

# Gaurav Sett

Contact | [gauravsett@icloud.com](mailto:gauravsett@icloud.com) | [linkedin.com/in/gauravsett](https://www.linkedin.com/in/gauravsett) | [gauravsett.com](https://gauravsett.com)

## Education

---

### Georgia Institute of Technology

Master of Science in Computer Science

Atlanta, GA

2023-01 to 2023-08

- Machine Learning specialization; 4.0 GPA

### Georgia Institute of Technology

Bachelor of Science in Computer Science

Atlanta, GA

2019-08 to 2022-12

- Artificial Intelligence and Human-Computer Interaction threads; 3.9 GPA, Honors Program

## Work

---

### RAND Corporation

Technology and Security Policy Fellow

Washington, DC

2023-11 to Present

- Developing novel research agenda on affirmative cases for AI system safety for industry practices and government policy
- Drafted ten policy recommendations for mitigating CBRN risks from AI systems to be sent to the President
- Provided technical expertise on machine learning for perspective pieces and Senate policy briefs
- Developed strategy for research focus, talent acquisition, and external coordination for the center's new technical hub
- Provided guidance on collecting information on the development practices of AI companies to Department of Commerce
- Wrote feedback on corporate responsibility policies and suggested best practices for leading AI company

### Center for AI Safety

Fundraising Contractor

San Francisco, CA

2023-09 to 2023-10

- Conducted research on funding opportunities and developed strategy for donor engagement

### Washington Post

Data Science Intern

Washington, DC

2022-05 to 2022-08

- Categorized article topics in breaking news stream with LSTM and BERT models using PyTorch
- Created API to handle continuous collection, processing, and storage of articles using AWS ECS, Lambda, and S3
- Visualized breaking news topics and API performance on dashboard for journalists using Datadog

### Federal Reserve Board

Economics Research Intern

Washington, DC

2021-05 to 2021-08

- Developed text analysis tool regularly used in policy meetings to analyze corporate earnings calls using SpaCy
- Automated collection, processing, and storage of 200K documents; used parallelization to improve speed 8x with Dask
- Presented to economists on NLP techniques such as word vectors and LDA topic modeling with interactive Django app

### Georgia Tech Research Institute

Undergraduate Research Intern

Atlanta, GA

2020-05 to 2020-12

- Created classification models identifying conspiratorial anti-vax Reddit comments with 80% accuracy using SciKit-Learn
- Analyzed the partisan differences in tweets about COVID-19 topics from members of Congress using SpaCy
- Built GUI application enabling social scientists to collect and analyze Twitter API data using PyQt5

## Leadership

---

### Supervised Program for Alignment Research

Founder & Program Manager

Remote

2023-07 to 2023-09

- Launched program facilitating AI safety research projects for 15 advisors and 80 students at 12 universities
- Recruited participants, managed applications, and organized 18 projects in alignment, engineering, and policy

### AI Safety Initiative at Georgia Tech

Founder & Director

Atlanta, GA

2022-08 to 2023-10

- Founded group hosting events, seminars, bootcamps, and research projects in AI alignment and governance

- Managed programs for dozens of students, leading team operations, and supporting career transitions
- Open Philanthropy University Organizer Fellowship: Awarded \$7,100 grant to support work as group organizer
- Open Philanthropy Group Support Grant: Secured \$5,000 grant to support research and education programs

## Research

---

### Georgia Tech College of Computing

Atlanta, GA

#### Graduate Projects

2023-01 to 2023-08

- Forthcoming Paper: Conklin, Bae, Sett, Hoffman, & Biddle (2023). Report on the Conference on Ethical and Responsible Design in the National AI Institute: A Summary of Challenges. *Forthcoming*.
- Conference Presentation: Conklin & Sett (2023). AI Safety, Governance, and Alignment Tutorial. *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*.
- PhilBench: Developed value alignment benchmark for reward modeling methods using text corpora
- Language Model Analysis: Measured the ability of GPT-3 to rationalize moral judgements in a chatbot setting
- Co-Citation Prediction: Measured likelihood of paper co-citation using graph neural networks and BERT

#### Undergraduate Projects

2019-08 to 2022-12

- Gentrification Forecasting: Built transformer model forecasting of Atlanta gentrification with tax data using PyTorch
- Expert Models Review: Wrote literature review on potential development of chatbots with subject matter expertise
- Portfolio Clustering: Analyzed investment portfolios to identify clusters of asset price time-series using K-means
- Employment Review: Wrote literature review on the likely effects of AI on wages and employment over this century
- Social Media Review: Conducted literature review on how social media amplifies misinformation in public discourse

## Teaching

---

### Introduction to Cognitive Science

Atlanta, GA

Graduate Teaching Assistant

Spring 2023

Graduate Teaching Assistant

Summer 2023

## Skills

---

**Machine Learning** | Python | R | SQL | PyTorch | TensorFlow | HuggingFace | SciKit-Learn | Pandas | NumPy | Matplotlib

**Software Engineering** | Java | C | JavaScript | HTML/CSS | Git | APIs | AWS | React | Django | Agile