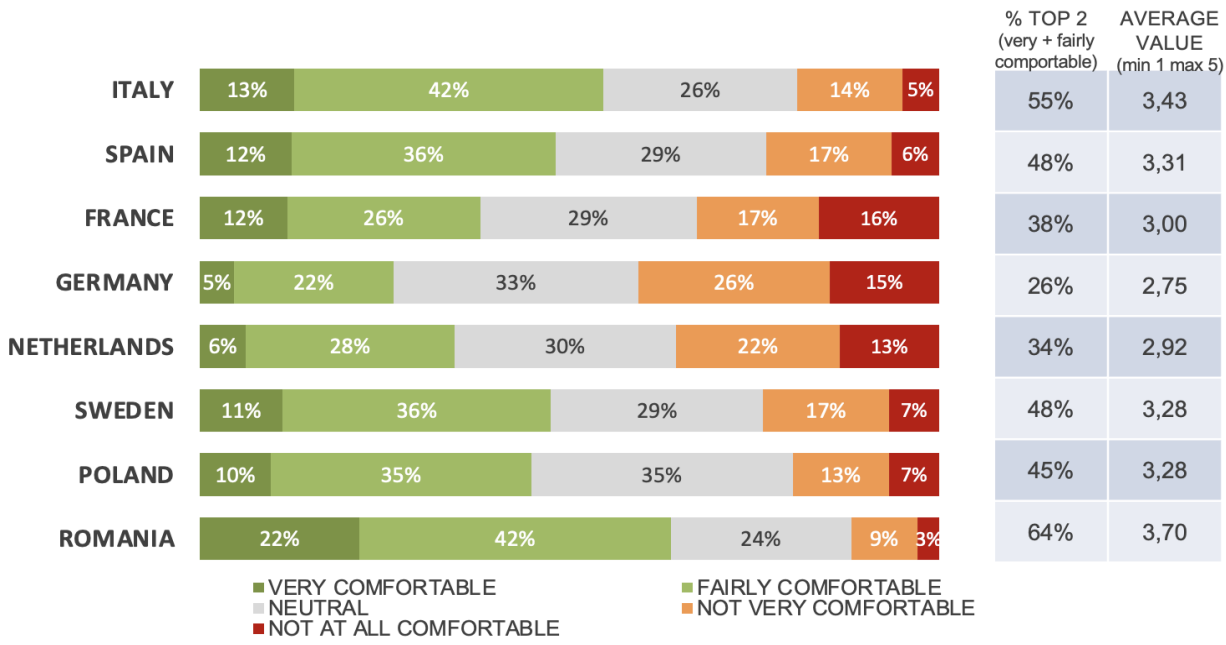

Figure 5: Comfortability with scenario 1 (AI used in job application) - For a full description of the scenario see question 9 of the citizen questionnaire in appendix

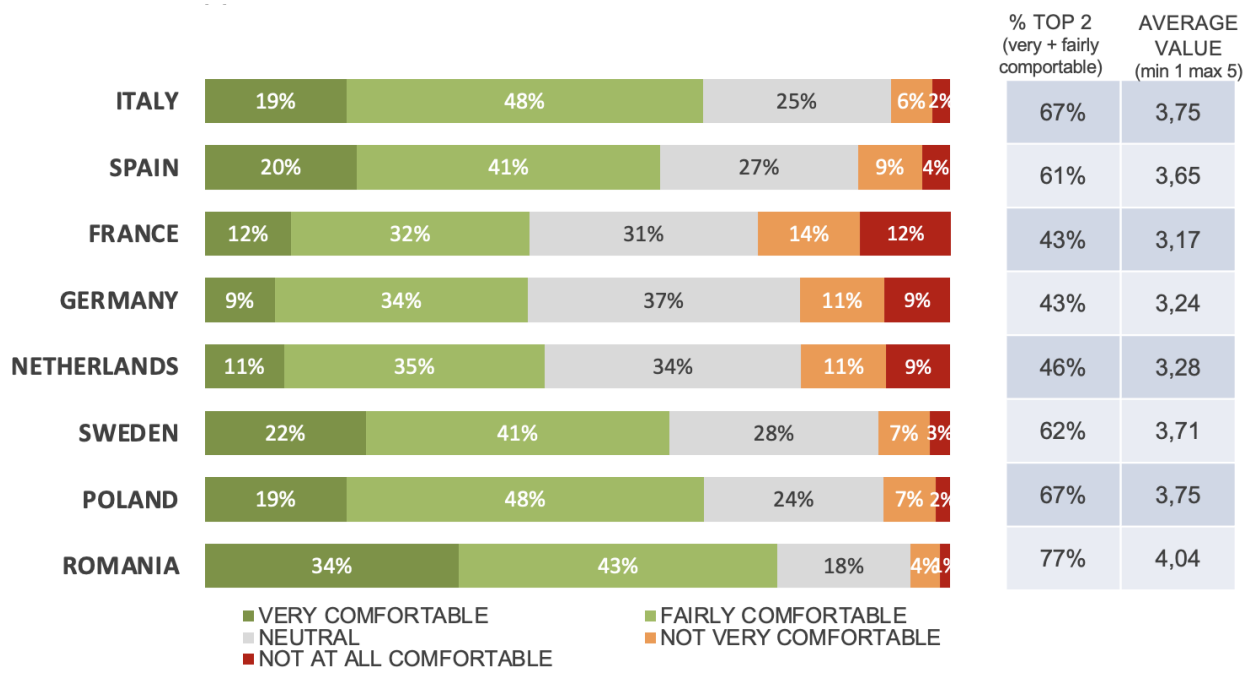

Figure 6: Comfortability with scenario 2 (AI used for energy consumption) - For a full description of the scenario see question 10 of the citizen questionnaire in appendix

Trust & AI

Among the seven ethical principles suggested by the HLEG privacy and data protection are considered as the aspects that should be prioritized to achieve Trustworthy AI. Less considered dimensions include the social and environmental impact of AI. To ensure the correct application of AI, participants report highest trust levels in universities and research centers (over 70% in Italy, Spain or Romania) and the lowest ones in social media companies (36,5%). Surprisingly, national governments and the EU (including the Commission and the Parliament) are trusted as much as private tech companies (about 54%).

All countries agree on the importance of an adequate education in AI, with 72% of approval. However, numbers slightly drop to 62% when considering the commitment to attending free educational courses on AI. Among the countries with higher enthusiasm for this opportunity there are Romania, Italy and Spain (81-74% of respondents interested to attend the course). While in other countries such as the Netherlands and Germany less than 50% of the respondents report to be interested in this opportunity.

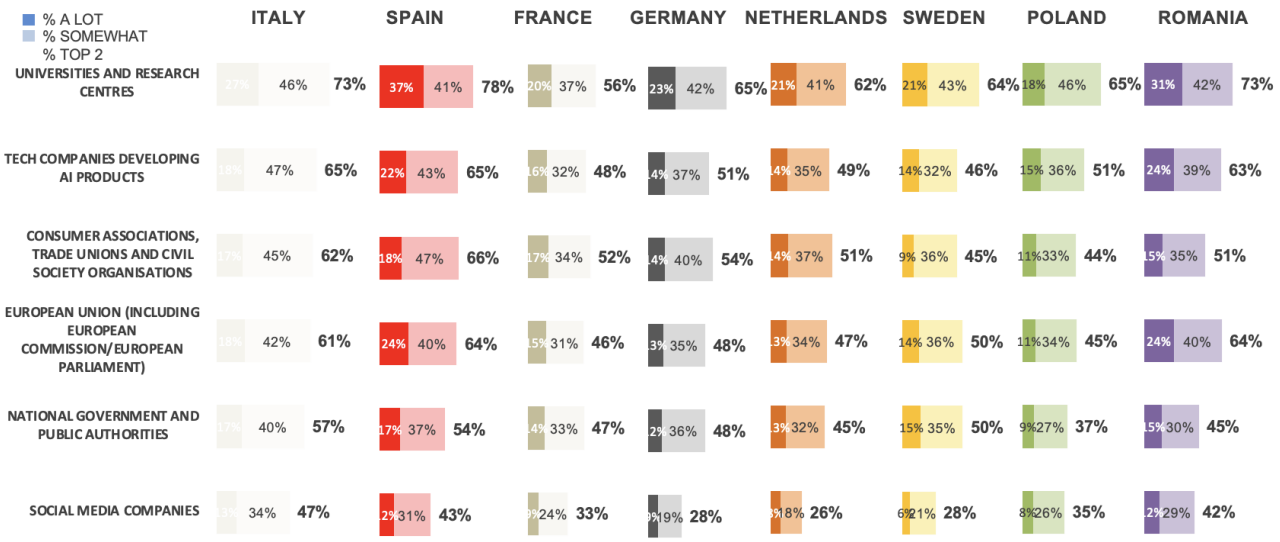

Figure 7: Trust in entities ensuring beneficial AI - "Question: How much do you trust the following entities in ensuring that AI is in the best interest of the public? A lot/Somewhat/So and so/Not so much/Not at all"

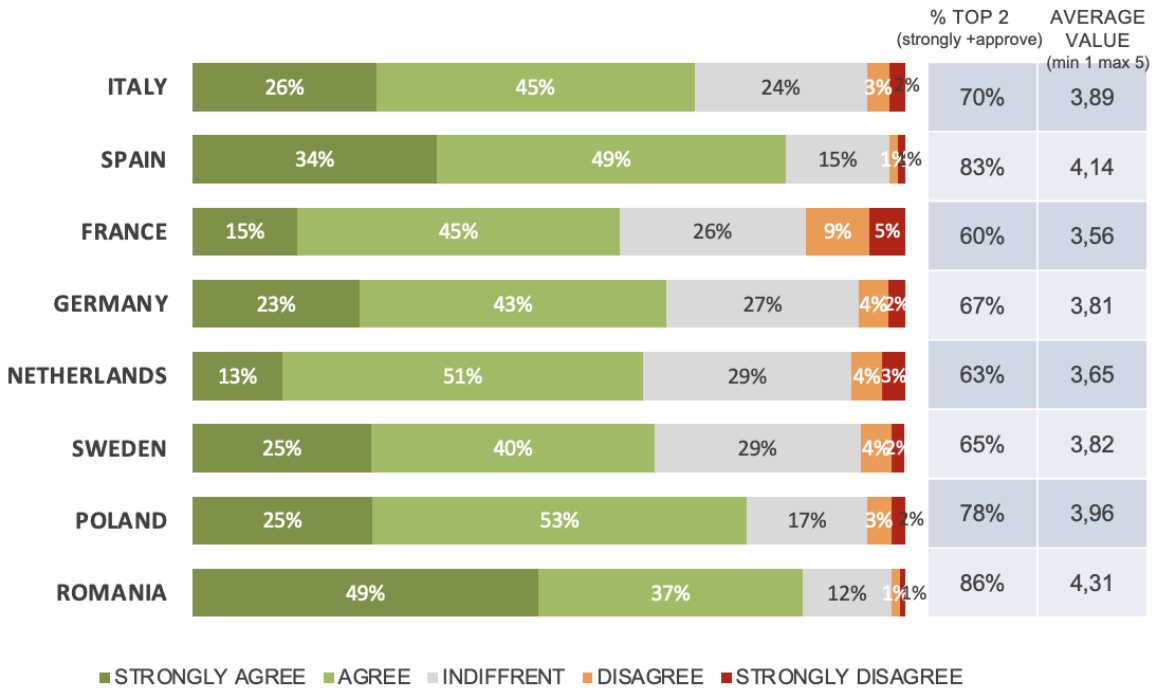

Figure 8: Importance of education to increase trust in AI - "Question: To what extent do you agree that having a better education on what AI is, as well as its current and future uses, would improve your trust in it? Strongly agree/agree/indifferent/disagree/strongly disagree"

In short, this survey contributes to the analysis of people's opinions about AI and its ethical impact within Europe. In particular, with the discussion of the collected results we want to offer insights on key factors that can deeply influence the development and uptake of AI across Europe, such as citizens' awareness and trust. In addition, the questionnaire provides citizens with new stimuli to reflect upon their interaction with new technologies and its possible impact making them more aware and curious towards AI and digital tools.

The rapid transformations introduced by AI into our life solicit a greater consideration of people's concerns and views by organizations involved in the development and deployment of AI systems. We believe that similar reflections are important to foster sustainable innovation and get closer to the Human-centric approach that the EU wishes to achieve.

4.2 Expert consultation

In the road towards Trustworthy AI, a major challenge is to take concrete actions to move from principles to practice and make the ideal of trustworthiness a reality. Not surprisingly, experts take a different stance on what building Trustworthy AI means and propose different strategies to achieve it. Some believe that the development of Trustworthy AI rests, first and foremost, on engineering ethical principles, as, for instance, software toolkits that can help the scrutiny of particular ethical requirements. Other more systematic approaches are concerned with the design of artificial moral agents, such as machine ethics[22]. Other experts hold that a purely engineering approach suffers from severe limitations. For example, Arvan[23] argued that existing methods to programming ethical AI are either too semantically strict, too semantically flexible or overly unpredictable.

While the debate between different conceptualizations of Trustworthy AI goes on in the background, the AI ethics community at large (academia, non-profit organisations, companies, etc.) has provided a multitude of methods and tools which vary in the strategy adopted and the purpose they want to achieve[24]. The abundance of methods, policy options and the recent EU proposal for a regulation[25] have added a layer of complexity to the debate about Trustworthy AI introducing further considerations with respect to firms and authorities, among others.

In the recent past, several surveys relating to AI and its social and ethical issues were launched. For example, Muller and Bostrom[26] examined experts' predictions on the development of high-level machine intelligence and the associated risk for humanity in the coming decades. Grace et al.[27] asked machine learning experts about their predictions on the progress in AI, with the aim to connect policymakers with the opinion of researchers. In 2020, after the publication of the White Paper on AI, the Commission made available a public consultation addressed to AI practitioners, public and private sectors, SMEs, academia and citizens. The aim was to collect feedback on the upcoming policy options presented in the document. The Ad-Hoc Committee on AI[28] launched a multi-

[22] Anderson, M. and Anderson, S. (2011), *Machine Ethics*, Cambridge University Press, https://doi.org/10.1017/CBO9780511978036

[23] Arvan, Marcus. Mental time-travel, semantic flexibility, and A.I. ethics. AI & Society, May 2018, https://link.springer.com/article/10.1007/s00146-018-0848-2

[24] For an extensive review see: Morley, J., Floridi, L., Kinsey, L. *et al.* From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics* 26, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5

[25] EC: European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM/2021/206 final (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[26] Müller, Vincent C. & Bostrom, Nick (2016). *Future progress in artificial intelligence: A survey of expert opinion.* In Vincent Müller (ed.), Fundamental Issues of Artificial Intelligence_. Springer. pp. 553-571.

[27] Grace, K., Salvatier, J., Dafoe, A., Zhang, B., Evans, O. (2018), Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts, Journal of Artificial Intelligence. https://doi.org/10.1613/jair.1.11222

[28] CAHAI: Ad-Hoc Committee on AI (2021), *Analysis of the Multi-stakeholder Consultation*, https://rm.coe.int/cahai-2021-07-analysis-msc-23-06-21-2749-8656-4611-v-1/1680a2f228

stakeholder consultation to identify the key elements of the legal framework based on the Council of Europe's standards on human rights, democracy and the rule of law. The CLAIRE association prepared a survey for the AI community and general audience regarding the proposal for regulation on AI. These two last initiatives share a main interest in the regulatory process of AI and are addressed to a wide spectrum of participants.

The expert questionnaire presented here is addressed to a broad range of experts who deal with Trustworthy AI from different angles and fields of study such as Computer Science, Engineering, Philosophy, Law and Political Science. The questionnaire targets only AI experts from different fields of study, aiming to understand their vision on the notion of Trustworthy AI as well as their familiarisation with the existing methods to implement ethical requirements into practice. It is up to date with the latest actions taken by the EU Commission on defining the regulatory framework for AI, but also includes broader aspects related to the culture of AI and ethics in Europe. Hence, the expert consultation aimed to contribute to the debate around the concept of trustworthiness and governance of AI and, in particular, to dig into the following themes:

1. *General approach to Trustworthy AI*: this set of questions aims to investigate experts' opinion about different approaches to the notion of Trustworthy AI. For example, is this a purely technical concept or are non-technical methods also important? In addition, this part addresses two important components of the European ethics guidelines, i.e. interdisciplinary work and stakeholder participation.
2. *The implementation of Trustworthy AI:* these questions explore the experience and best practices in the field. This includes, for example, the evaluation of feasibility in achieving AI principles (AI HLEG, 2018), the level of confidence with existing methodologies / tools and experts' opinion on the utility and relevance of such tools. Also, there are questions related to the importance of promoting collective discussion and ethical reflection among engineers and computer scientists.
3. *The governance of AI:* this group of questions aims to gain information and opinions about governance mechanisms to achieve Trustworthy AI. This includes, among others, the evaluation of the Proposal for a Regulation on AI and opinions on soft law mechanisms.

The questionnaire was distributed across different channels (social media, AI4EU community and platform, European AI networks, etc...) to reach a large sample of experts including those working in other AI-related H2020 projects. The questionnaire is available online at this web address: https://www.consultationai4eu.eu/ and some screenshots of the website are presented in Figure 9 and Figure 10.

Figure 9: Screenshot of the expert questionnaire homepage

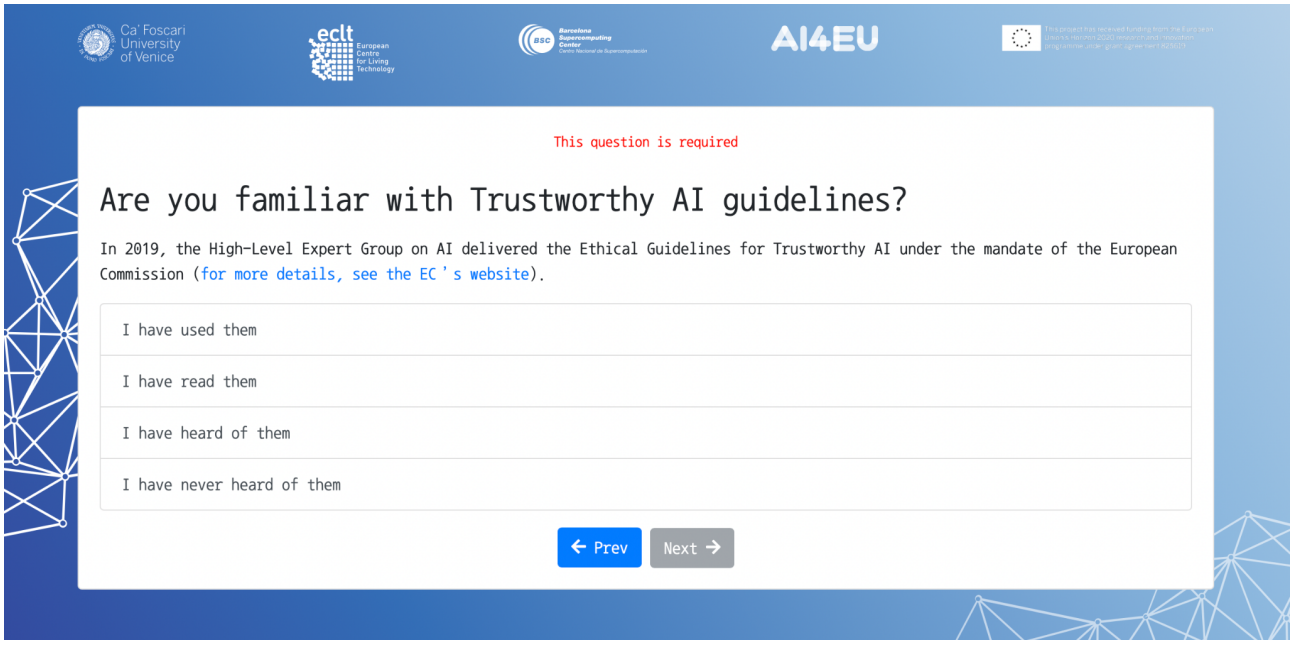

Figure 10: Screenshot of a question of the expert consultation

At the present time the questionnaire attracted 254 respondents but only half of these completed the full questionnaire. The OSAI team is analysing the data collected for a scientific publication. A preliminary analysis suggests that about half of the respondents (56%) claim to be familiar with the HLEG AI's guidelines for Trustworthy AI and the 20% of these state that they have used them (see the first chart in Figure 11). There is nevertheless 44% of our respondents who expressed less confidence claiming that they have only heard of or never heard of the guidelines (the last group represents the 20% of the population). On what concerns stakeholder participation, the 52% of the respondents claim that it never occurs or is little and this result may suggest further work to do in the

field. 75% of the respondents (strongly) agree that there should be more time and space allocated for reflection and collective discussions on responsible AI.

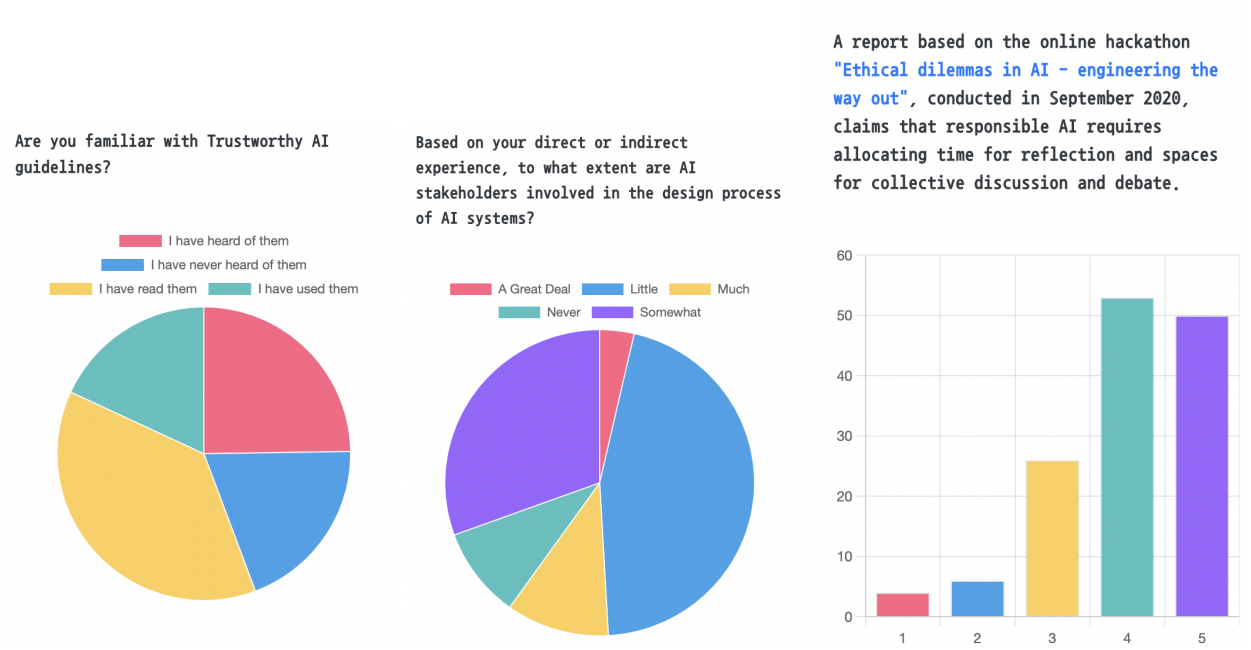

Figure 11: Charts reporting answers to questions n. 1, 3a (based on a 5-point scale where 1= I do not know and 5 = to a great extent) and 9a (based on a 5-point scale where 1=strongly disagree and 5 = strongly agree)

In Figure 12 we report further data collecting opinions about the governance of AI. In particular, with respect to the European Commission's proposal for an AI regulation we observed similar proportions of answers found in the first question (regarding the Ethics Guidelines), with 57% of the population claiming that they read carefully or part of the proposal. On a specific measure introduced in the proposal for a regulatory framework of AI (*i.e.* an EU database of stand-alone AI systems), half of our respondents (strongly) agree with this action but the 30% is undecided. When it comes to assess the contribution of soft law mechanisms (such as no-binding principles) the respondents express law confidence in the efficacy of such methods since 58% of the population say these would contribute not at all or not too much.

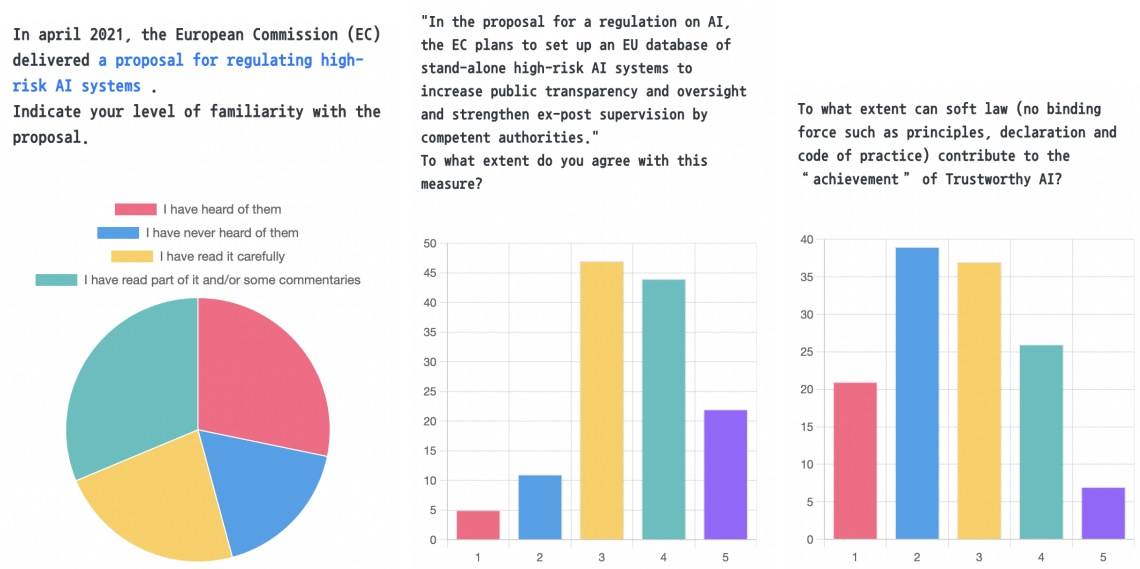

Figure 12: Charts reporting answers to questions n. 11, 12 (based on a 5-point scale where 1=strongly disagree and 5 = strongly agree) and 13 (based on a 5-point scale where 1= I do not know and 5 = to a great extent)

5. Workshops

The OSAI team organized three workshops between 2020-2021. These initiatives served to support and organize the activities of working groups and open the discussion of ELSEC AI issues to the broader community. The following sections provide brief descriptions of the contents addressed and the results achieved during these events.

Online Workshop "Trustworthy AI made in Europe: From Principle to Practice" (November 2020)

On November 13th 2020, The AI4EU Observatory on Society and AI organised an online workshop on "Trustworthy AI made in Europe: from Principles to Practices". The workshop saw an active participation of a total of 184 participants, 70 through the Zoom platform (out of 175 registrations received) and 156 watching live from Youtube.

The event focused on current ethical, legal, and technical challenges raised in the design and deployment of AI systems and the impact these have over society. Discussions and reflections on each panel session put forward how these challenges should be addressed to be aligned with the European human-centric approach of AI. The workshop included speakers from different backgrounds and sectors (see Figure 13 and Figure 14 with the poster and the agenda of the event). A summary of the results achieved by the workshop with main takeaways of keynote talks and panel discussions is offered in an OSAI article[29].

[29] "Lessons learnt on Trustworthy AI made in Europe": https://www.ai4europe.eu/node/319

Figure 13: Poster of the Workshop in November 2020

Figure 14: Agenda of the Workshop in November 2020

Hybrid Workshop "Archives and AI: Coping with Climate Change" (May 2021)

The workshop was held on 22-24 May 2021 in a dual format some participants attended in person, others joined the event online. The physical meeting took place in San Servolo Island (Venice, Italy), at the Venice International University. The event has been organized in collaboration with other European initiatives, in particular the EU FET project MUHAI, the EU FET project ODYCCEUS and Science Gallery Venice. The workshop used the recently created Aqua Granda Digital Community Memory[30] as a source of concrete case studies and thus addressed opportunities as well as theoretical and practical issues in the use of AI for unlocking archives.

Hybrid Workshop on "The Culture of Trustworthy AI (September 2021)

The workshop on "Trustworthy AI: Public Debate, Education and Practical Learning" was successfully held in Venice on 2-3 Sept 2021 gathering 40 physical participants and 40+ online attendees joining the event through the zoom platform and the Youtube channel of the European Centre for Living Technology (ECLT).

The event had three keynote speakers of Prof Raja Chatila, who presented hot AI issues at the venue, while Prof Mireille Hildebrandt and Prof Luc Steels delivered their speeches on key AI definitions & AI thought leaderships over live streams that went smooth without much technical glitch.

The AI4EU Working Groups (WGs), drawing volunteer researchers and professionals in AI from various backgrounds, presented their research findings and progress so far in their own pilots and case studies. The WGs were a truly multidisciplinary approach in dealing with complex research issues (some WG participants attending the event are presented in fFgure 15)

Figure 15: Some WG members attending the AI4EU workshop in September 2021

[30] To consult the archive on "Aqua Granda" see the website http://aquagrandainvenice.it/it/welcome

The workshop also listened to nine contributed talks and three panel discussions on the ethical and legal challenges of European AI, on the governance of AI and on ethics and AI education. Representatives of different European projects including AI4Media, Tailor, Humane-AI net and StairwAI demonstrated their research priorities and focused in the panel on the ethical and legal challenges of AI in Europe.

On the governance of AI, scholars, and professionals from CAHAI, fAIr LAC, A+ Alliance for Inclusive Algorithms, African Digital Rights Hub shared their opinions on the ongoing priorities and relevant AI issues in Europe, Africa, and Latin America.

Presenters from Umea University, Ethics of AI MOOC (Finland) and Embedded EthiCS at Harvard (USA) shared their experiences and best practices at the AI and ethics in higher education panel.

Recordings of the event with links to each presentation are available on the ECLT YouTube channel (see day 1 and day 2)

The event allowed the AI4EU WGs members to meet in person after 10 months of tight collaboration online, people from different EU projects and initiatives to know each other, strengthen relations and discuss possible future collaboration.

For example, we would like to highlight interesting activities such as AI & Equality toolbox[31] and the ethical and legal framework for AI applications in the public sector under development by the EU project ETAPAS[32].

The workshop was organized by the European Centre for Living Technology at Università Ca' Foscari University of Venice, in collaboration with the Barcelona Supercomputing Center, Universitat Politècnica de Catalunya and Umea University (see the poster of the event in Figure 16).

Figure 16: The poster and the agenda of the event

[31] For more details visit the website of the initiative: https://aiequalitytoolbox.com

[32] More information are available on the website of the project: https://www.etapasproject.eu

6. Conclusions

The Observatory on Society and AI was born with an ambitious objective to become a reference at European level as a center for learning and debate around Trustworthy AI in Europe. The creation and curation of the Observatory has been a continuous learning process, having to adapt to the evolution of the conceptual design and development of the AI4EU platform.

During its period of activity, the Observatory has reached two significant outcomes. On the one hand, the team has created a multidisciplinary network of European experts interested in different topics related to the notion of Trustworthy AI and its implementation. As a result, the working groups have been working during the last year in an environment open to dialogue and debate and have also been able to generate tangible outcomes like the two consultations. On the other hand, and despite the difficulties of the last years due to the Covid-19 situation, the Observatory has organised different events able to attract AI stakeholders and discuss relevant topics such as the implementation of ethical guidelines, the regulation of AI or the strategies to promote education and literacy of AI and ethics.

Unfortunately, the Observatory has not achieved the expected results regarding the creation of content in the platform. Most of the published content came from solicited contributions, under the request of the OSAI team, and the number of spontaneous inputs was limited. Factors such as the dependence and limitations of the design during the first version of the platform and the interruptions on the editorial process during the migration of the platform have certainly affected the user engagement. The Observatory team has also learned from the experience and gathered new ideas and motivations for a sustainability plan that is defined in D5.5 “The AI4EU Observatory as a service”.

The coordination of the Observatory and the working groups have enabled the interaction with experts of different domains with a common interest: the promotion and implementation of responsible practices for the research and development of AI made in Europe. The workshops brought the opportunity to learn from initiatives coming from research centers, start-ups and high-level organisations such as the OECD or AI Watch. We have observed a general trend towards producing assessment tools for AI systems, all based on similar guidelines based on European values, fundamental rights and bioethical principles. In addition, the AI community is contributing to progress on the notion of transparency, putting efforts in the research of explainable methods applied to different applications and sectors.

Our preliminary overview to the expert consultation shows that there is no consensus among AI practitioners on which is the best method (technical or non-technical) to implement responsible AI. At the same time, results show that there is still a significant part of the community that is not enough familiarised with the AI Act nor have a strong position on whether AI should be enforced through regulation or monitored via soft law. There is a need of cooperation among AI stakeholders to raise awareness towards the ELSEC aspects of AI and create a literacy to adopt a new way of design, develop, use and evaluate AI systems.

Similarly, we have observed a general agreement among citizens to promote educational material and training of AI to improve the adoption towards the technology and increase trustworthiness. Indeed, we believe that transparency and trustworthiness will be enhanced by including the society and domain experts in the AI system life cycle to identify and mitigate biases, understand user requirements and adapt the capabilities of the technology to the normative environment at European, national and sectoral level.

Annex1: Publications

1. Scantamburlo T., Cortés A., and Schacht M., *Progressing Towards Responsible AI*, ECAI 2020 Workshop on Evaluating Progress in AI (2020). arXiv:2008.07326
2. Scantamburlo, T., Cortés, A., Dewitte, P., Van der Eycken, D., Billa, V., Duysburgh, P., Laenens, W., *Covid-19 and Contact Tracing Apps: A review under the European Legal Framework*. https://arxiv.org/abs/2004.14665
3. Scantamburlo, T., Cortés, A., Dewitte, P., Van der Eycken, D., De Wolf, R., & Martens, M. (2021). *Covid-19 and tracing methodologies : a lesson for the future society.* HEALTH AND TECHNOLOGY. https://doi.org/10.1007/s12553-021-00575-1
4. Scantamburlo, T. Non-empirical problems in fair machine learning. *Ethics Inf Technol*(2021). https://doi.org/10.1007/s10676-021-09608-9 (open access)
5. Foffano, F., Scantamburlo, T., Cortés, A., Bissolo, C. *European Strategy on AI: Are we truly fostering social good?* IJCAI 2021 Workshop on AI for Social Good (2020) https://crcs.seas.harvard.edu/publications/european-strategy-ai-are-we-truly-fostering-social-good
6. Cortés U, Cortés A, Barrué C, Sánchez A, Moya-Sánchez EU, Garcia-Gasulla D. *To Be fAIr or Not to Be: Using AI for the Good of Citizens.* IEEE Technology and Society Magazine, vol. 40, no. 1, pp. 55-70, March 2021, doi: 10.1109/MTS.2021.3056173 (2021).
7. Cortés U, Cortés A, Pérez R, Álvarez-Napagao S, Garcia-Gasulla D, *The ethical use of high-performance computing and artificial intelligence: fighting COVID-19 at Barcelona Supercomputing Center.* AI & Ethics Journal, 10.1007/s43681-021-00056-1 (2021)
8. Garcia-Gasulla D, Cortés A, Álvarez-Napagao, Cortés U, *Signs for Ethical AI: A Route Towards Transparency*. https://arxiv.org/abs/2009.13871 (2020)
9. Foffano F, Scantamburlo T, Cortés A,. *Investing in AI for Social Good: An analysis of European National Strategies* (submitted at AI & Society for the special issues on AI for people, 2021).
10. Cortés U, Cortés A, Gibert K. *The use of Artificial Intelligence for Citizen Services and Government. GAVIUS: a case in Catalonia* (submitted at AI & Society for the special issues on AI for people, 2021).
11. Scantamburlo T. and Grandi G., *Apprendimento Automatico e Decisione Umana* in Fossa F., Schiaffonati V., Tamburrini G. "Automi e Persone", Carocci editore 2021

Annex 2: Communication and dissemination

7.1 Organization of seminars and events

The observatory organized a number of seminars (physical and online) for the research community and a book presentation open to the large public. The events include:

29 March 2019
Hey, Merry Men! Robin-Hood Artificial Intelligence is Calling You! by Fabio Massimo Zanzotto (University of Rome Tor Vergata), DAIS, Ca'Foscari University, Venice

09 October 2019
Gradient Institute: the Science and Practice of Ethical AI by Tiberio Caetano (Gradient Institute, University of New South Wales), ECLT, Ca'Foscari University, Venice

11 November 2019
Book presentation: "En attendant les robots" by Antonio Casilli (Paris School of Telecommunications), ECLT, Ca'Foscari University, Venice

9 December 2019
Is technology neutral? by Silvia Crafa (University of Padova), ECLT, Ca'Foscari University, Venice

16 December 2019
Reproducibility in Artificial Intelligence: Experimentation in the Artificial by Viola Schiaffonati (Politecnico di Milano), ECLT, Ca'Foscari University, Venice

29 April 2020
Online seminar: The A4EU Observatory on Society and AI by Teresa Scantamburlo (ECLT, Ca'Foscari University)

29 June 2020
Online Special Session: “Fairness in Algorithms” organized in collaboration with the Commission for the History and Philosophy of Computing, in the context of Computability in Europe 2020

2 July 2020
Online seminar: Z-Inspection: A Holistic Analytic Process to Assess Ethical AI by Roberto Zicari (Goethe University Frankfurt)

12 November 2020
AI4EU working group on Ethical and Legal AI - (online) launch event

13 November 2020
AI4EU Workshop “Trustworthy AI Made in Europe: From Principles to Practice” (online event)

24 May 2021
AI4EU-MUHAI-ODYCCEUS Workshop “Archives and AI: Coping with Climate Change” (hybrid event)

2-3 September 2021
AI4EU Workshop “The Culture of Trustworthy AI. Public Debate, Education and Practical Learning, 2-3 September 2021, Island of San Servolo, Venice (hybrid event)

7.2 Participation and presentations at conferences and workshops

Dagstuhl Seminar, “Ethics and Trust: Principles, Verification and Validation” (Dagstuhl Seminar 19171), 22-26 April 2019, Dagstuhl, Germany

Invited presentation, “The AI4EU Observatory on Society and AI”, 4th European Conference on AI in Finance and Industry, ZHAW, Winterthur, Switzerland, 5 September 2019

Invited presentation, workshop “La responsabilità delle macchine", University of Padova & Associazione Giovani Avvocati - sezione Treviso, 9 October 2019, Treviso

Speaker in “AI Governance Forum”. June 8th 2020 (virtual event).

Co-organisers and speaker in the “AI and Human Rights: Ombudsmanship, Challenges, Roles and Tools” workshop. 2nd-3rd March 2020, Barcelona (Spain)

Speaker in “Driving AI in Europe”, parallel workshop at Transfiere 2020. February 12th 2020, Málaga (Spain)

Luc Steels, Online symposium “Empathic AI. Art shapes industry”, 2 July 2020

Co-organisers of the EU Challenges panels in ECAI 2020 “H2020 came to an end: What is next? The European Strategy for AI” and “Challenges for European Research in AI“, September 2nd and 3rd 2020

Participation in a roundtable “Are Algorithm Sexist?”, 19° International Film Festival and Forum on Human Rights, Geneva, 8 March 2021 (virtual / physical event),

Participation in a roundtable, EU-Canada cooperation workshop on AI: “Equity, diversity and inclusion in Artificial intelligence”, organized by the Europe-Canada Programme Level Cooperation Task Force, 26 April 2021 (virtual event),

Moderator in a Roundtable on AI and education, workshop “Regulation: What we need to talk about when we talk about AI”, 3 June 2021

Participation in a roundtable at the ICML 2021 Workshop on “Deploying and Monitoring ML systems”, panel on open problems – application in the legal systems, virtual event, 23 July 2021,

Invited presentation. “Surveying the Opinion of European Citizens on AI”, *Artificial Intelligence in Industry and Finance*, 6thEuropean COST Conference on Mathematics for Industry, ZHAW, 9 September 2021 – online event

7.3 Dissemination Activities

Participation in AI4EU Webcafé with the topic “Covid-19 contact tracing apps” along with the authors of the aRxiV paper and with participants from the EU funded project HELIOS and guest speaker Steen Rasmussen, from the University of Southern Denmark.

Speaker at the panel discussion on “Etica, algoritmi e IA”, organized by Parole Ostili, 31 May 2019, Trieste

Public seminar “Macchine intelligenti e vita quotidiana”, Caffè della Scienza, organized by Associazione “Mestre mia”, 19 September 2019

Public lecture, “Intelligenza Artificiale: opportunità e rischi”, Scuola socio-politica, diocesi di Milano, 7 February 2020, Villa Cagnola, Gazzada, Varese

Speaker at the panel discussion, “L’UE punta sull’Intelligenza Artificiale”, programma “Modem”, radiotelevisione Svizzera Italiana, 20 February 2020

Public seminar “The AI4EU Observatory on Society and AI” organised by the Catalan Observatory of Ethics and AI (OEIAC), 13 September 2021

Speakers at the panel discussion “The Great Global Data Divide” organised by Science Business, 15 September 2021

● Annex 3: Working Groups participants

Alessandro Fabris is a PhD student with the University of Padua, where he works to make algorithms more accountable and fair. His work focuses on the understanding and mathematical formalization of fairness criteria that are relevant to specific contexts, including search engines and car insurance premiums.

Dr **Atia Cortés** (she/her) is a computer science engineer with a MsC and a PhD in Artificial Intelligence by the Universitat Politècnica de Catalunya. She is currently a post-doctoral researcher at the Social Link Analytics unit of the Barcelona Supercomputer Center, where she is also part of the Bioinfo4Women programme. For over a decade, she has participated in several European and national funded projects related to the design and deployment of AI solutions applied to healthcare. Her main research interest focuses on the ethical and social impact of AI, the assessment of AI, and in particular the identification of sex and gender biases in AI, and the promotion of social awareness and responsible AI practices.

Angeliki Dedopoulou is Senior Manager of EU Public Affairs at Huawei, responsible for the policy area of Artificial Intelligence, Blockchain, Digital Skills and Green-related policy topics. Before joining Huawei’s EU Public Affairs team, she was an adviser for the European Commission for over 5 years (through everis, an NTT Data Company) on DG Employment, Social Affairs and Inclusion. Her main focus during this period was the European Classification of Skills, Competences, Qualifications and Occupations (ESCO) and the Europass Digital Credential project. Ms Dedopoulou is a Member of the Board of the Hellenic Blockchain Hub and a Member of the Beltug Blockchain Taskforce. She studied Political Science and History in Greece, Sociology in France and European Governance in Luxembourg. She also regularly writes articles and has travelled across Europe delivering speeches to policymakers, governments and industry summits, on topics ranging from the digital labour market to Blockchain in education and employment.

Ana Chubinidze is a founder and CEO of Adalan AI, consulting firm on AI Governance and Policy; also founder and director of non-profit organization AI Governance International. She is an invited founding editorial board member of Springer Nature’s AI and Ethics journal and member of the European AI Alliance. She often speaks at AI forums and conferences internationally and contributes to the work of several AI-related associations.

Dr. **Andrea Aler Tubella** (PhD, Computer Science, female) is a Senior Research Engineer at Umeå University with focus on the the design of formalisms and systems, and their applications as tools for the responsible design and monitoring of intelligent systems. Her research expertise includes formal logic, proof theory, as well as the use of logical modeling to describe reasoning and behaviour and its applications in AI.

Bárbara Urban Gonzalez (Castelló de la Plana, 1981) is a Spanish researcher whose works are oriented towards the relationship between robotics and human beings.She graduated in Social and Cultural Anthropology (UNED) and Master in Ethics and Democracy (UJI). At this moment, she is finishing out her doctoral thesis on Roboethics.She is a lecturer at the National Distance Education University and collaborates with the Jaume I University.Her publications and her participation in congresses have tried to highlight the need to investigate the coexistence between humans and robots, especially in relation to transhumanism and the cyborg phenomenon.

Christoph Heitz is professor at School of Engineering, Zurich University of Applied Sciences, Switzerland. He has been working in the field of data-based decision making, developing approaches and algorithms that harvest data for improving business processes, customer interaction, and service co-creation. In the last years, he has been heavily engaged in developing new approaches for addressing ethical challenges of commercial data-based value creation. He is one of the authors of the “Code of Ethics for Data-Based Value Creation” which has been developed in a joint effort of Swiss companies and universities, for supporting companies in creating ethical data-based business. He also leads several research projects on algorithmic fairness (e.g. https://fair-ai.ch/).

Dario Garcia-Gasulla is a senior researcher at the Barcelona Supercomputing Center. He leads research on the High Performance Artificial Intelligence group, in topics such as deep neural representations and AI for medical imaging. He coordinates and teaches the Deep Learning course at the Master's on AI offered by the UPC, UB and URV universities. Occasionally he contributes to fields like characterization of misinformation, and transparent and accessible AI.

Evert F. Stamhuis (LLM, PhD) holds a chair for Law and Innovation at Erasmus School of Law since 2017 and is Senior Fellow of the Jean Monnet Centre of Excellence on Digital Governance. Previously he held a chair in criminal law and procedure at the Open University (NL). His research is on the interaction between law, governance and new technologies, with a special focus on the public domain, health care and regulated markets. As a researcher Stamhuis is affiliated to the International Centre for Financial Law & Governance, the Centre for Law and Economics of Cybersecurity and the Erasmus Initiative Dynamics of Inclusive Prosperity. Other current affiliations are the University of Aruba and the Court of Appeal of ‘s Hertogenbosch (NL).

Fabio Fossa (PhD, University of Pisa) is a researcher at the Department of Mechanical Engineering of the Politecnico di Milano. His main research areas are applied ethics, philosophy of technology, robot and AI ethics, and the philosophy of Hans Jonas. His current research deals with the philosophy of artificial agency and the ethics of autonomous driving. He is Editor-In-Chief of InCircolo – Rivista di filosofia e culture, a steering committee member of the META Research Group, and a founding member of the Zetesis Research Group.

Francesca Foffano is a researcher at the European Centre for Living Technology, Ca' Foscari University of Venice working at the AI4EU project. She holds a Master in Human-Computer Interaction at the University of Trento and previously she obtained her Bachelor in Psychology at the University of Padua. During her studies, she collaborates with the CADIA research centre at Reykjavik University and in the industry. Her research interest focuses on the user' understanding and perception of AI, social and ethical influences, and a definition of more human-centric design approaches.

Joris Krijger works as an Ethics & AI specialist at the Dutch bank de Volksbank while also holding a PhD position at the Erasmus University Rotterdam on Ethics & AI. He has a background in Philosophy, Economic Psychology and Media Studies. During his studies Joris was awarded a Dutch national prize for both his high-tech startup Condi Food (Rabobank Wijffels Innovation Award 2014) as well as for his Philosophy thesis on technology, ethics, and the financial crisis of 2008 (Royal Holland Society of Sciences and Humanities, 2017). He presently works on bridging the gap between principles and practice in AI Ethics by studying the operationalization of ethical principles from an academic and practical perspective. Additionally, Joris holds positions as a.o. Advisory Board Member at the Frankfurt Big Data Lab, Subject Matter Expert for CertNexus' 'Certified Ethical Emerging Technologist' and Founding Editorial Board Member of Springer Nature's AI and Ethics Journal.

Long Pham is the Community Manager of AI4EU, a €20M project that won funding from the European Union's Horizon 2020 research and innovation program. She manages regular communications with a community of 400+ members from the 80 project partners, 5000+ users on the AI4EU Platform, nearly 10K followers on AI4EU social media channels. She supports dissemination activities and ecosystem development of European AI via collaborations with a series of European AI initiatives and winning projects. In her research, Long focuses on citizen engagement aspects of smart city programs, local policy development, and policy and regulation for technology adoption in the development of smart and sustainable cities.

Manuela Battaglini is a specialist in strategic digital marketing, a law graduate and an independent researcher studying the social impact of automated decision-making processes and personal profiling. She works on Digital Ethics (data ethics, security ethics, algorithm ethics and ethics in practice) She is also CEO of Transparent Internet, a consulting firm that helps organizations make their AI systems ethical, transparent and trustworthy. Due to her research activity, Manuela Battaglini was called by the Spanish Government, together with another governmentally appointed group of experts, she was called to help define the Spanish Charter of Digital Rights, where she leads the 'Ethical Considerations' working group.

Dr. **Ricardo Vinuesa** is an Associate Professor at the Department of Engineering Mechanics, at KTH Royal Institute of Technology in Stockholm. He is also a Researcher at the AI Sustainability Center in Stockholm and he is Vice Director of the KTH Digitalization Platform. He received his PhD in Mechanical and Aerospace Engineering from the Illinois Institute of Technology in Chicago. His research combines numerical simulations and data-driven methods to understand and model complex wall-bounded turbulent flows, such as the boundary layers developing around wings,

obstacles, or the flow through ducted geometries. Dr. Vinuesa's research is funded by the Swedish Research Council (VR) and the Swedish e-Science Research Centre (SeRC). He has also received the Göran Gustafsson Award for Young Researchers. Research Group Web: www.vinuesalab.com

Risto Uuk is a PhD Researcher in Economics at Tallinn University of Technology focusing on the impact of AI on the labor market. In addition, Project Manager at the World Economic Forum's Global AI Council working on a white paper putting forward positive visions for a future economy driven by AI. Previously did research on trustworthy AI for the European Commission.

Teresa Scantamburlo is a post-doc researcher at the European Centre for Living Technology, Ca' Foscari University of Venice (Italy) and before that has worked at the University of Bristol (UK). Her main research interests lay at the intersection of Computer Science and Philosophy and include the impact of Artificial Intelligence (AI) on human-decision making, the role of data and algorithms in social regulation, and the ethical assessment of AI systems. She is also interested in studying AI from the point of view of epistemology and the philosophy of science (e.g. some topics of interest include the problem of induction, the problem-solving approach and the notion of progress).

Steven Umbrello currently serves as the Managing Director at the Institute for Ethics and Emerging Technologies. His primary research interests are on value sensitive design (VSD) and its application to transformative technologies like AI, nanotechnology, and industry 4.0 technologies.

Xin Chen is Executive Director European Lead on AI & Data Governance Policy & Standards & Industry Digitization and Corporate Strategy Department at Huawei Technologies He jointed Huawei in 2005 in the UK. Since then He held various leadership roles within Huawei's Carrier Business Group and Enterprise Business Group. At Enterprise BG, he has played a key role in building a significant Enterprise CPE business in the convergent communication sector with some carrier partners and helped to grow the strategic partnership and business with verticals in Europe. He recently joined Huawei's Corporate Strategy Department leading the European standards and policy related activities including industry enablement on AI & Data and Health Care. He has a number of industry engagements including being a member of TechUK AI & Big Data Leadership Committee, AI4EU Trustworthiness & Legal AI WG and Digital Europe AI & Data and eHealth WG. Prior to joining Huawei, he worked in Lucent Bell Lab in the UK (2000) and Fujitsu Laboratory of Europe (2003).He held a BSC in Communication Engineering from Beijing Jiaotong University and a MSC in Data Communications from The University of Sheffield in the UK.

Zahoor ul Islam is currently working as a PhD Student in Responsible Artificial Intelligence group at Umeå University, Sweden. Zahoor received his MS degree from the University of Goteborg, Sweden in Software Engineering and Management, and has been working as a Software Engineer in multiple organizations. His research focuses on addressing and integrating ethical, legal and social values in the design and development life-cycle of AI systems and ensuring that engineering of AI systems is carried out in a responsible manner while complying with established set Software Engineering practices, methodologies, and standards. To know more, visit https://www.umu.se/en/staff/zahoor-ul-islam/.

Ulises Cortés is a Full-Professor and Researcher of the Universitat Politècnica de Catalunya (UPC) since 1982 (tenured since 1988 and habilitated as Full-Professor since 2006) working on several areas of Artificial Intelligence (AI) in the Computer Science (formerly Software Department) including knowledge acquisition for and concept formation in knowledge-based systems, as well as on machine learning and in autonomous intelligent agents.

Luc Steels studied linguistics at the University of Antwerp (Belgium) and computer science at the Massachusetts Institute of Technology (USA). His main research field is Artificial Intelligence covering a wide range of intelligent abilities, including vision, robotic behavior, conceptual representations and language. In 1983 he became a professor of computer science at the University of Brussels (VUB) and in 1996 he founded the Sony Computer Science Laboratory in Paris and became its first director. Currently he is ICREA Research Professor at the Institute for Evolutionary Biology (CSIC,UPF). Steels has been PI in a dozen large-scale European projects and almost 40 PhD theses have been granted under his direction. He has produced over 300 articles and edited 15 books directly related to his research.

● Annex 4. Questionnaire for citizens

Consider a simple definition: Artificial intelligence (AI) refers to computer systems that can perform tasks that usually require intelligence (e.g. making decisions, achieving goals, planning, learning, reasoning, etc.). AI systems can perform these tasks based on objectives set by humans with a few explicit instructions.

1. When it comes to Artificial Intelligence (AI) and its impact on society, I feel my competency on the subject would be: [AI awareness, self-assessed] [AI impact awareness, self-assessed]

- Expert knowledge
- Advanced knowledge
- Intermediate knowledge
- Basic knowledge
- Almost no knowledge

2. How would you describe your attitude towards Artificial Intelligence (AI) and its applications? [AI attitude]

- strongly approve
- approve
- Indifferent
- disapprove
- strongly disapprove

3. To what extent do you feel Artificial Intelligence (AI) and its applications impact your daily life already? [AI impact awareness, self-assessed]

- A lot
- Somewhat

- So and so
- Not so much
- Not at all

4 Have you ever heard about the following European initiatives regarding AI?
[AI awareness, self-assessed]

- General Data Protection Regulation (GDPR) Yes / No
- Ethics Guidelines for Trustworthy AI Yes / No
- Proposal for a Regulation on AI Yes / No

5. How often are you aware of interacting with a product/service based on or including AI? [AI awareness, self-assessed]

- Always
- Often
- Sometimes
- Seldom
- Never
- I don't know

6. Consider the following list of applications. Please select which ones you think may incorporate AI. [AI awareness]

- ride sharing apps (e.g. Uber, Lyft, Blabla car)
- calculators
- contents and products product recommendations (e.g. Youtube, Amazon, Netflix)
- accommodation booking sites (e.g. Tripadvisor, Trivago, Airbnb)
- phone camera
- messaging apps (e.g. WhatsApp, Telegram)
- email spam filters
- search engines (e.g. Google, Bing)
- drones
- social media (e.g. Facebook, Twitter)
- traffic navigation apps (e.g. Google Maps, Waze, TomTom)
- facial recognition apps (e.g. face unlock in phones)
- text editor (e.g. Word, Open Office)
- calendar app (e.g. Google Calendar, iCal)
- internet browser (e.g. Chrome, Firefox)
- teleconferencing app (e.g. Zoom, Skype, Google Meet)
- others_____
- None of the above

7. To what extent do you think AI is used in each of the following sectors in Europe? Please use the five-point scale to plot your answer. [A lot, Somewhat, So and so, Not so much, Not at all] [AI awareness]

- Healthcare (e.g. diagnostic support, personalised medicine)
- Insurance (e.g. fraud detection, personalized risk assessment)
- Agriculture (e.g. robotic harvesting, crop optimization)
- Finance (e.g. fraud detection, loan decision support systems)
- Military (e.g. automated weapons, cybersecurity for data protection)

- Law enforcement (e.g. predictive policing to forecast areas where crime is likely and dispatch police units, face recognition in public places)
- Environmental (e.g. climate prediction, energy harvesting forecast)
- Transportation (e.g. self-driving vehicles)
- Manufacturing industry (e.g demand forecasting, robotics)
- Human resource management (e.g. CV screening, workforce planning)

8. How would you describe your attitude towards the use of AI in the following sectors in Europe? Please use the five-point scale to plot your answer. [5-point scale: I strongly approve it, I approve it, Indifferent, I disapprove it, I strongly disapprove it] [AI attitude]

- Healthcare (e.g. diagnostic support, personalised medicine)
- Insurance (e.g. fraud detection, personalized risk assessment)
- Agriculture (e.g. robotic harvesting, crop optimization)
- Finance (e.g. fraud detection, loan decision support systems)
- Military (e.g. automated weapons, cybersecurity for data protection)
- Law enforcement (e.g. predictive policing to forecast areas where crime is likely and dispatch police units, face recognition in public places)
- Environmental (e.g. climate prediction, energy harvesting forecast)
- Transportation (e.g. self-driving vehicles)
- Manufacturing industry (e.g demand forecasting, robotics)
- Human resource management (e.g. CV screening, workforce planning)

9. Read carefully the following scenario: [AI attitude]

Imagine that you are applying for a job in a large company and the recruitment process consists of two steps. The first step is based on an AI software which scans your resume and your answers to a set of questions on strategic competencies. The software assigns you a score which is used to select those candidates who can move on to the second stage (the interview). The company claims that the software makes the process faster and more objective. Also, the company says that the data is anonymised, and no personal information is used. To what extent would you feel comfortable or uncomfortable with this process?

- Very comfortable
- Fairly comfortable
- Neutral
- Not very comfortable
- Not at all comfortable

10. Read carefully the following scenario:[AI attitude]

Imagine that you are looking for a smart meter to reduce energy consumption in your house, cut the cost of utilities, and adopt a more sustainable lifestyle. You are offered a smart meter that uses AI to analyse home energy consumption and make recommendations for more efficient usage. Among functionalities, this system can give you the opportunity to receive personalised offers from energy suppliers which can help you save money.
The company producing the smart meter says that your data is anonymised, and no personal information is shared with third parties without your consent. To what extent would you feel comfortable or uncomfortable with this application?

- Very comfortable

- Fairly comfortable
- Neutral
- Not very comfortable
- Not at all comfortable

11. With respect to the previous scenarios, which of the following aspects should an organisation developing or using AI consider more? Please select three items and rank them. [Trust in AI]

- Security and accurate results
- Fair treatment and equitable access to the AI application for all members of society
- Privacy and data protection
- Human supervision over the AI outcome and process
- Clear communication about the AI application's purpose and limitations
- Risk management and identification of responsibility
- Societal and environmental impact of the AI application

12 How important are the following measures to increase your trust in AI? Please use the five-point scale to plot your answer. [5-point scale: Very Important, Important, Moderately Important, Of Little Importance, Not important at all] [Trust in AI]

- A set of laws enforced by a national authority which guarantees ethical standards and social responsibility in the application of AI.
- Voluntary certifications released by trusted and competent agencies which guarantee ethical standards and social responsibility in the application of AI.
- Having independent expert entities that monitor the use and misuse of AI in society, including the public sector, and inform citizens.
- The adoption and application of a self-regulated code of conduct or a set of ethical guidelines when developing or using AI products
- The provision of clear and transparent information by the provider that describes the purpose, limitations and data usage of the AI product
- The creation of design teams promoting diversity and social inclusion (e.g. gender wise, different expertise, ethnicity, etc) and the consultation of different stakeholders throughout the entire lifecycle of the AI product

13 To what extent do you agree that having a better education on what AI is, as well as its current and future uses, would improve your trust in it? [Trust in AI]

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

14 How much do you trust the following entities in ensuring that AI is in the best interest of the public? Please use the five-point scale to plot your answer [5-point scale: A lot, Somewhat, So and so, Not so much, Not at all] [Trust in AI]

- National Governments and public authorities
- European Union (including European Commission/European Parliament)
- Universities and research centres
- Consumer associations, trade unions and civil society organisations

- Tech companies developing AI products
- Social media companies

15 Now that you have answered several questions about AI, to what extent do you feel AI and its applications impact your daily life already? [AI impact awareness, self-assessed]

- A lot
- Somewhat
- So and so
- Not so much
- Not at all

Would you be interested in attending a free course on AI to improve your knowledge?
16 [AI attitude] [AI awareness]

- Yes
- No

Profiling

Please indicate your job:

- Entrepreneur / employer
- Self-employed / freelance professional
- Manager, officer
- White collar / employee
- Craftsman
- Shop owner, retailer
- Teacher, professor, writer, journalist, artist
- Manual or technical worker
- Student
- Retired
- Homemaker
- Unemployed
- Other_____

What is your highest level of formal education?
[Eurostat-compatible / answers can vary depending on country]

1. Lower secondary education or lower education
2. Upper secondary education
3. Post-secondary non-tertiary education
4. Short-cycle tertiary education
5. Bachelor's or equivalent level
6. Master's or equivalent level
7. Doctoral or equivalent level

Please indicate where you live in your country:
[rephrase answers in terms of population size]

8. City
9. Suburb near city
10. Small town not near a city

11. Rural area
12. Not sure

If you were to describe your digital skills, how would you define yourself:

- Not at all expert: I use digital tools only if it is strictly necessary (e.g. email, messages)
- Not very expert: I'm not sure of my skills and I have to get someone to help me with new things I don't understand
- Enough expert: I'm not entirely sure of my skills, but I manage to do the best I can when I need to do something online and I try to learn new skills when I need them.
- Expert: I am quite sure of my digital skills, I try to exploit the potential it can offer and to be updated on the news.
- Very expert: I am sure of my digital skills, I am always attentive to innovation, I have no difficulty in moving in the digital world for everything I need, and I am interested in.

Annex 5. Questionnaire for experts

Addressing Trustworthy AI

1. In 2019, the High-Level Expert Group on AI delivered the Ethical Guidelines for Trustworthy AI under the mandate of the European Commission (for more details, see the EC's website)
Are you familiar with Trustworthy AI guidelines?

- I have used them
- I have read them
- I have heard of them
- I have never heard of them

2. To what extent do you agree with these statements? [5 scale: strongly disagree, disagree, undecided, agree, strongly agree] [reflect better on statements]

- Trustworthy AI should be framed in precise terms to avoid ambiguity (e.g. by using mathematical/logical tools)
- Trustworthy AI should be translated into engineering practices
- Trustworthy AI needs to be a combination of technical (e.g. software tools) and non-technical methods (e.g. governance mechanisms)
- Trustworthy AI is a mindset that needs education and practical learning
- Trustworthy AI is a misleading notion that should be avoided (a machine cannot be trusted as we do with humans)
- Trustworthy AI is out of reach

3. The European AI strategy emphasises stakeholders participation and, in particular, those who are part of vulnerable groups such as women, persons with disabilities, ethnic minorities and children.

a. Based on your direct or indirect experience, to what extent are AI stakeholders involved in the design process of AI systems?

- A Great Deal
- Much
- Somewhat

- Little
- Never

b. In your opinion, which category of stakeholder is not considered enough in the topic of trustworthy AI?

- Women
- People with disabilities
- Ethic minorities
- Children
- Others [please specify]_____

4. The European AI strategy adds the following: “Interdisciplinarity should also be supported (by encouraging joint degrees, for example in law or psychology and AI). The importance of ethics in the development and use of new technologies should also be featured in programmes and courses.” (AI for Europe, 2018 p 13)

Do you have interdisciplinary collaborations? Yes / No
4 bis If so, in which fields are your collaborators trained? [multiple answers are possible]

- Generic programmes and qualifications
- Education
- Arts and humanities
- Social sciences, journalism and information
- Business, administration and law
- Natural sciences, mathematics and statistics
- Information and communication technologies
- Engineering, manufacturing and construction
- Agriculture, forestry, fisheries and veterinary
- Health and welfare
- Services

5. In your opinion, which fields currently do not have sufficient influence on the topic of trustworthy AI?

- Generic programmes and qualifications
- Education
- Arts and humanities
- Social sciences, journalism and information
- Business, administration and law
- Natural sciences, mathematics and statistics
- Information and communication technologies
- Engineering, manufacturing and construction
- Agriculture, forestry, fisheries and veterinary
- Health and welfare
- Services

6. Please, tell us your positive or negative experience with interdisciplinary work in the field of AI?(pls avoid personal details) [open] _____

Implementing Trustworthy AI

7. Below we list several requirements to implement Trustworthy AI. How challenging is their implementation on a scale from 1 (very easy) to 5 (very difficult)?

- AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights.
- AI systems should be designed to be safe, reliable and secure, preventing risks and unintentional and unexpected harms.
- AI systems should guarantee privacy and data protection, including the data they gather or process
- AI systems should be transparent: humans should always be aware that they interact with a product/service empowered with AI, its purpose, limitations and data usages.
- It should be possible to demand an explanation of the AI system's outcomes adapted to the user expertise.
- AI systems should facilitate inclusion and diversity. It should also be ensured that all society members have equal access and equal treatment in using or interacting with an AI system.
- AI systems should be sustainable, and their design should take into account the impact on society, the environment and future generations
- AI systems should ensure the identification of responsibility and, if required, be open to public scrutiny. If something goes wrong, adequate redress should be ensured.

8.a Which of the following methods are you most familiar with?

- Questionnaire (like ALTAI) and checklists
- Algorithmic tool (specific fairness metrics, explanation methods, privacy enhancing technology, adversarial attacks...)
- Impact assessment (risk assessment analysis / data and algorithm impact assessment /...)
- Code of conducts (e.s. ACM, IEEE..) / guidelines & requirements (EU, OECD…)
- Standards (standard ISAE 3402, security standards…)
- Protocols and governance framework
- Red teaming
- Stakeholders participation
- Flag mechanisms
- Others_____
- None of the above

8.b To what extent would the methods mentioned above facilitate the implementation of Trustworthy AI? [5 scale answers: A Great Deal, Much, Somewhat, Little, Never + I'm not familiar with]

- Questionnaire (like ALTAI) and checklists
- Algorithmic tool (specific fairness metrics, explanation methods, privacy enhancing technology, adversarial attacks...)
- Impact assessment (risk assessment analysis / data and algorithm impact assessment /...)
- Code of conducts (e.s. ACM, IEEE..) / guidelines & requirements (EU, OECD…)
- Standards (standard ISAE 3402, security standards…)
- Protocols and governance framework
- Red teaming
- Stakeholders participations and user
- Flag mechanisms

- Others_____
- None of the above

9.a A report based on the online hackathon “Ethical dilemmas in AI - engineering the way out”, conducted in September 2020, claims that responsible AI requires allocating time for reflection and spaces for collective discussion and debate (https://standards.ieee.org/initiatives/artificial-intelligence-systems/ethical-dilemmas-ai-report.html)
To what extent do you agree with this claim?

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

9.b Which of the following would increase opportunities for generative discussion within organisations?

- Development of a common language between AI researchers and experts in ethics/law/sustainability/policy/IT/engineering
- Multidisciplinary events with concrete agendas and case studies
- Multidisciplinary boards supporting designers and managers with ethical and legal issues
- Seminars and professional courses on topics regarding Trustworthy AI and responsible innovation
- Joint initiatives with other entities such as trade unions, civil society organisations, manufacturers, insurance companies, policy and military, and academia
- others_____

10.a Also, the report recommends additional ethics training in engineering and computer science courses (academic and corporate levels). To what extent do you agree with this recommendation?

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

10.b Could you please suggest topics for the ethical training of engineers? [open] _____

Governance of AI

11. In april 2021, the European Commission (EC) delivered a proposal for regulating high-risk AI systems (https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206). Indicate your level of familiarity with the proposal.

- I have read it carefully
- I have read part of it and/or some commentaries
- I have heard of them
- I have never heard of them

11 bis . In april 2021, the European Commision (EC) delivered a proposal for regulating high-risk AI systems (https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206).

- Highly satisfied
- Satisfied
- Neutral
- Dissatisfied
- Highly dissatisfied

For those (highly) satisfied and (highly) dissatisfied: could you please tell us why you are satisfied / dissatisfied

Optional: in general, are you in favour of an AI regulation? Yes / No

12. In the proposal for a regulation on AI, the EC plans to set up an EU database of stand-alone high-risk AI systems that the EC will manage to increase public transparency and oversight and strengthen ex-post supervision by competent authorities. To what extent do you agree with this measure?

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

13. To what extent can soft law (no binding force such as principles, declaration and code of practice) contribute to the “achievement” of Trustworthy AI?

- To a great extent
- Somewhat
- Not much
- Not at all
- I do not know

14. Consider the following soft law mechanisms. How effective are they on a scale from 1 (highly ineffective) to 5 (highly effective)?

- Voluntary labelling
- Certification by standards organisations (e.g. IEEE’s Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) program. Link: https://standards.ieee.org/industry-connections/ecpais.html)
- Economic compensation
- Reviewing mechanisms
- Third-party audits
- AI license to restrict the use of AI systems (see Responsible AI license https://www.licenses.ai/)
- Whistleblowers protection
- Others_____

Demographics

Could you please tell us in which country you work?(multiple answers in case the interviewee collaborates with more governments)
EU and non-EU countries

Could you please specify your field of education?(multiple answers might be possible) [based on International Standard Classification of Education]

- Generic programmes and qualifications
- Education
- Arts and humanities
- Social sciences, journalism and information
- Business, administration and law
- Natural sciences, mathematics and statistics
- Information and Communication Technologies
- Engineering, manufacturing and construction
- Agriculture, forestry, fisheries and veterinary
- Health and welfare
- Services
- Others_____

In which sector do you work?

- Public sector (e.g. government)
- Private sector (e.g. most businesses and individuals)
- Not-for-profit sector
- Academia
- Other_____

What is your age?
- 18-34
- 35-50
- 51-69
- 70+

How do you identify yourself?
- Female
- Male
- Other

Which best describes your role?
- Researcher / Professor
- Manager
- Administrative staff
- Student
- Trained professional
- Consultant
- Civil servant
- Others_____

Main dedication of your Institution

- Education
- Innovation
- Research
- Development
- Production
- Commercialization (?)
- Services
- Others

Your professional Area

- Marketing
- Research
- Management
- Production
- Innovation
- Technology development
- Human resources
- Legal Department
- Others

Area of application (optional ?)

- Health
- Life Science
- Human Resources
- Education
- Finance
- Marketing
- Cybersecurity
- Insurance
- Automation
- Energy
- Agriculture and Livestock
- Food
- Mobility
- Others

How long have you been working in your field?

- 0-9 years
- 10-19 years
- More than 20 years

How would you rate your level of expertise in your field?

- Basic knowledge
- Novice
- Intermediate
- Advanced
- Expert

Annex 6: Ethical training for AI4Media pilot (WP6)

Project: AI4EU - AI4Media Pilot

Moderators

- **Teresa Scantamburlo** (Ca' Foscari University - AI4EU)
- **Atia Cortés** (Barcelona Supercomputing Center - AI4EU)
- **Francesca Foffano** (Ca' Foscari University - AI4EU)

Participants

- **Philippe Henry Gosselin** (Principal Scientist - Interdigital)
- **Siegfried Loeffle** (Director of Business Development - Interdigital)
- **Slim Ouni** (Associate Professor - University of Lorraine)
- **Arnaud Gotlieb** (Chief researcher scientist - Simula)

Introduction

As a result of an agreement between the Observatory on Society and AI (WP5) and the AI4MEDIA pilot of the AI4EU project (WP6), the partners decided to collaborate towards **a deeper understanding of the ethical practices of AI**. This document focuses on reporting the activities done with the partners.

After a preliminary discussion aimed to evaluate the needs and expectations for the pilot, the Observatory team prepared **two activities**: **a seminar to introduce the concept of ethics** and present the European requirements for trustworthy AI, and a **practical activity to understand and apply the requirements**. The activities have been reported as valuable for the future development of the pilot and the Observatory Team states their complete availability to support further discussions or collaboration to the partners involved.

2 Activities
The Observatory team organized a first meeting to define the needs and expectations of the project. In the previous report we collected the following needs:

1) Improving the work process using biometric data. According to GDPR Article 4.1, the definition of "personal data" includes any data that can be used to identify directly or indirectly one or several specific properties unique to physical, physiological identity (among others). Biometric data, such as facial and voice recognition, are protected by this article.
2) Understanding ethical implication using biometric data
3) Understanding how to build a high-performing product with social and legal issues in mind
4) Building expectations on future risks.

The result was the design of two sessions: a seminar and a practical activity. The former aimed to introduce the partners to the concept of ethics in engineering disciplines and the European approach

for a human-centric, trustworthy AI. The latter proposed the application of the ethical requirements to real case studies.

2.1 Seminar

On August 31st 2020, the Observatory team held the first of the activities to inform and reinforce the knowledge of ethics and its application. **In the first part of the seminar, the historical background of ethics and current applications in technological fields were presented**. In the second part of the seminar, the partners attended a presentation focused on the Ethics guidelines for trustworthy AI, with particular attention to each requirement.
The seminar concluded with an open discussion on the topics and examples introduced during the two presentations. Participants expressed a particular interest in the European guidelines and future expectations from the European Commission.

2.2 Practical activity

The second part of the activity took place on September 10th 2020, and **aimed to understand the application of the European requirements to real case studies.** As preparatory task, each partner received a set of questions in order to understand their relationship with the European regulation and their opinion on the requirements.

- Are you familiar with any existing regulation, best practice or international standard related to the Trustworthy AI guidelines? If so, could you mention a few?
- Do you have internal processes to (partially) cover some of the requirements? If so, could you provide an example?
- From the seven requirements presented (see the presentation on Trustworthy AI guidelines), could you select the three most relevant ones for your field? Could you explain why?

During the activity, the partners were invited to answer one of the questions presented in the preparatory task regarding which requirements were more relevant in their job. This aimed to understand the importance assigned to the requirements based on the case study under consideration. The aim was not to evaluate the ethical requirements from a company perspective but to obtain their personal opinions and interpretations of these. One of the criteria that were used for the selection of requirements was the final purpose of the AI system, as the requirements and priorities might be different for a research enabling technology than a final product in use. The requirements reported were: **Privacy, Transparency, Accountability, Technical Robustness and Diversity**.

- **Privacy** has been reported unanimously as the most critical value for their case study. For partners this requirement should be included by default
- **Transparency** has been reported essential to ensure the documentation of AI development.
- **Accountability** resulted necessary for an organization to ensure that actions have been taken to avoid the misuse of the AI systems.
- **Robustness** and **diversity** ensure an application to be safe and versatile to be applied to a whole range of scenarios without create discrimination.

After this initial debate, the Observatory team **presented some case studies that the Observatory considered that could be directly or indirectly related to the pilot use case and its possible future applications.** Each case study was contextualized to the participants within their use case

with the aim to **identify both the associated risks (technical and social) and the ethical requirements** that should be taken into account. In particular, the following examples were presented during the activity:

- **Face recognition**: gender and racial bias in existing tools (ex: Gender Shades project) and how social reactions in civil protests have triggered technical solutions that respect and protect human rights.
- **Video and voice manipulation** as a research enabling technology that has proven to be used for fraudulent usage (ex: deep fake news)
- **Image manipulation**: users' awareness to raise mental health issues and law enforcement to promote transparency in the usage of technology in images of celebrities or models.
- **Videoconferencing** and **Privacy**: an investigation suggested security vulnerabilities (see e.g. malicious people joining Zoom calls and broadcasting porn or shock videos), privacy breaches (e.g. see data sold to third parties without users' consents or meeting hosts tracking attendees) and misleading information about security measures to the users

The Observatory team provided inspiration for group discussion and engaged the participants to share ideas and points of view. The discussion was focused on the similarity and differences of the pilot case study. From the discussion emerged the following considerations:

- Concerning transparency and accountability requirements, **documenting** the design process is a crucial task. This helps keep track of the algorithm and the process used for the development of an AI system. A proper documentation can also support safety and organization's transparency, especially when collaborating with various stakeholders. This is a significant research issue that could even need **funded projects to find technical solutions** and the level of appropriate disclosure in the documentation.

- **Privacy resulted fundamental** to respect and protect the confidentiality of the data collected for training a system. It is also important to ensure **governance mechanisms** during the whole process to guarantee the quality and integrity of the data, the controlled access to the data and the traceability of any decision related to training data and models used. These aspects are strongly related to the broader requirements of transparency and accountability, previously mentioned.

- There is a concern about what can be done on the technical side in order to **anticipate and, possibly, prevent harmful situations**. An issue is the ability to identify these situations at the early stages of a project or when AI is part of a larger application. This is also true for AI applications sprung from specific contexts or businesses (e.g. movie production) and then moved in a different setting with larger statements of users. In this situation it is difficult to predict interactions and the effects on more vulnerable users (see for example addictive behaviors favored by large recommending systems like YouTube).

Scenarios and Risks

In line with the preliminary meeting, the system proposed by the pilot can be applied for different applications, mainly online media: TV, YouTube, video-game, with the aim to translate content in many languages.

However, depending on the domain of the application and the use case at hand it will apply, the AI system can encounter different risks both during data collection (face/voice recording) and processing. Potential risks connected to the pilot are:

- **Fake news**: Spreading fake and misleading information through the media. This can be done for fun but may also have a huge political and social impact (see, for example, attempts to damage the credibility of public figures)
- **Voice phishing:** A telephonic fraud used to enter in possession of personal data, as financial or sensitive information. In the use case, this practice may be used to create new harmful contents
- **Data / privacy breach**: Inappropriate use of users' data and disclosure of personal data without users' consent
- **Social Impact:** there are several consequences associated with the exposure of voice and face manipulations that can be foreseen in mid or long term. One example is the social rejection of new technologies, which can affect the Society's trust in public and democratic institutions. It can also cause the loss of capabilities to determine what is real content from manipulated one, leading to delusion, or subordination among others.

Tools recommendations

As a conclusion of the activities, the Observatory team suggests a set of additional materials that can help the partner to evaluate and continuously reflect on the ethical risk of the pilot.
These tools are recommended to be used during the life-cycle of the AI system to improve the trustworthiness and safety of the system entering the market.

Appendix material

- Ethical Explorer : Decks of card created to identify, anticipate and limit risks.
- Datasheets for Datasets: Assessment list to evaluate the data-set and the data contained
- Data Ethics Canvas: Canvas created to identify and mitigate ethical issues
- Compass- Responsible Innovation: Self- questionnaire focus on company creation, definition and deliver on the market of innovation to evaluate sustainable innovation
- Ethical OS: Assessment to evaluate future risk zones through the use of scenarios.
- Human rights impact assessment guidance and toolbox: A toolbox in 5 phases to evaluate the impact of a product on human rights.
- Aequitas:An open-source bias audit toolkit for data-set
- Data Protection Impact Assessment (DPIA): Assessment based on the GDPR
- Assessment List for Trustworthy AI: produced by the High-Level Expert Group of the European Commission, this checklist aims to translate ethical principles into practices.