Víctor Abia Alonso
victor.abia-alonso@citystgeorges.ac.uk
(website)

EDUCATION

IES Villablanca (Madrid)	Compulsory Secondary Education	Final grade: **9.9/10**
IES Ramiro de Maeztu (Madrid)	International Baccalaureate	Final grade: **43/45**
University College London	Bsc Mathematics with Economics	Final grade: **2:1**
City St George's, University of London	MSc Artificial Intelligence	Final grade: **1:1**
City St George's, University of London	PhD, AI Ethics	Ongoing until 2027

ACADEMIC ACTIVITY

- November 2024 — **Graduate Teaching Assistant** — City St George's, University of London
  Served as a GTA of Object-Oriented Analysis and Design; marked coursework. Obtained Advance HE Associate Fellowship (AFHEA), a UK-recognised teaching qualification.
- May 2025 — **Dissertations (BSc&MSc), 1st marker** — City St George's, University of London
  Spent several weeks reading and analysing dissertations to provide exhaustive feedback.
- June 2025 — ***How to bypass the Turing Trap?* Summer School** — Aranjuez, Spain
  Attended a two-day Spanish-language summer school on technical approaches for value alignment problem, myths of the AI promise and the limits to computational cognition.
- July 2025 — **Cooperative AI Summer School** — Marlow, UK
  Engaged in this five-day program discussing frontier research by the Cooperative AI Foundation. Participated in lectures by Vincent Conitzer, Nora Ammann, Zarinah Agnew.
- August 2025 — **DemocrAI @ IJCAI2025** — Montreal, Canada
  Attended IJCAI2025 and presented at the DemocrAI workshop; received the best student paper award.
- October 2025 — **AI, Ethics and Society (AIES) 2025** — Madrid, Spain
  Served as reviewer for four papers and took part in conference discussions on multicultural LLM values, machine ethics, algorithmic fairness and power dynamics of AI.
- November 2025 — **Formal and Ethical Agents and Robots (FEAR)** — Manchester, UK
  (to come)

TECHNICAL EXPERIENCE

❖ **Applied Researcher, NATO's SAPIENCE Project** (January 2024 - August 2024)
Initiative aimed at advancing autonomous multi-agent drone technology for disaster response, structured as an inter-university competition. Lead the Mathematical and AI sub-team, with a strategic focus on constructing and leveraging AirSim-based RL environments for enhanced drone decision-making capabilities. Also, developing Deep Learning architectures for precise territorial mapping and identification of hazards and individuals requiring aid. Part-time, flexible paid position. Our team won the competition.

❖ **Research Intern, AI Safety Camp** (January 2024 - April 2024)
Worked in a 3 month part-time online paid research group supervised by Jett Janiak aimed to develop a repo to automatize for Small Language Models interpretability by studying the emergence of linguistic capabilities during training time. My tasks involved developing scripts for PyTorch models, managing datasets and writing-up results. Repo.

❖ **Research Intern, AI Safety Hub Labs** (July 2023 - September 2023)
Participated in a 12-week research program, partially conducted in Oxford, investigating inductive biases of Language Models after finetuning via Reinforcement Learning, under

the supervision of Bogdan-Ionut Cirstea. Duties involved replicating paper results, conducting literature review (utilizing Zotero software), drafting a preprint (with LaTeX), managing GPU allocation across local machines and v100s (using SSH keys), version control (with GitHub), and overall project organization. Paper published at SoLar 2023.

- ❖ **ARENA Virtual Program** (May 2023 - June 2023)
  Completed an intensive 6-week program designed to develop research engineering skills in AI Safety. Gained exposure to Mechanistic Interpretability, based on Anthropic's work on transformer circuits. Practiced causal interventions (path-patching and probing) on the activations of Large Language Models (LLMs) to decipher internal reasoning processes. Introduced to the principles of Reinforcement Learning. Read and replicated AI Safety papers.

- ❖ **Machine Learning Safety Scholars** (June 2022 - August 2022)
  Participated in an 8-week online program centered on enhancing undergraduate proficiency in technical Machine Learning and introducing participants to AI Safety. Developed neural networks using PyTorch, tackled assigned problems, participated in weekly discussions with fellows, and summarized Machine Learning papers.

OUTREACH & LEADERSHIP

- ❖ **Community Builder in London** (February 2022 - October 2022)
  Launched and facilitated 2 Effective Altruism (EA) Introductory Workshops and social events for EA UCL members, and organized a fundraising race which raised £3550 for the Against Malaria Foundation. Co-founded the London Existential Risk Initiative to foster x-risk communities at four London universities through reading groups, speaker events, and Machine Learning Upskilling Bootcamps. Recipient of a grant by Open Philanthropy.

OTHER RESEARCH and EARLY ACADEMIC WORK

- ❖ **Second Year Project on Game Theory and cancer cells** (June 2022)
  Chosen to do a 2-week interdisciplinary research project with 4 UCL Maths students focused on mathematical modelling techniques from economics applied to biological systems.

- ❖ **Red Teaming Challenge: "WELLBY calculation"** (April 2022 - June 2022)
  Selected by Training for Good to critically analyze the subjective well-being (SWB) estimates of the Happier Lives Institute related to malaria prevention via net distribution. Conducted a comprehensive review of SWB research papers.

- ❖ **Essay on Robert Nozick's ideas.** Mark: 33/34 (October 2019 - April 2020)
  I chose philosophy as the subject for my IB Extended Essay (4,000 words) to write about the implicit utilitarian foundation of Nozick's libertarian moral principles.

HIGH SCHOOL HONOURS

- **Silver medal** (10th place) at the *2020 Spanish Mathematical Olympiad.*
- **Bronze medal** (44th place) at the *2020 Spanish Physics Olympiad.*
- **Merit** on the *2020 Spanish Chemistry Olympiad.*
- **Winner** of the *2020 Joaquín Hernández Intercentros Mathematical Contest*
- **Winner** of the *2018 Spring Mathematical Contest of Madrid (Level III).*
- **1st prize** in *2018 CienciaShow national scientific monologue contest.*

[References are available upon request]