

Anirudh Ajith

☎ +1 (609) 375-4637 ✉ anirudh.ajith@princeton.edu ✉ anirudh.ajith@gmail.com

🌐 anirudhajith ✉ anirudhajith.github.io 🎓 Anirudh Ajith

Education

M.S. Computer Science, Princeton University, GPA: 4.00/4.00 Aug 2022 - May 2024

B.Tech. Computer Science, Indian Institute of Technology, Madras, GPA: 9.52/10.00 2018 - 2022

Relevant Coursework: Natural Language Processing, Large Language Models, Computer Vision, Reinforcement Learning, Deep Learning, Machine Learning, Computational Models of Cognition, Long Term Memory for AI

Publications & Preprints

[Detecting Pretraining Data from Large Language Models \[blog\]](#) ICLR 2024; NeurIPS RegML 2023 (Oral)

A. Ajith*, W. Shi*, M. Xia, Y. Huang, D. Liu, T. Blevins, D. Chen, L. Zettlemoyer

- Co-developed *Min-K% Prob* and *WikiMIA* to perform black-box membership inferences on LLM pretraining data, leading to a 7.4% improvement over previous SOTA methods.
- Led case studies using *Min-K% Prob* to detect leakage of 1) ICL benchmark data and 2) copyrighted content into LLM pretraining corpora demonstrating practical applications, generality and effectiveness.

[InstructEval: Systematic Evaluation of Instruction Selection Methods](#) NAACL 2024 (Findings)

A. Ajith*, C. Pan*, M. Xia, A. Deshpande, K. Narasimhan NeurIPS R0-FoMo 2023 (Spotlight)

- Led a project designing and implementing a suite to holistically assess instruction selection methods for LLM in-context learning across various models and tasks, utilizing 3 accuracy and 2 sensitivity metrics.
- Introduced *mean relative gain* to enable holistic comparisons, and isolated heuristics for optimizing in-context learning performance, revealing that expensive instruction-selection techniques are often unnecessary.

[Adapting Language Models to Compress Contexts](#) EMNLP 2023

A. Chevalier, A. Wettig, A. Ajith, D. Chen

- Contributed to creating *AutoCompressor* models, augmenting off-the-shelf autoregressive LLMs to process extended context lengths via innovative compression of token sequences into soft prompts.
- Designed evaluations for AutoCompressors, demonstrating enhanced performance in language modeling, in-context learning, and retrieval-augmented tasks.

[Performance Trade-offs of Watermarking Large Language Models](#) arXiv:2311.09816

A. Ajith, S. Singh, D. Pruthi Under submission at ACL 2024

- Investigated the effects of LLM output watermarking on diverse downstream tasks, identifying key impacts (and causal factors) on performance, enhancing the understanding of watermarking's broader implications.
- Explored techniques for mitigating LLM performance drops while maintaining watermark detection efficacy.

Work Experience

Indian Institute of Science — Research Intern (Bangalore, India) Jun 2023 - Aug 2023

- Systematically evaluated (previously unstudied) performance trade-offs of watermarking LLM generations.
- Developed strategy for recovering up to 92% of lost performance without sacrificing watermark signal strength.
- Our work culminated in the release of a preprint that is under submission at ACL 2024.

AI4Bharat — Researcher (Chennai, India) Nov 2021 - May 2022

- Worked on creating *Samanantar 2.0*, the largest ever public collection of parallel corpora for Indian languages.
- Leveraged FAISS, language-agnostic embeddings to mine 50M bitext pairs b/w English and 11 Indic languages.

Microsoft R&D — Data Science Intern (*Hyderabad, India*)

May 2021 - Jul 2021

- Created a semantic deduplication pipeline for answers to technical queries for Bing search using deep learning models and clustering techniques leading to significantly increased click-through rates during A/B testing.
- Received a pre-placement offer for full-time employment for demonstrating exceptional skills and competence.

Flutura Decision Sciences & Analytics — Summer Intern (*Bangalore, India*)

May 2020 - Jul 2020

- Created end-to-end deep-learning-based pipelines for computer-vision products for multiple clients.
- Fine-tuned CNNs and leveraged pretrained object-detection models to perform 1) defect detection in die-casted components, 2) cell-phone usage detection and 3) defect detection in printed circuit boards with > 96% accuracy

Centre for Innovation, IIT Madras — Project Member (*Chennai, India*)

April 2019 - April 2020

- Created an autonomous attendance system pipeline for classrooms that requires only 5 photographs per enrolled student and no additional training to achieve 0.93 F1 @ 1 over a classroom of size 50.
- Leveraged convolutional neural networks MTCNN and FaceNet, KMeans, and other machine learning algorithms.

Technical Skills

Languages	Python, Bash, C, C++, C#, Java, JavaScript, Julia, OCaML, SQL, TypeScript
Technologies	PyTorch, TensorFlow, Keras, Linux, Slurm, Git, Docker, L ^A T _E X, GNU Octave, GIMP
Development	HTML, CSS, JavaScript, nodeJS, ReactJS, Angular, Android

Academic Achievements

IAS Fellowship	Recipient of Indian Academy of Sciences Summer Research Fellowship	2020
Flipkart GRiD 2.0 Hackathon	Declared semi-finalist from a pool of 30k participants	2020
KVPY	Secured rank 108 out of 50k students in Kishore Vaigyanik Protsahan Yojana (SA)	2016
NTSE	Secured National Talent Search scholarship (awarded to top 1000 students among 900k applicants)	2016
Indian National Olympiads	National finalist in Computing, Astronomy, Physics and Chemistry olympiads achieving national top 1% positions as a top-ranker in the state.	2015-18
National Mathematics Talent Contest	Secured rank 9 among 66k applicants in Ramanujan contest	2017
Regional Mathematics Olympiad	Selected for Indian National Mathematics Olympiad Training Camp	2016-17