

# Sheikh Abdur Raheem Ali

http://abdur-raheem.com

Email : ali@abdur-raheem.com

Mobile : +1 (604) 790 5369

## EDUCATION

---

- **Stanford University** Palo Alto, CA, USA  
*Advanced Software Systems, Dropped out of the Graduate Program after one Quarter* 2021 – Present
- **American University of Sharjah** Sharjah, UAE  
*Bachelor of Science in Computer Science; Minor in Applied & Computational Math* Sep 2017 – Dec 2020

## EXPERIENCE

---

- **Microsoft** Vancouver, BC, Canada  
*Software Engineer II* Apr 2021 – Present
  - Lead of Technical AI Alignment at Microsoft.
  - Youngest engineer ever hired to my level in Canada. Typically requires a PhD or equivalent experience.
  - Co-owned impersonation protection in Defender for Office 365. Worked with the FBI, NYPD, U.S DoD to protect 250m+ people against phishing threats. Cut mailbox intelligence false positive rates by 55%.
  - Reduced download time for Substrate (the #2 largest monorepo in MS) from 2-3 hours to  $\approx$  20 minutes. Used by over 10% of all developers at the company and saves at least \$1m per annum.
  - Climbed to rank #14/36,210 members on our internal bug reporting platform. My feedback has been used by 86 teams (including Github Copilot, Bing Chat, Azure OpenAI) to fix issues before customer impact.*Software Engineering Intern* Jul. 2020 – Apr 2021
  - Built priority accounts. Was asked to stay on contract after the end of my internship until I graduated.
- **SWVL** Swvl Headquarters, Dubai, UAE  
*Quantitative Finance Intern* Feb 2020 – Apr 2020
  - Rewrote the system that prices how much a ride-sharing startup unicorn charges for each transaction.
- **Dubai Electronic Security Center** Cyber Intelligence Lab, University of Dubai, Dubai, UAE  
*ML Research Intern* Dec 2019 – Jan 2020
  - Recruited by the Government's Head of Research & Innovation. Did a 5-week project in my first weekend.
- **American University of Sharjah** Academic City, Sharjah, UAE  
*Research and Teaching Assistant* Sep 2019 – Dec 2019
  - Some ML interpretability work (generalizing Sobol Sensitivity Indices for models with correlated inputs).

## PROJECTS

---

- **Projecting Attack Vectors across LLM Layers with Activation Streaming:** Developing manipulation resistant multi-agent systems for scalable content filtering. Mentored a team of student researchers under Berkeley AI Safety's Supervised Program in Alignment Research. Poster presentation at Technical AI Safety Conference (TAIS 2024) in Tokyo, Japan. Featured speakers include Dan Hendrycks and researchers from Anthropic and DeepMind.
- **Interpretable Forecasting with Transformers:** Eliciting accurate, well-calibrated, and honest predictions from gpt-4-base. Supported by independent grants and mentored by Nuño Sempere. External partnership with SERI MATS (Cyborgism stream) building high-bandwidth interfaces for human-LLM collaboration. More details at manifund page.

## AWARDS & SCORES

---

- **2<sup>nd</sup> place (USD 35,000) @ HITB+ CyberWeek AI Challenge:** Captain of a team including students from Harvard and UToronto that used genetic algorithms to modify existing malware to defeat virus detection agents.
- **Full Scholarship @ European Summer Program on Rationality:** Invited to the UK to attend seminars on topics related to applied mathematics, AI safety, and cognitive science. Taught by ex-Google employees and MIT graduates.
- **Perfect Score @ ACT:** First attempt. Out of 2.03m students who took the 2017 ACT, about 2000 got a full 36/36.

## SKILLS

---

- Programming, Prompting, Platform Scaling, Probability, Profiling, Post-Production, and Personalization