Anirudh Ajith

anirudhajith.github.io anirudh.ajith@gmail.com github.com/anirudhajith linkedin.com/in/anirudhajith

Education

Princeton University **2022 – (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*

Research Projects

Using LLM interpretability techniques for prompt evaluation *Dr. Karthik Narasimhan* | **Oct 2022 – (ongoing)**

- Conducting experiments testing if quantities computed using recent work in large language model interpretability can be used to predict the quality (measured as the test set accuracy obtained when using the prompt) of a candidate LLM prompt.

Dataset distillation for NLP datasets *Dr. Danqi Chen* | **Dec 2022 – (ongoing)**

- Working on adapting and applying dataset distillation techniques developed for computer vision datasets to textual datasets.
- Applying bilevel optimization and backpropagation through time frameworks to optimize input embeddings to create a synthetic distilled dataset.

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-neighbour search with the goal of improved bitext mining.
- Adapted an existing differentiable product quantization formulation to create a differentiable formulation of IVFPQ quantization that outputs quantized representations and codes required for IVFPQ search which is e2e trainable in a neural network.
- Devised and tested training paradigms to allow the model to optimise sentence embeddings it generates to make them more suitable for high-recall bitext mining (when using IVFPQ indexing).

Sample-specific attention-head masks in BERT models *Dr. Pratyush Kumar, Dr. Mitesh Khapra* | **Feb – Apr 2021**

- Performed experiments testing possible applications of trained sample-specific attention-head masks in BERT models.
- Developed a technique to detect adversarial inputs during test-time using their sample-specific masks using mask-inversion, layer-wise predictions, etc.
- Achieved accuracies of between 0.8055 and 0.9027 accuracy on adversarial input detection on four GLUE datasets.

Internships/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) autonomous defect detection in printed circuit boards.

Selected Course Projects

CLIP-based histology annotation predictor | *PyTorch* *Professor Jia Deng* | **Dec 2022**

- Prepared a dataset of tissue images and associated pathology notes by downsampling, segmenting and cropping 25k images from a database maintained by the Genotype Tissue Expression Consortium.
- Trained a model using the CLIP framework to rank tissue pathology note diagnoses based on images of prepared biological tissue specimen slides.
- Found via qualitative analysis that non-obvious diagnoses could be carried out using nearest-neighbor search in the embedding space shared by pathology note captions and tissue images.

automated B/W portrait colorization | *PyTorch* *Professor 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.

image2image translation | *PyTorch* — *Professor Anurag Mittal* | **Dec – Jan 2021**

- implemented, tested and benchmarked a unified framework proposed by a CVPR paper on Image to Image Translation for Domain Adaptation

σ-promoter classification | *PyTorch* — *Professor Manikandan Narayanan* | **Nov – Dec 2020**

- augmented a SOTA model for σ-promoter classification in *E. coli* by introducing attention layers and residual connections to increase accuracy by 1.6%.

device driver | *C, RISC-V* — *Professor Chester Rebeiro* | **Nov – Dec 2020**

- Wrote a UART device driver for *ZephyrRTOS* for the RISC-V *Shakti E-class Parashu* SOC.
- Performed testing on a physical SOC unit.

C compiler | *C, x86 assembly, Lex, Yacc* — *Professor Rupesh Nasre* | **Jul – Nov 2020**

- Wrote an compiler for a slightly stripped-down version of C using the tools Lex and Yacc.
- Wrote an LR(1) context free grammar for C and encoded it into Yacc, designed logic to carry out code generation and implemented 6 parse-tree level optimizations.

Personal Projects

automated attendance system | *TensorFlow, Keras* — **May – Jul 2020**

- Created an autonomous attendance system pipeline for classrooms using the popular neural networks *MTCNN* and *FaceNet*.
- Wrote a KNN-like algorithm to match faces from a PTZ camera feed to personal identities using a database containing ∼4 photographs each of students' faces.

process wallpaper | *Python, Bash* — **Aug - Sep 2019**

- Wrote a set of Python and bash scripts which periodically set the desktop wallpaper to a wordcloud of the most resource-intensive processes running.
- This project became semipopular on GitHub and was mentioned on an episode of a podcast called *Linux Unplugged*.

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; Statistical Foundations of Data Science

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

Positions of Responsibility

Assistantship in Instruction — **Fall 2022**
Teaching Assistant for COS240: Reasoning About Computation — *Princeton University*

Computer Vision and Intelligence Group — **2019**
Project Member — *Indian Institute of Technology Madras*

Developmental Operations Team, Saarang 2020 — **2019**
Coordinator — *Indian Institute of Technology Madras*

Computer Science Association — **2013 – 2014**
President — *National Public School, Rajajinagar*