

Colaborações entre a Indústria, Academia e Governo: Uma Plataforma Inteligente

Denis de Gois Marques (Doutorando)¹, Wylliams Barbosa Santos (Orientador)¹,
Cleyton Mario de Oliveira Rodrigues (Coorientador)¹

¹ Departamento de Engenharia da Computação
Universidade de Pernambuco (UPE) – Pernambuco, PE – Brasil

Data Inicio Doutorado: Setembro de 2022

Data Qualificação Doutorado: Julho de 2025

Data Final Doutorado: Setembro de 2026

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

Abstract. *One of the most impactful factors in the insertion of new technologies in industry and government is the implementation of Collaborative Practices between Industry and Academia (IAC), where through Research and Development (RD) it is possible to carry out digital transformations and the development of new applied technologies. This work proposes the development of a Systematic Literature Review (SLR) and the development of an intelligent platform that uses Natural Language Processing (NLP) and Large Language Models (LLM) practices to enhance these collaborations. Currently, we have developed an extensive SLR, a first version of an MVP of the project has been developed and we are organizing the new stages for improvements in the MVP, the system and the thesis. Future Results will allow sharing, interconnection with institutions and researchers and assist in better decision-making by participants.*

VIDEO LINK: <https://youtu.be/Gt80d00oTEg>

Resumo. *Um dos fatores impactantes nas inserções de novas tecnologias na indústria e governo são as realizações de práticas Colaborativas entre Indústria e Academia (IAC), onde através das Pesquisas-Desenvolvimento (P&D) é possível realizar transformações digitais e o desenvolvimento de novas tecnologias aplicadas. Esse trabalho propõe o desenvolvimento de uma Revisão Sistemática da Literatura (RSL) e o desenvolvimento de uma plataforma inteligente que utiliza práticas de Processamento de Linguagem Natural (PLN) e Large Language Models (LLM) para potencializar essas colaborações. Atualmente, desenvolvemos uma RSL extensa, foi desenvolvido uma primeira versão de um MVP do projeto e organizamos as novas etapas para melhorias no MVP, do sistema e da tese. Os Resultado Futuros vão permitir o compartilhamento, interconexão com as instituições e pesquisadores e auxiliar nas melhores tomadas de decisões dos participantes.*

LINK DO VÍDEO: <https://youtu.be/Gt80d00oTEg>

1. Introdução

As desconexões de conhecimentos entre a indústria e a academia são amplamente discutidas na literatura, onde um dos grandes desafios são as divergências entre os objetivos das instituições. A indústria tem como foco a

construção e venda do produto e a academia em desenvolver novos conhecimentos [Sandberg and Crnkovic 2017][Marques et al. 2022a][Garousi et al. 2019].

As práticas de Colaboração entre a Indústria e Academia (IAC) apresentam impactos muito positivos, onde as comunidades podem identificar as necessidades uma das outras e desenvolver estratégias de cooperação para sanar as demandas das comunidades. As aplicações de práticas de IAC na indústria e academia podem ser encontradas em artigos acadêmicos, como [Garousi et al. 2016] [Dallegrave and Santos 2023] [Barbosa et al. 2020] [Wohlin et al. 2012] e reafirmam a importância e valores dessas colaborações entre a indústria e a academia.

Um dos pontos principais das colaborações entre a indústria e academia são aspectos de financiamento, contratos e privacidade dessas colaborações. Esses aspectos não são detalhados de forma concisa e abundante na literatura científica, como por exemplo no artigo de [Marques et al. 2023], que descreve uma Revisão Sistemática da Literatura (RSL), limitando-se a poucas citações sobre aspectos de financiamento e contratos de colaboração.

Nesse sentido, a pesquisa de tese de doutorado tem como objetivo a difusão de conhecimentos sobre as práticas de financiamento e colaborações entre indústria, academia e governo, a partir do desenvolvimento de uma plataforma inteligente que potencialize as colaborações entre a indústria, academia e o governo (IAC) e auxilie em tomadas de decisão. Um dos pontos poucos explorados em artigos científicos, nessa temática, são as fontes de financiamento e a colaboração direta com o governo, sendo um contexto deficitário quando se refere a IAC, sendo necessária outras formas de abordagem na temática.

2. Referencial Teorico

Para garantir fluidez em nossas discussões, é importante definir alguns conceitos, temáticas e terminologias importantes para entender a discussão da pesquisa, dessa forma, trazemos as conceituações de Colaboração entre a Indústria, Academia e Governo e Aprendizado de Máquina e Processamento de Linguagem Natural.

2.1. Colaboração entre Indústria e Academia

As Colaboração Indústria, Academia e Governo é um mecanismo robusto de inovação e transferência de conhecimento para a indústria de desenvolvimento de software [Carver and Prikładnicki 2018] [Garousi et al. 2019] [Marijan and Gotlieb 2021]. Estabelecer e conduzir uma IAC eficiente é necessário que os parceiros de colaboração (pesquisadores e profissionais) entendam as motivações, os objetivos e necessidades envolvidas no projeto de pesquisa, sendo esses objetivos alinhados entre as comunidades [Wohlin et al. 2012] [Garousi et al. 2019].

Os desafios da integração entre as comunidades se estendem até o presente, onde diversos artigos descrevem as dificuldades nas contratações de recém-formados em engenharia de software devido os desalinhamentos de habilidades e do desalinhamento entre as habilidades requeridas pela indústria e pela apresentada nos currículos acadêmicos [Garousi et al. 2019][Barbosa et al. 2020][Dallegrave et al. 2021].

Diante do contexto, há diversos artigos apresentando as perspectivas de inclusão do pesquisador na indústria e que demonstram resultados positivos nessa prática. Como

exemplo, [Petersen et al. 2014] aplica duas pesquisas-ações na Ericsson, seguindo cinco etapas cíclicas de desenvolvimento da pesquisa-ação, sendo as seguintes etapas: i) Diagnóstico; ii) Planejamento e Design de Ação; iii) Tomada de Ação; iv) Evolução; e v) Especificação de Aprendizado. A pesquisa sendo aplicada partindo dos princípios de pesquisa-ação, realizando transformações na indústria de software.

O trabalho de [Marques et al. 2023] foi a condução de uma revisão abrangente da literatura em IAC em agilidade de software, empregando Revisão Sistemática da Literatura (SLR) e metodologias de bola de neve em cinco bancos de dados. Este estudo incluiu uma análise de 8.460 artigos. Como resultado, dez categorias foram identificadas que especificam desafios e 14 que descrevem práticas eficazes relevantes para projetos colaborativos. Além disso, a respectiva pesquisa forneceu uma descrição detalhada de sete modelos de colaboração entre a indústria e a academia.

2.2. Aprendizado de Máquina e Processamento de Linguagem Natural

O objetivo do Processamento de Linguagem Natural (PLN) é capacitar os computadores a lidar com a linguagem humana de forma eficaz. O contexto de processamento é a incorporação de diversas habilidades humanas nas máquinas, como entender, gerar, extrair conhecimento útil e comunicar-se, entre outras [Mcshane and Nirenburg 2021].

A extração de dados e informações a partir de textos é uma das aplicações de Processamento de Linguagem Natural (PLN), um campo dentro da Ciência da Computação que busca converter a linguagem natural humana em uma representação formal, de modo que se torne mais facilmente manipulável por máquinas [Manning et al. 2014].

3. Plano de Projeto

O objetivo da pesquisa é realizar a difusão de conhecimentos sobre as práticas de financiamento e colaborações entre indústria, academia e governo, a partir do desenvolvimento de uma plataforma inteligente que potencialize as colaborações entre a indústria, academia e o governo (IAC) e auxilie em tomadas de decisão. Esta plataforma auxiliará na qualificação de projetos e editais de pesquisadores, profissionais e Instituições Científicas e Tecnológicas (ICTs). A plataforma realizará a coleta automática de editais de fomento (através de Web Scraping) e analisará o conteúdo desses editais utilizando técnicas de Processamento de Linguagem Natural (PLN), com o intuito de proporcionar um direcionamento mais adequado para profissionais e pesquisadores e validações de propostas. A pesquisa tem as seguintes objetivos específicos:

- Unificação de Editais de Fomento em uma única plataforma (realizada de forma automática);
- Realização de um mapeamento ou revisão sistemática da literatura sobre práticas de processamento de linguagem natural em editais de fomento de pesquisa;
- Construção de algoritmos de aprendizado de máquina, baseados em técnicas de processamento de linguagem natural, para treinamento de classificadores de texto em editais de fomento;
- Construção e validação de uma plataforma inteligente, que a partir da construção do algoritmo de aprendizado de máquina, possa inferir se as propostas dos projetos são adequadas ao edital de fomento (nas perspectivas da Academia, Governo e Indústria) e o direcionamento para o edital mais adequado para a proposta da empresa (na perspectiva da Indústria);

- Validação da proposta com especialistas da área.

3.1. Metodologia

Este trabalho apresenta a temática de Colaborações entre Indústria, Academia e Governo (IAC), com foco na qualificação de distribuições e análises de Editais de Fomento a Pesquisa. A principal metodologia de pesquisa utilizada no projeto será o Design Science Research (DSR), e segue uma estrutura para o desenvolvimento, na Figura 1, é possível observar a modelagem arquitetural da plataforma, já na Figura 2 é possível observar as etapas de execução do projeto.

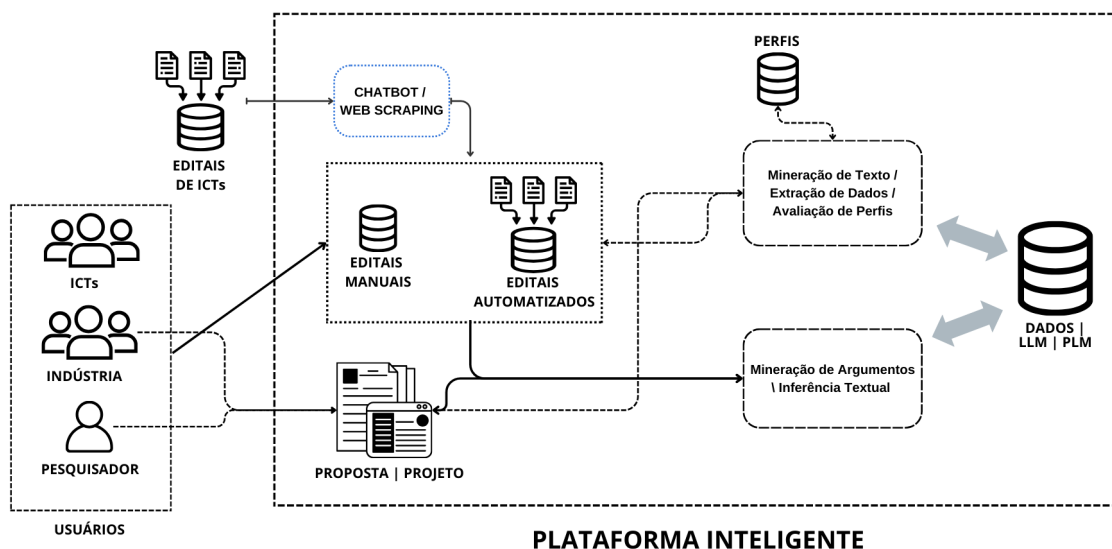


Figura 1. Desenho Arquitetural

A **Etapa 0** é constituída as ideias e problemáticas do projeto de pesquisa, referenciado no ciclo de DSR como “Ciclo da Relevância”. A proposta do projeto é derivada de estudos anteriores já formatados na temática de IAC, onde foram realizadas Revisões Sistemáticas da Literatura (RSL) sobre a temática, visualizando lacunas de conhecimento e difusão de informações entre as instituições [Marques et al. 2022b][Marques et al. 2022a][Garousi et al. 2019].

A **Etapa 1** envolve a validação das propostas de pesquisa e a criação de um Mínimo Produto Viável (MVP) para o projeto em questão. Essa etapa tem como objetivo validar as ideias e gerar novas oportunidades de desenvolvimento por meio de colaborações entre academia e governo. Atualmente, o desenvolvimento do primeiro MVP está concluído, estamos em processo de validação do MVP, e já é possível observar um projeto de alta qualidade e inovação. Além disso, nessa etapa, vamos reavaliar as ideias do projeto, refinando, discutindo e definindo o escopo adequado para o desenvolvimento do projeto. É importante destacar que no momento da escrita deste artigo estamos no período final dessa fase, onde poderá ter mudanças de escopo.

A **Etapa 2** consiste na formulação e execução de uma Revisão da Literatura (seja a execução de um Mapeamento Sistemático da Literatura, uma Revisão Sistemática da Literatura ou uma Literatura Cinzenta - *Grey Literature*), com finalidades de obter uma compreensão aprofundada nos estados atuais de pesquisa na temática do projeto, seja na

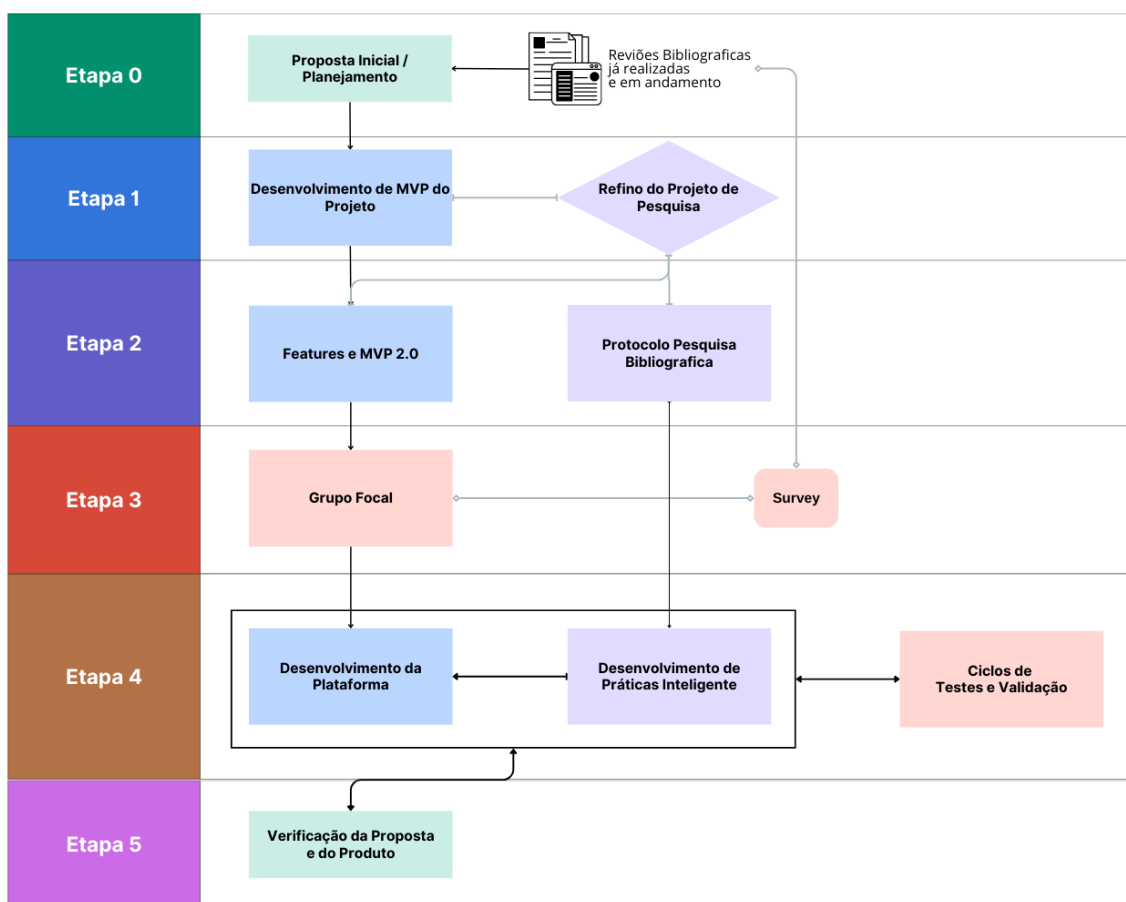


Figura 2. Etapas do Projeto

perspectiva de avaliação editais de fomento à inovação, na indicação de perfis de para públicos específicos ou em práticas de processamento de linguagem natural (PLN). Além dos pontos elucidados vai ser possível observar práticas e projetos semelhantes ao projeto atual.

Além do desenvolvimento da Revisão da Literatura vamos avançar em uma segunda rodada de desenvolvimento do MVP do projeto, adequando novas features de desenvolvimento, com foco na divisão e particularidades dos usuários finais da plataforma (indústria, academia e governo) e na perspectiva de adição de técnicas de inteligência computacional.

Na **Etapa 3** é caracterizada como Diagnóstico Organizacional, que é caracterizada por uma análise detalhadas da situação das organizações e identificar oportunidades, nesse diagnóstico vamos realizar consultas com especialistas e profissionais que devem utilizar a plataforma e possuem conhecimentos para detalhar novos insights e informações sobre as práticas de inovação. Esse processo pode ser realizado através de Surveys com especialistas, Grupos Focais ou Entrevistas com esses profissionais. As propostas iniciais para essa etapa é saber como avaliar um edital de fomento e como avaliar as propostas obtidas (na perspectiva de Governo) e se a plataforma é adequada ao uso (em ambas as perspectivas), diante desse cenário é importante a colaboração de participantes do meio de inovação, como participantes e profissionais das ICTs como CAPES, as FAPs, FINEP

e as Secretarias de Ciência e Tecnologia dos Estados.

Na **Etapa 4** é a construção da plataforma, passando de MVP para uma plataforma digital utilizável e o desenvolvimento de inteligências computacionais para análises de editais de fomento. A inserção de inferências computacionais e de inteligência computacional passam por três segmentos principais, sendo: Em primeiro segmento são as extrações e análises de informações dos editais de fomento, sendo geridas e extraídas por Mineração de Texto, Grandes Modelos de Linguagem (LLM) e PLN dos Editais disponibilizados. O segundo segmento estão as análises de compatibilidade de editais com os envolvidos, ou seja, realizar um processo de classificação e alinhamento com os perfis, dessa forma selecionando os pesquisadores, as pesquisas e as indústrias que podem ser relacionadas a partir de determinadas temáticas. E o terceiro segmento é a adequação da proposta do edital, com a proposta submetida. Importante destacar que as práticas dessa etapa são cíclicas, onde vão passar por fases de desenvolvimento e validação da proposta.

Na **Etapa 5** são as validações do produto com especialistas, para obter feedbacks serão realizadas entrevistas com especialistas e possíveis usuários da plataforma. Com base nas análises e no feedback coletado podem haver alterações no processo, além de identificar oportunidades de melhores *features* para o projeto.

3.2. Estado Atual e Resultados Iniciais

Atualmente, estamos trabalhando na melhoria da proposta inicial da pesquisa. Estamos preparando as propostas e o MVP (Produto Mínimo Viável) para discussão com acadêmicos e profissionais da área, com o intuito de refinar a proposta e gerar novas ideias. O MVP foi desenvolvido em apenas 5 sprints de uma semana cada, dentro de um projeto de inovação aberta. Apesar da falta de experiência prévia dos estudantes e profissionais envolvidos em pesquisa acadêmica e desenvolvimento de software, conseguimos entregar um MVP de alta qualidade.

O MVP desenvolvido já realiza a coleta automática de dados (*Web Scraping*) em páginas de Organizações e Agências de Financiamento (ICTs), como a FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) e a FINEP (Financiadora de Estudos e Projetos). Esses dados são disponibilizados em uma plataforma digital, que permite o cadastro de usuários e o gerenciamento de submissões de projetos. Para usuários (em perspectivas gerais), é possível submeter editais fora do escopo da plataforma e gerenciar todos os editais disponíveis e favoritos dentro da plataforma.

4. Conclusões

A principal função da pesquisa é divulgar informações sobre os processos de incentivo e fomento à pesquisa, conectando todos os vetores da transformação digital e inovação (Governo, Indústria e Academia). Nossa plataforma tem como objetivo facilitar essa disseminação de conhecimento e tornar o processo mais ágil e inteligente para a tomada de decisões, utilizando práticas de inteligência computacional. As etapas descritas nas seções anteriores ainda estão sujeitas a discussão e podem levar a alterações nos escopos e nas atividades.

5. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

- 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.
- Dallegrave, T., Vasconcelos, G., Alves, G., and Santos, W. (2021). Are we ready? identifying the gap between academia and software industry in the context of agile methodologies. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 35–47, Porto Alegre, RS, Brasil. SBC.
- 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.
- 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. d. G., Dallegrave, T. L. D. d. A., Barbosa, L. E. L., Rodrigues, C. M. d. O., and Santos, W. B. (2022a). 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., Barbosa, L. E. L., Rodrigues, C. M. d. O., and Santos, W. B. (2022b). 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.