Trustworthy AI

Trustworthy AI Project Newsletter – Issue 3, April 2022

CONTENTS

Co-funded by the Erasmus+ Programme of the European Union

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.

Co-funded by the Erasmus+ Programme of the European Union

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

Co-funded by the Erasmus+ Programme of the European Union

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

Co-funded by the Erasmus+ Programme of the European Union

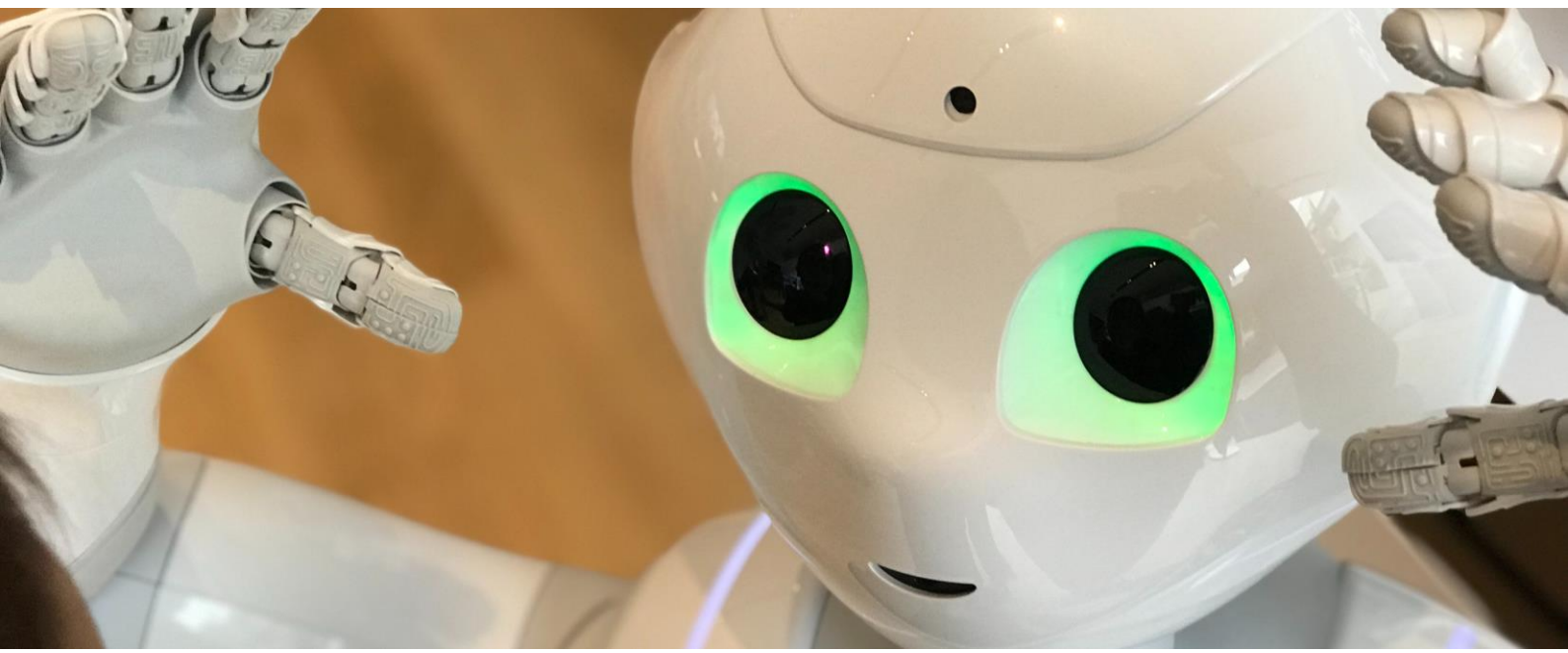
WELCOME TO OUR NEWSLETTER

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 this second issue of theTrustworthy AI newsletter!



Co-funded by the Erasmus+ Programme of the European Union

News from the project

Trustworthy AI

Co-funded by the Erasmus+ Programme of the European Union

An update from IO3

The Hackathon Guide

This Guide for Conducting Ethical AI Hackathons is in development; both UAH and UIIN, as part of the project Intellectual Output 3, are working together to provide a document that aims to introduce the concept of Ethical AI Hackathons to both higher education students and teachers through the multiple steps to organise one as a complementary strategy to their courses. In our guide, we conceptualise hackathons (and, in particular, Ethical AI hackathons) and we provide an in-depth overview of the steps and processes to run one, either in an onsite, online or hybrid setting due to these uncertain times for organisation of large-scale events.

The Guide for Conducting Ethical AI Hackathons stands out as an innovative element of teaching Trustworthy AI because despite being a widely used strategy in the technology sector and in civil society, it is still little known in higher education. It will also be innovative because hackathons are typically used by STEM (Science, Technology, Engineering, and Mathematics) students to solve technical problems; in this version we pioneer how to adapt the format to generate solutions that strike a balance between technological and civic-social aspects.

We expect the medium- and long-term impact to be evidenced by the increased capacity of teachers to develop digital skills and transversal competences in students, and the development of competences such as problem solving, teamwork and creativity among students, all with a process of consolidation of ethical and civic values. For universities, the exercise of a hackathon will allow them to better contribute to the formation of students prepared to be responsible citizens and contribute to European values.

The hackathon model has a high potential for transferability in itself and we are confident that - thanks to the very practical aspect of this guide - it will inspire teachers and HEI leaders to evaluate how it can be used in various aspects of higher education. We foresee a high potential for its use in strengthening ethical and trust issues in other advanced technologies and even in other types of community engagement projects and social issues.

The Guide for Conducting Ethical AI Hackathons will also be put to test in three pilot hackathons in UAH, UMU and NUIM, while NUIM will lead the user tests to provide a report on the experience that will complement the Guide as the main output of the IO3 of our project. We look forward to your participation!

Author: Marçal Mora-Catallops, University of Alcalá

Co-funded by the Erasmus+ Programme of the European Union

Key Learnings from IO1

EUEI experts' interview insights

With the goal of exploring the state-of-the-art of Trustworthy AI in Higher Education, 11 Expert interviewees were selected for their involvement in HE, whether through governance, program management or teaching. The experts, with affiliations in 5 different countries, brought use cases spanning medicine, law, computer science and social sciences. The specific goals of these interviews were to obtain expert feedback on the following topics:

1. General awareness of the Guidelines amongst stakeholders in HE.
2. Inclusion of the Requirements in current educational programs.
3. Current educational practices for Trustworthy AI (topics, learning outcomes, evaluation).
4. Incentives to facilitate the inclusion of Trustworthy AI topics in HE
5. Risks and opportunities.

Key Learning 1 - State of the art

Firstly, there is a need for training and availability of materials for educators of different disciplines to become aware of and gain knowledge on Trustworthy AI from the perspective of the High-Level Expert Group known as "HLEG" guidelines. Secondly, many interviewees feel like the expertise to teach about different requirements is spread across disciplines and topics. In this sense, rather than having a single course focused on the Guidelines, most consider it more natural to include teaching on the Requirements in already existing courses where they are relevant.

Key Learning 2 - Learning outcomes and assessment

The interviews revealed the need for two different levels of expertise. The first is the call for educating on how to recognise whether a requirement is being followed, and how. This competence corresponds to understanding what a requirement means in the context of a certain application. In fact, this type of question universally applies to students as citizens, as it allows for identifying and adopting trustworthy technology. In addition, it provides an initial maturity level in terms of understanding the HLEG Requirements.

The second competence identified across requirements corresponds to technical methods for trustworthy AI development. There is consensus across interviewees about the need to teach concrete methods for explainability, traceability, data collection, impact assessment, etc.

Co-funded by the Erasmus+ Programme of the European Union

Key learning 3 - Educational Resources

Uniformly across the expert interviews, experts mention that they do not use any specific resources related to Trustworthy AI. Rather, some mention the use of current topical examples, case studies, and relevant literature. This lack of resources is emphasised in several of the interviews by an added emphasis on the lack of training and time to get acquainted with Trustworthy AI and find available teaching resources.

Many of the interviewees coincided in asking for use cases. Interestingly, there was significant consensus on the type of use cases deemed necessary: they should be realistic and implementable. Indeed, using real cases brought directly from the industry that mimic situations where graduating students may find themselves in is seen as important for the usefulness of these scenarios. In contrast with the literature, where use cases are often used for reflection and debate, several interviewees suggested that use cases should be used for practical exploration, where they can implement and “play with” different solutions.

Another frequent mention is a need for material to aid in evaluation, i.e. exercises or assignments with a grading guide that can be directly used for assessing students. Indeed, several interviewees shared the difficulty of evaluating knowledge of abstract concepts. A final shared theme was the need for resources for teachers.

Author: Canice Hamill, EUEI

For more information on the Trustworthy AI Project, please visit https://www.trustworthyaiproject.eu/

Teaching and learning Trustworthy AI

Educational resources

Artificial Intelligence (AI) impacts and concerns everyone: from software engineers creating AI-systems, to organisations using them, to politicians and lawmakers making policy for them. AI has the potential to support humans and bring many advantages to society. However, its potential is double-sided as it comes with great risks to human rights, health and safety and raises ethical concerns. To tackle this problem, the EU's High Level Expert Group on Artificial Intelligence developed Ethics Guidelines for Trustworthy AI which hold 7 requirements that will guide the development, deployment and use of AI towards Trustworthy AI.

Trustworthy AI starts with education

It is highly important that the new generation of AI designers and developers but also of AI de-ployers and users gains skills and knowledge regarding Trustworthy AI, to ensure that AI operates in a lawful, ethical and robust manner. The Trustworthy AI Project has developed educational re-sources to help students from STEM but also from business administration and political and social sciences fields familiarise themselves with the Ethics Guidelines for Trustworthy AI and develop the ability to understand, critically assess, discuss and apply Trustworthy AI in a practical manner. The resources include 8 short knowledge clips, a 42-part card deck and a 7-step exercise.

Trustworthy AI knowledge clips

The knowledge clips are meant to introduce the concept of Trustworthy AI, and the 7 require-ments that underpin it, to the students. The clips can be used in a lecture, as an independent edu-cation resource, or can be integrated into an existing course. Apart from the first introductory vid-eo, each video introduces a separate requirement and includes an explanation of it, the principles it entails, a real-life example of its application, and relevant questions that can be asked when as-sessing AI against the requirement.

Trustworthy AI card deck

The purpose of the card deck is to create and drive meaningful and practical discussions on the 'trustworthiness' of AI applications, involving legal, ethical and societal elements. These discussions will help students understand the complexity 'AI ethics' from a practical perspective. It will encour-age students to look at the implications of AI from the point of view of different stakeholders and seek a balance between conflicting interests. The cards provide a general explanation of AI tech-niques that can be mixed and matched onto different domains to create use cases, but also in-cludes some pre-set use cases for inspiration and understanding. The deck is built up of 5 different sets of cards. With these different sets multiple games can be played that let students critically an-alyze AI techniques and use cases, guided by the 7 Requirements for Trustworthy AI.

Exercise: 7 Steps towards Trustworthy AI

Lastly, the "7 Steps towards Trustworthy AI" exercise functions as a thinking framework that teaches students how to approach AI use cases from the perspective of Trustworthy AI, and find solutions in a practical, problem-based, and systematic manner. In the exercise, students will ana-lyze a problem that can potentially be solved by AI, following a thinking process of 7 steps, with the 7 Requirements for Trustworthy AI integrated as part of the process. They will learn to identify, apply and balance ethical, moral and social elements and dilemmas relating to AI.

The exercise is a flexible teaching method that can be applied to any case study and can be inte-grated in different ways in class; in the form of group-work, an individual assignment or a class ac-tivity.

In the coming months all materials will be tested in education settings at 3 Universities. They will become available for wider use towards the end of this year.

ALLAI. nd

Co-funded by the Erasmus+ Programme of the European Union

Featured articles

Trustworthy AI

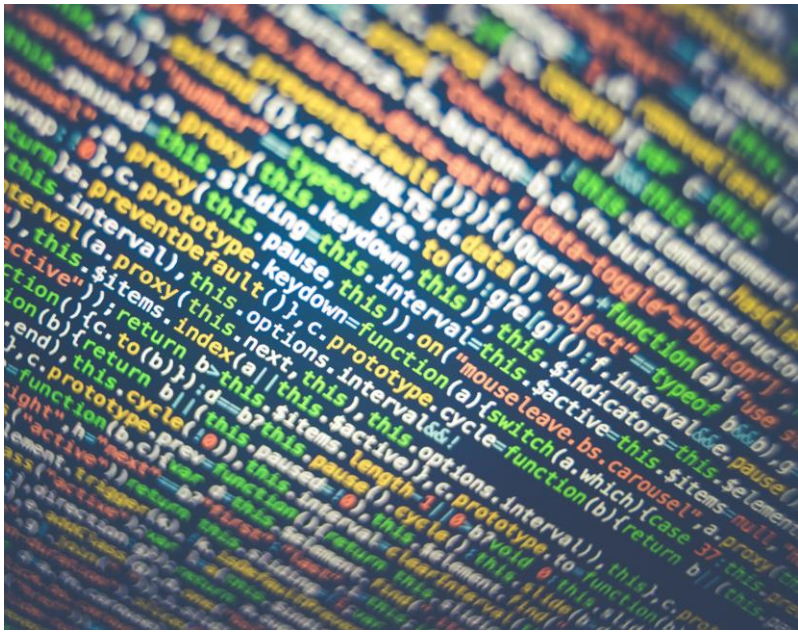
Co-funded by the
Erasmus+ Programme
of the European Union

Our partner's research

Maynooth University: Key Capabilities to drive Digital Transformation Paths underpinned by Trustworthiness and Data

Digital technology has a critical role in all aspects of our life. Indeed, the COVID-19 pandemic highlighted not only how much we rely on technology, but also how important it is to communicate, exchange information and enable businesses and commerce across boundaries (organisational, geographical, time). Digital Transformation is of ***significant strategic importance to Europe, Ireland and Internationally***, and it is embedded in many Strategies and Policies. As we advance our digital technologies, a major challenge over the next decade is the shaping of a **digital society**, a key priority of the European Commission. Trust, digital rights and core principles are key of realising the digital society. The new **EU Digital Strategy**, "**Shaping Europe's Digital Future"** sets out ambitious goals, with the subsequent requirement to establish new practices and approaches to digital transformation. As digital transformation continues across all sectors, many organisations are faced with strategic re-positioning of value chains and business ecosystems, driven by the use of Data and advanced technologies such as Artificial Intelligence (AI). The World Economic Forum (WEF) highlights the importance of Digital Culture and Skills and estimates significant growth with technology, governance and skills key drivers of the business transformation over the next decade. The Irish Digital Skills and jobs coalition highlights the importance of the human capital and research capacity. Similarly, the **EU's digital agenda** includes priority areas on Digital Skills, Technology, Ecosystems and Society. However, many organisations require advice, knowledge and research to address those significant challenges along their **Digital Transformation Paths** as best practices and validated frameworks are still not fully developed.

A skills revolution is required to ensure people can thrive in the digital transitions in all aspects of business and society. Understanding strategic questions and developing sustainable options in the form of **Digital Transformation Paths** are key. Economic, societal and business implications of digital transformation need to be discussed with all stakeholders. The wider impacts and constraints of digitalisation needs to be understood in order to **co-create Digital Transformation Paths** to drive a greener and digital society that benefits all.



This requires a different approach to both research and skills development that brings together multiple stakeholders across disciplines and sectors and moves the lecture-based skills education to real world setting or real-world replica. Indeed, the opportunities to provide digital replica of the real world are immense, as the examples around Buildings with accurate and complex digital twins demonstrate. This concept can be expanded to many other forms, digital replicas that represent digital humans, digital cities and also digital enterprises. Training needs to incorporate skills, technology, organisational processes and customer expectations, operation models and business models as well as an understanding of the entire ecosystem. We need strategic thinking, brining multiple stakeholders together to form innovative solutions, evidence-based insights how to best transform and to maximize the value of digitalization for both industry and wider society.

Co-funded by the Erasmus+ Programme of the European Union

Informed by scientific challenges and aligned with industry and societal needs, we have developed a capability framework along 7 key skills areas. The framework is continuously adjusted to reflect the demand and opportunities on an ongoing basis and include new technologies. It builds the foundation of dynamic capabilities around digital transformation and incorporates strategic relevance of ethics and trustworthiness. The 7 capability areas include

- **Sustainable and Resilient Ecosystem:** Leveraging the capabilities of, and creating synergies amongst, participants involved in the value chains and business models supporting digital business objectives and sustainability goals for regions and business ecosystems
- **Real-Time Enterprise Execution Management:** Establishing coherent direction by actionable analytics and insights regarding how digital can assist the enterprise to compete, thrive and grow. Digital operating models.
- **Architecture:** To advance the understanding of structure and behaviours in both technical and social aspects to plan, manage and guide digital transformation paths.
- **Information Value Management:** Investigate how to manage data as an asset and maximise the Information value to leveraging data to improve decision-making and business outcomes
- **Cybersecurity, Risk Management and Privacy:** Mitigating threats to digital business objectives, and enforcing regulatory obligations, standards, policies and guidelines
- **Ethics and Data Governance:** Responsible, trustworthy and transparent data management
- **Supply Chain Management:** Providing the digital enablement across the organization
- **Talent Development and Organizational Design:** Aligning leadership, skills and management structures in support of the organization's digital business objectives
- **Financial Planning and Optimization:** Improving the return on investment from technology-related resources. Adapting financial planning for an agile world.

With the aim to provide guidance and insights into key digital transformation capabilities, the 7 capability areas are detailed with multi-stakeholders working groups following an open innovation approach. Further detail on the approach and underpinning capability maturity model, can be found at: https://ivi.ie/

Author: Markus Helfert, Director Innovation Value Institute, Maynooth University

Co-funded by the Erasmus+ Programme of the European Union

Variable autonomy for Trustworthy AI

Many governmental and non-governmental institutions have published guidelines to describe the development of AI systems that adhere to our legal, ethical, and social values. Indeed, the HLEG's Guidelines for Trustworthy AI, the basis for the Trustworthy AI project, are only one of many documents produced to that effect. Often, guidelines are written at a high level, providing guiding principles that reflect the publishing organisation's values.

One of the important values set out by the HLEG guidelines is that of "Human oversight" as a means to support human autonomy and agency. Human oversight has many dimensions, including human autonomy and agency, control of critical decisions, as well as oversight in the development process. Several technical approaches have been proposed to implement human oversight, such as human-in-the-loop and human-on-the-loop. However, a more flexible approach can be given by variable control, also known as variable autonomy.

Within autonomous systems, variable autonomy denotes those systems for which the level of autonomy can be varied dynamically. The level of autonomy can be set by the system itself, deferring to an operator for complex or sensitive tasks, or by an operator that determines which level of autonomy is appropriate. In addition, variable autonomy approaches vary in terms of which aspects of autonomy are adjusted, from permissions and tasks to capabilities or sensor activation.

The implementation of variable autonomy comes with many challenges that are currently being tackled. These include design decisions on which aspects of autonomy are adjustable, which have to be deliberated on and made apparent, boosting accountability and transparency. Additionally, context-awareness is key in the design of such systems, where decisions on when to change autonomy levels are taken. This aspect automatically aligns with the idea of context-awareness demanded by the idea of human control set out in the HLEG guidelines. Finally, designers of variable autonomy systems must carefully consider how to display the right amount of information so that the humans involved in the loop know when to take control and what to do, without overwhelming them. This aspect makes them aligned with the intersection of explainability and human control.

For the reasons outlined above, variable autonomy systems by their nature align with many aspects highlighted in the HLEG guidelines for Trustworthy AI. They can therefore provide a new avenue for research into how to develop trustworthy intelligent systems.

For more detail, see: Methnani, L., Aler Tubella, A., Dignum, V., & Theodorou, A. (2021). Let Me Take Over: Variable Autonomy for Meaningful Human Control. Frontiers in Artificial Intelligence, 133.

Author: Andrea Aler Tubella, UMEA University

Co-funded by the Erasmus+ Programme of the European Union

Trustworthy AI

CONTACT US

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

Mario Ceccarelli
Dissemination & Outreach, UIIN
ceccarelli@uiin.org

LinkedIn name

Facebook name

trustworthyaiproject.eu