

# Arjun Yadav

arjun@arjunyadav.net

## Education

---

### The University of Manchester

BSc (Hons) Computer Science

Sep 2024 – Present

Manchester, United Kingdom

### Delhi Private School, Sharjah

Physics, Chemistry, Mathematics, Computer Science and English  
Activities and Societies:

Apr 2010 – Apr 2024

Sharjah, United Arab Emirates

- Co-founded and co-facilitated three editions of the Society of Logic and Rationality in Students - the UAE's premier high school logic and rationality club.
- Active member of the MUN school team.
- Leader of the school's Python club (PyTech) after 2023; taught Python to middle schoolers and organised the school's Inter-House Hackathon for 2023.
- Active member of the school's environment club (EcoGen).

Student Council and Leadership Roles: Vice Head Boy for 2023-24 and Head of Events - EcoGen for 2023-24. Formerly Coordinator - Environment during my junior year.

## Experience

---

### Member

Apr 2024 – Present

*Horizon Omega*

- Member of Horizon Omega, a registered non-profit that aims to be a platform that accelerates AI safety research by facilitating high-impact projects and connecting experts, researchers, and professionals.
- Further part of the Horizon Events team and previously a host of the Provable AI Safety Seminars.

### Fellow

Feb 2024 – Present

*AI Safety Fundamentals*

- Selected to the AI Safety Fundamentals: Governance Course in April 2024.
- Selected to the AI Safety Fundamentals: Alignment Course in February 2024.

### Developer

Dec 2023 – Present

*AI Plans*

- Part of the AI Plans development team, working primarily with the front-end and database management.

### Research Collaborator

Dec 2023 – Present

*Dioptra*

- Part of the general intelligence benchmark team, helping on projects related to machine learning safety and interpretability.
- Led to my first publication as a co-author!

### Team Member

Nov 2023 – Present

*Offline Streaming Systems*

- Member of Offline Streaming Solutions, an organisation that aims to provide offline streaming capabilities to the masses.
- Selected to the Middle East Deep Tech Hacker Space to work on this further.

**Blog Writer and Editor** Jul 2023 – Present

*Ocean Busters*

- Currently working with an international non-profit by writing and proofreading posts relating to the issue of ocean pollution.

**Committee Member and Volunteer** Jan 2023 – Present

*OptX*

- Volunteered for three events for Opportunities to Expand: a symposium between a surgeon + astrophysics professor and students, UAE AID volunteering for victims of the 2023 Turkey-Syria Earthquake, an Iftar distribution campaign in Dubai.
- Played a major role in research and development and in organising said events.

**Co-Founder** Dec 2022 – Present

*EA UAE*

- Co-founded the UAE's national EA group in late 2022 - helping to bridge the gap between EA AUS and EA NYUAD.

**Inductee** Dec 2023 – May 2024

*AI Safety Camp*

- Became an inductee to AI Safety Camp 2024! Working at Project 29: Organise the next Virtual AI Safety Unconference.
- The project itself led to the Virtual AI Safety Unconference (VAISU) 2024, which had over 40 in-depth and intriguing talks with speakers all across the world, with hundreds of participants also joining in internationally!

**Research Assistant** Jun 2023 – Dec 2023

*American University of Sharjah*

- Currently working with Mr Ali Reza Sajun and Dr Imran A. Zuolkernan at the CSE department on the issue of ghost images using efficient transformers.
- Learnt about transformers, semi-supervised learning and more!

**Student Moderator - Metaverse Committee** Sep 2023 – Oct 2023

*Project: Unboxed*

- Moderated the metaverse committee of Project: Unboxed, a youth mental-health conclave hosted at Delhi Private School, Sharjah to a successful plan-of-action.

**Event Organiser** May 2023 – Sep 2023

*TEDxYouth@DPSS 2023*

- Worked extensively with our media and outreach teams to organise the successful second edition of TEDx at Delhi Private School, Sharjah!

**Co-Resource Head - Earth and Climate Change Council** Dec 2022 – May 2023

*Youth Global Solutions Summit*

- Helped to develop the agenda for the council, review pre-summit papers and, in general, oversee the flow of the two days of fruitful debate.
- Won best council for our work.

**Beta Reader** Jan 2022 – Jan 2023

*Cold Takes*

- Beta reader for Cold Takes, a blog by the co-founder and co-CEO of Open Philanthropy, Holden Karnofsky.
- Learnt more about the art of giving constructive criticism and got the privilege to see exemplary philosophy and AI safety content early.

**Design and Research** Jul 2022 – Sep 2022

*EAGxIndia*

- Created mock-ups for the event's landing page.
- Worked on mapping document highlighting EA communities in Asia.

- Learnt "soft skills" such as using Slack effectively and how Asana works
- Overall, became a much better communicator and designer.

### Facilitator

Jul 2022 – Sep 2022

#### *Effective Altruism Virtual Programs*

- Facilitated the In-depth fellowship offered by EA Virtual Programs. Earned an average rating of 8.4 out of 10 from my cohort.

### Front-end Developer and Data Visualizer

Oct 2021 – Apr 2022

#### *Effective Altruism Data, Legal Priorities Project*

- Worked with a data scientist from New Zealand to help redesign an EA data visualizer.
- Worked with Legal Priorities Project for data visualization for their paper, "Protecting future generations".
- Learnt about Dash and Plotly with Python, as well as got more familiar with the Git workflow.
- Learnt more about matplotlib from work with Legal Priorities Project.

### *Awards & Honors*

---

#### **Outstanding Delegate - United Nations Office on Drugs and Crime**

*DXBMUN'23*

*Dec 2023*

#### **Best Delegate - United Nations Special Committee on Peacekeeping Operations**

*NMSMUN'23*

*Oct 2023*

#### **Open Philanthropy Consolation Prize**

*Open Philanthropy*

*Oct 2023*

#### **AP Scholar**

*CollegeBoard*

*Jul 2023*

#### **Best Speaker - United Nations Office for Outer Space Affairs**

*GFSMUN III*

*Jun 2023*

#### **Best Speaker - For-side - Inter-House Debate**

*Delhi Private School, Sharjah*

*May 2023*

#### **Second Rank - Euclid Mathematics Contest**

*University of Waterloo*

*May 2023*

#### **Youth Global Solutions Summit - Best Council**

*Delhi Private School, Sharjah*

*May 2023*

#### **Second Round - 10th National Math Contest**

*Abu Dhabi University, Al Ain*

*Mar 2023*

#### **First Position - Essay Writing Competition**

*Delhi Private School, Sharjah*

*Jan 2023*

#### **Team Leader of Winning Hackathon Team - Senior Category**

*Sharjah and N. Emirates Chapter of Code Battle Hackathon*

*Nov 2022*

#### **First-place Senior Category**

*ATLAB Well-being App Development Competition*

*2020*

#### **Best and Outstanding Delegate**

*Stay Home Model UN 4.0, Woodllem Summit'20 and Stay Home Model UN 8.0*

*2020, 2020, 2021*

#### **KENKEN - 2x Silver Medal Holder**

*KENKEN UAE*

*2019, 2020*

### *Skills and Knowledge*

---

**Programming and Markup Languages:** Knowledgeable in Python, JavaScript, Java and LaTeX.

**Python Modules:** Knowledgeable in Dash, Plotly, NumPy and matplotlib.

**JavaScript Frameworks:** Knowledgeable in React.js and Next.js.

**Mathematics:** Certified in Linear Algebra and Multivariate Calculus through the Mathematics for Machine Learning course offered by Imperial College London.

**AI Safety:** Taught AI safety to my fellow seniors during the Summer of 2023, and I possess certification for my skills via the Intro to ML course by the Center for AI Safety

### *Other Interests*

---

**Public Speaking and Debating:** Experienced in Model UN procedures, 2x Best Delegate, 1x Best Speaker, 1x Outstanding Delegate and 1x Honourable Mention.

**Effective Altruism:** Proficient in knowledge relating to Effective Altruism. Facilitated Effective Altruism Virtual Programs' In-Depth EA Program and attended EAGxOxford 2022.

**AI Safety:** I have been learning independently about AI safety since the Summer of 2023, I write about my new-found knowledge at <https://arjunyadav.net/notebook>.

**Environmental Awareness:** Helped to conduct and coordinate events relating to raising awareness on climate change, combating climate denialism, etc.

### *Publications*

---

#### **GameBench: Evaluating Strategic Reasoning Abilities of LLM Agents**

*Anthony Costarelli, Mat Allen, Roman Hauksson, Grace Sodunke, Suhas Hariharan, Carlson Cheng, Wenjie Li, Arjun Yadav*