

# The Center for Responsible AI Project

**Maria Ana Henriques**

Unbabel

maria.henriques@unbabel.com

**Catarina Farinha**

Unbabel

catarina.farinha@unbabel.com

**Nuno André, António Novais, Sara Guerreiro de Sousa, Bruno Prezado Silva, Ana Oliveira, Helena Moniz, André Martins, Paulo Dimas**

Unbabel

{nuno.andre, antonio.novais, sara.guerreiro, bruno, ana.oliveira, helena, andre.martins, pdimas}@unbabel.com

## Abstract

This paper describes the project “NextGenAI: Center for Responsible AI”, a 39-month Mobilizing and Green Agenda for Business Innovation funded by the Portuguese Recovery and Resilience Plan, under the Recovery and Resilience Facility (RRF). The project aims to create a new Center for Responsible AI in Portugal, capable of delivering more than 20 AI products in crucial areas like “Life Sciences”, many of which use generative AI, particularly NLP models such as those for Machine Translation, contributing to translating into legislation the European Law included in the EU AI Act, and creating a critical mass in the development of responsible AI technologies. To accomplish this mission, the Center for Responsible AI<sup>1</sup> is formed by an ecosystem of start-ups and research institutions driving research in a virtuous way by addressing real market needs and opportunities in Responsible AI.

## 1 Introduction

McKinsey estimates that Generative AI could add the equivalent of \$2.6 trillion to \$4.4 trillion annually in value to the global economy — by comparison, more than 10x the Portuguese GDP. This would increase the impact of all artificial intelligence by 15 to 40 percent (Chui et al., 2023). The Center for Responsible AI aims to create the next generation of AI products, delivering business impact in this global market.

However, AI has inherent risks that are limiting its full potential, particularly in critical fields such as

Life Sciences, where AI's wrong decisions could endanger lives. Moreover, AI can cause harm to humanity, notably by jeopardizing privacy, exacerbating carbon emissions due to its demanding computational needs, and operating opaquely, impeding understanding of its operations, decision-making processes, and ethical alignment. Responsible AI technologies and principles are the key to bringing all the benefits of AI to humanity while ensuring sustainability and preventing harm.

To accomplish this mission, the Center for Responsible AI is formed by an ecosystem of AI start-ups that will create synergies with top research centers, driving research in a virtuous way by addressing real market needs. To ensure real-world impact, the consortium includes leading companies in key markets as technology adopters. The Center aims to create or leverage over 20 Responsible AI products<sup>2</sup>, many of which use generative AI, particularly NLP models such as those for Machine Translation while creating a critical mass of knowledge and talent in Responsible AI technologies. This initiative seeks to establish Europe as a global leader in Responsible AI, influencing principles and regulations in this domain, and contributing to EU AI legislation.

## 2 Project overview

The Center for Responsible AI's proposal received approval for a total investment of around €78 million that will be executed from October 2021 to December 2025. The Center's overall mission is to revolutionize the AI landscape responsibly and to develop the next generation of AI products that are explainable, fair, trustworthy, and sustainable. The project's main goals can be summarized as follows:

- Create the next generation of AI products driven by Responsible AI technologies by promoting a

© 2024 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

<sup>1</sup> <https://centerforresponsible.ai/>

<sup>2</sup> <https://centerforresponsible.ai/products/>

virtuous cycle between start-ups, research centers, and industry leaders. These products will have a real impact on society and solve new use cases for AI, opening new business opportunities in sensitive domains like Life Sciences, but also in crucial sectors like Retail and Tourism.

- Spearhead R&D efforts focused on practical solutions for universal challenges. From R&D teams at leading tech start-ups to world-class academic researchers, the Center focuses on applying cutting-edge research to everyday problems, tackling tangible issues using a responsible AI framework.

- Address crucial AI use cases through advanced machine learning and natural language processing approaches, including Machine Translation.

- Position Europe as a world leader in principles and policies of Responsible AI to influence European regulation which will be key to the future development of AI products.

- Retain and attract the best AI talent worldwide, the key ingredient to invent the future. This talent circulates from academia to start-ups transferring the knowledge that drives innovation into products.

## 2.1 Key partners and people

The Center is made up of ten start-ups (Unbabel, Automais, Emotai, NeuralShift, Priberam, Visor.ai, YData, Youverse, and two unicorns - Feedzai and Sword Health), eight research centers (Champalimaud Foundation, CISUC, FEUP, Fraunhofer Portugal AICOS, INESC-ID, IST, IST-ID/ISR and IT), a law firm (Vieira de Almeida) and five industry leaders in Life Sciences, Tourism and Retail (BIAL, Centro Hospitalar de São João, Luz Saúde, Grupo Pestana and SONAE).

## 2.2 Methodology

New mechanisms were put in place to facilitate the contacts between research centers and start-ups which significantly contribute to shortening the time required to transfer technologies from low TRLs<sup>3</sup> to a commercialization stage. To this end, the project was organized in 2 main streams:

1. The “Product Pods”, *i.e.*, smaller groups within the Consortium that put together research centers, start-ups, and technology takers centered on developing a specific product/technology, promoting the creation of a virtuous Research & Innovation circle. The Consortium architecture promotes this virtuous circle as it was designed for start-ups to bring product-driven research challenges to research centers. In this way, the Center’s top AI research

groups will be inspired to solve hard problems that will create market value, giving these products a competitive advantage globally.

2. Research Projects, in which the Center’s highly qualified research teams tackle complex research challenges that may not have a direct market application yet. However, the ultimate goal of the project is to promote an ecosystem in which the fundamental research being pursued lead to future product innovation in AI companies. The ongoing research projects can be grouped into 5 main areas: (i) Energy-Efficient and Sustainable AI; (ii) Privacy-Preserving AI Systems; (iii) Transparent, Fair, and Explainable AI; (iv) Language Technologies and Embodied Human-AI Interaction; and (v) Multilingual and Contextualized Conversational AI.

## 2.3 Current outcomes

The project has already accomplished the following outcomes:

- 18 product pods were created to develop AI products powered by Responsible AI technologies. An example is Unbabel’s “Translation for High Risk Content” which aims to offer a responsible way to produce AI-enabled translations in domains in which critical translation errors cannot be tolerated;
- 163 highly qualified jobs have been created;
- 40 PhDs and MScs are currently in progress;
- More than 20 scientific papers were published (e.g. Guerreiro et al., 2024 - which recently got accepted to TACL);
- 10 patent applications have been submitted or are under submission;
- €19 million has been invested in AI;
- A Position Letter<sup>4</sup> with recommendations for the EU AI Act was released.

## 3 Acknowledgements

This work was supported by the Portuguese Recovery and Resilience Plan (PRR) through project C645008882-00000055, Center for Responsible AI.

## References

- Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023, June 14). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.
- Guerreiro, N. M., Rei, R., van Stigt, D., Coheur, L., Colombo, P., & Martins, A. F. (2023). xcomet: Transparent machine translation evaluation through fine-grained error detection. arXiv preprint arXiv:2310.10482.

<sup>3</sup> Technology Readiness Level

<sup>4</sup> <https://centerforresponsible.ai/eu-ai-act-position-letter/>