Anirudh Ajith

anirudh.ajith@gmail.com | anirudhajith.github.io | Anirudh Ajith | +1 (609) 375 4637

Education

Princeton University — **2022 – 2024 (ongoing)**
Master of Science in Engineering (M.S.E.), Computer Science; GPA: 4.00/4.00 — *Princeton, New Jersey, USA*

Indian Institute of Technology Madras — **2018 – 2022**
B. Tech, Computer Science and Engineering; CGPA: 9.52/10.00 — *Chennai, India*

Publications + Preprints

2023 Detecting Pretraining Data from Large Language Models
Anirudh Ajith*, Weijia Shi*, Mengzhou Xia, Yangsibo Huang, Daogao Liu, Terra Blevins, Danqi Chen, Luke Zettlemoyer
Accepted at ICLR 2024; Accepted (Oral) at RegML Workshop @ NeurIPS 2023;

2023 InstructEval: Systematic Evaluation of Instruction Selection Methods
Anirudh Ajith*, Chris Pan*, Mengzhou Xia, Ameet Deshpande, Karthik Narasimhan
Accepted (Spotlight) at R0-FoMo Workshop @ NeurIPS 2023; *Received ACL Rolling Review meta review score: 4/5; Committed to NAACL 2024;*

2023 Performance Trade-offs of Watermarking Large Language Models
Anirudh Ajith, Sameer Singh, Danish Pruthi
arXiv preprint arXiv:2311.09816

2023 Adapting Language Models to Compress Contexts
Alexis Chevalier, Alexander Wettig, **Anirudh Ajith**, Danqi Chen
Accepted at EMNLP 2023;

Research Projects

Benchmarking Instruction Selection methods for in-context learning — *Dr. Karthik Narasimhan* | **Feb – June 2023**

- Led the *InstructEval* project, designing and implementing a suite to systematically assess instruction selection methods for LLM in-context learning across various models and tasks, utilizing three accuracy metrics and two sensitivity metrics.
- Developed and introduced the *mean relative gain* metric for a comprehensive comparison of instruction selection methods, overcoming limitations of previous aggregation metrics.
- Conducted thorough comparative analyses of automated versus simple baseline instruction-selection techniques in ICL.

Detecting pretraining data from large language models — *Dr. Danqi Chen, Dr. Luke Zettlemoyer* | **July – Sept 2023**

- Co-led the development and testing of a novel method called *Min-K% Prob* for performing membership inference attacks on LLM pretraining data, and *WikiMIA*: a benchmark for testing MIA attacks on pretraining data.
- Spearheaded the conceptualization and execution of a case study demonstrating the efficacy of *Min-K% Prob* for detecting leakage of ICL benchmark data into LLM pretraining corpora.
- Co-led execution of a case-study into detecting the presence of copyrighted content in `text-davinci-003`'s pretraining data.

Assessing trade-offs of watermarking large language models — *Dr. Danish Pruthi, Dr. Sameer Singh* | **June – Oct 2023**

- Comprehensively investigated the impact of a popular technique for watermarking LLM outputs on downstream tasks including classification, multiple-choice question-answering and generation.
- Conducted thorough analyses to understand the underlying reasons for the performance impacts on each task type.
- Experimented with possible augmentations to this watermarking strategy to help recoup lost performance.

Extending effective context-window lengths in large language models — *Dr. Danqi Chen* | **April – June 2023**

- Collaborated in the development of a technique for converting off-the-shelf autoregressive LLMs into *AutoCompressor* models that are capable of processing larger context lengths by compressing discrete token sequences into short soft prompts.
- Instrumental in the design and execution of experiments to validate the utility of AutoCompressors in language modeling and various downstream tasks like in-context learning and retrieval-augmented language-modeling.

Tuning sentence-embeddings for high-recall IVFPQ search — *Dr. Mitesh Khapra, Dr. Pratyush Kumar* | **Aug – May 2022**

- Devised a method to improve recall of IVFPQ approximate nearest-neighbor search with the goal of improving bitext mining, and explored multiple training paradigms to enable this.
- Adapted an existing differentiable product quantization formulation to create an e2e trainable, differentiable formulation of IVFPQ quantization that outputs quantized representations and codes required for IVFPQ search.

Professional Experience

AI4Bharat | *Python, MongoDB* **Nov – May 2022**

- Worked on creating *Samanantar 2.0*: the largest ever publicly available parallel corpora for Indian languages.

Microsoft India (R&D) Pvt Ltd | *C#, Python, Microsoft COSMOS, other internal tools* **May – Jul 2021**

- Created a troubleshooting-snippet disambiguation pipeline for Microsoft's *Bing* search-engine.
- The pipeline takes a set of solution snippets (scraped from various websites using existing *Bing* infrastructure) to a tech-related troubleshooting search query and filters it down to a set of semantically unique solutions for direct display on the *Bing* SERP.

Flutura Decision Sciences & Analytics | *Python, TensorFlow, Keras* **May – Jul 2020**

- Developed computer vision models based on *YOLOv4* and *Retinanet*.
- Created computer-vision products for multiple clients from scratch on problems including 1) autonomous defect detection in die-casted components, 2) autonomous cell-phone usage detection and 3) defect detection in printed circuit boards.

Other Relevant Projects

CLIP-based histology annotation predictor | *Computer Vision* *Dr. Jia Deng* | **Dec 2022**

- Prepared a dataset of 25k tissue images and pathology notes by downsampling and cropping images from the GTEC database.
- Utilized CLIP framework to train a model for ranking tissue pathology diagnoses from biological tissue specimen slide images.
- Determined through qualitative analysis that nearest-neighbor search in the shared embedding space of pathology notes and tissue images enables non-obvious diagnoses.

automated B/W portrait colorization | *Computer Vision* *Dr. Sukhendu Das* | **Sep – Nov 2021**

- Created a pipeline which performs image restoration, colorization and enhancement using multiple published methods for photo-realistically converting B/W historical to color using few training samples.

σ-promoter classification | *Computational Biology, Deep Learning* *Dr. Manikandan Narayanan* | **Nov – Dec 2020**

- Improved a SOTA *E. coli* σ-promoter model with attention layers and residual connections, increasing accuracy by 1.6%.

automated attendance system | *Computer Vision* **May – Jul 2020**

- Created an autonomous attendance system pipeline for classrooms using the popular neural networks *MTCNN* and *FaceNet*.
- Utilized KNN to match faces from PTZ camera feed with identities using a database of approximately 4 photos per student.

Scholastic Achievements

2020 **IAS Fellowship** Recipient of Indian Academy of Sciences Summer Research Fellowship

2020 **Flipkart GRiD 2.0 Hackathon** Declared National level Semi-Finalist

2016 **KVPY** Secured All India Rank 108 in Kishore Vaigyanik Protsahan Yojana (SA)

2016 **NTSE** Secured National Talent Search scholarship

2015-18 **Indian National Olympiads** National Finalist in Computing/Informatics every year from 2015 to 2018, in Astronomy in 2017 & 2018 (State rank 1, National top 1%), Physics in 2018 (State rank 4, National top 1%) and Merit Certificate for State top 1% in Chemistry

2017 **National Mathematics Talent Contest** Secured All India Rank 9 in Ramanujan contest

2016-17 **Regional Mathematics Olympiad** Selected for Indian National Mathematics Olympiad Training Camp

Relevant Coursework

AI-related: Understanding Large Language Models (seminar); Natural Language Processing; Advanced Computer Vision; Reinforcement Learning; Foundations of Deep Learning; Pattern Recognition and Machine Learning; Long-Term Memory for AI Systems; Computational Models of Cognition; Statistical Foundations of Data Science

mathematics: Multivariable Calculus; Discrete Math for CS; Series and Matrices; Basic Graph Theory; Probability, Stochastic Processes and Statistics; Differential Equations; Linear Algebra

Positions of Responsibility

Reviewer | *PERSONALIZE Workshop @ EACL 2024* *Malta* | **March 2024**
Teaching Assistant | COS240: Reasoning About Computation *Princeton University* | **Fall 2022, Spring 2023, Fall 2023**
Project Member | Computer Vision and Intelligence Group *IIT Madras* | **2019**
Coordinator | Developmental Operations Team, Saarang 2020 *IIT Madras* | **2019**
President | Computer Science Association *National Public School, Rajajinagar* | **2013 – 2014**

Technical Skills

languages: C, C++, Python, Julia, JavaScript, Bash
software: Linux, Git, Docker, GNU Octave, LaTeX, GIMP, Google Sketchup
development: HTML, CSS, JavaScript, nodeJS, ReactJS, Angular