ACTIVITY REPORT
(1ST MARCH 2025 - 30TH SEPTEMBER 2025)

Claudio Di Sipio
Dipartimento di Ingegneria, Scienze Informatiche e Matematiche
Università degli studi dell'Aquila
claudio.disipio@univaq.it

September 30, 2025

1 General Information

Title: "Ingegnerizzazione della fairness in sistemi software complessi"

Prot. n. 5339 del 15.12.2023

SSD: INF/01

Responsabile della ricerca: Prof. Davide Di Ruscio

2 Conducted research activities

This section gives an overview of the research activities conducted during the period 1st March 2025 - 30th September 2025.

In the context of the FRINGE project, we propose a novel approach for automating the fairness assessment process, leveraging the dataset's structural features as *bias symptoms* [J1]. To this end, we collect 24 existing datasets in the literature to detect early signals of bias under three different bias definitions, i.e., Statistical Parity, Equal Opportunity, and Average Odds. This approach can have different practical advantages, from reducing the training time of machine learning (ML) models to assisting practitioners in the fairness assessment process. In addition, we collaborated with the University of Sannio to investigate the presence of biases in issue reports and pull requests mined from popular GitHub repositories. In particular, ChatGPT was first used to obtain the summary and run a binary classification of the selected pull requests and issues. Afterward, I contributed to the manual checking and an internal discussion involving other authors of the paper.

Concerning the other research activities, I started to conceive a taxonomy for supporting the development of multi-agent systems (MAS) based on Large Language Models (LLMs). In particular, we conducted a mixed-methods study that involved the analysis of both scientific literature and existing frameworks. From the selected materials and tools, we elicit a first version of the taxonomy that describes MAS core features, behavioral model, governance rules, agents monitoring, and reusability features. In addition, I co-supervised three master's students who conducted a technical comparison of the most promising frameworks to support the summarization of GitHub README files. Even though these activities are not part of any publication, the content has been used to contribute to the deliverables D2.1 and D3.1 of the MOSAICO EU project[1]. Concerning the application of LLM to software engineering tasks, we explore security issues by using retrieval augmented generation (RAG) methodology [C1] and code obfuscation techniques [C2]. In the former, we discovered that RAG does not contribute significantly to detecting malicious PyPi packages, even though further experiments are needed to confirm this finding. Concerning the latter, our experiment involving three main proprietary models, i.e., GPT4, Gemini, and Claude, reveals the counterintuitive "obfuscation by simplification"

[1] https://cordis.europa.eu/project/id/101189664

phenomenon where models consistently reduce rather than increase cyclomatic complexity, highlighting promising directions for leveraging LLMs in software security.

Furthermore, we investigate the combination of prompt engineering and hyperparameter tuning for supporting domain modeling in the Model Driven Engineering (MDE) domain [C3]. In particular, we show that hyperparameter tuning performed with the NSGA-II evolutionary algorithm is not generalizable for different modeling domains, even though the accuracy can be enhanced with structured prompting, e.g., chain-of-thought (CoT) strategy. Finally, we proposed CodeXHug, a curated dataset of pre-trained models (PTMs) mined from Hugging Face with their actual usage on GitHub public repositories [C4]. In addition, the mined data have been used to support the generation of enhanced model cards, shepherding non-expert users in employing PTMs in their daily tasks.

Concerning collaboration with other institutions, I continued to investigate the usage of LLMs for generating modeling operations in the context of the AIDoARt EU project with the University of Teramo [J2]. In addition, we report our experience in running hackathon events in the context of large collaborative ICT projects like AIDoARt to foster partners' collaboration and engagement. This work is currently under revision for the ACM Transactions on Software Engineering and Methodology (TOSEM) [R3]. I continued the collaboration with the Gran Sasso Science Institute (GSSI) to develop an MDE approach based on weaving modeling to support the semi-automated configuration of recommender systems in different domains [C5]. We applied this methodological framework to develop MoToRev, an approach to support the development of tourism recommender systems using modeling artifacts to produce actual source code. This work is currently under review in the special issue entitled "New Directions of Recommender Systems in Tourism" to appear in the ACM Transactions on Recommender Systems (TORS) journal [R4].

3 List of publications

3.1 Journals

J1 *d'Aloisio G.,* ***Di Sipio, C.****, Di Marco, A., and Di Ruscio, D.*. **Towards Early Detection of Algorithmic Bias from Dataset's Bias Symptoms: An Empirical Study.** *Information and Technology Journal, Special Issue on "Are you ready for the age of fairness in software systems?*. Accepted for publication

J2 *Muttillo V.,* ***Di Sipio C.****, Rubei R., and Berardinelli L.* (2025). **Leveraging synthetic trace generation of modeling operations for intelligent modeling assistants using large language models**. *Information and Software Technology*, 186, 107806. ISSN 0950-5849. DOI

3.2 Conferences

C1 *Ibiyo M., Louangdy T., Nguyen P.T.,* ***Di Sipio C.****, and Di Ruscio D.* (2025). **Detecting Malicious Source Code in PyPI Packages with LLMs: Does RAG Come in Handy?**. In: *Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering (EASE 2025)*. `[pre-print]`.

C2 *De Tomasi L.,* ***Di Sipio C.****, Di Marco A., and Nguyen P.T.* (2025). **Simplicity by Obfuscation: Evaluating LLM-Driven Code Transformation with Semantic Elasticity**. In: *Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering (EASE 2025)*. `[pre-print]`.

C3 *Bulhakov V., d'Aloisio G.,* ***Di Sipio C.****, Di Marco A., and Di Ruscio D.* (2025). **Investigating the Role of LLMs Hyperparameter Tuning and Prompt Engineering to Support Domain Modeling**. *The 51st Euromicro Conference Series on Software Engineering and Advanced Applications (SEAA)*. `[pre-print]`.

C4 *Palombo S.,* ***Di Sipio C.****, Di Rocco J., and Di Ruscio D.* (2025). **Generate with CodeXHug: A Dataset to Enhance Model Cards with Code Usage Patterns**. *The 51st Euromicro Conference Series on Software Engineering and Advanced Applications (SEAA)*.

C5 *Pereira R.S.,* ***Di Sipio C.****, De Sanctis M., and Iovino L.* (2025). **Automated Recommender System Integration in Model-based Ecosystems**. In: Krogstie J., Rinderle-Ma S., Kappel G., Proper H.A. (eds) *Advanced Information Systems Engineering. CAiSE 2025. Lecture Notes in Computer Science*, vol 15702. Springer, Cham. DOI

3.3 Under review/revision

R1 ***Di Sipio C.****, Eramo E., Muttillo V., and Rubei R.* **Foster the use of Hackathons in Collaborative Research Projects: Methodology, Experience Report and Lesson Learned**. Under review at *Transactions on Software Engineering and Methodology Journal*

R2 *Pereira R.S., **Di Sipio C.**, De Sanctis M., and Iovino L.* (2025). **MoToRev: Leveraging Interactive Modeling Artifacts to Assist Tourism Recommender Systems Development**. Under review in *Transactions on Recommender Systems Journal*.

4 Participation In National And International Research Projects

From June 2022 PRIN 2022 FRINGE (context-aware FaiRness engineerING in complex software systEms) project. The overarching goal of the FRINGE project is to provide software engineers, data scientists, and ML experts with a comprehensive set of methodologies, approaches, and software engineering (SE) solutions to improve the development, monitoring, and design of fairness-related properties of ML-intensive systems. FRINGE will help to sustain the digital transformation by empowering advanced artificial intelligence applications through instruments that can lead to the trustworthy and ethical exploitation of large amounts of data.

From January 2025 MOSAICO is an ambitious research and innovation project that will develop a comprehensive portfolio of AI-powered tools and techniques to revolutionize software development. This will involve creating a vibrant ecosystem of intelligent agents that can seamlessly collaborate with developers, enhancing efficiency, improving software quality, and accelerating innovation.

June 2022 - June 2025 PRIN 2021 research project consortium "EMELIOT (Engineered MachinE Learning-intensive IoT systems)." The EMELIOT project aims to study solutions for engineering highly dependable, ML-intensive IoT systems. EMELIOT foresees systems where different sensors collect data from physical systems, which in turn can get their status changed by actuators. Collected data are filtered, aggregated, and used to manage ML models locally, on the edge, by dedicated IoT gateways, or on remote cloud servers.

5 Professional Services

This section reports my professional activities.

- Reviewer for ACM Transactions on Software Engineering and Methodology (TOSEM)
- Reviewer for ACM Transactions on Autonomous and Adaptive Systems (TAAS)
- Program Committee member of the NIER track in the 28th International Conference on Model-Driven Engineering Languages and Systems (MODELS 2025)
- Program Committee member of the Artifact evaluation track in the 28th International Conference on Model-Driven Engineering Languages and Systems (MODELS 2025)
- Program Committee member of the Artifact Evaluation Track and ROSE Festival in the 41th International Conference on Software Maintenance and Evolution (ICSME 2025)
- Reviewer for Elsevier Expert System with Application Journal (ESWA)
- Reviewer for Springer Automated Software Engineering Journal (ASE)
- Reviewer for Nature Scientific Reports Journal
- Reviewer for Springer Automated Software Engineering Journal (ASE)
- Reviewer for Springer Empirical Software Engineering Journal (EMSE)
- Reviewer for Elsevier Journal of System and Software (JSS)
- Reviewer for Elsevier Information and Software Technology Journal (IST)
- Reviewer for IEEE Transaction on Software Engineering Journal (TSE)
- Program Committee member of the Artifact Evaluation Track and ROSE Festival in the 41th International Conference on Software Maintenance and Evolution (ICSME 2025)
- Reviewer for Springer International Journal on Software and Systems Modeling (SoSyM)

6 Organizing activities

- Guest editor of the Special Issue on Evaluation of Qualitative Aspects of Intelligent Software Assistants (EQUISA) Co-Guest Editor: Valeria Pontillo, Riccardo Rubei, and Pablo Gómez-Abajo. website

- Executive Guest Editor of the Special issue on Next-Generation Model-Based Software Engineering with Foundation Models, Information Technology Journal (IST). Co-Guest Editors: Silvia Abrahão, Martin Weyssow, and Fabio Palomba. website
- Co-organizer of the 7th Workshop on Artificial Intelligence and Model-driven Engineering (MDEIntelligence 2025), co-located with the 28th ACM/IEEE International Conference on Model-Driven Engineering Languages and Systems (MODELS 2025). Co-organizers: Lola Burgueño and Domink Bork. website
- Co-organizer of the 1st Workshop on Generation of Synthetic Datasets for Information Systems (GenSyn 2025), co-located with the 37th International Conference on Advanced Information System Engineering (CAiSE 2025), Co-organizers: Arianna Fedeli, Riccardo Rubei, and Eduard Kamburjan. website
- Co-organizer of the 1st Workshop on Evaluation of Qualitative Aspects of Intelligent Software Assistants (EQUISA), co-located with the 29th International Conference on Evaluation and Assessment in Software Engineering (EASE 2025), Co-organizers: Valeria Pontillo, Riccardo Rubei, and Pablo Gómez-Abajo. website

7 Thesis co-supervision

I co-supervised the following bachelor's and master's theses:

Bachelor Lorenzo Baldacci, *"Modelli Ai per l'Identificazione, Classificazione e Sintesi di Pipeline Data Science"*, A.Y. 2024-2025. Thesis Supervisor: Juri Di Rocco

Bachelor Antonio Addario, *"Generating Pull Request Messages with LLMs Using Metadata and Code Changes"*, A.Y. 2024-2025. Thesis Supervisor: Juri Di Rocco

Bachelor Giacomo Calcaterra, *"Progettazione e implementazione di un sistema di raccomandazione per data workflow"*, A.Y. 2024-2025. Thesis Supervisor: Juri Di Rocco

Bachelor Stefano Palombo, *"Utilizzo degli LLM per Generare Hugging Face Model Card da esempi di codice."*, A.Y. 2024-2025. Thesis Supervisor: Juri Di Rocco

Master Mariama Celi Serafim De Oliveira, *A Comparative Analysis of LLM-Based Multi-Agent Frameworks*. A.Y. 2024-2025.Thesis Supervisor: Prof. Davide Di Ruscio, University of L'Aquila, Italy.

Master Marco Gianrusso, *Sistemi multi-agente e modelli linguistici nell'ingegneria del software: analisi e confronto*. A.Y. 2024-2025. Thesis supervisor: Prof. Davide Di Ruscio, University of L'Aquila, Italy.

Master Motunrayo Osatohanmen Ibiyo, *Large Language Models based Multi-Agent Systems for Detecting Malicious PyPI Packages*. A.Y. 2024-2025. Thesis supervisor: Prof. Phuong T. Nguyen, University of L'Aquila, Italy.

Responsabile della ricerca
Davide Di Ruscio

L'Aquila, 30/09/2025
Claudio Di Sipio