Iván Arcuschin Moreno

iarcuschin.com • FlyingPumba • iarcuschin • publications

Education

2018 - 2024 **Computer Science PhD**, *University of Buenos Aires*, Argentina
Thesis title: Random Espresso Test Case Generation for Android. Supervisor: Dr. Juan P. Galeotti.

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

Publications

2025 **Arcuschin I.***, Janiak J.*, Krzyzanowski R.*, Rajamanoharan S., Nanda N., Conmy A., *Chain-of-Thought Reasoning in the Wild is not Faithful. Workshop on Reasoning and Planning for LLMs (ICLR 2025)*, *Equal contribution

2025 Venhoff C.*, **Arcuschin I.***, Torr P., Conmy A., Nanda N., *Understanding Reasoning in Thinking Language Models via Steering Vectors. Workshop on Reasoning and Planning for LLMs (ICLR 2025)*, *Equal contribution

2025 Mueller A.*, Geiger A.*, Wiegreffe S.*, Arad D., **Arcuschin I.**, Belfki A., Chan Y.S., Fiotto-Kaufman J.F., Haklay T., Hanna M., Huang J., Gupta R., Nikankin Y., Orgad H., Prakash N., Reusch A., Sankaranarayanan A., Shao S., Stolfo A., Tutek M., Zur A., Bau D., Belinkov Y., *MIB: A Mechanistic Interpretability Benchmark. Accepted at 42nd International Conference on Machine Learning (ICML 2025)*, *Equal contribution

2024 Gupta R.*, **Arcuschin I.***, Kwa T., Garriga-Alonso A., *InterpBench: Semi-Synthetic Transformers for Evaluating Mechanistic Interpretability Techniques. 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks.*, *Equal contribution

2024 **Arcuschin I.**, Di Meo L., Auer M., Galeotti J., Fraser G., *Brewing Up Reliability: Espresso Test Generation for Android Apps. 17th IEEE International Conference on Software Testing, Verification and Validation (ICST 2024)*

2024 Auer M., **Arcuschin I.**, Fraser G., *WallMauer: Robust Code Coverage Instrumentation for Android Apps. ACM/IEEE 5th International Conference on Automation of Software Test (AST 2024)*

2022 **Arcuschin I.**, Ciccaroni C., Galeotti J., Rojas J.M., *On the feasibility and challenges of synthesizing executable Espresso tests. ACM/IEEE 3rd International Conference on Automation of Software Test (AST 2022)*

2021 **Arcuschin I.**, Galeotti J., Garbervetsky D., *An Empirical Study on How Sapienz Achieves Coverage and Crash Detection. Journal of Software: Evolution and Process. Volume 35, Issue 4, 2023 (JSEP)*

2020 **Arcuschin I.**, Galeotti J., Garbervetsky D., *Algorithm or Representation? An empirical study on how SAPIENZ achieves coverage. ACM/IEEE 1st International Conference on Automation of Software Test (AST 2020)*

2020 **Arcuschin I.**, *Search-Based Test Generation for Android Apps. Doctoral Symposium at 42nd International Conference on Software Engineering (ICSE 2020)*

AI Safety Research & Educational Programs

Jan 2025 - ongoing **Research Scholar**, *ML Alignment & Theory Scholars*, (MATS)
Mentorship: *Arthur Conmy* (Google DeepMind).
The project focuses on developing an approach for discovering instances of unfaithful chain-of-thought reasoning in existing open-source and closed-source models, without injecting any external biases into the prompt.

Oct-Nov 2024 **Neel Nanda's training program**, *ML Alignment & Theory Scholars*, (MATS)
I paired up with Jett Janiak on a research project that involved running patching and probing experiments to understand/detect unfaithful chain-of-thought. We were **accepted** as scholars to the MATS program under **Arthur Conmy**'s mentorship, on the Neel Nanda/Arthur Conmy stream.

Jan-Jul 2024 **Research Scholar**, *ML Alignment & Theory Scholars*, (MATS)
Mentorship: *Dr. Adrià Garriga-Alonso* (FAR AI).
The project involved building InterpBench: a collection of 86 semi-synthetic transformers with known circuits, designed to evaluate mechanistic interpretability techniques. This work was accepted at NeurIPS 2024.

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

Industry Experience

2022 **Applied Scientist Intern**, *Amazon Web Services (AWS)*, New York, U.S
Working in the Automated Reasoning Group on the CodeGuru Reviewer project.

Jan 2017 - Dec 2022 **Founding Backend developer** (part-time), *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.

Jan - Feb 2016 **Android Developer**, *Restorando*, Buenos Aires, Argentina
Development of Android app for restaurant reservations.

May 2013 - Mar 2015 **Android Developer** (part-time), *Ekumen Labs*, Buenos Aires, Argentina
Development of Android app to control drones using the ROS framework.

Software Engineering Research Experiences

2020 Student Volunteer participation, *42nd International Conference on Software Engineering (ICSE 2020)*

2019 **BEHAPI School**, *Leicester, UK*, July 2019
Participated in the Behavioural Approaches for API-Economy with Applications School at Leicester, UK. This school was supported by the EU H2020 RISE programme.

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"*.

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

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, *39th International Conference on Software Engineering (ICSE 2017)*

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*.

Mentoring

2024 **Lisandro Díaz Di Meo**, *University of Buenos Aires*, Argentina
MSc thesis title: View Ranking: A model-based algorithm for automatic test generation of mobile applications.

2023 **Joaquín Arribas & Nicolas Walter**, *University of Buenos Aires*, Argentina
MSc thesis title: An empirical study on the effectiveness of automatic code repair tools for Android applications.

2020 **Gustavo Giráldez**, *University of Buenos Aires*, Argentina
MSc thesis title: An empirical study on the effectiveness of coverage measurement tools in the Android platform.

2020 **Christian Ariel Ciccaroni**, *University of Buenos Aires*, Argentina
MSc thesis title: Improving Espresso test case generation for Android applications.

Reviewing services

2025 **Workshop on Reasoning and Planning for LLMs (ICLR 2025)**
2024 **Journal of Open Source Software (JOSS)**
2024 **International Symposium on Software Testing and Analysis (ISSTA)**
2023 **European Symposium on Programming (ESOP/FoSSaCS)**
2023 **International Symposium on Software Testing and Analysis (ISSTA)**
2022 **European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)**
2022 **ACM Transactions on Evolutionary Learning and Optimization (ACM TELO)**
2020 **International Conference on Software Engineering (ICSE)**
2019 **Formal Methods – Fun for Everybody (FMFun)**
2019 **Automated Software Engineering (ASE)**