PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Cristian Mascia

Cycle: XL

Training and Research Activities Report

Year: First

Cristian Mascia

_____

Tutor: prof. Stefano Russo

Co-Tutor: Roberto Pietrantuono

Date: October 31, 2025

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: **Author:**

_____

1. Information:

- ➢ **PhD student: Cristian Mascia**
- ➢ **DR number: 999878**
- ➢ **Date of birth: 04/03/1997**
- ➢ **Master Science degree: Computer Engineering   University: University of Naples Federico II**
- ➢ **Doctoral Cycle: XL**
- ➢ **Scholarship type: UNINA**
- ➢ **Tutor: Stefano Russo**
- ➢ **Co-tutor: Roberto Pietrantuono**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Shaping robustly control loop: look into stability margins & uncertainties	Seminar	1	0.2	2024-12-05	Prof. Ciro Visone	**Y**
QUIC: the secure protocol shaping the future of real-time communication over the Internet	Seminar	2	0.4	2024-12-09	Prof. Simon Pietro Romano	**Y**
Strutture basate su regole e strutture basate su approssimazioni	Seminar	1.5	0.3	2024-12-10	Prof. Francesco Cotugno	**Y**
QUIC: the secure protocol shaping the future of real-time communication over the Internet	Seminar	2	0.4	2024-12-11	Prof. Simon Pietro Romano	**Y**
5G & DIGITAL TRANSFORMATION: A VIEW FROM AN UNCONVENTIONAL PERSPECTIVE	Seminar	4	0.8	2025-03-14	5G Academy	**Y**
Safety Assessment of Autonomous Vehicles: Approaches and Challenges	Seminar	1	0.2	2025-03-24	Prof. Stefano Russo	**Y**
Sovranità digitale: cos'è e quali sono le principali minacce al cyberspazio nazionale	Seminar	2	0.4	2025-06-23	Prof. Marcello Cinque	**Y**
IEEE Authorship and Open Aces Symposium: Tips and Best Practices to Get Published from IEEE Editors	Seminar	2	0.4	2025-10-15	IEEE	**Y**
Guardians or Threats? AI at the Frontlines of	Seminar	4	0.8	2025-10-17	Prof. Antonia Maria Tulino	**Y**

Cybersecurity						
Advanced Course on Data Science & Machine Learning	Doctoral School		8	2025-06-9 - 2025-06-13	Giuseppe Nicosia & Panos Pardalos	**Y**
AI For System Engineering	Course		6	2025-04-23	Prof. Roberto Pietrantuono	**Y**
How to boost your PhD	Course		5	2025-02-19	Prof. Antigone Marino	**Y**
Using Deep Learning Properly	Course		4	2025-02-19	Andrea Apicella	**Y**

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	**Seminars**	**Research**	**Tutorship**	**Total**
Bimonth 1	**0**	**1.3**	**9**	**0**	**10.3**
Bimonth 2	**9**	**0**	**2**	**0**	**11**
Bimonth 3	**6**	**1**	**4**	**0**	**11**
Bimonth 4	**8**	**0.4**	**2**	**0**	**10.4**
Bimonth 5	**0**	**0**	**10**	**0**	**10**
Bimonth 6	**0**	**1.2**	**8**	**0**	**9.2**
Total	**23**	**3.9**	**35**	**0**	**61.9**
Expected	**30 - 70**	**10 - 30**	**80 - 140**	**0 – 4.8**	

3. Research activity:

During this first year, the main focus of the research was performance testing for microservice architectures, which have become a popular software architectural style. In particular, I investigated techniques for the automatic generation and execution of performance test cases based on past execution data.

Performance testing in microservice-based systems is essential for understanding how business workloads affect system performance and resource utilization. However, despite its importance, performance testing of microservice architectures remains a relatively underexplored research area. The literature concentrated on evaluating performance aspects under varying load conditions and diverse deployment configurations. Such evaluation is valuable for all applications running in modern data centers, as it supports capacity planning (making informed decisions about infrastructure scaling), prevention (defining proactive measures to avoid performance issues), and resource usage analysis.

A particular focus of my work was on identifying minimal configurations that trigger performance degradation, even under relatively low loads. These critical configurations, often referred to as "*knee points*" of the system's load–performance curve, are crucial for performance engineers to determine the minimum workload at which the system still operates without degradation. This information is highly valuable for capacity planning.
Identifying such configurations is typically an expensive process, as it requires multiple test sessions to explore a wide space of workload conditions.
Therefore, I focused on detecting the workload configurations most likely to lead to performance issues.

To this end, I proposed a methodological framework called **microWave**, which leverages Artificial Intelligence (AI) techniques to identify workloads that expose performance failures in microservice systems. The

framework assists testers in automatically generating workloads that cause the system to behave improperly, particularly by identifying minimal critical workloads.

In this study, I compared three different AI strategies: a Deep Neural Network, a Causal Model, and a Large Language Model (LLM). The experimental results showed that models can learn from historical execution data containing performance issues, causal reasoning outperforms the other approaches. Conversely, when data with performance issues is unavailable, a random predictor (used as a baseline) performs better, though at a higher testing cost due to the need for more test runs.

Despite the recent success of LLMs across various research areas, my findings suggest that LLMs alone cannot effectively act as "*performance testing engineers*". Based on these results, I then focused on methods to enhance the capabilities of LLMs in software engineering tasks, primarily through their integration with reasoning models such as causal models. As part of this effort, I proposed **CALLMIT**, a strategy to automate the generation of critical workload configurations and corresponding test cases. **CALLMIT** aims to reduce testing effort by executing only those test cases most likely to reveal performance degradation. It combines Causal Reasoning with LLMs and introduces a novel Retrieval-Augmented Generation (RAG) approach that enhances LLM prompts using automatically inferred causal relationships between microservice performance metrics. This allows the model to propose workload configurations with a higher likelihood of causing system failures.

CALLMIT was experimentally evaluated in multiple variants and compared with a conventional RAG technique across three widely used benchmark systems. The results demonstrate that causal models can significantly improve the ability of LLMs to accurately identify performance-critical workload configurations.

In the last part of the year, my research expanded to investigate additional strategies for integrating reasoning models with LLMs. While LLMs have shown remarkable performance across numerous research domains, their outputs often remain unpredictable and difficult to interpret. Consequently, I focused on developing methods to incorporate explicit reasoning models into LLM-based workflows, aiming to enhance their reliability, transparency, and effectiveness in software engineering applications.

4. Research products:

– *Microservices Performance Testing with Causality-enhanced Large Language Models - C. Mascia, A. Guerriero, L. Giamattei, R. Pietrantuono, S. Russo* – 2nd ACM international conference on AI Foundation Models and Software Engineering (FORGE*), published,* 2025

– *Learning-based Automated Generation of Critical Workload Configurations for Microservices Performance Testing – C. Mascia, L. Giamattei, A. Guerriero, R. Pietrantuono, S. Russo* - IEEE International Conference on Web Services (ICWS), *published*, 2025

5. Conferences and seminars attended

– 2nd ACM international conference on AI Foundation Models and Software Engineering (FORGE), Ottawa (Canada), 26 April - 04 May 2025 – the paper "*Microservices Performance Testing with Causality-enhanced Large Language Models*" was presented.

– International Conference on Web Services (ICWS), Helsinki (Finland), 7-12 July 2025 – the paper "*Learning-based Automated Generation of Critical Workload Configurations for Microservices Performance Testing*" was presented.