RELAZIONE SECONDO ANNO RTD/B (PERIODO OTTOBRE 2024 – SETTEMBRE 2025)

Juri Di Rocco
Dipartimento di Ingegneria, Scienze Informatice e Matematiche
Università degli Studi dell'Aquila
juri.dirocco@univaq.it

October 7, 2025

ABSTRACT

This report provides an overview of the research, teaching, service, and dissemination activities carried out by Dr. Juri Di Rocco (RTD/B) at DISIM during the period October 1, 2024 – September 30, 2025 (A.Y. 2025/2026). The research activities focused on advancing software engineering and model-driven engineering through the application of large language models (LLMs), empirical methodologies, and innovative modeling platforms. The main scientific contributions include: novel approaches for the detection and classification of Self-Admitted Technical Debt using NLP and LLMs; an empirical analysis of the energy consumption of ATL transformations; a comprehensive investigation into the role of LLMs within MDE ecosystems; the development and dissemination of *jjodel*, a cloud-based reflective modeling platform; and several studies leveraging LLM-based multi-agent systems for the summarization of software documentation and mobile app descriptions. Additional contributions addressed challenges in code usage mining, model card enrichment, popularity bias in library recommendation, and declarative constraint modeling in SaaS environments. These activities resulted in multiple journal articles, international conference and workshop publications, and one extended journal submission currently under review. Alongside these research efforts, Juri actively participated in national and international projects, coordinated and contributed to teaching activities at both bachelor and master levels, supervised several theses, served on editorial boards and program committees, and co-organized international workshops and conference tracks.

1 Research activities

During the past year, Juri's research focused on advancing software engineering and model-driven engineering through the use of large language models (LLMs), empirical methods, and novel modeling platforms. The following subsections summarize the main research contributions, each corresponding to published or under-review papers.

1.1 Detecting Self-Admitted Technical Debt with LLMs

Technical debt (TD) arises when developers make deliberate trade-offs between short-term efficiency and long-term maintainability. Self-Admitted Technical Debt (SATD) refers to textual comments where developers explicitly acknowledge such trade-offs. This work investigated to what extent binary and multi-class classification of SATD has been addressed in existing research and introduced *SALA*, a dual-purpose tool that leverages NLP techniques, neural networks, and LLMs to detect SATD in source code comments. Through a literature review and empirical evaluation, the study showed that while binary classification is well explored, multi-class classification remains under-investigated. SALA outperformed state-of-the-art baselines in both settings, demonstrating the potential of LLMs for more fine-grained SATD detection **[J1]**.

1.2 Energy Consumption of ATL Transformations

Model transformations are core operations in model-driven engineering, yet their energy footprint remains poorly understood. This work conducted an empirical study of ATL transformations, analyzing correlations between transformation characteristics, model size, and energy consumption. The study revealed measurable and systematic relationships between energy usage and structural properties of transformations, offering practical guidelines for designing greener transformation pipelines **[J2]**.

1.3 The Role of LLMs in Model-Driven Engineering

This research provided a comprehensive analysis of the opportunities and challenges of applying large language models in Model-Driven Engineering (MDE). By systematically examining key areas such as model repository classification, co-evolution, recommender systems, and collaborative modeling, the paper proposed a structured research agenda for integrating LLMs into MDE workflows. It emphasized how LLMs can support language engineering tasks and accelerate traditionally manual processes **[J3]**.

1.4 Jjodel: Bridging Usability and Complexity in MDE

Traditional MDE tools are often limited by installation and maintenance overhead, hindering their adoption in education and industry. *Jjodel* is a cloud-based, low-code modeling platform designed to overcome these limitations by offering a zero-setup experience and a reflective architecture that adapts dynamically to metamodel and model changes. This work described the conceptual and architectural foundations of Jjodel, illustrating how it reconciles usability with the expressiveness required for advanced modeling tasks **[J4]**.

1.5 LLMs-based Multi-Agent Systems for README Summarization

With the increasing number of GitHub projects, README files have become central for developer onboarding, yet they often lack effective summaries. This work proposed *MetaGente*, a multi-agent system leveraging LLMs to generate structured summaries of README.md files. By coordinating multiple specialized agents, the approach achieves higher quality summaries than single-agent baselines, improving the accessibility of software documentation **[C1]**.

1.6 Mining Code Usage Patterns to Enrich Model Cards

This work introduced *CodeXHug*, a dataset and approach for enriching Hugging Face model cards with real-world code usage patterns mined from GitHub. By clustering and synthesizing code snippets through an LLM, the method identified recurring usage patterns that differ from official model card examples, suggesting ways to automatically enrich model documentation. The dataset covers thousands of models and hundreds of thousands of code snippets **[C2]**.

1.7 Modeling Smart Ecosystems

In collaboration with several leading researchers in the MDE community, this work articulated a research vision positioning modeling as the heart and soul of engineering smart ecosystems. It discussed the role of modeling in ensuring trustworthiness, composability, and sustainability in complex socio-technical systems, setting directions for future research **[C3]**.

1.8 Mitigating Popularity Bias in Library Recommendations

This workshop paper explored how LLMs can be leveraged to address popularity bias in third-party library recommendation systems. By analyzing and generating recommendations that go beyond the most frequently used libraries, the approach aims to improve fairness and diversity in library adoption, supporting developers in discovering relevant but less popular dependencies **[W1]**.

1.9 Declarative Constraint Modeling in SaaS Tools

This contribution proposed an approach for bringing declarative constraint modeling—traditionally expressed in OCL—into modern Software-as-a-Service (SaaS) modeling tools by means of JSX-based representations. It demonstrated how this integration enables the expression of validation logic directly within web-based low-code environments, improving accessibility and maintainability of constraints **[W2]**.

1.10 LLM-based Summarization of Mobile App Descriptions

Finally, a journal article currently under second-round review at the *Automated Software Engineering (ASE)* journal extends the MetaGente framework to the mobile application domain. The work focuses on summarizing large-scale mobile app descriptions, combining multi-agent LLM architectures with tailored prompt engineering strategies to improve coverage and accuracy. The approach aims to support automated analysis of app marketplaces and software ecosystems **[R1]**.

2 Research products

2.1 Journals

- **J1** Francesca Arcelli Fontana, Juri Di Rocco, Davide Di Ruscio, Amleto Di Salle, Phuong T. Nguyen, *Binary and multi-class classification of Self-Admitted Technical Debt: How far can we go?* Information and Software Technology, Volume 187, 2025, 107862. https://doi.org/10.1016/j.infsof.2025.107862.
- **J2** Riccardo Rubei, Juri Di Rocco, Davide Di Ruscio, *On the Energy Consumption of ATL Transformations*. Software: Practice and Experience, 2025. (Wiley)
- **J3** Juri Di Rocco, Davide Di Ruscio, Claudio Di Sipio, Phuong T. Nguyen, Riccardo Rubei, *On the use of large language models in model-driven engineering*. Software and Systems Modeling, 24(3), 923–948, 2025.
- **J4** Bucchiarone, A., Di Rocco, J., Di Vincenzo, D. et al. Modeling in Jjodel: towards bridging complexity and usability in model-driven engineering. Softw Syst Model (2025). https://doi-org.univaq.idm.oclc.org/10.1007/s10270-025-01324-y

2.2 Conferences

- **C1** Duc S. H. Nguyen, Bach G. Truong, Phuong T. Nguyen, Juri Di Rocco, Davide Di Ruscio, *Teamwork makes the dream work: LLMs-Based Agents for GitHub README.MD Summarization*. In Proc. of the 33rd ACM Int. Conf. on the Foundations of Software Engineering (FSE Companion '25), Trondheim, Norway, June 23–28, 2025, pp. 621–625.
- **C2** Stefano Palombo, Claudio Di Sipio, Juri Di Rocco, Davide Di Ruscio, *Generate with CodeXHug: A Dataset to Enhance Model Cards with Code Usage Patterns*. In Proc. of Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2025), Springer Nature, pp. 327–344, September 2025.
- **C3** Antonio Bucchiarone, Benoit Combemale, Alfonso Pierantonio, *et al.* (incl. J. Di Rocco), *Modeling: The Heart and Soul of Engineering Smart Ecosystems*. In Proc. of SAM 2025 (to appear).

2.3 Workshops

- **W1** Claudio Di Sipio, Juri Di Rocco, Davide Di Ruscio, Vladyslav Bulhakov, *Addressing Popularity Bias in Third-Party Library Recommendations Using LLMs*. Fairness Workshop co-located with RENE 2025.
- **W2** Antonio Bucchiarone, Juri Di Rocco, Damiano Di Vincenzo, Alfonso Pierantonio, *From OCL to JSX: Declarative Constraint Modeling in Modern SaaS Tools*. OCL Workshop 2025, co-located with STAF 2025, Koblenz, Germany.

2.4 Under review

- **R1** Duc S. H. Nguyen, Minh T. Nguyen, Phuong T. Nguyen, Juri Di Rocco, Davide Di Ruscio, *Automated Summarization of Software Documents: An LLM-based Multi-Agent Approach*. Revised version under second-round review at *Automated Software Engineering (ASE) Journal*. (Extended Metagente for mobile app description summarization)

3 Participation In National And International Research Projects

From June 2022: PRIN 2021 research project consortium "EMELIOT (Engineered MachinE Learning-intensive IoT systems)." The project investigates methods for engineering dependable, ML-intensive IoT systems spanning edge, gateways, and cloud orchestration.

From January 2025: **MOSAICO** project funded by the European Union under the Grant Agreement No. 101189664. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the

European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

4 Teaching Activities - Academic year 2024/2025

- Laboratorio di programmazione ad oggetti. Università degli Studi dell'Aquila, Corso di Laurea in Informatica (full coordinator – 6 CFU, 60 hours).
- Model-Driven Engineering (Prof. Alfonso Pierantonio), Università degli Studi dell'Aquila, Corso di Laurea in Informatica (guest lectures).
- Introduction to Programming [DT0712], PRE-MASTER'S FOUNDATION PROGRAMME IN APPLIED MATHEMATICS (3 CFU, 18 hours).
- Laboratorio di didattica della programmazione al PFA (1 CFU, 12 hours).

5 Thesis supervision

Juri has supervised and co-supervised the following theses at the Università degli Studi dell'Aquila:

- **Vinay Sanga** (Master) — *Automatic evaluation of critical cargo: A Computer Vision approach*. EDISS (Erasmus Mundus), in collaboration with Ahola Digital. Co-supervised with Phuong T. Nguyen and Toni Penttinen.
- **Lorenzo Baldacci** (Bachelor) — *Modelli AI per l'Identificazione, Classificazione e Sintesi di Pipeline Data Science*.
- **Stefano Palombo** (Bachelor) — *Utilizzo degli LLM per Generare Hugging Face Model Card da esempi di codice*.
- **Ercole Rutolo** (Bachelor) — *Sistema multi-agente per code refactoring basato su attributi di qualità*.
- **Antonio Addario** (Bachelor) — *Automatic generation of Pull Request messages using Large Language Models*.

6 Institutional, organizational and university services

- **Department Delegate for Placement and Orientation** (DISIM).
- **Member of the CAD Orientation Committee**.
- **Member of the University Orientation and Tutoring Committee**.
- **Member of the PFA Exam Committee**.

7 Organization And Participation Of International Conferences And Workshops

7.1 Event Organization

- **Co-chair** of the 19th edition of the International Workshop on Models and Evolution (ME 2025).
- **Workshops and Tutorials Co-Chair** of the IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER 2026).
- **Co-chair** of the 6th International Workshop on Modeling in Low-Code Development Platforms (LowCode 2025).

7.2 Program Committee Member

Juri has served as Program Committee member for the following international events:

- The IEEE/ACM International Conference on Software Engineering (ICSE 2026) — Program Committee Member.
- The 28th International Conference on Model Driven Engineering Languages and Systems (MODELS 2025) — Artifact Evaluation Track.

- The 28th International Conference on Model Driven Engineering Languages and Systems (MODELS25) - AE track, 2024.
- The IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER 2025).
- The 6th International Workshop on Modeling in Low-Code Development Platforms (LowCode 2025), co-located with MODELS 2025.
- The 4th International Workshop on Foundations and Practice of Visual Modeling (FPVM 2025), co-located with MODELS 2025.

7.3 Contributed presentations at scientific conferences

- **Presenter** of the paper *From OCL to JSX: Declarative Constraint Modeling in Modern SaaS Tools* at the **OCL Workshop 2025**, co-located with STAF 2025, Koblenz, Germany.
- **Journal-First Presentation** of the paper *A Framework for Evaluating Tool Support for Co-Evolution of Modeling Languages, Tools and Models* (published in *Software and Systems Modeling, SoSyM*) as a journal-first talk in 2025.

8 Journal boards and reviewing

- **Assistant Editor** of the *Journal of Object Technology* (JOT).
- Reviewer for the following journals:
  - *IEEE Transactions on Software Engineering* (TSE)
  - *Software and Systems Modeling* (SoSyM, Springer)
  - *Empirical Software Engineering Journal* (EMSE)
  - *Journal of Object Technology* (JOT)
  - *Journal of Industrial Information Integration* (JII)
  - *Journal of Computer Languages* (COLA, Elsevier)
  - *Science of Computer Programming* (SCICO, Elsevier)

L'Aquila, October 7, 2025

Juri Di Rocco