

Gaurav Sett

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

Education

MS in Computer Science — Georgia Institute of Technology (*Atlanta, GA*) 2023-01 to 2023-08

- Machine Learning specialization; 4.0 GPA

BS in Computer Science — Georgia Institute of Technology (*Atlanta, GA*) 2019-08 to 2022-12

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

Experience

Technology and Security Policy Fellow — RAND Corporation (*Washington, DC*) 2023-11 to Present

- Conducting technical analysis and policy research on AI risk management and regulation
- Writing research reports on safety cases for AI deployment and supervision of AI development
- Providing analysis to policymakers, hosting workshops to coordinate industry, government, and society

Fundraising Contractor — Center for AI Safety (*Remote*) 2023-09 to 2023-10

- Conducting donor research and analysis to identify potential supporters for leading AI safety non-profit

Data Science Intern — Washington Post (*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

Economics Research Intern — Federal Reserve Board (*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

Undergraduate Research Intern — Georgia Tech Research Institute (*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

Projects

Executive Director — AI Safety Initiative at Georgia Tech (*Atlanta, GA*) 2022-08 to 2023-10

- Founded group hosting events, seminars, bootcamps, and research projects in AI alignment and governance
- Managed programs for over 100 students, leading team operations, and supporting career transitions

Program Manager — Supervised Program for Alignment Research (*Remote*) 2023-07 to 2023-09

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

Graduate Researcher — Georgia Tech College of Computing (*Atlanta, GA*) 2023-01 to 2023-08

- Developing benchmark to measure ability for fine-tuning methods to learn human values described in text data
- Collecting philosophy papers and survey of philosopher views, measuring alignment between model and philosophers

Undergraduate Researcher — Georgia Tech College of Computing (*Atlanta, GA*) 2019-08 to 2022-12

- Language Model Analysis: Conducted experiment to measure GPT-3's ability to rationalize moral judgements
- News Classification: Measured news story salience based on topic similarity across outlets using fine-tuned BERT
- Gentrification Forecasting: Built transformer model forecasting of Atlanta gentrification with tax data using PyTorch

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