Evolving Practices in Distributed R&D&I Projects: Bridging Academia and Industry Through Lightweight Collaboration

Lincoln Rocha
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
lincoln@dc.ufc.br

Paulo Antônio Rêgo
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
paulo@dc.ufc.br

Fernando Trinta
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
trinta@dc.ufc.br

Emanuele Santos
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
emanuele@dc.ufc.br

César Mattos
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
cesarlincoln@dc.ufc.br

Karla Souza
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
karla.soares@pdi.dc.ufc.br

Camilo Almendra
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
camilo.almendra@ufc.br

Maria Alice Macedo
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
alice.souza@pdi.dc.ufc.br

Isabela Carvalho
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
isabelacarvalho670@gmail.com

Julia Nobrega
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
julia.bastos@pdi.dc.ufc.br

Maria Eduarda Vieira
Alan Turing Lab
Federal University of Ceará
Fortaleza, Ceará, Brazil
eduardasv@pdi.dc.ufc.br

Eder Soares
Centro de Pesquisa, Desenvolvimento e Inovação Dell (LEAD) – Dell Technologies
Fortaleza, Ceará, Brazil
eder.soares@dell.com

ABSTRACT

This presentation reports the best practices and methods adopted by the Alan Turing Lab (ATLab) research group in the past years, running distributed research, development, and innovation (R&D&I) projects with a global industry partner. This work highlights the software engineering processes that were iteratively refined through years of collaboration. We also share insights in two domains where our researchers stand out: Visualization and Artificial Intelligence. This report may be useful for research groups and organizations beginning to engage in innovation partnerships.

KEYWORDS

Academy/Industry Partnership, Research, Development, Innovation, Experience Report

1 Introduction

The ATLab[1] and Dell Technologies collaboration launched at the onset of the COVID-19 pandemic. Although the initial team comprised researchers with experience in software engineering, the partnership confronted immediate remote work challenges. This scenario prompted new approaches to coordinating and delivering software research projects across distributed teams. Since 2020, ATLab has led more than twenty distributed development projects with Dell, involving team members in many cities in the state of Ceará, other states in Brazil, the United States, India, and Ireland. This multi-site setup demanded strong coordination, adaptability, and agile practices to ensure project alignment and success. This paper aims to document this evolving model, highlight lessons learned, and provide insights into how agile governance, cross-organizational alignment, and distributed collaboration matured over time within this unique academic-industry partnership.

2 Distributed Collaboration and Project Management

The ATLab can be described as a hybrid, dynamic ecosystem where academic teams work directly with corporate structures on projects across multiple cities and countries. Project teams mix undergraduate and graduate students, beginners and senior researchers, and junior and senior contractors. This environment fosters mutual learning, with academia gaining exposure to industry standards and infrastructure, and industry benefiting from innovative methods and talent pipelines. The team dispersion, compounded by the remote work model imposed by the COVID-19 pandemic, required flexible coordination strategies and adaptations to time zone differences and cultural contexts [3].

The continuous onboarding of new team members and the risk of staff turnover further contributed to instability, prompting iterative practices like ongoing legal regulatory report documentation and a shift from synchronous to asynchronous daily communications. Early onboarding practices also transitioned into a more curated

[1] https://www.atlab.ufc.br

and self-guided format to accommodate diverse schedules and team structures.

As project complexity increased, we sought more structured and responsive approaches. We sought to establish regular stakeholder alignment, early design validation, and focus on continuous development, to help bridge the differing cultures and paces of academia and industry. Agile methodologies, especially Scrum, were central to managing this geographically distributed work [8]. ATLab and Dell collaboration faced complex challenges due to the intersection of academic flexibility and corporate structure. Key difficulties included balancing financial and HR coordination with technical oversight, often amid unpredictable variables like staff turnover, delayed hiring, and shifting budgets. The evolving scope and technical uncertainty also demanded significant adjustments to standard agile practices. Core practices like continuous documentation and asynchronous updates became essential. Ongoing efforts focus on improving integration between agile planning and exploratory work, enabling earlier risk identification and greater stakeholder engagement in key decisions—critical for managing complex, innovation-driven environments.

3 Requirements Engineering

We faced significant challenges due to the uncertain scope and the lack of clearly identified target users in the early stages. This uncertainty made traditional methods ineffective, as requirements evolved with the team's growing understanding of the problem. Continuous elicitation, iterative refinement, and adaptable processes became essential. Communication issues, inconsistent terminology, and vague acceptance criteria increased the risk of misalignment between outcomes and business goals.

To manage these challenges, the teams adopted agile prototyping, stakeholder interviews, and short validation cycles for both functional and non-functional requirements. The selection of stakeholders for interviews is challenging, and we observed that a top-down approach fosters rapid convergence of scope. Often, R&D&I projects originate from top management, so mid-manager and technical personnel may not be aware of the research's high-level goals. Indeed, they are key to understand the current practices and to validate solutions, and the scope previously aligned with top stakeholders helps navigating in the their knowledge corpus. Requirements were documented succinctly, and tied to the backlog for easy updates. Rigid documentation was replaced with concise, evolving artifacts focused on sprint-relevant needs. Nowadays, we still need to improve acceptance criteria and deal with non-technical stakeholder representatives.

4 DevOps and Testing

The testing process needed to be more structured for both manual and automated approaches. In a distributed environment, it is easy to get used to share too much information and artifacts by messaging, thus making later recollection harder. To improve this, major changes were implemented across the testing workflow, Wiki-like tools were adopted as the central reporting tool, evolving into a shared QA database. Test documentation moved from scattered artifacts to structured spreadsheets. Beyond that, [10] details the current testing framework, developed specifically for an ATLab project, including practices in development of automated tests and its impacts, the use of GitLab's CI/CD to perform DevOps, managing tests and product quality. More specific aspects of the improvement of the testing process can be found there. We devised our current testing framework to align with the distributed setting and typical requirements of the projects, while fostering solid test management practices [10]

Other practices were adapted to fit the hybrid, distributed context. Developer communication became more focused, and technical checkpoints outside standard agile ceremonies helped synchronize testing and development. Yet, challenges persist, including scaling the framework to other ATLab projects, improving QA documentation, and aligning testing with evolving, often undocumented requirements.

5 Design and Visualization Experience

Visualization research and real-world design often diverge. While academic work aims to produce generalizable models and validated techniques, practical design is shaped by constraints like time, organizational dynamics, and domain-specific needs. As [7] notes, practitioners tend to rely on intuition and context-driven decisions rather than prescriptive frameworks. This disconnect is especially visible in collaborative projects, where shifting goals and unclear requirements demand flexibility from researchers to foster meaningful partnerships.

We often encounter vague objectives and evolving data sources. Design decisions unfold iteratively through conversations, sketches, and prototypes. Instead of rigid methodologies, we prioritize contextual exploration and collaboration, using tools like user journeys and stakeholder interviews to clarify expectations. This flexible approach parallels other adaptations in our process, such as asynchronous agile rituals and evolving requirement artifacts that accommodate change.

Another challenge involves integrating visualizations into existing BI tools, like Power BI or Tableau, which often lack necessary customization. Rather than replacing them, we extend their functionality using D3.js and React, bridging the gaps between user needs and technical systems. Ensuring long-term sustainability also requires attention to maintenance, onboarding, and evolving organizational practices. Effective visualization extends beyond innovation – it involves continuity, care, and integration into everyday workflows [1].

6 Artificial Intelligence

It is well known that, in the context of practical tools, the challenges of planning, developing, deploying, and maintaining a useful real-world AI system are even greater than designing the underlying machine learning (ML) algorithms [6]. This scenario resulted in the growth of the field called Machine Learning Operations (MLOps), which aims to systematize and automate most of the ML pipeline [2, 4].The classical CRISP-DM methodology [11] has also been revisited, for example, to focus on ML and quality assurance, which leads to the CRISP-ML(Q) process [9]. Despite the general MLOps recommendations and frameworks, the routine of R&D&I projects demand additional fine-tuning, considering characteristics of both the project team and the client.

The first challenges arise right at the beginning of the project. In the initial meetings and rounds of requirements elicitation, the team needs to translate the client demands, which are often operational, into a properly defined learning task that can be addressed from data. This phase presents important questions: (i) to what degree can the targeted operation be automated? (ii) how can the business knowledge embedded in the client be leveraged? (iii) is there data available to inform the ML models to be developed? The latter frequently involves a laborious data discovery step and the definition of a coherent labeling procedure, which can convert business rules into automated rules or, sometimes, require an active learning phase for semi-automated annotation of data patterns [5]. After the business and data understanding steps, the research team needs to investigate the state-of-the-art in the literature, while also gathering the available tools and software frameworks. The choice of which tools will comprise the solution stack must be taken with care, since it must meet, at the same time, the technical requirements, the intended use licenses, and the working experience of both the client and the team.

The development and prototyping of the ML workflow might be, at first, concentrated in the project team. However, we have found out that frequent rounds of assessment with the client, especially to coordinate the model evaluation metrics with the final application, is important. For example, in standard benchmarks, classifiers are evaluated in terms of accuracy, i.e., rate of correctly classified examples. However, in practice, it is common for the application to require the AI model to return only the top-K instances that most likely belong to a given class, with the amount K being defined by the context that the provided information will be handled.

Also, an ongoing effort has been devoted by us to better deal with the technology transfer and appropriation phase by the client. Even after surpassing the perilous road to deploy the proposed workflow in a production environment [6], we have noticed that a successful conclusion for the project is highly dependent on the client being capable of appropriating themselves with the AI-based solution. This stage is not only fundamental for the sake of continuous maintenance, but also in terms of overall understanding of the implemented methodology, including its strengths and limitations. We envision that such challenging tasks should be addressed along all the project duration and focus on the education of the client on the main concepts behind the designed solution.

7 Conclusions

This presentation is geared toward practitioners involved in the planning, execution, and support of software projects in distributed collaborative settings. Project managers and product owners may benefit from insights into cross-organizational coordination and agile governance. Requirements analysts will find value in the structured interface between stakeholders and compliance teams. Software developers and architects—particularly those working remotely or across time zones—will relate to the challenges of technical alignment in global teams. QA testers and support analysts may gain perspective on integration and delivery processes within a shared infrastructure.

Practitioners may gain insights into cross-organizational coordination, agile governance, stakeholder alignment, and the challenges of technical integration and delivery in distributed, collaborative environments. Additionally, educators and researchers will find the real-world context useful for teaching or studying agile practices in geographically distributed environments.

REFERENCES

[1] Derya Akbaba, Devin Lange, Michael Correll, Alexander Lex, and Miriah Meyer. 2023. Troubling collaboration: Matters of care for visualization design study. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15.

[2] Josu Diaz-De-Arcaya, Ana I Torre-Bastida, Gorka Zárate, Raúl Miñón, and Aitor Almeida. 2023. A joint study of the challenges, opportunities, and roadmap of MLOps and AIOps: A systematic survey. *Comput. Surveys* 56, 4 (2023), 1–30.

[3] Christof Ebert, Marco Kuhrmann, and Rafael Prikladnicki. 2016. Global software engineering: Evolution and trends. In *2016 IEEE 11th International Conference on Global Software Engineering (ICGSE)*. IEEE, 144–153.

[4] Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirschl. 2023. Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE access* 11 (2023), 31866–31879.

[5] Robert Munro Monarch. 2021. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI.* Simon and Schuster.

[6] Andrei Paleyes, Raoul-Gabriel Urma, and Neil D Lawrence. 2022. Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys* 55, 6 (2022), 1–29.

[7] Paul Parsons. 2021. Understanding data visualization design practice. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 665–675.

[8] Anna Börjesson Sandberg and Ivica Crnkovic. 2017. Meeting industry-academia research collaboration challenges with agile methodologies. In *39th Intl. Conf. on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*. IEEE, 73–82.

[9] Stefan Studer, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters, and Klaus-Robert Müller. 2021. Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine learning and knowledge extraction* 3, 2 (2021), 392–413.

[10] Maria Vieira, Vitor M. de Lima, Windson Viana, Michel Bonfim, and Paulo Rego. 2025. Enhancing Continuous Integration Workflows: End-to-End Testing Automation with Cypress. In *27th Intl. Conf. on Enterprise Information Systems (ICEIS)*. 160–167. doi:10.5220/0013230200003929

[11] Rüdiger Wirth and Jochen Hipp. 2000. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Vol. 1. Manchester, 29–39.