Trustworthy AI

Trustworthy AI Project Newsletter – Issue 1

CONTENTS

A Few Words From the Project Coordinator

ALLAI and UIIN, Our Two Partners From the Netherlands

Umea University Leads the First Intellectual Output of Trustworthy AI

Developing a Learning Framework for Trustworthy AI: The Beginning

Why Educate Trustworthy AI?

Who's Accountable When AI Fails?

Our APPROACH

Integrating AI with Ethics and Trust

Trustworthy AI is a pioneering project that integrates the teaching ethics and trust into the AI curricula, following the EU High-Level Expert Group guidelines about the 7 elements of trustworthy AI.

Transversal Teaching of AI

The project brings added value by raising awareness, for the first time, of the potential, opportunities and risks of AI amongst teachers and students of all backgrounds.

Innovative Teaching Strategies

Trustworthy AI makes a significant contribution in enhancing commitment and capacity of HEIs to innovate in their teaching, not only from the content perspective, but also with regard to the methodologies.

Trustworthy AI

Our **OBJECTIVES**

Produce 3 new resources to enhance the capacity of HEIs to introduce trustworthy AI teaching in their curricula.

Rigorously test the resources with more than 48 teachers and 200 students to optimise their relevance and effectiveness.

Strategically disseminate the resources produced, reaching at least 240 teachers that will integrate the latter in their teaching.

Trustworthy **AI**

THE CONSORTIUM

The Trustworthy AI project unites 7 partners from universities, businesses, start-ups, and networks from 5 EU Member States, whose experience and expertise provide an ideal foundation to achieve the project's objectives.

University of Alcalá – Project Coordinator
Madrid, Spain

National University of Ireland Maynooth
Maynooth, Ireland

European E-learning Institute
Copenhagen, Denmark

Umeå University
Umeå, Sweden

Stichting ALLAI Nederland
Amsterdam, The Netherlands

University Industry Innovation Network
Amsterdam, The Netherlands

Momentum Consulting
Leitrim, Ireland

WELCOME TO OUR NEWSLETTER

In recent years, Europe has witnessed an increasing development and use of Artificial Intelligence (AI). This has been applied to a huge variety of fields that range from health care and farming to production systems. However, the rapid spreading of AI has shed light on several challenges and risks that are connected to this technology. More specifically, AI brings into play a complex array of challenges that undermine fundamental Human Rights, such as privacy, social discrimination and security, just to give an example. Yet, several Member States within the European Union (EU), as well as other countries in the world, still have a limited regulatory framework when it comes to AI. This is also connected to the transnational dimension of the challenge that knows no borders.

With the Erasmus+ project **Trustworthy AI,** we aim to sensitise students of Science, Technology, Engineering, and Mathematics (STEM) and students of all disciplines of the scopes, opportunities and most importantly the risks that are connected to AI. To do so, we introduce a new methodology for AI teaching, enabling Higher Education Institution (HEI) teachers to act as catalysts towards all students, who will gain knowledge and real-life examples of trustworthy AI. The interdisciplinary approach is core to our project, as AI raises multidisciplinary challenges that stem from STEM topics to policymaking, philosophy, history and many more.

To keep our audience up to date with the developments within the project as well as a wider AI education landscape, we are happy to present you the Trustworthy AI newsletter series. The newsletter issues will feature the news from the partnership and highlight relevant articles on the topics of trustworthy AI, education that features ethical aspects of AI, and exceptional examples from the partner regions and beyond.

We hope you enjoy reading our first Trustworthy AI newsletter!

News from
the project

A FEW WORDS FROM THE PROJECT COORDINATOR

The University of Alcalá (UAH) is honoured to be the coordinator of the Trustworthy AI project and of such an outstanding partner consortium. Our institution is a leading university in Spain. It is located in the city of Alcalá de Henares, 35 km northeast of Madrid, which was named a World Heritage site by UNESCO in December 1998, in recognition of the fact that Alcalá was the first planned university town in the World. The university comprises 15 faculties and schools teaching over 20,000 undergraduate students in 40 degree programmes and 4,000 post-graduate students in 86 masters and doctorate programmes. The university employs 1712 permanent teaching and research staff and 834 non-teaching staff. Ongoing education is a strong feature of our work: we have 8000+ students who currently participate in expert and continuous education. Across all our students, approximately 60% are women.

The university will be represented in the project by the Intelligence, innovation, internet and information (I4) group, previously known as the Information Engineering Research Unit (IERU), within the Computer Science Department. It is one of the most active groups in the University of Alcalá, having produced weighty research activity in several fields in the last 15 years, with significant output in terms of scientific knowledge and funding. Given its extensive research and transfer history, we have not only the experience and background in coordinating and participating in projects at both national and international scope, but also strong links with the technology transfer office at UAH, a key link between the university research groups and the industry. The group has an intense activity in innovation in teaching and learning, we have a long record of research in learning technology and on-line instructional design, having coordinated and participated in many European and national projects in these topics. Members of the I4 group also teach at courses for public servants at higher levels at INAP, the National Institute in charge of the official training of employees of the public administrations and are currently consulting for the National Agency of Data Protection (AEPD, Agencia Española de Protección de Datos) on blockchain and GDPR.

Through these projects, we have developed a keen understanding that a digital divide exists and could widen if we continue to support a division between STEM subjects and the worlds of business and employment. Digital skills must not be confined to computer science and engineering faculties, but must be seen as transversal technical and non-technical abilities that enable individuals from all walks of life to flourish in an increasingly digital world. To that intent, we will not only coordinate the whole project, but we will specifically lead the last IO, IO3, developing a guide to conducting Ethical Artificial Intelligence hackathons that will guide teachers on how to use this innovative methodology to help consolidate students' skills based on real life business and social problems.

Looking forward to the results of the Trustworthy AI project and to be able to share all the outputs with the Higher Education community as soon as possible!

ALLAI AND UIIN, OUR TWO PARTNERS FROM THE NETHERLANDS

ALLAI

ALLAI is proud to be a partner in this project, and will bring expertise on concept of Trustworthy AI and the Ethics Guidelines for Trustworthy AI to this project. ALLAI already built significant expertise and experience in teaching Trustworthy AI in (public) organisations and will build on this experience by developing practical tools and resources for HEI teachers.

ALLAI is an independent international organisation that advocates responsible AI. ALLAI promotes AI that is safe, sustainable, ethically aligned, lawful and socio-technically robust. ALLAI was founded in 2018 by the three Dutch members of the High Level Expert Group on AI to the European Commission (Catelijne Muller, Virginia Dignum and Aimee van Wynsberghe), who each are considered the pioneers in setting and driving the agenda towards responsible AI, in particular at EU and international level. The motivation to found ALLAI was to make sure that the broad impact of AI on society remains at the top of the policy agenda's, but also to promote that the high level initiatives, guidelines and policies around AI actually find their way into society. To achieve this, ALLAI's work focusses on several pillars:

- AI policy: advocating and advising on AI policy making and regulatory developments around AI at EU, European and global level, through various activities and collaborations and roles with European and global institutions;

- Knowledge building and awareness raising among policy makers, public institutions, companies and civil society organisations on the opportunities and challenges of AI, AI policy developments and responsible AI practices;

- Translating responsible AI principles into practice through various programs and projects;

- Performing research activities on practical elements of responsible AI and translating existing responsible AI research through dissemination, knowledge and awareness projects;

- Developing educational tools and resources for teaching responsible AI;

- Specific responsible AI projects aimed at tackling particular AI challenges or advancing particular AI applications for the benefit of society.

UIIN

The University Industry Innovation Network is honoured to be part of this consortium and this project, which will help us gain a better understanding of the ethical aspect of Artificial Intelligence from the educational perspective, and how to ensure that this topic develops in a way that's both safe and useful for everyone by ensuring and encouraging the respect to the guidelines of the High Level Expert Group on Trustworthy AI.

UIIN is a dynamic international network of academics, practitioners and business professionals focused on establishing and improving relationships between education and industry. UIIN is a leading organisation in the field of university-industry interaction, collaborative innovation and entrepreneurial universities. With its consultancy, professional training and events, UIIN actively converts research to practice and supports universities, business and government in developing stronger and more strategic relationships, more future-oriented institutions and ultimately building a knowledge society for a better tomorrow.

Within the Erasmus+ Trustworthy AI project, UIIN is in charge of dissemination. This entails all those activities that are aimed at promoting the intellectual outputs and knowledge material that is produced throughout the project. More specifically, our dissemination tasks includes the production of a Dissemination Strategy, four newsletters and one special issue, a multiplier event with the participation of several international stakeholders, a hackathon with the participation of both students, and professors, branding material and much more.

UMEA UNIVERSITY LEADS THE FIRST INTELLECTUAL OUTPUT OF TRUSTWORTHY AI

Picture by Mattias Pettersson extracted from https://mediabank-umu.qbank-mediaportal.se/

Umeå University - UMU (Sweden) - is honored to be a partner of the Trustworthy AI project and to lead its first intellectual output. UMU will contribute to the project's aim to expand the teaching of AI in higher education, especially from the perspective of the principles and values laid out in the EU's guidelines for Trustworthy AI.

UMU was founded in 1965 and is Sweden's fifth oldest university, and today a strong international and multicultural environment with more than 32 000 students and 4300 employees. The university conducts research on a broad range of topics in the area of Computing Science and offers courses and study programs at bachelor, master and doctoral level. Each year, over 2000 students take at least one course at the department of Computing Science.

UMU joined the Trustworthy AI project through the Responsible AI research group established at the Department of Computing Science, which has over 120 employees with a background from more than 20 countries. The Responsible AI group focuses on the ethical and societal impact of Artificial Intelligence (AI) through the development of tools and methodologies to design, monitor, and develop trustworthy AI systems and applications. In the Trustworthy AI project, UMU is tasked with leading

the first Intellectual Output: developing a Teaching Framework for AI in Higher Education. This framework will be based on the framework created by the European Commission and other highly relevant stakeholders on what reliable and ethical AI should look like. This framework will present an integrated approach to teaching AI, including its social, ethical, and economic aspects, but also (and perhaps more importantly) the principles and learning strategies that students need to obtain all the necessary competencies for the development of Trustworthy AI.

Through the Trustworthy AI project, UMU aims both to obtain a deep understanding of the educational aspects crucial for Trustworthy AI development, as well as participating in the production of high-quality educational material that can be incorporated into our own education programmes. Trustworthy AI is a significant project for UMU at the intersection between AI and its ethical, social and legal implications, which is a core focus of our research. Some other initiatives include:

- AI4EU (https://www.ai4eu.eu/): Europe's first on-demand AI platform and ecosystem (Horizon 2020 Project), where UMU's contribution is the design of a development methodology for AI to create systems aligned with European values;

- Humane AI net (https://www.humane-ai.eu/): European human-centered AI initiative (Horizon 2020 Project), empowering European member states to build human-centered artificial intelligence and achieve the goals set by the European Commission in its European approach to artificial intelligence;

- WASP-AS (https://wasp-sweden.org/): Sweden's largest strategic recruitment project for researchers, educators and faculty in autonomous systems and software development (Knut and Alice Wallenberg Foundation Project).

We are looking forward to the collaboration within the project and to the development of a comprehensive Teaching Framework for Trustworthy AI in Higher Education!

DEVELOPING A LEARNING FRAMEWORK FOR TRUSTWORTHY AI: THE BEGINNING

In the Trustworthy AI project, our team at Umeå University (UMU) is tasked with leading the first Intellectual Output: developing a Teaching Framework for AI in Higher Education. This framework will be based on the [framework created by the European Commission and other highly relevant stakeholders](#) on what reliable and ethical AI should look like. This framework will present an integrated approach to teaching AI, including its social, ethical, and economic aspects, but also the principles and learning strategies needed for higher-education students to obtain all the necessary competencies for the development of Trustworthy AI.

An important goal of this framework is to give a strong basis for the development of high-quality educational material in the remainder of the project. For this reason, we are taking a highly focused approach around three specific research questions:

1. What competencies need to be acquired by students in higher education so that our future workforce is equipped to implement the European Guidelines for Trustworthy AI?

2. How to teach and evaluate these competencies in the context of higher education?

3. What is the current focus of AI education with respect to Trustworthy AI? Which resources are available, and which aspects are missing?

In our approach to answer these questions, we aim to take a broad perspective with a multi-disciplinary outlook. We believe an inclusive perspective is fundamental, as the Trustworthy AI principles endorsed by the EU include many inter-disciplinary concepts that require insights into how to develop social and ethical reasoning skills tackled in fields such as medicine or the social sciences. Furthermore, the EU includes a variety of perspectives and educational systems brought by its multiculturality: it is therefore important to capture this richness in the developed educational framework.

With these factors in mind, our UMU team is undertaking two main tasks that will inform the Teaching Framework. The first, a systematic literature review, aims to reveal which competences are needed to equip higher education students with skills enabling the development of Trustworthy AI.

This review has the additional goal of studying which educational approaches are effective for the teaching and assessment of such competences. The second task is dedicated to obtaining a multi-cultural and multi-disciplinary perspective into the current state and needs of education with respect to Trustworthy AI. For this purpose, we developed an interview protocol to gather the insights of 15 to 20 educators, policymakers and other stakeholders around the EU. With the help of our wonderful partners, who will each conduct a significant portion of the interviews, in the next few months we will gather the invaluable insight of stakeholders of a variety of backgrounds and locations. This information will then be aggregated and analysed, resulting in a comprehensive outlook into the status of Trustworthy AI education in the EU.

As this work progresses, we are looking forward to creating a comprehensive Learning Framework grounded in established knowledge and answering current needs. We will keep you updated!

Our partners' vision

WHY EDUCATE TRUSTWORTHY AI?

Trustworthy AI has become a 'household name' in Europe. Fist coined by the EU High Level Expert Group on AI in 2019, it reflects the European view and strategy for AI that is 'lawful, ethically aligned and socio-technically robust'. But what does this mean exactly? And, more importantly, how do we make sure that trustworthy AI moves form a 'narrative' into actual practice.

What is Trustworthy AI?

As said, the High Level Expert Group on AI to the European Commission set the standard for what Trustworthy AI entails.

It rightfully recognised that AI is not an unregulated technology and that many existing laws and regulations apply to AI in quite the same way that they apply to the development and use of any other (software) tool. The first pillar of Trustworthy AI therefore is: lawfulness. The development, deployment and use of AI should comply with existing (and future) legislation. That is not to say that all existing regulation is 'fit for purpose' for a world with AI, being the reason that both the European Commission and the Council of Europe are working on a legal frameworks for AI, in order to fill the gaps in existing legislation, including human rights laws.

The second pillar of Trustworthy AI is ethical alignment. Ethical reflection on AI technology can serve multiple purposes. First, it can stimulate reflection on the need to protect individuals and groups at the most basic level. Second, it can stimulate new kinds of innovations that seek to foster ethical values, such as those helping to achieve the UN Sustainable Development Goals, which are firmly embedded in the forthcoming EU Agenda 2030. Trustworthy AI can improve individual flourishing and collective wellbeing by generating prosperity, value creation and wealth maximisation. It can contribute to achieving a fair society, by helping to increase citizens' health and well-being in ways that foster equality in the distribution of economic, social and political opportunity. Third, ethical reflection often leads to legislation, thus acting as a precursor for rules for which the time had not come yet. Finally, ethical reflection can serve the purpose of interpreting, explaining, valuing and correctly applying (or not) rules that already exist.

The third pillar of Trustworthy AI is socio-technical robustness. Even if lawfulness and an ethical purpose is ensured, individuals and society must also be confident that AI systems will not cause any unintentional harm. Such systems should perform in a safe, secure and reliable manner, and safeguards should be foreseen to prevent any unintended adverse impacts. It is therefore important to ensure that AI systems are robust. This is needed both from a technical perspective (ensuring the system's technical robustness as appropriate in a given context, such as the application domain or life cycle phase), and from a social perspective (in due consideration of the context and environment in which the system operates). Ethical and robust AI are hence closely intertwined and complement each other.

Achieving Trustworthy AI through Education

ALLAI believes that one of the main avenues to bring trustworthy AI into practice is through education. Education of those who develop AI, but also of those who might not necessarily develop AI, but will work with AI in their future professions, for example in decision making or management positions. The decision if and how to deploy AI in a trustworthy manner in a company, a government institution, or a city for example, requires broad reflection and substantive knowledge of: 1. Why Trustworthy AI is important; 2. What Trustworthy AI actually entails and; 3. How it can be effectively achieved within any given setting.

Education is also important for those that might come to work with AI, such as for example legal, medical or HR professionals, journalists, teachers, law enforcement officers and so on. They need to understand the technical, ethical, legal and societal implications of the AI application they are working with. They need be empowered to maintain their professional autonomy, to appreciate the capabilities and limitations of the technology, so that an appropriate and responsible human-machine cooperation is reached.

The Trustworthy AI project aims to equip higher education professionals throughout the European Union with the tools and resources to teach all elements of Trustworthy AI to these future professionals.

Trustworthy AI

WHO'S ACCOUNTABLE WHEN AI FAILS?

Although AI is not a new field by any means (its history dates back to the 1950s) it hasn't been until recently when it has definitely taken off, thanks to the advances in technology, computation and electronics, among others. Its increasing presence in all areas of our lives shows how what once could be discarded as a hyped technology can now be considered a reality. AI can be used and adapted to execute a myriad tasks in almost any domain or sector and, in general, improving the efficiency of the process. AI's ability to generate business advantages, economic benefits and social good is hard to question; thus, basically all countries worldwide have adopted plans and strategies to lead the AI revolution in the coming years. Europe is not an exception, with its "Communication on Artificial Intelligence for Europe" in 2018.

The development and widespread use of AI is not without risks, however. As it becomes increasingly common and used even for the most mundane tasks, it also becomes subject to more public scrutiny; researchers, practitioners, journalists and civil organizations often bring to light AI-related scandals and controversies. To give a few examples, in 2018 the Cambridge Analytica scandal was uncovered, demonstrating the risks of AI in Social Media. Amazon, who had been using an AI tool to assist in the recruiting process, realized that it didn't like women; the tool penalized resumes that included the word "women's," as in "women's chess club captain", as the tool had been trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period - mostly coming by men. And only a few days ago, to give a more ordinary example, YouTube's AI blocked a channel after allegedly mistaking discussions about chess pieces being *black* and *white* for racist slurs. Evidently, a mistake on temporarily banning a user from YouTube might not have the same impact as not hiring women or as giving an incorrect treatment to a patient, but what is clear is that they are all derived from asymmetries of information or power or from not involving minorities or vulnerable groups. The victims are always the same: human rights and democratic values are at stake.

What is interesting, however, is that the blame is always put on the tool itself. We read statements such as "Amazon's system taught itself that male candidates were preferable", or we say that

YouTube's AI made a mistake. Little attention is set on the human beings (yes, actual people) who wrote, designed and devised those systems. Algorithms don't have values; they reflect the values and biases of their developers. And, so far, we have mostly been focusing in preparing young people with advanced programming skills, but its ethical and values-based components have been inadvertently left out; the consequences of doing so start to surface in our daily life.

It is in this context where Higher Education must play an important role in this area of strategic importance for the economic and social development of the European Union, contributing to cutting-edge, safe, ethical AI, with the European values at its core. It is indeed up to us to prepare young people with advanced programming skills, but also to prepare all students to understand the implications of AI and ensure it is put to good use and always for the benefit of the community. Empowering our students to be able to identify what we should do rather than what we (currently) can do with technology is, thus, the only way to ensure the fundamental rights and their underlying values are respected today and in the future.
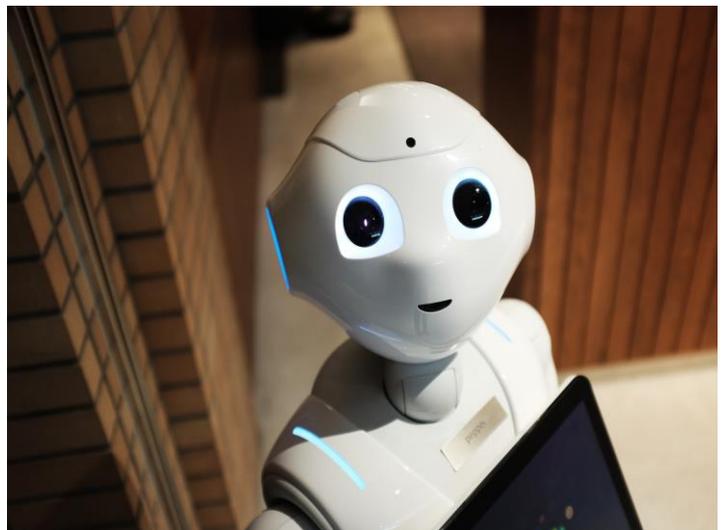
CONTACT US

Marçal Mora Cantallops
Project Coordinator, UAH
marcal.mora@uah.es

Jose Villagran Polo
Dissemination & Outreach,
UIIN
villagranpolo@uiin.org

Trustworthy AI project

Trustworthy AI project

trustworthyaiproject.eu