

LISA ALAZRAKI

lisa.alazraki20@imperial.ac.uk • [Website](#) • [Scholar](#) • [GitHub](#) • [Linkedin](#) • [EU Citizen, UK Settled Status](#)

Third-year PhD student in the NLP Group at Imperial College London. Research interests: LLM reasoning and planning, model robustness.

EDUCATION

Imperial College London , <i>PhD Computing</i>	Supervisor: Marek Rei	2027
Imperial College London , <i>MSc Computing (AI and Machine Learning)</i>	Classification: Distinction	2021
The Open University , <i>Grad. Cert. Theoretical Statistics and Probability</i>	Classification: Distinction	2020
The Open University , <i>BSc (Hons) Computing & IT and Mathematics</i>	Classification: 1st Class	2019

Scholarships and awards: Imperial Computing Conference 2024 Poster Competition First Prize • IET Research Awards 2024 – Postgraduate Prize Alan Turing Institute Enrichment Placement Award 2024/25 • Sir Richard Stapley Trust Annual Grant 2024 • Imperial College Trust Grant 2023 IET Travel Award 2023 for International Travel • Sir Richard Stapley Trust Annual Grant 2023 • IEEE CogMI 2022 Best Student Paper Award EPSRC Doctoral Scholarship 2022 • Imperial College London Distinguished MSc Dissertation Award 2021 • DeepMind MSc Scholarship 2020/21 Open University Official Commendation from the Faculty of Maths, Computing and Technology 2017 • Leslie Walshaw Award 2016 in Mathematics

EXPERIENCE

Meta, *Research Scientist Intern* • London, UK Jun - Dec 2025

- Manager: Akhil Mathur. Team: Llama Reasoning and Planning.

Cohere, *Research Intern* • London, UK Jun - Dec 2024

- Manager: Max Bartolo. Team: Command Post-training.
- Developed a reinforcement learning pipeline for reverse engineering human preferences that boosts LLM-as-a-judge evaluation.
- Investigated implicit learning from mistakes, showing LLMs attain higher accuracy when not shown explicit corrective feedback.
- Completed two distinct research projects at the same time, both resulting in first-author papers.

Google, *Research Intern* • Amsterdam, Netherlands Jun - Sep 2023

- Manager: Thomas Mensink. Team: Perception.
- Developed a model-ensembling framework for knowledge-intensive VQA that beats SOTA by 5% on Encyclopedic-VQA.
- Presented the resulting publication at ICBINB at NeurIPS 2023.

Google, *Student Researcher* • London, UK Oct - Dec 2022

Research Intern • Zurich, Switzerland Jun - Sep 2022

- Manager: Hamza Harkous. Team: Applied Privacy Research.
- Developed a new pipeline for retrieval-augmented generation of user issues that was deployed to production.
- Improved recall of existing issues by 10x over the previous model, with comparable semantic accuracy for new issue generation.
- Granted a global patent as co-inventor of the overall system for navigating user feedback.

SELECTED PAPERS

How to Improve the Robustness of Closed-Source Models on NLI, *In review*. 2025

Joe Stacey, [Lisa Alazraki](#), Aran Ubhi, Beyza Ermis, Aaron Mueller, Marek Rei

Reverse Engineering Human Preferences with Reinforcement Learning, *In review*. 2025

[Lisa Alazraki](#), Yi Chern Tan, Jon Ander Campos, Maximilian Mozes, Marek Rei, Max Bartolo

No Need for Explanations: LLMs Can Implicitly Learn from Mistakes In-context, *In review*. 2025

[Lisa Alazraki](#), Maximilian Mozes, Jon Ander Campos, Yi Chern Tan, Marek Rei, Max Bartolo

Enhancing LLM Robustness to Perturbed Instructions: An Empirical Study, *BuildingTrust @ICLR'25*. 2025

Aryan Agrawal*, [Lisa Alazraki](#)*, Shahin Honarvar, Marek Rei (*Equal contribution)

How can representation dimension dominate structurally pruned LLMs?, *SLLM @ICLR'25*. 2025

Mingue Xu, [Lisa Alazraki](#), Danilo Mandic

Meta-reasoning Improves Tool Use in Large Language Models, *NAACL'25 Findings*. 2024

[Lisa Alazraki](#), Marek Rei

How (not) to ensemble LVLMs for VQA, *Proceedings on ICBINB @NeurIPS'23*, pp. 1-20. *PMLR*. 2023

[Lisa Alazraki](#), Lluís Castrejon, Mostafa Dehghani, Fantine Huot, Jasper Uijlings, Thomas Mensink

SKILLS

Programming languages Python, TypeScript, JavaScript, Java, Lua, MATLAB/Octave, Maxima, Solidity, Prolog, Unix/Bash, HTML, CSS
Libraries / frameworks PyTorch, TensorFlow, Keras, NumPy, Pandas, Scikit-learn, Transformers, NLTK, Jinja2, Matplotlib, React, Flask