Giuseppe Colavito

Bari

+39 350 547 8689

giuseppe.colavito@uniba.it

https://www.linkedin.com/in/peppocola/

Current Position

November 2022 - Present

PhD Student

Currently enrolled in 3nd year of the National PhD in Artificial Intelligence

- Research Project: "Generative AI For Automatic Labeling in Software Engineering". The research project focuses on the design, implementation, and performance monitoring of text categorization approaches for automatic issue classification.
- Supervisors: Prof. Nicole Novielli, Prof. Filippo Lanubile

Work Experience

August 2024 - November 2024

Research Scholar @ NASA

NASA Goddard Space Flight Center - Maryland, USA

- Project: Large Language Model based models for Safety and Mission Assurance Data Analysis
- Supervisor: Dr. Ying Shi

May 2024 - August 2024

Research Assistant

Università degli Studi di Bari "Aldo Moro"

Research Assistant at Collaborative Development Research Group

- Monitoring the Performance of Text Categorization Models after Deployment as Cloud Services
- Supervisor: Prof. Nicole Novielli

July 2023 – Sept 2023

Research Assistant

Università degli Studi di Bari "Aldo Moro"

Research Assistant at Collaborative Development Research Group:

- Deploying Text Categorization Models as Cloud Services
- Supervisor: Prof. Filippo Lanubile

March 2022 – May 2022

Curricular Internship (Master's Degree)

Università degli Studi di Bari "Aldo Moro"

Internship at Collaborative Development Research Group:

- Development, implementation and evaluation of a supervised classifier for issue label classification for automatic classification of Github issues. Use of state-of-the-art models for text classification, including pre-trained language models such as BERT.
- Supervisors: Prof. Filippo Lanubile, Prof. Nicole Novielli

October 2021 – July 2022

Reserch Assistant

Università degli Studi di Bari "Aldo Moro"

Research Assistant at Collaborative Development Research Group:

- Development of classifiers for emotion recognition through NLP
- Supervisors: Prof. Filippo Lanubile, Prof. Nicole Novielli

March 2020 – June 2020

Curricular Internship (Bachelor's Degree)

Università degli Studi di Bari Aldo Moro

Internship at Collaborative Development Research Group:

- Development of a text classifier for documentation decluttering for Java source code. Use of NLP techniques and Supervised Learning.
- Supervisor: Prof. Nicole Novielli

Education

2022 **Master's Degree in Computer Science**

Università degli Studi di Bari "Aldo Moro" - Dipartimento di Informatica

- Master's degree (2 years)
- Computer Science - Artificial Intelligence
- Thesis title: Issue Report Classification Using BERT
- Mark: 110 cum laude
- Date: July 15th 2020
- Supervisors: Prof. Filippo Lanubile, Prof. Nicole Novielli

2020 **Bachelor Degree in Computer Science**

Università degli Studi di Bari "Aldo Moro" - Dipartimento di Informatica

- Bachelor's degree (3 years)
- Informatica
- Thesis title: Development of a text classifier for documentation decluttering
- Mark: 110 cum laude
- Date: July 20th 2022
- Supervisors: Prof. Nicole Novielli, Prof. Pierpaolo Basile

Publications

2025 **Issue Classification with LLMs: an Empirical Study of the NASA Flight Software Systems**

Colavito, G. and Lanubile, F. and Novielli, N. and Arreza, C. and Shi, Y., Submitted to ICSE SEIP 2026

2025 **The NLBSE'25 Tool Competition**

Al-Kaswan, A. and Colavito, G. and Stulova, N. and Rani, P., Intl. Workshop on Natural Language-based Software Engineering (NLBSE 2025), co-located with ICSE 2025, Ottawa, Canada, April 27 2025,

DOI: `https://doi.org/10.1109/NLBSE66842.2025.00012`

2025 **Foundation Models for Automatic Issue Labeling**

Colavito, G., 47th International Conference on Software Engineering (ICSE) - Doctoral Symposium Track,

DOI: `https://doi.org/10.1109/ICSE-Companion66252.2025.00038`

2025 **Benchmarking large language models for automated labeling: The case of issue report classification**

Colavito, G. and, Lanubile, F. and Novielli, N., Information and Software Technology Journal, Scimago Q1,

DOI: `https://doi.org/10.1016/j.infsof.2025.107758`

2024 **The NLBSE'24 Tool Competition**

Kallis, R. and Colavito, G. and, Al-Kaswan, A. and Pascarella, L. and Chaparro, O. and Rani, P., Intl. Workshop on Natural Language-based Software Engineering (NLBSE 2024), co-located with ICSE 2024, Lisbon, Portugal, April 20 2024,

DOI: `https://doi.org/10.1145/3643787.3648038`

2024 **Large Language Models for Issue Report Classification**

Colavito, G. and Lanubile, F. and Novielli, N. and Quaranta, L., Convegno Nazionale CINI sull'Intelligenza Artificiale (Ital-IA), Workshop "AI Generativa", CEUR Proceedings, vol. 3762,

[Online]: `https://ceur-ws.org/Vol-3762/500.pdf`

2024 **Leveraging GPT-like LLMs to Automate Issue Labeling**

Colavito, G. and Lanubile, F. and Novielli, N. and Quaranta, L., The Mining Software Repositories Conference (MSR24), Core Ranking A,

DOI: `https://doi.org/10.1145/3643991.3644903`

2024 **Continuous Quality Improvement of AI-based Systems: the QualAI Project**

Novielli, N. and Oliveto, R. and Palomba, F. and Calefato, F. and Colavito, G. and De Martino, V. and Della Porta, A. and Giordano, G. and Guglielmi, E. and Lanubile, F. and Quaranta, L. and Recupito, G. and Scalabrino, S. and Spina, A. and Vitale, A., Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '24), Barcelona, Spain, DOI: `https://doi.org/10.1145/3674805.3695393`

2024 **QualAI: Continuous Quality Improvement of AI-based Systems**

Novielli, N. and Oliveto, R. and Palomba, F. and Calefato, F. and Colavito, G. and De Martino, V. and Della Porta, A. and Giordano, G. and Guglielmi, E. and Lanubile, F. and Quaranta, L. and Recupito, G. and Scalabrino, S. and Spina, A. and Vitale, A., In Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement

[Online]: `https://ceur-ws.org/Vol-3674/RP-paper3.pdf`

2024 **The social adaptation as an emergent property of LLMs**

Curci, A. and Colavito, G., Giornale Italiano di Psicologia, Volume 51, Number 3, Pages 591-595, DOI: `https://doi.org/10.1421/114434`

2023 **Impact of Data Quality for Automatic Issue Classification Using Pre-trained Language Models**

Colavito, G. and Lanubile, F. and Novielli, N. and Quaranta, L., Journal of Systems and Software, Scimago Q1 for Software,

DOI: `https://doi.org/10.1016/j.jss.2023.111838`

2023 **Few-Shot Learning for Issue Report Classification**

Colavito, G. and Lanubile, F. and Novielli, N., Intl. Workshop on Natural Language-based Software Engineering (NLBSE 2023), co-located with ICSE 2023, Melbourne, Australia, May 20 2023,

DOI: `https://doi.org/10.1109/NLBSE59153.2023.00011`

2022 **Issue Report Classification Using Pre-trained Language Models**

Colavito, G. and Lanubile, F. and Novielli, N., Intl. Workshop on Natural Language-based Software Engineering (NLBSE 2022), co-located with ICSE 2022, Pittsburgh, USA, May 8 2022,

DOI: `https://doi.org/10.1145/3528588.3528659`

2022 **Lexicon Enriched Hybrid Hate Speech Detection with Human-Centered Explanations**

M. Polignano, G. Colavito, C. Musto, M. de Gemmis, G. Semeraro. In: UMAP '22 Adjunct. Barcelona, Spain: Association for Computing Machinery, 2022, pp. 184–191,

DOI: `https://doi.org/10.1145/3511047.3537688`

2020 **Leveraging Textual and Non-Textual Features for Documentation Decluttering**

Colavito, G. and Basile, P and Novielli, N., 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME), 2020, pp. 862-863, Core Ranking A,

DOI: `https://doi.org/10.1109/ICSME46990.2020.00113.`

Professional Service

2026 **PC Member @ FORGE 2026 (co-located with ICSE 2026)**

Conference on AI Foundation Models and Software Engineering

`https://conf.researchr.org/home/forge-2026`

2026 **Web Chair @ ICPC2026 (co-located with ICSE 2026)**

34th IEEE/ACM International Conference on Program Comprehension (ICPC 2026)

`https://conf.researchr.org/home/icpc-2026`

2026 **Program Co-Chair @ NLBSE2026 (co-located with ICSE 2026)**

I'm organizing the NLBSE2026 workshop, co-located with ICSE2026.

5th International Workshop on Natural Language Based Software Engineering

`https://nlbse2026.github.io/`

2025 **Tool Competition Co-chair @ NLBSE2025**

I organized the Code Comment Classification Tool Competition, contributing to curating code, evaluating metrics, and reviewing submissions.

4th International Workshop on Natural Language Based Software Engineering (NLBSE2025)

`https://nlbse2025.github.io/`

2025 **Junior PC Member @ MSR 2025 (co-located with ICSE 2025)**

22nd Mining Software Repositories Conference

`https://2025.msrconf.org/track/msr-2025-junior-pc`

2025 **PC Member @ FORGE 2025 (co-located with ICSE 2025)**

Conference on AI Foundation Models and Software Engineering

`https://conf.researchr.org/track/forge-2025/forge-2025-papers#Call-for-Papers`

2024 **Guest editor for Science of Computer Programming Journal (SCP)**

NLBSE'24 Special Issue in the Software Track of the Journal of Science of Computer Programming

`https://www.sciencedirect.com/journal/science-of-computer-programming/about/call-for-software`

2024 **Tool Competition Co-chair @ NLBSE2024**

I organized the Issue Report Classification Tool Competition, for which I have curated code, created the dataset, and reviewed submissions. 3rd International Workshop on Natural Language Based Software Engineering (NLBSE2024)

`https://nlbse2024.github.io/`

2024 **Student Volunteer @ ICSE2024**

The IEEE/ACM International Conference on Software Engineering

2023 **Student Volunteer @ ICSE2023**

The IEEE/ACM International Conference on Software Engineering

Reviewer

2025 **Automated Software Engineering**

2025 **Journal of Systems and Software**

2024 **Information and Software Technology Journal**

Reviewer for the Special Issue on Generative AI

2024 **The ACM International Conference on the Foundations of Software Engineering (FSE) 2025**

Co-reviewer

2024 **The IEEE/ACM International Conference on Software Engineering (ICSE) 2025**

Co-reviewer

2024 **The IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)**

Co-reviewer

2024 Empirical Software Engineering Journal (EMSE)

2024 The International Conference on Software Maintenance and Evolution (ICSME) 2024

Co-reviewer

2024 Journal of Software: Evolution and Process

2024 The 1st ACM international conference on AI Foundation Models and Software Engineering (FORGE 2024)

Co-reviewer

2024 International Conference on Evaluation and Assessment in Software Engineering (EASE) 2024

Co-reviewer

2023 PeerJ Computer Science

2023 The Mining Software Repositories (MSR) conference 2024

Co-reviewer

Talks

April 28th 2025 ICSE 2025 - Doctoral Symposium

I presented my PhD research on the use of foundation models to automate labeling tasks in software engineering, focusing on issue classification as a representative case. I discussed the challenges of applying these models to domain-specific tasks and how techniques such as prompt engineering and few-shot learning can improve their performance. I explored how to adapt foundation models to accurately classify issues in tracking systems, reducing manual labeling effort while maintaining high quality. I also examined strategies for integrating these automated approaches into real-world development workflows.

October 17th 2024 IBM Research - Almaden

The classification of issue reports is a critical task in modern software development, facilitating effective project management and decision-making. Early approaches to automating this task focused on BERT-based models, which have set the standard for supervised classification. However, the rise of generative models like GPT has introduced new possibilities, enabling issue classification even in the absence of labeled data. This talk will present findings from a comprehensive benchmark study comparing BERT-based and GPT-like models, highlighting trade-offs in performance, resource consumption, and deployment costs. Based on these findings, guidelines will be provided for effectively leveraging both types of models, depending on the available computational resources, as well as the quality and quantity of data. While each approach offers unique strengths, careful consideration is needed in scenarios with limited data or hardware constraints.

April 15th 2024 MSR 2024

I presented the ”Leveraging GPT-like LLMs to Automate Issue Labeling” paper during the International Conference on Mining Software Repositories (MSR) 2024, co-located with ICSE 2024

July 12th 2023 SATToSE 2023

I presented a summary of my work on Issue Report Classification using Pre-Trained Language Models and Few-Shot Learning, during the 15th Seminar on Advanced Techniques & Tools for Software Evolution (SATToSE)

May 20th 2023 NLBSE 2023, co-located with ICSE 2023

I presented the ”Few-Shot Learning for Issue Report Classification” paper during the International Workshop on Natural Language-based Software Engineering (NLBSE) 2023, co-located with ICSE 2023

May 8th 2022

NLBSE 2022, co-located with ICSE 2022

I presented the "Issue Report Classification Using Pre-trained Language Models" paper during the International Workshop on Natural Language-based Software Engineering (NLBSE) 2022, co-located with ICSE 2022

September 29th 2020

ICSME 2020, co-located with ICSE 2020

I presented the "Leveraging Textual and Non-Textual Features for Documentation Decluttering" paper during the DocGen2 workshop, at the IEEE International Conference on Software Maintenance and Evolution (ICSME) 2020, co-located with ICSE 2020

Schools

June 12th - 15th 2023

16th International Summer School on Software Engineering (ISSSE)

Organized by the University of Salerno.

June 5th - 9th 2023

2nd AI & Society Summer School

Organized by the National PhD Course in Artificial Intelligence.

April 3rd - 7th 2023

9th International Spring School on Deep Learning (DeepLearn)

Organized by the University of Bari, Universitat Rovira i Virgili, and IRDTA.

Personal skills

Coding C, C++, Java, Haskell, Prolog, Python and scientific packages (Pandas, NumPy, sklearn, Tensorflow, PyTorch, HuggingFace Transformers, langchain)

Machine Learning Supervised and Unsupervised Learning, Neural Network Architectures, Natural Language Processing, Large Language Models, Deployment and Performance Monitoring of Text Classifiers, Sentiment Analysis, Computer Vision.

Databases Cassandra, MySQL, Neo4j, Oracle, MongoDB

Developing environments Git, GitHub, PyCharm, Anaconda, Jupyter Notebook, Google Colab, Visual Studio Code, Cursor, IntelliJ IDEA

Spoken languages English (Professional)

- Level B2 certification issued by the teacher responsible for the English language courses for the Degree in Computer Science of the University of Bari.

Italian (Native)