Iván Arcuschin Moreno

✉ iarcuschin@gmail.com • 🌐 iarcuschin.com • FlyingPumba • in iarcuschin
Argentine citizen

About

I'm a close-to-be-graduated **PhD candidate in Computer Science at the University of Buenos Aires**, Argentina. I'm passionate about **Machine Learning** and **Software Engineering**, two topics which I've studied in-depth during my PhD. Right now, I'm fascinated by the way AI models are becoming incredibly common and making a big impact in society, even though *we still don't fully understand how they work*. I believe that in order to **ensure that AI is safe and aligned with humanity's values**, we need to achieve a comprehensive understanding of the inner workings of these models. This has motivated me to dive deep into the field of **Mechanistic Interpretability**, taking part in the **ML Alignment & Theory Scholars** (MATS) program, where I worked on a project to build a **benchmark** for evaluating circuit discovery techniques in compressed models. My PhD experience, along with my extensive teaching experience, has taught me how to **break down complex problems** into smaller parts, **analyze them thoroughly**, and communicate the results in a **clear and concise way**. It has also given me the skills to conduct **empirical studies** using proper statistical analysis. I'm a **fast learner** and **proficient programmer**, skills that I further honed during my **AWS Applied Scientist internship**, allowing me to effectively solve a wide range of challenging problems.

Education

2018 - 2024 **PhD candidate in Computer Science**, *University of Buenos Aires (UBA)*, Argentina
Thesis title: "Random Espresso Test Case Generation for Android". Supervisor: Dr. Juan P. Galeotti.
Expected graduation date: April 2024 (only pending dissertation defense).

2013 - 2018 **Licenciate (BSc + MSc) in Computer Science**, *UBA*, Argentina
Thesis title: "An Empirical Evaluation of Sapienz Approach for Automatically Generating Test Cases for Android Applications"

AI Safety & Alignment Research Experience

Jan-Mar 2024 **Research Scholar**, *ML Alignment & Theory Scholars*, (MATS)
Project title: *In Pursuit of Superposition: A Benchmark for Mechanistic Interpretability in Compressed Models*.
Mentorship: *Dr. Adrià Garriga-Alonso*, Research Scientist at FAR AI.
This project aimed to build a benchmark of synthetic transformers with known circuits for evaluating mechanistic interpretability techniques. We leverage the transformers generated by the Tracr tool, and propose an algorithm for non-linearly compressing the residual stream size of these models, making them more realistic and challenging, while at the same time preserving the ground truth circuit.
A public repository for this benchmark is available online at github.com/FlyingPumba/circuits-benchmark.

Nov-Dec 2023 **Auditor Participation**, *Neel Nanda's training program @ ML Alignment & Theory Scholars*
Completion of several ARENA tutorials on Mechanistic Interpretability and participation in reading groups for papers on the same topic.

Software Engineering Research Experience

My PhD's research focuses on improving the techniques for **automatic test case generation on Android apps**, specifically using Espresso, a popular UI testing framework for Android. I've also studied the effectiveness of **search-based algorithms**, such as genetic and evolutionary, for test generation in this context. My work has been published at international conferences: **ICSE** and **AST**. Additionally, I served as a mentor to several undergraduate students working on their Master's theses, helping me develop my leadership capabilities.

Published papers

2024 **Arcuschin I.**, Di Meo L., Auer M., Galeotti J., Fraser G., *Brewing Up Reliability: Espresso Test Generation for Android Apps*. *To appear in International Conference on Software Testing, Verification and Validation* **(ICST)**

2022 **Arcuschin I.**, Ciccaroni C., Galeotti J., Rojas J.M., *On the feasibility and challenges of synthesizing executable Espresso tests*. *International Conference on Automation of Software Test* **(AST)**

2021 **Arcuschin I.**, Galeotti J., Garbervetsky D., *An Empirical Study on How Sapienz Achieves Coverage and Crash Detection*. *Journal of Software: Evolution and Process* **(JSEP)**

2020 **Arcuschin I.**, Galeotti J., Garbervetsky D., *Algorithm or Representation? An empirical study on how SAPIENZ achieves coverage*. *International Conference on Automation of Software Test* **(AST)**

2020 **Arcuschin I.**, *Search-Based Test Generation for Android Apps*. *Doctoral Symposium at International Conference on Software Engineering* **(ICSE)**

Research visits

2019 **Visiting researcher**, *ERATO Metamathematics for Systems Design*, Tokyo, Japan
Visited **Prof. Dr. Fuyuki Ishikawa**. Duration: 3 months. I worked on the problem of **generating realistic test scenarios**, aiming to assess the quality of **self-driving vehicle's control software** in the presence of **unreliable ML components**.

2019 **Visiting researcher**, *University of Leicester*, Leicester, UK
Visited **Prof. Dr. José Miguel Rojas**. Duration: 2 months. This visit laid the foundations for the paper *"On the feasibility and challenges of synthesizing executable Espresso tests"*, in which we analyze the feasibility of leveraging state-of-the-art **Android generation tools** for producing Espresso tests by **translating their output to the Espresso API**.

Scientific Experiences & Awards

2020 Student Volunteer participation, *International Conference of Software Engineering* **(ICSE)**

2017 & 2018 Google Latin America Research Award, **Google**
Awarded for the research project "*EVOLUTIZ: Multi-objective Test Generation for Testing Evolving Android Applications*". This project was accepted for an extension in 2018.

2017 Student Volunteer participation, *International Conference of Software Engineering* **(ICSE)**

Industry

2022 **Applied Scientist Intern**, *Amazon Web Services (AWS)*, New York, U.S
Member of the Automated Reasoning Group. Developed a methodology to minimize **confidential software artifacts**, discarding sensitive information while preserving **security defects** for automated analysis on tools such as CodeGuru Reviewer. The methodology was implemented as a tool and deployed in production.

2017 - 2022 **Backend developer (Freelance)**, *Dubbing Digital*, Buenos Aires, Argentina
Designed and implemented RESTful APIs and DB models. Development of extensive test suite for backend. Technologies: **JavaScript**, **TypeScript**, **Node.js** and **PostgreSQL**.

Coursework in PhD curriculum

Artificial Intelligence and Machine Learning

- **Data Science with R: Statistical Foundations**, *Lecturers: Dr. Ana Bianco and Dr. Mariela Sued*
  Final project: Design, prepare and record a short lesson on how to implement a **Naive Bayes** classifier for **sentiment analysis** using **R**.
- **Ethics on AI**, *Lecturers: Dr. Vanina Martinez and Dr. Ricardo Rodriguez*
  Final project: Written essay analyzing the **ethical implications** of using **AI agents** for finding missing children in the context of **warfare**, based on the G20's OECD framework for the classification of AI systems.
- **Introduction to Machine Learning**, *Lecturers: Dr. Pablo Brusco and Dr. Matías Lopez-Rosenfeld*
  Final project: Oral presentation of the paper *"Competition-Level Code Generation with* ***AlphaCode***", and an evaluation of different ML algorithms (e.g., Decision Trees, SVM, etc.) on a simple classification problem.
- **Introduction to Natural Language Processing**, *Lecturer: Dr. Luciano del Corro*
  Topics: Word embeddings (word2vec), N-grams, LSTM models (ELMo), Transformer models (BERT, GPT).

Automated Software Engineering

- **Meta-heuristics**, *Lecturer: Dr. Irene Loisseau*
- **Development and Automated Testing of RESTful APIs**, *Lecturer: Dr. Andrea Arcuri*
- **Advanced Analysis and Automatic Synthesis of Programs**, *Lecturer: Dr. Diego Garverbetsky*
- **Models and Algorithms for Systems Analysis**, *Lecturer: Dr. Víctor Braberman*

Self-learning projects

- **Mechanistic Interpretability walkthroughs**
  Completed Neel Nanda's walkthrough for the paper *"Progress measures for grokking via mechanistic interpretability"* and tutorial for implementing a GPT-2 style transformer from scratch in PyTorch.
- **Software Engineering for AI (SE4AI) study group**
  Participated in a weekly reading group about SOTA techniques for **Automated testing and verification of ML-enabled systems**, organizated by the Software Engineering & Formal Methods laboratory at UBA.
- **Presentation of Mechanistic Interpretability paper at SE4AI study group**
  Oral presentation of the paper *"Progress measures for grokking via mechanistic interpretability"*, with a focus on the **mechanistic interpretability** research framework.

Teaching

2023, 2018 - 2019 **Head Teaching Assistant**, *University of Buenos Aires*, Argentina
Courses: *Software Engineering II*, *Computer Architecture II*, and *Operating Systems*.

2016 - 2018 **Teaching Assistant**, *University of Buenos Aires*, Argentina
Courses: *Introduction to Programming*, *Programming Paradigms*, and *Algorithms and Data Structures II*.