Intelligent Platform for Industry-Academia Collaboration in Research Project and Researcher Prospecting

Denis de Gois Marques[1], **Cleyton Mario de Oliveira Rodrigues**[1], **Wylliams Barbosa Santos**[1]

[1]Universidade de Pernambuco (UPE
Recife – PE – Brasil

{denis.marques, cleyton.rodrigues, wbs}@upe.br

Abstract. *One of the main factors influencing the adoption of new technologies in industry and government is the collaboration between Industry, Academia, and Government. Through Research and Development, it is possible to drive digital transformations and create new applied technologies. This paper proposes conducting Systematic Literature Reviews (SLR) and developing an intelligent platform that uses Natural Language Processing techniques and Large Language Models to enhance these collaborations. So far, we have completed two comprehensive SLR, developed two versions of a project prototype, and the goal is to enable information sharing and interconnection with institutions and researchers.*

1. Introduction

The knowledge gaps between industry and academia are frequently addressed in academic literature, standing out as one of the biggest challenges for collaboration between these two sectors. One of the main factors contributing to this disconnect is the fundamental differences in the goals and priorities of each institution. Industry focuses on the production, improvement, and commercialization of products, seeks practical and immediate results, aiming for profitability and market competitiveness. On the other hand, academia's primary mission is research and the creation of new knowledge, emphasizing scientific advancement and a deep understanding of specific phenomena. This disparity in objectives can hinder the exchange of information and the application of academic discoveries in the industrial context, while also making it challenging to adapt the needs of the industry to the pace and format of academic research [Sandberg and Crnkovic 2017][Garousi et al. 2019][Marques et al. 2022].

Collaboration practices between Industry and Academia (IAC) have very positive impacts, as communities can identify each other's needs and develop cooperation strategies to address the demands of both. The application of IAC practices in industry and academia can be found in academic papers such as [Garousi et al. 2016, Dallegrave and Santos 2023, Barbosa et al. 2020, Wohlin et al. 2012], which reaffirm the importance and value of these collaborations between industry and academia.

One of the main aspects of collaborations between industry and academia involves issues related to funding, contracts, and privacy within these partnerships. However, these aspects are not often addressed in a detailed and comprehensive manner in the scientific literature. Examples of this can be found in the paper by [Marques et al. 2022] and in

[Marques et al. 2023], which present Systematic Literature Reviews (SLR). These works have limitations in terms of citations and information on funding and contract issues in industry-academia collaborations.

The doctoral thesis research aims to spread knowledge about collaboration, funding, and partnership practices between industry, academia, and government (IAC), through the development of an intelligent platform for information systems and data that enhances these collaborations and supports decision-making. One aspect that is rarely explored in scientific papers on this topic is funding sources and direct collaboration with the government, which creates a deficit when it comes to industry, academia, and government (IAC) interaction, highlighting the need for new approaches to address this issue.

2. Theoretical Framework

It is essential to establish some key concepts, themes, and terminologies that are crucial for understanding the analysis of the research. In this way, we present the definitions of Collaboration between Industry, Academia, and Government (IAC), as well as Machine Learning (ML) and Natural Language Processing (NLP).

2.1. Collaboration between Industry, Academia, and Government

Collaboration between Industry, Academia, and Government is a robust mechanism for innovation, development, and knowledge transfer to the software development industry [Carver and Prikladnicki 2018] [Garousi et al. 2019] [Marijan and Gotlieb 2021]. To establish and conduct an efficient IAC, it is essential that the collaboration partners (researchers, professionals, and funding agencies) understand the motivations, goals, and needs involved in research projects, ensuring that these objectives are aligned across the different communities [Wohlin et al. 2012][Garousi et al. 2019].

Research on this topic is extensive and points to challenges in the integration between the parties involved in the collaboration, which still persist, with several articles highlighting the difficulties in hiring recent graduates in software engineering due to a misalignment of skills, both between industry requirements and the competencies addressed in academic curricula. Additionally, challenges related to the teaching-learning process for individuals transitioning into the field of Computing are also observed [Garousi et al. 2019] [Barbosa et al. 2020] [Dallegrave and Santos 2023] [Marques et al. 2024].

In this context, several articles present perspectives on the inclusion of researchers in industry (from an applied research perspective) and demonstrate positive results from this practice, ranging from process improvement to product development [Marques et al. 2023]. As an example, [Petersen et al. 2014] conducts two action-research studies at Ericsson, following five cyclical stages of action research development, which are: i) Diagnosis; ii) Action Planning and Design; iii) Action Execution; iv) Evolution; and v) Learning Specification. The research is applied based on the principles of action research, driving transformations in the software industry.

From an academic perspective, aiming to observe the contexts of these works, the research by [Marques et al. 2023] involved conducting a comprehensive literature review on IAC in software agility, employing a Systematic Literature Review (SLR) across five databases, and performing snowballing (both forward and backward). This study included

an analysis of 8,460 articles. As a result, ten categories were identified that specify challenges, and 14 that describe effective practices relevant to collaborative projects. Furthermore, the research provided a detailed description of seven collaboration models between industry and academia.

2.2. Machine Learning and Natural Language Processing

Natural Language Processing (NLP) aims to enable computers to understand and interact with human language effectively. This field seeks to incorporate various human skills into machines, such as the ability to understand, generate, extract relevant knowledge, and communicate, among others [Mcshane and Nirenburg 2021].

Data and information extraction from texts is one of the key applications of Natural Language Processing (NLP), a field within Computer Science aimed at converting human natural language into a formal representation, making it more easily processable by machines [Manning et al. 2014].

3. Project Plan and Methodology

The goal of this research is to disseminate knowledge about the possibilities of research partnerships, with or without funding prospects, and direct collaborations between industry, academia, and government (IAC), through the development of an intelligent platform that enhances collaborations between these sectors, using an information system that supports decision-making. This platform will assist in evaluating projects and calls for proposals from researchers, professionals, and Scientific and Technological Institutions (ICTs). The platform will automatically collect funding calls and analyze the content of these calls using Natural Language Processing (NLP) techniques, aiming to provide more appropriate guidance for professionals and researchers and validate proposals. The research has the following specific objectives:

- Unification of funding calls into a single platform (automatically performed);
- Conducting a mapping or systematic literature review on the use of natural language processing practices in research funding calls;
- Development of machine learning algorithms, based on natural language processing techniques, to train text classifiers for funding calls;
- Design and validation of an intelligent platform that, through the development of machine learning algorithms, can infer whether project proposals align with the funding call (from the perspectives of Academia, Government, and Industry) and guide the proposal to the most suitable funding call for the company's proposal (from the industry's perspective);
- As a form of proposal validation, continuous analyses by domain experts will be conducted.

3.1. Systematic Literature Review

A systematic literature review is a means of identifying, evaluating, and interpreting all available research on a particular phenomenon of interest, research question, or area of interest [Kitchenham et al. 2015]. This phase involved the development and execution of a Systematic Literature Review (SLR), aimed at identifying primary articles on "How are collaboration procedures between industry and academia occurring in the context of

Software Engineering?" [Marques et al. 2022, Marques et al. 2023]. The SLR presents a comprehensive review and snowballing approach regarding the collaboration between industry and academia in the context of Software Engineering. The study analyzed 8,460 articles from five academic databases (Scopus, IEEExplore, Science Direct, ACM and Springer) and identified 76 best practices and 37 challenges associated with this collaboration.

As a result of the SLR, seven collaboration models (methodologies for applying IAC) were described, along with ten categories and dozens of subcategories expressing challenges in IAC integration. Additionally, fourteen categories with numerous subcategories outlining best practices in collaborative projects were identified. In this way, connections were made between the challenges and mitigation best practices in collaborative projects.

3.2. Methodology

This work addresses the theme of collaborations between Industry, Academia, and Government (IAC), with an emphasis on the qualification of distributions, data processing, and the analysis of Research Funding Calls, as well as providing a platform capable of performing these processes. The main research methodology adopted in the project will be Design Science Research (DSR), which presents an iterative and dynamic process of continuous improvements to the created artifact.

Stage 1 consists of the ideas and issues of the research project, referenced at the beginning of the DSR cycle as the "Relevance Cycle". The project proposal is derived from previous studies already structured around the IAC theme, where Systematic Literature Reviews (SLR) were conducted on the topic, identifying knowledge gaps and information dissemination between institutions, as described in the previous section [Marques et al. 2022][Marques et al. 2023][Garousi et al. 2019].

Stage 2 involves the validation of the research proposals and the creation of a Minimum Viable Product (MVP) for the project in question. The goal of this stage is to validate the ideas and generate new development opportunities through collaborations between academia and government. Currently, the development of the first MVP has been completed, and we are in the process of validating the MVP, with a high-quality and innovative project already observable. Additionally, in this stage, we will reevaluate the project ideas, refining, discussing, and defining the appropriate scope for the project's development.

Stage 3 involves the progress of a new systematic review, currently in the final stages, with the aim of gaining a deeper understanding of the current research states in the project's theme. In addition to this development, a second round of MVP development was conducted, incorporating new features, focusing on the division and particularities of the platform's end-users (industry, academia, and government), bug fixes from the previous MVP version, the addition of user profiles, and initial perspectives on computational intelligence techniques.

In **Stage 4**, the construction of the platform takes place, transitioning from MVP to a usable digital platform, along with the development of computational intelligence for analyzing funding calls. The integration of computational inferences and intelligence goes through three main segments: the first segment involves the extraction and analysis

of information from funding calls, managed and extracted through Text Mining, Large Language Models (LLM), and Natural Language Processing (NLP) from the available calls, mainly using the transformers architecture. The second segment focuses on analyzing the compatibility of funding calls with the parties involved. This means performing a classification and alignment process with profiles, thereby selecting researchers, research projects, and industries that can be associated with specific themes. The third segment addresses the alignment of the funding call proposal with the submitted proposal. It is important to highlight that the practices in this stage are cyclical, involving phases of development and validation of the proposal.

In **Stage 5**, the product validation process with experts takes place to gather feedback. Interviews will be conducted with specialists and potential users of the platform. Based on the analysis and the collected feedback, adjustments may be made to the process, as well as the identification of opportunities for improving features of the project.

3.3. Current Status and Initial Results

The current phase involves refining and improving the research proposal, enhancing the theoretical framework, and adding new features to the MVPs already developed. Additionally, we are preparing a validation process with academics and industry professionals to further refine the proposal and generate new ideas. The MVP has already gone through two short development cycles (each lasting 5 weeks), following an open innovation and software factory development approach.

The developed MVP already performs automatic data collection (Web Scraping) from the websites of Organizations and Funding Agencies (ICTs), such as FACEPE (Foundation for the Support of Science and Technology of the State of Pernambuco), CAPES, CNPQ, and FINEP (Funding Agency for Studies and Projects). This data is made available on a digital platform that allows user registration and the management of project submissions. For users (in general terms), it is possible to submit funding calls outside the platform's scope and manage all available and favorited funding calls within the platform, among other features.

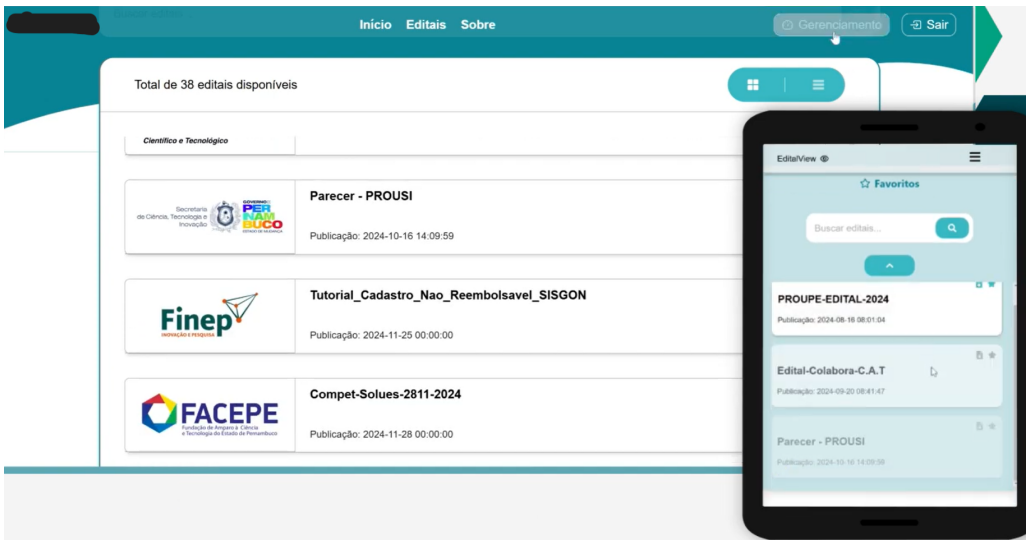

Figure 1. Minimum Viable Product version 2.0

In Figure 1, a frame of the application can be seen, listing some of the funding calls added to the platform. Currently, there is an ongoing effort to develop a new version of the MVP to introduce Artificial Intelligence practices and process automation (in an automatic and intelligent manner).

4. Conclusions

The main purpose of the research is to disseminate information about research funding and incentive processes, connecting all the vectors of digital transformation and innovation (Government, Industry, and Academia). Our platform aims to facilitate this knowledge dissemination and make the process more efficient and intelligent for decision-making, using computational intelligence practices. The stages described in the previous sections are still subject to discussion and may lead to changes in the scopes and activities.

5. Acknowledgements

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 001.

References

Barbosa, A., Galindo, G., Lencastre, M., Fagundes, R., and Santos, W. (2020). Fostering industry-academia collaboration in software engineering using action research: A case study. In *Anais do XIX Simpósio Brasileiro de Qualidade de Software*, pages 411–419, Porto Alegre, RS, Brasil. SBC.

Carver, J. and Prikladnicki, R. (2018). Industry–academia collaboration in software engineering. *IEEE Software*, 35:120–124.

Dallegrave, T. and Santos, W. B. (2023). Action research for industry academia collaboration : A replication study. In *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6.

Garousi, V., Petersen, K., and Ozkan, B. (2016). Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review. *Information and Software Technology*, 79:106–127.

Garousi, V., Pfahl, D., Fernandes, J., Felderer, M., Mäntylä, M., Shepherd, D., Arcuri, A., Coşkunçay, A., and Tekinerdogan, B. (2019). Characterizing industry-academia collaborations in software engineering: evidence from 101 projects. *Empirical Software Engineering*, 24.

Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). *Evidence-Based Software Engineering and Systematic Reviews*. Chapman & Hall/CRC.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In Bontcheva, K. and Zhu, J., editors, *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Marijan, D. and Gotlieb, A. (2021). Industry-academia research collaboration in software engineering: The certus model. *Information and Software Technology*, 132:106473.

Marques, D., Rocha, R., Santos, B., Pacheco, F., Rodrigues, C., and Santos, W. (2024). Industry academia and government collaboration to reduce gaps in software engineering: Applications for students and professionals in career transition. In *Anais do XXIII Simpósio Brasileiro de Qualidade de Software*, page 657–664, Porto Alegre, RS, Brasil. SBC.

Marques, D. d. G., Dallegrave, T. L. D. d. A., Barbosa, L. E. L., Rodrigues, C. M. d. O., and Santos, W. B. (2022). Industry-academy collaboration in agile methodology: a systematic literature review. In *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6.

Marques, D. d. G., Dallegrave, T. L. D. d. A., Rodrigues, C. M. d. O., and Santos, W. B. (2023). Successful practices in industry-academy collaboration in the context of software agility: A systematic literature review. In Filipe, J., Śmiałek, M., Brodsky, A., and Hammoudi, S., editors, *Enterprise Information Systems*, pages 292–310, Cham. Springer Nature Switzerland.

Mcshane, M. and Nirenburg, S. (2021). *Linguistics for the Age of AI*.

Petersen, K., Gencel, C., Asghari, N., Baca, D., and Betz, S. (2014). Action research as a model for industry-academia collaboration in the software engineering context. WISE '14, page 55–62, New York, NY, USA. Association for Computing Machinery.

Sandberg, A. B. and Crnkovic, I. (2017). Meeting industry-academia research collaboration challenges with agile methodologies. In *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*, pages 73–82.

Wohlin, C., Aurum, A., Angelis, L., Phillips, L., Dittrich, Y., Gorschek, T., Grahn, H., Henningsson, K., Kågström, S., Low, G., Rovegard, P., Tomaszewski, P., Van Toorn, C., and Winter, J. (2012). The success factors powering industry-academia collaboration. *IEEE Software*, 29:67–73.