

# LISA ALAZRAKI

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

2nd-year PhD student in the NLP Group at Imperial College London. Research interests: LLM reasoning and planning, model robustness.

## EDUCATION

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

<b>Meta</b> , <i>Research Scientist Intern</i> • London, UK	Jun - Dec 2025
<ul style="list-style-type: none"><li>Manager: Akhil Mathur. Team: Llama Reasoning and Planning.</li></ul>	
<b>Cohere</b> , <i>Research Intern</i> • London, UK	Jun - Dec 2024
<ul style="list-style-type: none"><li>Manager: Max Bartolo. Team: Command Post-training.</li><li>Developed a reinforcement learning pipeline for reverse engineering human preferences that boosts LLM-as-a-judge evaluation.</li><li>Investigated implicit learning from mistakes, showing LLMs attain higher accuracy when not shown explicit corrective feedback.</li><li>Completed two distinct research projects at the same time, both resulting in first-author papers.</li></ul>	
<b>Google</b> , <i>Research Intern</i> • Amsterdam, Netherlands	Jun - Sep 2023
<ul style="list-style-type: none"><li>Manager: Thomas Mensink. Team: Perception.</li><li>Developed a model-ensembling framework for knowledge-intensive VQA that beats SOTA by 5% on Encyclopedic-VQA.</li><li>Presented the resulting publication at ICBINB at NeurIPS 2023.</li></ul>	
<b>Google</b> , <i>Student Researcher</i> • London, UK	Oct - Dec 2022
<i>Research Intern</i> • Zurich, Switzerland	Jun - Sep 2022
<ul style="list-style-type: none"><li>Manager: Hamza Harkous. Team: Applied Privacy Research.</li><li>Developed a new pipeline for retrieval-augmented generation of user issues that was deployed to production.</li><li>Improved recall of existing issues by 10x over the previous model, with comparable semantic accuracy for new issue generation.</li><li>Granted a global patent as co-inventor of the overall system for navigating user feedback.</li></ul>	

## SELECTED PAPERS

<b>AgentCoMa: A Compositional Benchmark Mixing Commonsense and Mathematical Reasoning in Real-World Scenarios</b> , <i>In review</i> .	2025
<a href="#">Lisa Alazraki</a> , Lihu Chen, Ana Brassard, Joe Stacey, Hossein A. Rahmani, Marek Rei	
<b>How to Improve the Robustness of Closed-Source Models on NLI</b> , <i>In review</i> .	2025
Joe Stacey, <a href="#">Lisa Alazraki</a> , Aran Ubhi, Beyza Ermis, Aaron Mueller, Marek Rei	
<b>Reverse Engineering Human Preferences with Reinforcement Learning</b> , <i>In review</i> .	2025
<a href="#">Lisa Alazraki</a> , Yi Chern Tan, Jon Ander Campos, Maximilian Mozes, Marek Rei, Max Bartolo	
<b>No Need for Explanations: LLMs Can Implicitly Learn from Mistakes In-context</b> , <i>EMNLP 2025 (Oral)</i>	2025
<a href="#">Lisa Alazraki</a> , Maximilian Mozes, Jon Ander Campos, Yi Chern Tan, Marek Rei, Max Bartolo	
<b>Enhancing LLM Robustness to Perturbed Instructions: An Empirical Study</b> , <i>ICLR 2025 BuildingTrust</i>	2025
Aryan Agrawal*, <a href="#">Lisa Alazraki</a> *, Shahin Honarvar, Marek Rei (*Equal contribution)	
<b>How can representation dimension dominate structurally pruned LLMs?</b> , <i>ICLR 2025 SLLM</i>	2025
Mingue Xu, <a href="#">Lisa Alazraki</a> , Danilo Mandic	
<b>Meta-reasoning Improves Tool Use in Large Language Models</b> , <i>NAACL 2025 Findings</i>	2024
<a href="#">Lisa Alazraki</a> , Marek Rei	
<b>How (not) to ensemble LVLMs for VQA</b> , <i>NeurIPS 2023 ICBINB</i>	2023
<a href="#">Lisa Alazraki</a> , Lluís Castrejon, Mostafa Dehghani, Fantine Huot, Jasper Uijlings, Thomas Mensink	

## SKILLS

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