What ASSISTANT brings you: Interview with Dr. Eduardo Vyhmeister on the ASSISTANT project and Ethical and human-centric toolbox

Dr. Eduardo Vyhmeister is a Postdoctoral Researcher at the Insight Centre of Data Analytics - University College Cork, actively participating in the ASSISTANT project with a focus on the development and deployment of AI components under the considerations of Responsible AI. Eduardo is a Chemical Engineer (Ph.D. from joint research between the University of Puerto Rico and the University of Arizona) that initially showed research interest in the modeling, simulation, and control domain. Given the trends of AI in the Academic and Industrial sectors, he started to focus more and more on different aspects of AI. Currently, his main contributions are in applying AI components in industrial cases and, at the same time, the development of them under the considerations of responsible AI.

Félicien: Eduardo, what is ethical by design, and why is it important?

Eduardo: Ethical by design is one of the different approaches that exist to implement ethical considerations within systems, processes, and, in general, any component that could be affected by a moral judgment or could confront ethical dilemmas. To understand this approach, it is

required to understand why it is crucial to include ethics within systems. Ethics can be seen as a standardized prescription of human behavior based on what is correct and incorrect under given circumstances. This "definition" clearly establishes the human as a centerpiece of the prescription and the behavior. When the systems include human participation or components that can be linked to human behavior, values, and trends, the incorporation of ethical frameworks within the system will allow it to be perceived positively by the humans interacting with the system. When a component is not designed under ethical considerations, it has a higher chance of failure during the deployment stage. For example, it is well document that in less than 24 hours, controversial concepts such as racist and misogynistic talks corrupted a chatbot. Could ethical frameworks have been deployed within this chatbot to reduce the chances of being corrupted by the media? Yes, it could. It is better to understand the importance of incorporating ethical consideration within systems; ethical-by design implies that structures, methods, or approaches are embedded in the system to perform based on ethical considerations.

Félicien: What is trustworthy AI, and why it matters?

Eduardo: For a human to develop trust, it is required that the agents or the systems interacting with the human "behave" or "perform" reliably over time. This concept also implies that risks need to be reduced to the minimum. At the same time, it is expected that this agent or system possesses the values (or ethical components) and trends in concordance to the human expectation. Just consider a bicycle, an aircraft, or a roller-coaster, which all have intrinsic risk when we use them but, how is can we develop trust in them? The same concepts apply to AI. AI is another technology, but the difference in AI is shown to impact society at a high level. So, we must develop trust over the components built under this "relatively new" technology. The trustworthy AI is a framework developed by the European High-Level Expert Group on AI That establishes the requirements that developed, deployed, and used AI components needed to fulfill to increase its reliability or trust over time. There are seven general requirements, some familiar to other science domains. This framework's importance

is to comply with ethical requirements and improve the acceptance of users of the developed AI components; this framework offers a first approach for establishing liability and responsibility of stakeholders that develop, deploy, and use AI components. There are no specific regulations on this matter at the current stage, but the need for them can be clearly seen.

Félicien: How can ethically designed and trusted AI be implemented in manufacturing? What will be the advantage of manufacturers to apply them?

Eduardo: The manufacturing sector is currently in the stage of incorporating AI components within its processes. These AI components can interact with humans within the manufacturing sector, such as robots, or perform pre-specified tasks or estimations and provide different industry stakeholders' desired feedback or estimations. Among the many benefits that the manufacturing sector can obtain from AI components, these include: improved quality of the products, reduce waste production, reduce lead times, control processes, interact collaboratively with workers to increase productivity, estimate market conditions, check raw material requirements, define/build system models, and reduce the number of sales lost. The advantages are well established, but understanding how to implement these tools, especially in industries that are not technologically ready for their implementation, makes this process cumbersome. In AI ethics, some of the most technified industries have already established some frameworks or ethical perspectives to develop and implement AI components. Nevertheless, these frameworks have been questioned since they are built based on the industry's interest and not society. Is this something negative and cannot be trusted? Is the industry sector not concerned about the outcomes of a poorly constructed AI? Well, that is why it important for the industry sector to use AI components based on generalized frameworks; they should use the different responsible AI frameworks, tools, guidelines, and approaches created by institutions that have a broader visualization on the limitations, benefits, and impact of AI components and their ethical considerations. To be more specific, the

manufacturing sector can develop AI components that can be trusted if the different responsible AI approaches are implemented during the development, deployment, and use of the AI components. These approaches involve both technical and non-technical components. For example, in terms of technical components, there are several tools under development to make the AI components outcomes explainable. These components can be used to provide users with an understanding of why a complex algorithm comes with a given solution. In terms of non-technical, there could be certification and standardization processes that would allow customers and users to develop trust in the AI components.

Félicien: What are the threats if ethical considerations on the application of AI tools are not considered in manufacturing environments?

Eduardo: To answer this question, let us create a simple scenario. Let us assume there is a child in a school and that this child steals a sandwich from a classmate. This action has "legal" components and ethical components. The legal components are established by society to regulate activities or actions. These legal components, which depend on the severity of the violation, are generally linked to ethical components since ethics deals with what is correct and incorrect. So, ethics can be seen as a framework for "expected" behaviors. One important consideration when we move back to AI is that AI is a component that could provide or perform activities like humans or interact with it in different ways. Therefore, AI could be seen as an agent that in case of failure, which can correspond to our case AI "steeling component", the responsibility of it should fall over the user, the developers, or in the deployers of the AI component.

Ethical considerations in AI can be seen as a general framework or method to establish AI components' requirements and behaviors. Currently, there is a considerable gap in terms of regulations of AI components and what is applied; therefore, ethical AI components can be implemented to safeguard and regulate the responsibility of each of the stakeholders involved in the AI.

So now, going back to your question. The problem of not considering the incorporation of ethical approaches in the manufacturing sector will be observed when an AI component fails. The implications involve brand name damage, loss of productivity, safety considerations, and legal considerations, among several other situations.

Félicien: Is AI so dangerous to have such a warning on the shoulders of manufacturers?

Eduardo: There are different perspectives of what AI can do and should do in the future. My view of AI is that by itself, is that it is not dangerous. What is dangerous are the users, its inappropriate intention, and a poor understanding of its strengths and limitations. There is a well-known saying in AI, garbage in garbage out. This is mentioned highly in the concept that AI is mainly driven by information. Therefore, depending on the quality and quantity of the data supplied to AI the results can be garbage. I would also add that these technology users, including developers, have the same level of responsibility. If there are no validation processes, a lack of understanding of the developed tools' implications, a poor ethical background of the company or institutions involved in the development and deployment, the results will also be garbage. I would say that the manufacturing sector must commit its level of responsibility in AI technology to the same or even higher level. This is because the higher the impact, the higher the responsibility of developing and deploying sound components.

Félicien: Great to know, Finally, what do you think of the ASSISTANT project?

Eduardo: My perspective is that ASSISTANT is a great project that could lead to incorporate responsible AI in the manufacturing sector and other domains. In ASSISTANT responsible AI is applied vertically on each of the approaches, tools, architectures, and frameworks that would have embedded AI components. The embedded AI components would participate in several tasks that include, among others:

decision-making, control (including process and robotic control), data cleaning, and modeling tasks. ASSISTANT would allow a better understanding of the limitations or strengths of different approaches that implement ethical considerations within AI development and deployment, but with a focus on the manufacturing domain. Additionally, ASSISTANT would facilitate tool components throughout the AI4EU portal that would allow stakeholders from the manufacturing domain and others to take advantage of the ASSISTANT developments to access pre-constructed components with responsible AI considerations. Finally, ASSISTANT would focus on defined frameworks for developing and deploying AI components specifically for the manufacturing sector.

Félicien: Thank you Eduardo.

Eduardo: Thank you for having me.