UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

DOTTORATO DI RICERCA / PHD PROGRAM IN
INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Activities and Publications Report

PhD Student: Cristina Improta

Student DR number: DR996616

PhD Cycle: XXXVIII

PhD Chairman: Prof. Stefano Russo

PhD program student's start date: 01/11/2022

PhD program student's end date: 31/10/2025

Supervisor: Domenico Cotroneo

e-mail: cotroneo@unina.it

PhD scholarship funding entity: Università Federico II

General information

Cristina Improta received in year 2022 the Master Science degree in Computer Engineering from the University of Napoli Federico II. She attended a curriculum in AI-based Software Engineering within the PhD program in Information Technology and Electrical Engineering. She received a grant from Università Federico II.

Study activities

Attended Courses

Year	Course Title	Type	Credits	Lecturer	Organization
1st	Using Deep Learning Properly	Ad hoc course	4	Dr. Andrea Apicella	ITEE
1st	IoT Data Analysis	Ad hoc course	4	Dr. Raffaele Della Corte	ITEE
1st	Virtualization technologies and their applications	Ad hoc course	5	Dr. Luigi De Simone	ITEE
2nd	Strategic Orientation for STEM Research & Writing	Ad hoc course	5	Dr. Chie Shin Fraser	ITEE
3rd	Innovation and Entrepreneurship	Ad hoc course	3	Prof. Pierluigi Rippa	ITEE
3rd	AI Code Generation: Foundations, Evaluation, and Security	Ad hoc course	3	Dr. Pietro Liguori	ITEE

Attended PhD Schools

Year	School title	Location	Credits	Dates	Organization
1st	2023 Spring School on Transferable Skills	Online	2	24/05/2023, 25/05/2023	Department of Pharmacy, University of Napoli Federico II
1st	3rd International Software Engineering Summer School (SIESTA2023)	Lugano, Switzerland	3	11–13/09/2023	Software Institute, Università della Svizzera Italiana, Lugano, Switzerland

Attended Seminars

Year	Seminar Title	Credits	Lecturer	Lecturer affiliation	Organization
1st	Connecting the dots: Investigating an APT campaign using Splunk	0.4	Dr. Antonio Forzieri	EMEA Cyber Security Specialization and Advisory Splunk Inc.	ITEE
1st	Cybercrime and Information Warfare: National and International	0.4	Dr. Pierluigi Paganini	Cibhorus S.r.l.	ITEE
1st	Privacy and Data Protection	0.4	Dr. Stefano Mele	Gianni & Origoni	ITEE
1st	Digital Forensics	0.4	Group-IB	Group-IB	ITEE
1st	Multi-robot Control of Heterogeneous Herds	0.2	Dr. Eduardo Montijano	Universidad de Zaragoza, Spain	Scuola Superiore Meridionale
1st	Analysis and control of functional brain networks	0.2	Fabio Pasqualetti	University of California at Riverside, United States	Scuola Superiore Meridionale
1st	How to Publish Under the CARE-CRUI Open Access Agreement with IEEE	0.3	Nino Grizzuti, Eszter Lukacs, Stefano Bianco	CARE-CRUI and UNINA, IEEE, CRUI-CARE and INFN	CARE-CRUI and IEEE
1st	Open-source software e sicurezza della software supply chain	0.4	Antonino Sabetta & Serena Ponta	SAP Security Research	ITEE
1st	Exploring Advanced Aerial Robotics: A Journey into Cutting-Edge Projects and Neural Control	0.2	Eng. Eugenio Cuniato	Autonomous Systems Lab (ASL), ETH Zurich	ITEE
1st	Models of human motor coordination – a critical assessment and some open problems	0.2	John Hogan	University of Bristol, UK	Scuola Superiore Meridionale
1st	BGP & Hot-Potato Routing: graceful and optimal convergence in case of IGP events	0.2	Prof. Pascal Merindol	University of Strasbourg, France	ITEE
2nd	Economic Fitness: Concepts, Methods and Applications	0.2	Dr. Luciano Pietronero	Enrico Fermi Research center (Rome, Italy)	Scuola Superiore Meridionale
2nd	Analytic center selection of optimization-based controllers for robot ecology	0.2	Prof. Gennaro Notomista	University of Waterloo, Waterloo, Canada	ITEE

2nd	Program Comprehension in the Era of Large Language Models: Achievements and Challenges	0.2	Gabriele Bavota	Università della Svizzera italiana, Lugano, Svizzera	International Conference on Program Comprehension
2nd	Mining Our Way Back to Incremental Builds for DevOps Pipelines	0.15	Shane McIntosh	Software REBELs at University of Waterloo	Mining Software Repositories Conference
2nd	Questioning the questions we ask about the impact of AI on software engineering	0.15	Margaret-Anne Storey	University of Victoria, Canada	International Conference on Program Comprehension
2nd	Open Source Software Digital Sociology: Quantifying and Managing Complex Open Source Software Ecosystem	0.15	Minghui Zhou, Yuxia Zhang, Xin Tan	Peking University, Beijing Institute of Technology, China, Beihang University	Mining Software Repositories Conference
2nd	Cohort Studies for Mining Software Repositories	0.15	Nyyti Saarimäki, Sira Vegas, Valentina Lenarduzzi, Davide Taibi	Tampere University, Universidad Politecnica de Madrid, University of Oulu	Mining Software Repositories Conference
2nd	Challenges and Opportunities in Model Checking Large-scale Distributed Systems	0.15	Rupak Majumdar	Max Planck Institute for Software Systems	International Conference on Software Engineering
2nd	Software Engineering in a World with Generative AI	0.15	Martin C. Rinard	Massachusetts Institute of Technology	International Conference on Software Engineering
2nd	AI is making us rethink everything, including software development	0.15	Soumith Chintala	Meta	International Conference on Software Engineering
2nd	Trustworthy by Design	0.15	Carol J. Smith, Margaret-Anne Storey	Carnegie Mellon University, University of Victoria, Canada	International Conference on Software Engineering
2nd	Regolazione in tema di intelligenza artificiale alla luce dell'AI act	1	Elvira Raviele	Dirigente Ufficio di Gabinetto MIMIT	5G Academy
2nd	IEEE Authorship and Open Access Symposium: Tips and Best Practices to Get Published from IEEE Editors	0.3	Dr. Petar Popovski, Eszter Lukács, Judy Brady	Aalborg University, Denmark, IEEE	IEEE Xplore Team

2nd	From ACE Technologies to Sustainable, Accessible and Equitable Urban Mobility:An Optimization Journey	0.4	Prof. Mauro Salazar	Eindhoven University of Technology, Eindhoven, Netherlands	ITEE
2nd	Learning in nonstationary environments	0.4	Prof. Cesare Alippi	Politecnico di Milano, Milano, Italy, niversità della Svizzera italiana, Lugano, Switzerland	ITEE
2nd	Value-Flow-Based Code Embedding for Software Vulnerability Detection	0.2	Yulei Sui	University of New South Wales, Sydney, Australia	International Symposium on Software Reliability Engineering
2nd	Quality Assurance of AI-Based Systems	0.2	Hiroshi Maruyama	Preferred Networks, Inc.	International Symposium on Software Reliability Engineering
2nd	Quantum circuit compilation and compression	0.2	Kae Nemoto	Okinawa Institute of Science and Technology	International Symposium on Software Reliability Engineering
2nd	Software Reliability in the Era of Large Language Models: A Dual Perspective	0.2	David Lo	Singapore Management University	International Symposium on Software Reliability Engineering
3rd	Perché l'Intelligenza Artificiale crede di fare a meno della teoria linguistica, ma in realtà non potrà farlo	0.2	Alessandro Lenci	University of Pisa	ITEE
3rd	AI and Enabling Technologies for Social Robots	0.3	Prof.ssa Silvia Rossi	PRISCA Lab, UNINA	ITEE
3rd	5G & Digital Transformation: A View From An Unconventional Perspective	0.8	Dr. Maurizio Irlando	Director of Information Technology Noovle, a TIM Enterprise brand	5G Academy
3rd	Safety Assessment of Autonomous Vehicles: Approaches and Challenges	0.2	Prof. Peter Popov	City University London	ITEE

3rd	On the Notion of Binary Equivalence in Software Supply Chain Security	0.2	Prof. Jens Dietrich	Victoria University of Wellington, New Zeland,	ITEE
3rd	On the Security of Semantic Watermarking to Detect AI-Generated Content	0.2	Dr. Erwin Quiring	Ruhr University Bochum, Germany, ICSI, UC Berkeley	ITEE
3rd	PhD Survival Strategies	0.3	Prof. Gabriele Bavota	Università della Svizzera italiana, Lugano, Svizzera	ITEE

Research activities

Cristina Improta's research advances the evaluation and enhancement of performance and robustness of AI-based code generators for offensive security. She developed an automated framework to judge the syntactic and semantic correctness of AI-generated exploits and built solutions to scale correctness assessment in security contexts. Her work explores robustness by testing AI code generators against natural-language perturbations and strengthening them via training data augmentation. It also explores context-aware exploit generation, showing how additional contextual information can provide more accurate generation, based both on the type of information, i.e., related or unrelated, and the tested model's architecture.

Complementing this, Cristina Improta's research investigates how the quality of training data influences the reliability, security, and maintainability of code generated by AI-based code assistants. Her work traces quality issues in AI-generated code back to both intentional and unintentional corruption in the training corpus, and proposes a static analysis-based data curation pipeline that improves output quality without compromising functional accuracy. In parallel, she conducts large-scale empirical comparisons between human-written and AI-generated code across multiple languages and models, revealing systematic differences in structure, defect profiles, and vulnerability types. Her findings contribute toward standardized methodologies for evaluating AI-generated code and inform practices for its trustworthy integration into real-world software.

Tutoring and supplementary teaching activities

The candidate has performed supplementary teaching activities ("didattica integrativa") for the course "Sistemi Operativi" held by prof. Domenico Cotroneo , within the BSc Degree in Computer Engineering at Federico II University, for the academic years 2022-2023.

Credits summary

PhD Year	Courses	Seminars	Research	Tutoring / Supplementary Teaching
1st	18	3.3	40	0.24

2nd	5	4.7	49	0
3rd	6	2.2	52	0

Research periods in institutions abroad and/or in companies

PhD Year	Institution / Company	Hosting tutor	Period	Activities
2nd	Università della Svizzera italiana, Lugano, Svizzera	Gabriele Bavota, Professor & Head of SEART research group	02/03/2024 - 12/03/2024	Start of collaboration and preliminary experimentation on the impact of the quality of training data on the quality of AI-generated code.
2nd	Università della Svizzera italiana, Lugano, Svizzera	Gabriele Bavota, Professor & Head of SEART research group	02/05/2024 - 31/07/2024	Research on mining open-source software repositories to collect data for training Large Language Models for code generation. Experimental evaluation on the impact of the quality of training data on the quality of AI-generated code. Joint scientific paper preparation and submission to the International Conference on Program Comprehension.

PhD Thesis

In the PhD Thesis, Cristina Improta assesses and enhances code quality in the era of AI-developed software. The rapid evolution of AI-based code assistants is reshaping the way software is developed, shifting the developer's role from manual implementation to high-level orchestration and review of AI-generated code. Large Language Models (LLMs) are now capable of producing entire software components from natural language prompts, with this code increasingly integrated into production systems. While productivity gains are clear, this shift raises critical concerns about the reliability, security, and overall quality of AI-developed software.

Although AI-generated code often appears functionally correct, it may still be prone to critical defects and security vulnerabilities, issues largely overlooked by popular benchmarks focused solely on correctness. Existing research tends to treat these as isolated symptoms without tracing them back to a root cause: the quality of the training data. Evaluation practices are also fragmented, relying on inconsistent tools, metrics, and taxonomies that hinder comparison and reproducibility.

Cristina Improta's thesis addresses both limitations by systematically linking code quality issues to training data integrity and introducing a standardized evaluation methodology to enable consistent comparisons of AI-generated and human-written code. First, it shows that AI-generated code differs systematically from human-authored code, being shorter, more predictable, and less lexically diverse. Second, it identifies training data corruption as a root cause for low-quality code

generation. Through controlled experiments, it demonstrates that both malicious data poisoning and unintentional collection of low-quality code can bias models, leading to low-quality outputs without degrading functional accuracy. As a mitigation, it introduces an automated data cleaning pipeline using static analysis, improving the quality of generated code.
Finally, her thesis proposes a standardized quality evaluation methodology that maps tool outputs to established taxonomies such as Orthogonal Defect Classification (ODC) and MITRE's CWE, enabling rigorous, cross-language and cross-author comparisons. Applied at scale, it reveals consistent disparities between human and AI-generated code in both defect density and vulnerability profiles, particularly in high-severity CWEs.

Research products

Research results appear in 4 papers published in international journals and 7 contributions to international conferences.

List of scientific publications

International journal papers

P. Liguori, **C. Improta**, R. Natella, B. Cukic, D. Cotroneo ,
Who evaluates the evaluators? On automatic metrics for assessing AI-based offensive code generators,
Expert Systems with Applications,
vol. 225, pp. 120073, 2023, DOI: 10.1016/j.eswa.2023.120073.

R. Natella, P. Liguori, **C. Improta**, B. Cukic, D. Cotroneo,
AI Code Generators for Security: Friend or Foe?,
IEEE Security & Privacy,
vol. 22(5), pp. 73-81, 2024, DOI: 10.1109/MSEC.2024.3355713.

D. Cotroneo, A. Foggia, **C. Improta**, P. Liguori, R. Natella,
Automating the correctness assessment of AI-generated code for security contexts,
Journal of Systems and Software,
vol. 216, pp. 112113, 2024, DOI: 10.1016/j.jss.2024.112113.

C. Improta, P. Liguori, R. Natella, B. Cukic, D. Cotroneo,
Enhancing Robustness of AI Offensive Code Generators via Data Augmentation,
Empirical Software Engineering,
vol. 30(1), pp. 7, 2025, DOI: 10.1007/s10664-024-10569-y.

International conference papers

P. Liguori, **C. Improta**, S. De Vivo, R. Natella, B. Cukic, D. Cotroneo,
Can NMT Understand Me? Towards Perturbation-based Evaluation of NMT Models for Code Generation,
2022 IEEE/ACM 1st International Workshop on Natural Language-Based Software Engineering (NLBSE),
Pittsburgh, PA, USA, May 2022, pp. 59-66, IEEE Computer Society, DOI: 10.1145/3528588.3528653.

C. Improta,
Poisoning programs by un-repairing code: Security concerns of AI-generated code,
2023 IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW),
Florence, Italy, Oct. 2023, pp. 128-131, IEEE, DOI: 10.1109/ISSREW60843.2023.00060.

D. Cotroneo, **C. Improta**, P. Liguori, R. Natella,
Vulnerabilities in AI Code Generators: Exploring Targeted Data Poisoning Attacks,
2024 IEEE/ACM International Conference on Program Comprehension,
Lisbon, Portugal, Apr. 2024, pp. 280-292, IEEE Computer Society, DOI: 10.1145/3643916.3644416.

P. Liguori, **C. Improta**, R. Natella, B. Cukic, D. Cotroneo,
Enhancing AI-based Generation of Software Exploits with Contextual Information,
2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE),
Tsukuba, Japan, Oct. 2024, pp. 180-191, IEEE, DOI: 10.1109/ISSRE62328.2024.00027.

C. Improta, R. Tufano, P. Liguori, D. Cotroneo, G. Bavota,
Quality In, Quality Out: Investigating Training Data's Role in AI Code Generation,
2025 IEEE/ACM 33rd International Conference on Program Comprehension (ICPC),
Ottawa, ON, Canada, Apr. 2025, pp. 454-465, IEEE Computer Society, DOI: 10.1109/ICPC66645.2025.00056.

D. Cotroneo, **C. Improta**, P. Liguori,
Human-Written vs. AI-Generated Code: A Large-Scale Study of Defects, Vulnerabilities, and Complexity,
2025 36th IEEE International Symposium on Software Reliability Engineering (ISSRE),
Sao Paulo, Brazil, Oct. 2025. Accepted, to appear.

C. Improta,
Detecting Stealthy Data Poisoning Attacks in AI Code Generators,
2025 3rd IEEE International Workshop on Reliable and Secure AI for Software Engineering (ReSAISE),
Sao Paulo, Brazil, Oct. 2025. Accepted, to appear.

Date 27/10/2025

PhD student signature Cristina Improta

Supervisor signature