Trustworthy **AI**

Trustworthy AI Project
Newsletter – Issue 2, September 2021

CONTENTS

- Expert Interviews at the University of Alcalá de Henares
- Teaching a holistic approach to AI
- Understanding the state of the art on Trustworthy AI education: the framework
- The role of NUIM in the Trustworthy AI project

- Will the EU be leader of (trustworthy) AI?

Trustworthy AI

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

Trustworthy AI

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!

Trustworthy **AI**

News from
the project

Trustworthy AI

# EXPERT INTERVIEWS AT THE UNIVERSITY OF ALCALÁ DE HENARES

The University of Alcalá (UAH) conducted three expert interviews to obtain feedback from key stakeholders for the first Intellectual Output of our project, which aims to build a Learning Framework for the teaching of AI in HE. For that matter, we interviewed three HE representatives that are not only experienced teachers in their respective fields (two in Computer Sciences and one in Communication & Education) but also take part in the development of policies and courses either at their institutions or at a national level in Spain.

As a result of the interviews, we noticed some common thoughts across all profiles. First, even if it seems to be a bit contradictory, all of them highlighted how the unification of the educational space in Europe (the European Higher Education Area, EHEA), which has brought some benefits, has also been detrimental to the chances of including transversal topics such as Trustworthy AI into the existing study programs. Most of them have been shortened and currently are barely able to fit the core subjects; adding additional subjects is, thus, unfeasible. Therefore, all three interviewees highlighted the need of a structural change, specially in Spain, where departments and research fields are trapped within a rather rigid institutional structure, to be able to adjust to the more "modular" needs of today's education. Otherwise, Trustworthy AI (together with any other transversal topic) will only fit in postgraduate courses (e.g. Masters Degrees) or other extracurricular activities. In any case, they all agreed in the need for introducing Trustworthy AI in education, not only to generate better professionals for the future, but also to improve our skills as citizens in the present and future AI-enhanced World.

A few aspects generated differences in their opinions, though. In particular, it is interesting to notice how differently each field approaches the Trustworthy AI requirements. Computer Science professionals (and STEM in general) seems to favour Technical Robustness over the rest of the requirements, understanding that the most important thing for Trustworthy AI is a reliable and safe AI, which would, in turn, positively impact all other requirements. Thus, from a technical point of view, having models that work consistently, accurately and safely is a way of also facilitating other requirements such as privacy, transparency or non-discrimination. On the other hand, the Communication & Education professional that was interviewed decided to put the Human in the center; for her, Human Agency was the keystone of Trustworthy AI, understanding that, even though AI can and must be a facilitator for many tasks and decisions, this should always be in support of human agency and human decision-making, with human oversight in all steps, to avoid confusing AI decisions with human decisions and to make sure AI is used to enable democratic and equitable societies. In the same line, she also placed strong relevancies to privacy and data governance and accountability.

All in all, conducting the interviews and having the chance to talk for a few hours with three experienced and relevant stakeholders was a very enriching experience, which we are sure will be reflected in the final framework for IO1.
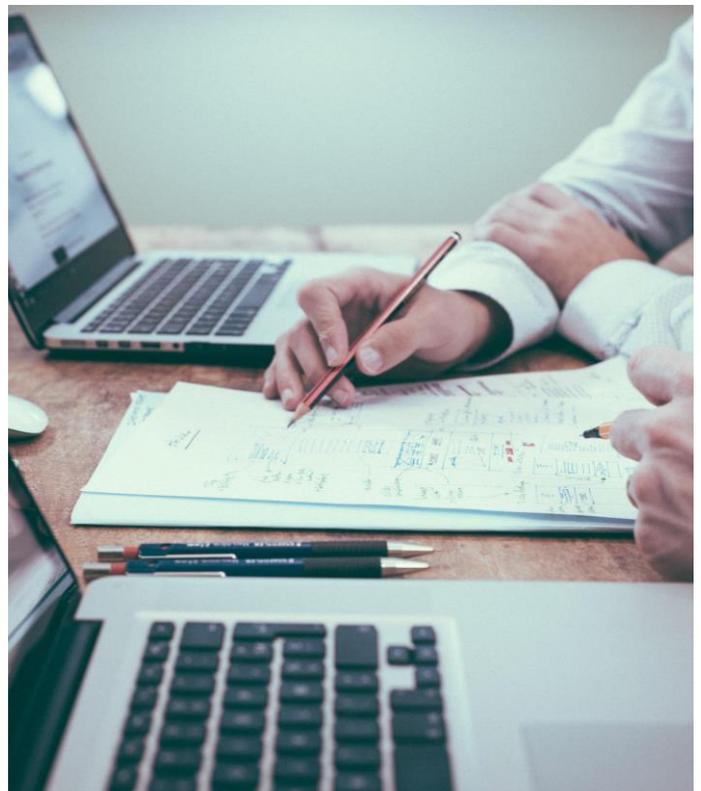
Photo by Scott Graham on Unsplash

TEACHING A HOLISTIC APPROACH TO TRUSTWORTHY AI

To explore the state-of-the-art of Trustworthy AI in higher education, we conducted interviews with experts who are well involved in teaching, program management or governance of AI-relevant academic subjects. The goal of the interviews was to inquire if and how the Ethics Guidelines for Trustworthy AI are included in higher education, what incentives and resources are needed to introduce the guidelines, and what sort of risks and opportunities shall be considered. What do you mean?

It was positive to notice that the majority of the interviewees already include parts of the requirements of the Trustworthy AI Guidelines in teaching of their respective subjects. What stood out was that the interviewees recognise that the question of AI and ethics surrounding it have become increasingly relevant in the public discourse. Hence, this raised awareness has pushed the interviewees to deliberate on how to include the ethics of AI in higher education.

However, due to the novelty of the subjects, there seemed not to be any sort of consensus across the experts on how to do this. Rather, the state-of-the-art of Trustworthy AI in higher education is that the teachers include few specific parts of the Trustworthy AI requirements in their education according to what they find relevant, and according to what resource they may even find. Hence, the application of the Trustworthy AI Guidelines in higher education in its present state is unsystematic, cherry-picked, incoherent and, if covered, their inclusion is often implicit rather than a deliberate effort to include the guidelines in education.

Despite such arbitrary and selective application of Trustworthy AI requirements in higher education, 100% of the interviewees found all seven requirements of the Trustworthy AI Guidelines relevant for higher education. Of course, the significance and importance of a certain requirement for a course may vary depending on the topic and the subject area.

For this reason, there was no agreement in the ordering of the requirements. Overall, the interviewees made it clear that there is a high demand for resources, teacher guidance and a structured approach for teaching AI ethics in higher education.
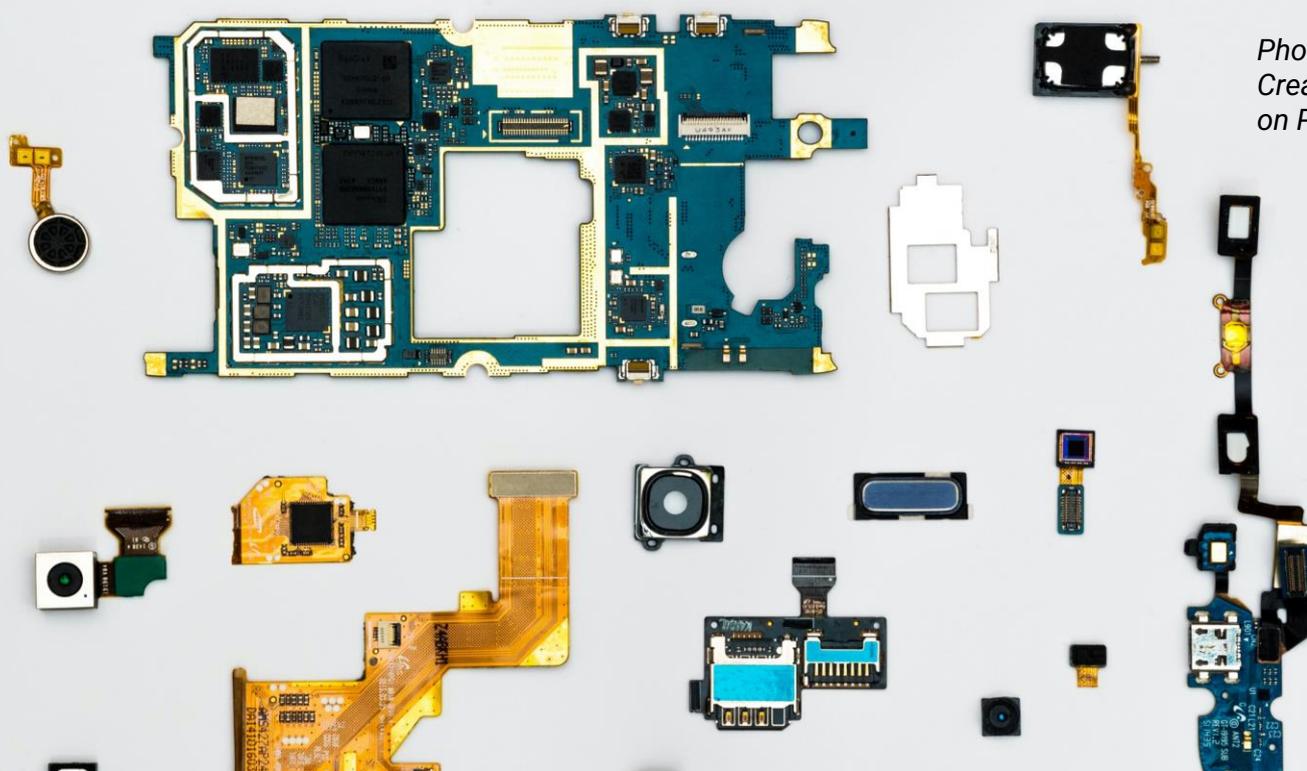
Photo by Free Creative Stuff on Pexels.

Trustworthy AI

Photo by Piixbay on Pexels.

To respond to this demand, some key opportunities and risks could be identified that require attention and emphasis to introduce the Trustworthy AI guidelines adequately in higher education. Firstly, there is a need for training and the availability of materials for educators of different disciplines to become aware of and gain knowledge on Trustworthy AI from the perspective of the Trustworthy AI Guidelines. . This need is echoed by many interviewees (5 out of 11), who mentioned that a barrier for incorporating Trustworthy AI into curricula may be a lack of experience or guidance for the educators, either because of the novelty of the topic, its inter-disciplinarity, or a simply lack of time to get familiar with these aspects. On the same note, these risks are perpetuated by a lack of relevant educational resources (case studies, examples, literature), and teachers having to put significant time and effort in finding relevant material for teaching AI ethics. Secondly, many interviewees feel like the expertise to teach about the different Trustworthy AI requirements is spread across disciplines and topics. Because different requirements have different levels of importance in each respective discipline, it is important to clarify this.

To alleviate the burden on educators, it is our task to provide them with resources that support them in teaching this important topic in several academic fields. For the coming period, we will be working on developing these resources for the academic fields of STEM, Law & IT and EPP (Economics, Politics, Law & Philosophy or equivalent). This could eventually lead to interdisciplinary teaching, fostering a holistic and diverse mindset among students to approach the issue of Trustworthy AI.

Authored by Christofer Talvitie and Catelijne Muller, ALLAI

UNDERSTANDING THE STATE OF THE ART ON TRUSTWORTHY AI EDUCATION: THE FRAMEWORK.

The Trustworthy AI project has the aim of helping introduce the European Union (EU) Guidelines for Trustworthy AI into Higher Education (HE). This important goal is part of the EU's digital strategy, which emphasises the need to train professionals that can "shape technology in a way that respects European values". In the first step of our project, we wanted to understand how Trustworthy AI is currently included in Higher Education and what are the needs to promote it further.

To get answers, we focused on two areas: the literature and the lived experience of those involved in higher education. We conducted a systematic literature review in which we analysed 24 papers about teaching ethical competence in higher education, and collected their shared insights. In parallel, we interviewed 11 experts with influence on higher education, through policy making, program management or teaching. The experts, with affiliations in 5 different countries, brought use cases spanning medicine, law, computer science and social sciences. We analysed the generous insight from the experts, and summarised three aspects of the current state of Trustworthy AI education: educator needs, resource needs, and policy needs.

The state of the art indicates that current education does not focus on the high-level expert group on artificial intelligence Guidelines, but rather that different aspects of trustworthy AI are tackled if and when they intersect with relevant course content. This finding points to a gap in HE, both for educators and students, in terms of awareness of the Guidelines and their content. 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 Guidelines. Most interviewees believe that the expertise to teach about the different requirements set down in the Guidelines is spread across disciplines and topics. Most consider, it is more natural to include trustworthy AI aspects in already existing courses where they are relevant, where some guidance on how to map the Guidelines to different courses would be very helpful.

Photo by Kindel Media on Pexels.

When interviewees were asked about what type of resources would be useful for integrating trustworthy AI in Higher Education, many coincided in asking for use cases that are realistic and implementable. In contrast with the literature, where use cases are often used for reflection and debate, several interviewees suggested that use cases should be deployed for practical exploration, where students can implement and experiment with different solutions. Several interviewees shared the difficulty of evaluating knowledge of abstract concepts, suggesting that exercises coupled with grading schemes would be a useful resource. A final shared theme was the need for resources for teachers: introductory material on the Guidelines, syllabi or evaluation Guidelines that educators could use to inform the design of their own courses. In addition, to ensure an interdisciplinary view on the Guidelines, a suggestion was to provide guidance on of what kind of experts could provide insight on the different topics or provide recorded lectures from scholars in different fields and the industry.

Trustworthy AI

Photo by Tara Winstead on Pexels.

On the other hand, most interviewees stated that they are not aware of any specific policy strategies to include aspects of trustworthy AI into education, either at the level of their institution or at a national level. Simultaneously, half of the interviewees mention that the topic of trustworthy AI is gaining importance in their organisation, and that they are actively considering how to include it in their programmes. This mismatch indicates that even though trustworthy AI is being introduced into HE, the effort is mainly driven by the educators themselves rather than by organisational or national strategies. This approach presents the risk of a mismatch in competences between programmes in different HE institutions, as the introduction of trustworthy AI into educational programmes is carried out independently rather than within a coordinated strategy. In terms of policy needs and incentives to boost the introduction of trustworthy AI in HE, interviewees delivered a big variety of suggestions. A big point of consensus between interviewees is the need for investing in expertise so that educators are equipped to teach these topics: this can take the form of investing into multidisciplinary training or boosting the hiring of experts in trustworthy AI

aspects to participate in education. This idea aligns with interviewees' mentions of lack of time to get acquainted with the topics in order to be prepared to introduce them in the classroom.

Overall, we received wonderful feedback: there was a general agreement on the importance of introducing trustworthy AI in HE, and many suggestions on necessary material to develop. As we move forward with the Trustworthy AI project, we will use the knowledge to develop and share resources to do precisely that.

Authored by UMU

THE ROLE OF NUIM IN TRUSTWORTHY AI PROJECT

The National University of Ireland Maynooth (NUIM) is a partner to the Trustworthy AI project. The key persons involved in this project are from Innovation Value Institute (IVI)- NUIM and their expertise are as follow. Prof. Markus Helfert is Professor in Digital Service Innovation at Maynooth University, Innovation Value Institute (IVI- Ireland) and the Director of IVI at Maynooth University. His research is focused on Digital Transformation and Data Governance and includes topics of data value and trustworthy information provisioning. Dr Zohreh Pourzolfaghar is Assistant Professor in Management Information System at School of Business, Maynooth University, and Research lead for Digital Construction at the Innovation Value Institute.

The team at NUIM have particular expertise related to the project in Curriculum development and Digital Technologies in the intersection between business and computing. This expertise is complemented with significant expertise in European Projects, both as partner and coordinator. Under the Erasmus+ scheme, Maynooth University has developed exchange partnerships with over 100 European universities.

As part of its interview tasks, NUIM has been conducting 8 interviews with professionals in academia and has prepared a report in an established format. Involving researcher Douglas Cirqueira, the team at IVI has conducted the interviews and analysed and structured the results. These have been reported to the project coordinator, providing a number of recognised challenges and recommendations.

Additionally, within the Trustworthy AI Learning Framework (developed by UMU and other partners), NUIM will be in charge of the coordination of all aspects of testing of the developed Learning Framework, in strict cooperation with users and with the leader of each deliverable.

Another task assigned to NUIM is in the phase of user testing. For this phase, a prototype competency framework will be developed and reviewed by other partners. Then each partner will share it with 3 - 4 representatives of the target groups by asking for their qualitative comments using a format established by NUIM. NUIM will compile the results and opinions in a report.

As the leader of the user tests, NUIM: 1) will share a guide for the organisation of the tests; 2) will coordinate with the other partners the selection and participation of 24 teachers and 120 of their students.

In conclusion, At the User Tests - First Hackaton, NUIM will be responsible to share guidelines for the coordination of the user testing phase. NUIM also will prepare a feedback questionnaire to obtain the feedback from the participants and then combine the results of the feedback to develop a user test report.

Photo by NUIM.

Trustworthy AI

Our Partners' Vision

Trustworthy AI

WILL THE EU BE LEADER OF (TRUSTWORTHY) AI?

Introduction

Artificial Intelligence (AI) is defined as the ability of computer systems to "to perform tasks normally requiring human intelligence, such as – yet not limited to – visual perception, speech recognition, decision-making, and translation between languages (Oxford Dictionary, 2021)." Undoubtedly, AI is an area of strategic interest for the European Union (EU), as it enables economic and social development. However, as we already discussed, AI does not come without risks. It is often argued in the literature (European Commission, 2021; Thomas, 2021) that AI can threaten individuals' privacy, increase socioeconomic inequalities, or spread fake news. All these risks, if not countered, will hinder fundamental EU values, such as democracy, human rights, equality, etc.

With this premise, in April 2018 through the joint Declaration of cooperation on Artificial Intelligence, Member States called on the EU for a common regulatory approach to trustworthy AI. Member States agreed to work together under the "EU umbrella" to collectively face AI ethical challenges, whilst ensuring the EU competitiveness in research and development of AI. Yet, what does this mean in practice, and where does the Bloc stand three and a half years later?

What has the EU achieved in terms of trustworthy AI?

Following the 2018 general declaration, several initiatives have taken place to achieve the abovementioned goal. Most notably, in June 2018, the EU appointed the High-Level Expert Group on AI (AI HLEG), which consists of 52 experts on AI from academia, business, and civil society across the 27 EU Member States. The group has been continuously working to make recommendations on mid-and long-term challenges and opportunities related to AI, based on the recent developments of the technology. The recommendations have fed into the EU policy development process, the legislative evaluation process and digital strategy, and have the goal to enable an environment for the research and development of trustworthy AI to thrive, whilst ensuring that fundamental rights are respected.

In April 2019, the Ethics Guidelines for Trustworthy AI were presented by the AI HLEG. The guidelines consist of a set of 7 key requirements that AI systems should meet for trustworthiness. These include:

• Human agency and oversight

• Technical robustness and safety

• Privacy and data governance

• Transparency

• Diversity, non-discrimination and fairness

• Societal and environmental well-being

• Accountability

The Guidelines support legislators, as well as businesses and society at large in the identification and definition of the challenges raised by AI. Additionally, the Guidelines set the foundation for the Final Assessment List on Trustworthy AI. Issued in 2020 by the AI HLEG, this EU tool consists of an accessible self-assessment checklist which helps developers and users of AI to evaluate the risks associated with the technology.

Finally, last April, the Commission proposed an EU regulation for the harmonisation of the rules concerning the use of AI. If this were to pass the Parliament and the Council of the European Union scrutiny, the regulation would set the base for the first EU-wide regulation that guarantees that AI-powered technology is fully respecting Union's fundamental values. To achieve this, the regulation categorises AI systems according to their potential risks, from minimal to unacceptable risk. The ultimate goal is to create "sustainable, secure, inclusive and human-centric artificial intelligence through proportionate and flexible rules (Liboreiro, 2021)"

Conclusion

Undeniably, the EU and its Member States have been acting as frontrunners both in Europe and globally in setting the rules for ethical use of AI. Through its expert groups, tools, and proposed regulation, the EU not only aims to ensure that AI will fully respect fundamental EU values, but also to become one of the main global players in the research and development of trustworthy AI. Yet, questions remain over the effectiveness of the EU action, especially when it comes to competitiveness with other global market players in the development of the technology, such as the United States or China, which might not necessarily comply with the EU standards of ethics and trustworthiness. A question that only time will answer.

Authored by Mario Ceccarelli, Junior Project Officer at UIIN.

Image representing the AI pyramid of risk of AI. Source: (Liboreiro, 2021)

Trustworthy AI

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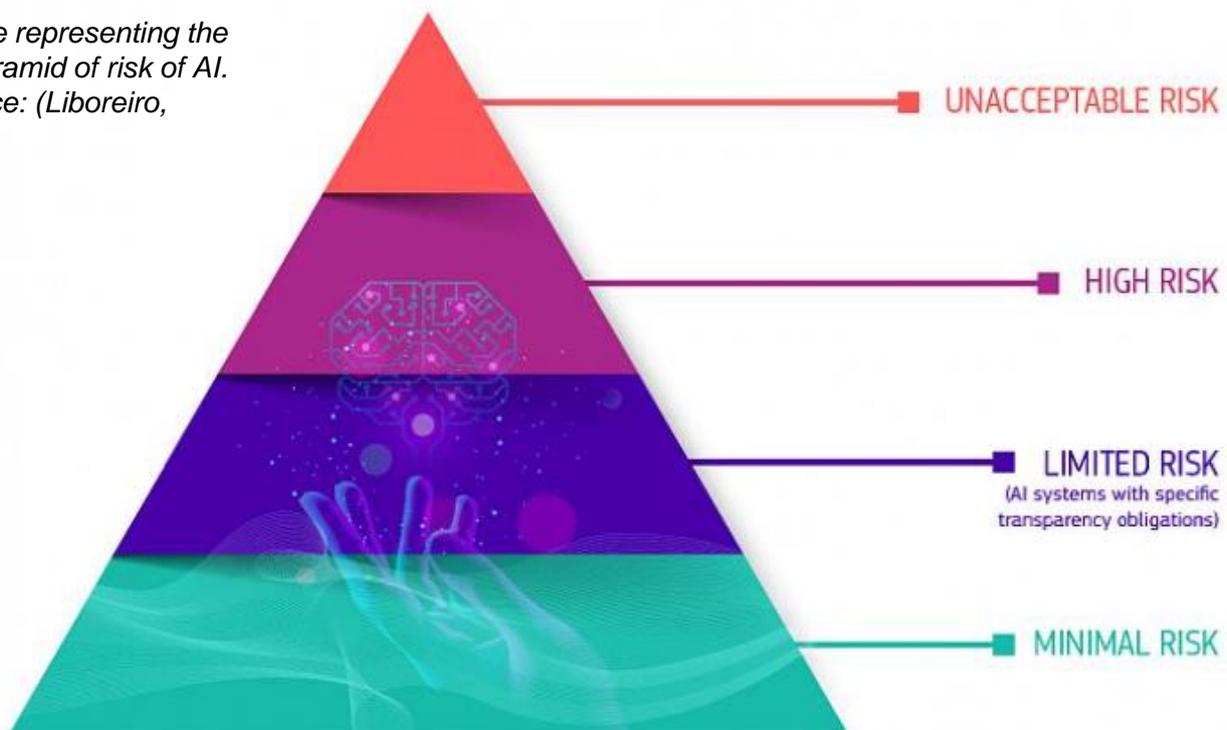
Trustworthy AI

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