Claudia **Mamede**

PhD student (AI for Software Security)

Porto, Portugal

cmamede@andrew.cmu.edu | claudiarmamede.github.io | claudiarmamede | claudiarmamede

My goal is to make developer's lives easier and software more secure.

Summary

PhD Student in Software Engineering and Computer Science at Carnegie Mellon University (USA) and University of Porto (PT). Active member of the QUASAR and SquaresLab research groups.

Interested in **vulnerability detection and repair** • **program analysis** • **large language models** • **machine learning for software engineering**

Education

Carnegie Mellon University

Pittsburgh, PA, USA

[Dual Degree] PhD in Software Engineering — *Sep. 2022 - Aug. 2028 (?)*

- Funded by FCT under the CMU Portugal Program
- Thesis topics: Large language models for security; interpretability and explainability; program analysis; program repair
- Current GPA: 4.0 out of 4.0
- *Advised by Dr Rui Abreu, Dr Claire Le Goues and Dr Jose Campos*

Faculty of Engineering of University of Porto

Porto, Portugal

[Dual Degree] PhD in Computer Science — *Sep. 2022 - Aug. 2028 (?)*

- Funded by FCT under the CMU Portugal Program
- Thesis topics: Large language models for security; program analysis; program repair; interpretability and explainability
- Current GPA: 18 out of 20
- *Advised by Dr Rui Abreu, Dr Claire Le Goues and Dr Jose Campos*

Faculty of Engineering of University of Porto

Porto, Portugal

M.Sc. in Informatics and Computing Engineering — *Sep. 2020 - Aug. 2022*

- Thesis: A transformer-based IDE plugin for vulnerability detection (graded 19 out of 20)
- *Advised by Dr Rui Abreu*

Faculty of Engineering of University of Porto

Porto, Portugal

B.S. in Informatics and Computing Engineering — *Sep. 2016 - Aug. 2020*

Publications

[Paper] Interpretable Vulnerability Reports

ICSE'26 (In submission; A conference)*

Claudia Mamede, Claire Le Goues, José Campos, Rui Abreu — *Feb 2025*

- We propose an interpretability convention for vulnerability reports, as well as the tools to generate and validate these reports. We also conduct a user study to validate the proposed standard considering the developer's perspective on report interpretability.

[Paper] Are Large Language Models Memorizing Bug Benchmarks?

LLM4Code'25 (co-located with ICSE)

Daniel Ramos, **Claudia Mamede,** Kush Jain, Paulo Canelas, Catarina Gamboa, Claire Le Goues — *Oct 2024*

- We systematically evaluate popular LLMs to assess their susceptibility to data leakage from widely used bug benchmarks. To identify potential leakage, we a study of benchmark membership within commonly used training datasets, negative log-likelihood and n-gram accuracy. Our findings show that certain models exhibit significant evidence of memorization in widely used benchmarks like Defects4J, while newer models trained on larger datasetse xhibit limited signs of leakage. **Won best paper award.**

[Poster] Interpreting Deep Learning Models Finetuned for Detection of Vulnerabilities Arising From Missing Code

Ciencia'24

Claudia Mamede, Claire Le Goues, José Campos, Rui Abreu — *Jun 2024*

- After identifying the vulnerabilities that can be explained, we shifted our focus to those that are typically harder to explain. In investigating the reasons for this, we found that vulnerabilities stemming from missing code cannot be addressed by traditional explainability methods, such as feature attributions. These methods rely heavily on input tokens, and when a critical input token is absent, they fail to provide meaningful explanations.

[Poster] An Empirical Evaluation of Explainable AI For Vulnerability Detection and Localization *CMU Portugal Doctoral Symposium'23*

Claudia Mamede, Claire Le Goues, José Campos, Rui Abreu *Dec 2023*

- We empirically evaluated the explanations for three different models to investigate whether or not these explanations could be used for vulnerability location. Our findings indicate that, while certain explanations may assist in locating vulnerabilities (depending on vulnerability type), they are generally not effective in most cases.
- Won best poster award

[Paper] Exploring Transformers for Multi-Label Classification of Java Vulnerabilities *QRS'22*

Claudia Mamede, Eduard Pinconschi, Rui Abreu *Dec 2022*

- We explore different BERT-based models for multi-label classification of vulnerabilities in Java on a synthetic dataset.

[Paper] A Transformer-based IDE Plugin For Vulnerability Detection *ASE'22 (A* Conference)*

Claudia Mamede, Eduard Pinconschi, Rui Abreu *Oct 2022*

- We built a transformer-based Visual Studio Code plugin to help developers identify vulnerabilities as quickly as possible.

Honors, Awards & Grants

Awards

2025	**Best Paper Award**, LLM4Code (co-located with ICSE'25)	*Ottawa, Canada*
2023	**Best Poster Award**, CMU Portugal Doctoral Symposium	*Pittsburgh, U.S.A*
2023	**1st place**, Green Hackathon 2023: Sentiment Analysis	*Denmark*
2022	**Best Master Thesis Award**, *Prémio Raul Vidal* for best master thesis in Software Engineering at University of Porto)	*Porto, Portugal*

Grants

2023	**CMU Portugal Fellowship for Dual Degree Program**, Awarded by the Fundação para a Ciência e a Tecnologia (FCT). It includes monthly stipend and tuition fees in both universities. Duration: 5 years.	*Portugal and USA*

Miscellaneous

2025	**Completed "Introduction to Safe and Secure AI: Attacks, Threats and Defenses" at FEUP**, Attended a three-day workshop covering core concepts and metrics of security vulnerabilities, adversarial attacks and defenses (including adversarial training and certifiable robustness), data poisoning and backdoor mitigation, model authentication techniques (model-stealing prevention and watermarking), and privacy attacks (membership inference, model inversion, and distributed-scenario defenses).	*Portugal*
2010 to 2022	**Basketball player**, team captain; won *Campeonato Académicos do Porto* (in college); won national and regional championships (federated); played in Liga Feminina de Basquetebol (premier women's basketball league in Portugal)	*Portugal*

Projects

SALVE – Securing Artificial Language Models Against Vulnerability Encoding *Portugal/Netherlands*

Researcher *ongoing*

- This project aims to improve code generation by optimizing LLM training to penalize the introduction of vulnerable code patterns. We are using contrastive learning and penalty terms to achieve this.

Leveraging data leakage to improve vulnerability detection and repair *Portugal/USA*

Researcher *ongoing*

- Building on our previous research in memorization and data leakage, we will adapt unknown samples into patterns that the model has already memorized. This will be achieved through a combination of genetic algorithms, adversarial learning, and code transformations, including AST-based modifications, to systematically manipulate and evaluate the model's recognition and generalization capabilities.

SecurityAware *Portugal/USA*

Researcher *2022 - 2023*

- Following the CodeAware vision, this project proposes a novel framework for the automatic and efficient detection of security issues, designed for seamless integration into CI pipelines. By implementing more fine-grained, unified, and faster approaches to CI static analysis, the framework enhances both accuracy and confidence in security assessments.
- **Ref:** Corina Pasareanu and Rui Abreu

Transformer-based VSCode plugin for vulnerability detection (Java)

Porto, Portugal

M.Sc. student / Researcher *2022*

- **VDet for Java** is a transformer-based VS Code extension that allows developers to locate software vulnerabilities at function-level. Our preliminary model evaluation presents an accuracy of 98% for multi-label classification and can detect up to 21 vulnerability types.

Work Experience

ARPA Elastic and FEUP

Porto, Portugal

MSc student / Researcher *Sep. 2021 - Jan. 2022*

- Built a XAML linter to detect and fix code smells in files created with RPA tools
- *Achieved goals:* the linter promotes early detection of code smells and issues in XAML files; it enforces the coding standards of the company across projects; reduces manual effort during code reviews.
- **Ref:** Joao Pascoal Faria

Fashable and FEUP

Porto, Portugal

MSc student / Researcher *Feb. 2021 - May. 2021*

- Intelligent mobile app to help customers shop for fashion items
- *Achieved goals:* data collection to train the models; created the logo of the company; developed the first prototype of the app.
- **Ref**: Rui Abreu

Onalytics and FEUP

Porto, Portugal

MSc student / Researcher *Sep. 2020 - Jan. 2021*

- Inventory and secure provisioning of IoT devices (for the beverage industry)
- *Achieved goals:* improved data collection and preprocessing process on ESP32 devices; integration with a cloud service (AWS)
- **Ref**: Filipe Correia

Community Services

Student Volunteer

- At ICSE'24, in Lisbon.
- At Talk A Bit'22, in Porto.

Miscellaneous

About Me

- Nationality: Portuguese (EU Citizen)
- Languages spoken: Portuguese (native), English (fluent), Spanish (fluent), French (elementary)
- Proficient in Python, Java, C++, C# and C
- LLM power-user; familiar with prompting stratxsegies and large language models in general.