The BridgeAI Project

Helena Moniz
University of Lisbon/INESC-ID
helena.moniz@campus.ul.pt

Joana Lamego
Champalimaud Foundation
joana.lamego@research.fchampalimaud.org

Nuno André, António Novais, Bruno Prezado Silva, Maria Ana Henriques, Mariana Dalblon, Paulo Dimas, Pedro Vale Gonçalves
Unbabel
{nuno.andre, antonio.novais, bruno, maria.henriques, mariana.dalblon.int, pdimas, pedro.goncalves}@unbabel.com

Abstract

This paper describes the project "BridgeAI: Boosting Regulatory Implementation with Data-driven insights, Global expertise, and Ethics for AI", a one-year science-for-policy research project funded by the Portuguese Foundation for Science and Technology (FCT). The project aims to provide decision-makers in Portugal with the best context to implement the EU Artificial Intelligence (AI) Act and bridge the gap between AI research and policy. Although not exclusively on machine translation, the project pertains to natural language processing in general and ultimately to each of us as citizens.

1 Introduction

World-leading researchers in artificial intelligence (AI) hold differing views on the potential risks of AI in the future and on the need and intensity of AI regulation (Novelli et al., 2023). These divergent views reinforce the need to align regulation and implementation with scientific evidence, informing decision-makers about real risks, opportunities, and future pathways.

Historically, the outcomes from science–policy interfaces have not proved to be straightforward and do not usually lead to the establishment of effective collaborations (Jagannathan et al., 2023). The process by which knowledge is transferred from scientists to decision-makers is usually considered ineffective, due to a lack of understanding.

Furthermore, the recent approval of the EU AI Act (AIA) by the European Parliament mandates swift implementation by Member States, presenting numerous societal and scientific challenges.

BridgeAI aims to respond to these challenges, by moving towards a context-based approach that facilitates the creation of actionable knowledge at the intersection of science and practical, ethical, social, legal, and political domains. BridgeAI's primary objective is to furnish decision-makers and relevant stakeholders with comprehensive contextual insights through the analysis of real-world case studies, and the collaboration between AI experts and decision-makers. By doing so, we aim to enhance the informed and effective implementation of the AIA in Portugal and empower stakeholders to transition from passive compliance with regulations to active participation in the responsible design of AI internationally (Floridi et al, 2018).

2 Project overview

BridgeAI's proposal for the science-for-policy project received approval in March 2024. Scheduled to initiate in April, the project entails six months of preparatory work followed by a three-day workshop in Lisbon. This workshop will convene experts from diverse fields organised into different working groups (WGs) to formulate recommendations for the application of the AIA by the Portuguese Public Administration. The concluding day of the workshop will feature presentations of findings and a roundtable discussion on broader topics, open to the public.

Subsequent to the workshop, a detailed analysis of discussions and recommendations will be conducted, culminating in the formulation of a positional paper to aid the Portuguese public administration in crafting a coherent strategy for the implementation of the AIA. The project is designed to impact beyond 2024, and should include broader activities topics such as AI literacy to the public domain.

Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2), pages 35–36
June 24-27, 2024 ©2024 European Association for Machine Translation

2.1 Key partners and people

Currently, BridgeAI counts with the following partners: Anacom, British Embassy Lisbon, Champalimaud Foundation, INESC-ID, Instituto de Telecomunicações, JLM&A, SGS, The Alan Turing Institute, Unbabel, and VdA. The team is investing a significant effort into engaging with more partners from the public administration.

2.2 Methodology

During the six-month preparatory work, distinct working groups (WG) will lay the groundwork for the workshop. Each WG, comprising approximately seven members, will focus on specific areas:

WG0 | AI technological case studies: Foundational and transversal WG that will create the case studies of AI products from the Center for responsible AI, serving as the basis for other WGs.

WG1 | Risk Assessment tools in AI products: Determine what should be in a practical AI risk assessment tool for public and private entities, based on tools already available to assess responsible AI principles (Morley et al., 2019).

WG2 | AI Ethics in Regulatory Processes: Taking into consideration the case-studies, the WG will define how we should address AI ethical concerns in the regulatory processes and how to provide ethical training at several levels.

WG3 | AI Act interface with other regulations, norms, audits and implementation metrics: Determine the key implementation initiatives that should arise to ensure the AI Act is effectively implemented and that all are conciliated (*e.g.*, certification, standards, audits and control).

WG4 | Advanced training and literacy: Define strategic measures for Portugal to increase levels of AI literacy and propose training programs to be developed. Based on actionable knowledge (Stern et al., 2020) methodologies and a diverse team, this WG will work to make responsible AI explicit to the public administration and citizens.

WG5 | AI ethics and regulatory efforts outside the EU: Point out best practices in AI regulation and ethics being developed outside the EU and understand how Portugal can learn from these or align best practices across legal frameworks.

To ensure the workshop is productive and insightful, preparatory work will feature periodic meetings and progress reports. Each WG will determine topics to be discussed in the workshop, and all members should be informed and aware of state-of-the-art scientific insights. Each WG will be led by a coordinator responsible for facilitating group efforts. Following the workshop, the project management team will distil the accumulated actionable knowledge to produce the expected outcomes and outreach to the core stakeholders.

2.3 Expected outcomes

The anticipated outcomes of the project encompass a positional paper to be submitted to the public administration, comprising well-founded, concrete, and actionable recommendations for the AIA implementation in Portugal, and an additional list of recommendations outlining a medium to long-term plan for the successful implementation of the AIA by the public administration, preparing the economy to the new paradigm of AI.

3 Acknowledgements

This project is funded by the Portuguese Science Foundation (FCT), under the science-for-policy programme, thematic area “Antecipar a regulação da Inteligência Artificial” reference 2023.10424.S4P23.

References

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schäfer, B., Valcke, P., & Vayena, E. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707.

Jagannathan, K., Emmanuel, G., Arnott, J., Mach, K. J., Bamzai-Dodson, A., Goodrich, K., Meyer, R., Neff, M., Sjostrom, K. D., Timm, K. M., Turnhout, E., Wong-Parodi, G., Bednarek, A. T., Meadow, A., Dewulf, A., Kirchhoff, C. J., Moss, R. H., Nichols, L., Oldach, E., … Klenk, N. 2023. A research agenda for the science of actionable knowledge: Drawing from a review of the most misguided to the most enlightened claims in the science-policy interface literature. *Environmental Science & Policy*, 144, 174–186..

Moniz, H. & Escartín, C. 2023. Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation. Springer, *Machine Translation: Technologies and Applications*, 4.

Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. 2019. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 26(4), 2141–2168.

Novelli, C. C., Casolari, F., Rotolo, A., Taddeo, M. & Floridi, L. 2023. Taking AI risks seriously: a new assessment model for the AI Act. *AI & SOCIETY*.

Stern, M. J., Briske, D. D., & Meadow, A. M. 2021. Opening learning spaces to create actionable knowledge for conservation. *Conservation Science and Practice*, 3(5), e378.