

# Gautam Mittal

mittal.ai | gautam@mittal.net | +1 (480) 648-8254 | github.com/gmittal

## EDUCATION

### STANFORD UNIVERSITY

M.S. Computer Science  
August 2023 | Stanford, CA  
GPA: 4.1 / 4.0

Advisor: Prof. Douwe Kiela  
Research with SAIL, CRFM, HAI

### UC BERKELEY

B.S. Electrical Engineering &  
Computer Science (EECS)

May 2022 | Berkeley, CA  
GPA: 3.9 / 4.0

Advisor: Prof. Ion Stoica  
Regents' and Chancellor's Scholar  
Accel Scholar, Kleiner Perkins Fellow  
Eta Kappa Nu (EECS Honor Society)  
Dean's List (3x), CS186 TA (2x)

## SKILLS

### LANGUAGES

Python • JavaScript • Swift • Java • C  
C++ • Go • SQL • Scheme • OCaml  
Ruby • RISC-V • x86 • Objective-C

### TOOLS

PyTorch • JAX/Flax • TensorFlow  
NumPy • Node.js • Flask • Rails • HTML  
CSS • React • Max/MSP • AWS • GCP  
UNIX • Git • MongoDB • PostgreSQL

## COURSEWORK

CS61C: Computer Architecture  
CS70: Discrete Math & Probability  
CS170: Algorithms & Intractability  
CS161: Computer Security  
CS162: Operating Systems  
CS164: Languages & Compilers  
CS184: Computer Graphics  
CS186: Database Systems  
CS188: Artificial Intelligence  
CS189: Machine Learning  
CS194-26: Computer Vision  
CS251: Blockchain & Cryptocurrencies  
CS285: Deep RL (Graduate)  
CS295: Software Eng. (Graduate)  
CS330: Meta Learning (Graduate)  
CS324: Foundation Models (Graduate)  
EECS16A: Linear Algebra & Circuits  
EECS16B: Diff. Equations & Control  
EECS126: Random Processes

## EXPERIENCE

### CONTEXTUAL AI | Research Engineer, Founding Team

April 2023 – June 2024 | Mountain View, CA

- Co-implemented distributed training, evaluation, and deployment frameworks for end-to-end retrieval (lexical & neural) and LM systems
- Project lead for **RAG 2.0**, co-implemented **LENS**, and led experiments across multimodal, retrieval, & language model pre- and post-training
- Benchmarking and scaling of HPC infrastructure to support training and inference of models up to 100B parameters
- Significant involvement in hiring and co-designed eng. loop (2→20+)

### RISELAB, UC BERKELEY | Student Researcher

September 2019 – May 2022 | Berkeley, CA

- Researching and developing Sky computing systems to enable large-scale multi-cloud ML workloads and deep RL techniques for query optimizers
- Co-implemented system for training a relational query optimizer without expert demonstrations and experimented with generative models, feature perturbation, and planning techniques to improve agent performance
- Published in **SIGMOD 2022** and **NSDI 2023**, core contributor to **SkyPilot** (5.7K+ stars)

### TESLA | Machine Learning Intern, Autopilot

May 2021 – August 2021 | Palo Alto, CA

- Engineering task owner for all offline 2D networks: supported new autolabeling, tracking, simulation, AutoHighbeam, and 3D network efforts
- Implemented SoTA panoptic segmentation, road semantics, and object detection models along with new data, training, evaluation, and visualization infrastructure from scratch
- Contributed to segmentation data engine, helping refine labeling ontology and implement system to improve label quality and diversity
- Presented model to Elon Musk and had internship work demoed by Autopilot leadership at AI Day 2021 (see 1:11:19 & 1:31:30 on livestream)

### GOOGLE | Research Intern, Google Brain

May 2020 – January 2021 | Mountain View, CA

- Researched deep energy-, score-, and diffusion-based generative models for symbolic music generation under the Magenta team
- Designed one of the first latent diffusion models, focused on unconditional and controllable generation
- Implemented and evaluated Transformer-based models, fast sampling mechanisms, and MusicVAE data pipelines with JAX, Flax, and TensorFlow
- First author paper (**ISMIR 2021**) on diffusion models for musical sequences

Earlier professional experience is available at [linkedin.com/in/mittalgautam](https://www.linkedin.com/in/mittalgautam).

Additional open-source work and projects available at [github.com/gmittal](https://github.com/gmittal).

All available publications and preprints can be viewed at [gbm.pw/gscholar](https://gbm.pw/gscholar).

## AWARDS

- 2022 Warren Y. Dere Design Award (top EECS senior for engineering design)
- 2022 Interact Fellow
- 2019 IEEE Eta Kappa Nu Member (top 25% of Berkeley EECS)
- 2019 Accel Scholar (run by Accel Partners & Berkeley EECS)
- 2019 Kleiner Perkins Engineering Fellow
- 2018 Regents' and Chancellor's Scholarship (top 2% of incoming class)
- 2015 Top 10 at MHacks, PennApps (international hackathons)