Kaustubh Ponkshe

kaustubhponkshe@iitb.ac.in | Personal webpage | Github link

Education

Indian Institute of Technology Bombay
Aug. 2019 - Jul. 2024

Interdisciplinary Dual Degree: Electrical Engineering and Artificial intelligence ***9.57/10***

- **Undergraduate Research Award (URA 02)** for one of the best bachelor's thesis in the entire batch
- **Ranked third** among the entire batch in the Artificial Intelligence and Data Science Programme
- Completed a minor degree in **Computer Science & Engineering Department**
- Secured **perfect grade (AA)** in every course undertaken from the Mathematics & Statistics Department

Publications & Preprints

Kaustubh Ponkshe*, Raghav Singhal, Eduard Gorbunov, Alexey Tumanov, Samuel Horvath, Praneeth Vepakomma, *Initialization using Update Approximation is a Silver Bullet for Extremely Efficient Low-Rank Fine-tuning* ***[Pdf][Page][Code]***

Kaustubh Ponkshe, Praneeth Vepakomma. *Power Learning: Differentially private embeddings for collaborative learning with tabular data* **Oral**,Trustworthy ML Workshop @ **CIKM 2024**, *under review* @ **SaTML'25** ***[Pdf][Page][Code]***

Raghav Singhal*, **Kaustubh Ponkshe***, Praneeth Vepakomma *FedEx-LoRA: Exact Aggregation for Federated and Efficient Fine-Tuning of Foundation Models* **Poster**,FITML Workshop @ **NeuRIPS 2024**, *ARR Meta Review Score:* **5/5*****[Pdf][Page][Code]***

Parjanya Prashant*, **Kaustubh Ponkshe***, Babak Salimi. *A Lightweight Method to Disrupt Memorized Sequences in LLM****[Pdf]***

Raghav Singhal*, **Kaustubh Ponkshe***, Rohit, Vartak, Praneeth Vepakomma *Fed-SB: A Silver Bullet for Extreme Communication Efficiency and Performance in (Private) Federated LoRA Fine-Tuning*, *under review* @ **ACL Rolling Review** ***[Pdf][Code]***

Kaustubh Ponkshe, Venkatapathy Subramanian, Ganesh Ramakrishnan and Natwar Modani. *StructFormer: Document Structure-based Masked Attention and its impact on LLM Pre-Training*, **Poster**, DocUI Workshop @ **AAAI 2025** ***[Pdf][Code]***

Priya Mishra*, Suraj Racha*, **Kaustubh Ponkshe**, Adit Akarsh Ganesh Ramakrishnan. *GUIDEQ: Framework for Guided Questioning for progressive informational collection and classification*, **Poster** @ **NAACL Findings** ***[Pdf][Code]***

Parjanya Prashant, **Kaustubh Ponkshe**, Chirag Garg, Ishan Pendse, Prathamesh Muley. *Crop Yield Prediction of Indian Districts Using Deep Learning*, International Conference on Image Information Processing **ICIIP 2021** ***[Pdf][Code]***

Research Experience

Researcher | Mohamed bin Zayed University of Artificial Intelligence
Abu Dhabi

Advisor: Prof. Praneeth Vepakomma, MIT & MBZUAI

LoRA-SB: Approximating Full Fine-tuning for Extremely Efficient LoRA *Nov 2024-present*

Co-Advisors: Prof. Alexey Tumanov, Georgia Tech & Prof. Samuel Horvath, MBZUAI

- Developed a novel initialization strategy for low-rank adaptation that optimally approximates the first step of full fine-tuning, enabling parameter-efficient model updates while preserving gradient update directions throughout training
- Formulated and proved theoretical guarantees for gradient optimization that eliminates hyperparameter tuning requirements by maintaining orthonormal bases, demonstrating consistent loss reduction during training
- Achieved up to 90x parameter reduction compared to standard LoRA while matching or exceeding its performance across mathematical reasoning, commonsense reasoning and language understanding tasks on models up to 9B parameters

FedEx-LoRA: Exact Aggregation for Federated Fine-Tuning of Foundation Models *Jun 2024 - Oct 2024*

- Formulated a systematic error residual assignment mechanism for federated LoRA addressing fundamental discrepancies in federated averaging, establishing **optimal** adapter updates while maintaining computational efficiency
- Implemented a streamlined communication protocol that integrates error residuals during aggregation stages, enabling model updates with **minimal** computational and communication overhead
- Demonstrated empirical superiority on diverse NLU and NLG tasks, validating consistent performance improvements over existing federated fine-tuning methods, substantiated by rigorous analysis of aggregation methodologies

Research Intern | Camera Culture Group, MIT Media Lab *Boston,MA*

Advisor: Prof. Ramesh Raskar, MIT Media Lab

Power-Learning: Differentially Private Embeddings for Distributed Learning *May 2023 - May 2024*

Co-Advisor: Prof. Praneeth Vepakomma, MIT & Dr. Peter Kairouz, Google Research

- Established theoretical conditions on transformation in normalizing flows to make data distribution private
- Devised a novel loss function to jointly learn embeddings of the input data by jointly optimizing for differential privacy and utility function in **few-shots** such that there is private data release between server and client
- Emperically proved that using these joint learning based private embeddings when passed on to server side results in very little drop in performance while also guaranteeing $\epsilon - \delta$ privacy of the client data (**5%** over DP-SGD)

Undergraduate Researcher | IIT Bombay *Mumbai*

Advisor: Prof. Ganesh Ramakrishnan, IIT Bombay

Document Structure-based sparse attention LM | Adobe Research India *Mar. 2021 - Apr. 2022*

Co-Advisor: Natwar Modani, Adobe Research

- Explored the impact of global attentions as a surrogate for **structure awarness** on Language Model pre-training
- Created a large corpus of structure-aware text for Pre-training a longformer model by using global attention tokens on 100K arxiv documents and fine-tuned the model for information extraction on the SciREX dataset
- Emperically established that using global tokens while pre-training transformers improves performance in language tasks as against using local sparse attention, by analyzing keywords and header attention patterns (**18%** over baseline)

GUIDEQ: Guided Generation using Explainable AI and LLMs | *Master's Thesis* *Jul. 2023 - Jul 2024*

- Developed **GUIDEQ**, a novel framework leveraging LLMs and explainable AI to generate guided questions for improving text classification accuracy in domains like healthcare and legal services
- Implemented an innovative occlusion-based method to identify label-significant keywords, integrating them with top-3 classifier outputs to create targeted prompting strategies for information gathering
- Demonstrated superior performance against LLM baselines with improved F1-scores through enhanced information collection, establishing better question quality metrics for QA systems

Professional Experience

AI Research Intern | Core-AI-Engine, AWL Inc. Japan *May. 2022 - Jul. 2022*

Advisor: Eduardo Narvaez

- Explored parallelizing the AI blocks like face head detection, head direction, age and gender classification using multi-threading for improving efficiency of the **AI-Core-Engine** pipeline for various vision tasks
- Implemented three graph-based algorithms namely Level-based, Kahn parallel and Branch-based algorithm

Research Intern | TCS Research *May 2021-Jul 2021*

Advisor: Arijit Ukil, Principal Scientist, TCS Research

- Compressed a Keras model capable of detecting heart arrhythmia using single lead ECG signals
- Applied Deep Neural Network model compression techniques like pruning, quantization and weight clustering to reduce model size by a **factor of 10** while keeping the model accuracy close to its original value
- Explored Lottery Ticket Hypothesis on MNIST data to achieve a **sparsity of 99.5%** without appreciable drop in performance and compared its results against pruning for which the accuracy dropped drastically

Scholastic Achievements

- Recipient of **Undergraduate Research Award (URA 02)** for one of the best bachelor's thesis among **80+ projects**
- Selected for the Indian National Math Olympiad, precursor to International Math Olympiad (**State rank 3**)
- Secured All India Rank **846** in JEE (Advanced) and All India Rank **993** in JEE (Mains) among **1.5 million+** students
- Finalist in the PRISM 2020 competition organized by KAIST for "Cropt Yield Prediction" (**top 20** international projects)

Key Projects

Crop Yield Prediction using Ensemble CNN-RNN model | ***PRISM 2020*** *Sep 2020*

Indo-Korean Science and Technology Center

- Designed a Big Data driven Deep Learning model to predict productivity of various crops at district level
- Collected and analyzed **200,000+** rows of data including district-wise monthly data of various geological and socio-economic parameters to build Ensemble CNN-RNN model with a correlation coefficient upwards of **0.9**
- Amongst **top 20** international projects selected for applications of Big Data and Artificial Intelligence

Learning Heuristics for Combinatorial Optimization | ***Optimization for ML*** *Jan. 2022 - Apr. 2022*

Course Instructor: Prof. Ganesh Ramakrishnan, IIT Bombay

- Implemented standard heuristic-based algorithms like nearest neighbour, nearest insertion and farthest insertion algorithms to solve the Salesman Problem and simulated results in python
- Explored deep reinforcement learning methods to learn heuristics using attention based transformer based architecture and the **REINFORCE** algorithm to compare results against the other heuristic algorithms

Generating Labels using human heursitics | ***Advanced Machine Learning*** *Jan 2022 - Apr. 2022*

Prof. Sunita Sarawagi, IIT Bombay

- Trained a generative model to learn weak labels for **twitter sentiment analysis** and salary prediction
- Applied a discriminative model on weak labels to establish that model generalizes beyond the heuristic labels

Extracurricular Activities

IIT Bombay Racing

Team of 70+ students that built a driverless race car to compete in FS Germany
First Indian team *to win the* ***Engineering Design*** *event in the history of FSUK*
Co-founded the Autonomous division and headed the Path Planning and Controls subsystem

Design Engineer *Jul 2022 - May 2023*

- Designed the autonomous controller for FSSIM simulation and achieved succesful lap completion
- Ideated automatic cruise control for bikes and won the Software Intelligence competition organized by **Ather Energy**
- Implemented Longitudinal Dynamic Control in Simulink using the bicycle model incorporating a PID controller

Junior Design Engineer *Jul 2021 - May 2022*

- Implemented Lateral Dynamic control to generate the required steering command using the Pure Pursuit controller
- Applied an amalgamation of Rapidly Exploring Random Trees Algorithm along with Delaunay Triangulation for Path
- Designed a State Machine Diagram and a Development Methodology as part of SI report for FSEV 2021

Teaching Assistant | ***Calculus I (MA 109)*** *Nov. 2020 - Jan. 2021*

Course Instructor: Ravi Raghunathan, Maths Department, IIT Bombay

- Responsible for conducting tutorial sessions for a **batch of 44 students** from the Computer Science and Electrical department and helping them clear conceptual doubts through personal interaction
- Conducted online exams for evaluation and assisted in the grading of exams and assignments

Institute Sports Convener *June 2020- May 2021*

Part of a 36-member council responsible for organizing all sports events in the institute

- Organized the **Blackcat Championship** to maintain the fitness of the Institute Sports Varsity teams
- Executed the **Virtual Run** event with a prize fund of **Rs.20,000** witnessing a participation of **1500+** people.