

# David Nader Palacio

Williamsburg, VA · 23185

Phone: +1 (757) 279-4265 | Email: danaderp@gmail.com | Webpage: danaderp.github.io/danaderp/

GitHub: github.com/danaderp | LinkedIn: linkedin.com/in/david-n-palacio-91b67423

## Summary

Ph.D. candidate specializing in deep learning and software engineering, with expertise in applying causal inference techniques to interpret large language models (LLMs), aimed at automating software maintenance tasks. Eager to leverage these skills as a Research Scientist focused on AI trustworthiness, alignment, and explainability.

## Skills and Interests

<b>Technical Skills</b>	Python, NumPy, Pandas, PL/SQL, Scala, Java, Docker, nbdev.
<b>Research</b>	Explainability, Interpretability, Trustworthiness, Code Generation, Software Automation.
<b>Machine Learning</b>	PyMC3, PyTorch, Keras, Tensorflow, FastAI, Scikit-learn, Spark.
<b>Deep Learning</b>	Foundation Models, Large Language Models, GPT, BART, BERT, T5, LSTMs, Unsupervised Models.
<b>Languages</b>	English (Fluent). Spanish (Native). German (Intermediate)
<b>Soft Skills</b>	Mentoring, Critical Thinking, Adaptability, Cooperation, Diversity Awareness, Flexibility, Versatility

## Education

<b>College of William &amp; Mary (W&amp;M)</b>	Williamsburg, VA
Ph.D., Computer Science [GPA: 3.75]	05/2024
• <b>Dissertation:</b> Towards an Interpretable Science of Deep Learning for Software Engineering: A Causal Inference View [paper link]	
<b>National University of Colombia (UNAL)</b>	Bogota, Colombia
M.Sc., Computer Engineering [GPA: 3.85]	12/2016
• Award: <i>Indian Government Scholarship</i> in Noida, India [C-DAC link]	
• Award: <i>Awarded Graduate Assistantship (Top 5%)</i>	
• BerkeleyX Certificate: Distributed Machine Learning with Apache Spark	
<b>Technische Universität München (TUM)</b>	Munich, Germany
M.Sc., Computer Engineering [Exchange Program]	10/2013
• <i>Exchange program (top 1%) in Machine Learning and Data Mining</i>	
<b>National University of Colombia (UNAL)</b>	Bogota, Colombia
B.Sc., Computer Engineering [GPA: 3.7]	12/2011
• Award: <i>Undergraduate Scholarship (Top 5%)</i>	

## Professional Experience

<b>Microsoft, Data &amp; AI Team</b>	Seattle, WA
Research Intern	10/2021 - 03/2022
• Published a patent for an innovative debugging tool designed to provide recourse to practitioners in explaining deep generative models trained on code as data. Worked in collaboration with four Microsoft Senior research scientists.	
• Developed a pioneering Python tool that leverages explainability rationales and Shapley values to facilitate debugging and bias detection for BART, BERT, and GPT-based models when applied to code generation.	
• Researched a local and global post hoc interpretability method to contextualize and understand Large Language Models used for code completion and test generation [project link].	
<b>Cisco Systems</b>	Research Triangle, NC
PhD Intern	05/2020 - 08/2020
• Investigated an information theory approach for interpreting and evaluating software retrieval techniques [project link].	
• Applied Natural Language Processing techniques on Natural Language and Code to address the problem of software traceability [ICSE'20 link].	
• Prototyped a Jenkins plugin in Python for analyzing security-related requirement traceability in the continuous integration pipeline.	
<b>KSMTI</b>	Bogota, Colombia
Software Engineer (Senior Back-end)	02/2016 - 12/2016
• Engineered reactive and functional programming architectures for enabling fast development of any type of marketplace business [project link].	
• Programmed automated pipelines in Scala for the construction and deployment of highly scalable software.	
• Reduced stakeholders' production costs by 45%.	

## Ministry of Education (SED)

Bogota, Colombia

Software Engineer (Full Stack)

03/2015 - 02/2016

- Maintained and refactored legacy ASP.NET software architecture of public high-school institutions in Bogota.
- Managed and instructed the adoption of software practices in the government institution optimizing 65% of the development process.

## Allianz AMOS SE

Munich, Germany

Software Tester (Intern)

04/2013 - 09/2013

- Created automatic test scripts in Java to integrate functional and non-functional reports reducing testing time by 8% on new releases.

## ITC Consultores SAS

Bogota, Colombia

Software Engineer (Leader)

01/2012 - 10/2014

- Managed the research team of seven computer scientists to optimize the software construction pipeline (achieved Level 4 CMMI).
- Refactored software business components in Java augmenting comprehensibility of the core financial system by 60%.
- Developed a critical PL/SQL back-end module of portfolio operations for the biggest financial entities in Colombia.
- Engineered the required architecture for a technology migration that impacts the core PLSQL system (helped client productivity by 40%).

# Research Experience and Projects

## Semeru Research Group (W&M)

Williamsburg, VA

Research Assistant

01/2017 - Present

- Worked at the SEMERU Research Group as a Lead Researcher under the advisement of Professor Denys Poshvanyk Ph.D, published 12 research papers across various top-tier Software Engineering conferences and journals.
- Defined a causal inference technique for interpreting large language models trained on code [TSE'24 link].
- Benchmarked causal inference datasets to Interpret Large Language Models for Source Code [ICSME'23 link].
- Curated code testbeds and datasets to evaluate and interpret LLMs for code generation [HugginFace Datasets].
- Surveyed the most prominent SE and DL conferences and journals (128 papers across 23 unique SE tasks) to propose general guidelines on the use of deep learning in software engineering [TOSEM'22 link].
- Investigated a T5 model to support four software-related tasks: automatic bug-fixing, assert statement generation, code summarization, and code mutant injection [ICSE'21 link].
- Proposed and developed a Bayesian probabilistic approach to improve the effectiveness of traceability links by 10% [ICSE'20 link].
- Designed and executed a convolutional neural net to identify security-related issues with 96% success rate [ICSME'19 link].
- Presented work at local and international software research conferences and journals such as ICSME, ICSE, TOSEM, and TSE.
- Reviewed and commented papers (over 100) for top-tier conferences in software engineering.

## Computer Science Department (W&M)

Williamsburg, VA

Teaching Assistant

08/2017 - 08/2019

- Assisted in teaching four undergraduate and graduate courses: Neural Networks & Deep Learning, Software Engineering, Software Development, and Reasoning Under Uncertainty.
- Supervised students in final projects, graded exams, and weekly homework.
- Mentored and trained three graduate students in research areas such as interpretability, deep learning, information theory, causal inference, statistics, and software maintenance and evolution.

## National University of Colombia

Bogota, Col

Graduate Researcher

06/2014 - 06/2022

- Developed a hybrid adaptive evolutionary algorithm in Java/Scala to detect and recommend feasible software refactorings [GECCO'18 link].
- Taught Computer Programming to two groups of 30 sophomores in weekly sessions for two semesters.
- Advised two graduates on master's theses concerning implementation and explainability of deep neural networks for code generation.

# Selected Publications

Patent

- **Palacio N. D.**, Tufano M, Clement C, Svyatkovskiy A. 2024. Debugging Tool For Code Generation Neural Language Models [MS 412268-US-NP].

Main Author

- **Palacio N. D.**, ..., Poshvanyk D. 2023. Evaluating and Explaining Large Language Models for Code Using Syntactic Structures.
- **Palacio N. D.**, Cooper N, Rodriguez A, Moran K, Poshvanyk D. 2023. Toward a Theory of Causation for Interpreting Neural Code Models.
- **Palacio N. D.**, McCrystal D, Moran K, Poshvanyk D. 2019. Learning to Identify Security-Related Issues Using Convolutional Neural Networks.
- **Palacio N. D.**, Rodriguez D, Gomez J. 2018. Assessing Single-Objective Performance Convergence and Time Complexity for Refactoring Detection. In Proceedings of the Genetic and Evolutionary Computation Conference.

Co-Author

- Mastropaolo A, Scalabrino S, Cooper N, **Palacio N. D.**, Poshvanyk D, Oliveto R, Bavota G. Studying the Usage of Text-To-Text Transfer Transformer to Support Code-Related Tasks. 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE).
- Moran K, **Palacio N. D.**, McCrystal D, Poshvanyk D. 2020. Improving the effectiveness of traceability link recovery using hierarchical bayesian networks. Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering Jun 2020.