Kaustubh Ponkshe

Efficient ML & Fine-Tuning · AI Safety & Security · Distributed Learning

ponkshekaustubh11@gmail.com | kaustubhp11.github.io

Education

2023 - 2024 **Indian Institute of Technology Bombay**
M. Tech. in Artificial Intelligence & Data Science *9.57 / 10*
Advisor: Prof. Ganesh Ramakrishnan

2019 - 2023 **Indian Institute of Technology Bombay**
B. Tech. in Electrical Engineering
Minor: Computer Science & Engineering

Research Experience

2024–2025 **Researcher, MIT & MBZUAI**, *Advisor: Prof. Praneeth Vepakomma*
2023–2024 **Research Intern, MIT Media Lab**, *Advisor: Prof. Ramesh Raskar*
2024–2025 **Research Collaborator, UC San Diego**, *with Prof. Babak Salimi*
2023–2024 **Masters Thesis Researcher, IIT Bombay**, *Advisor: Prof. Ganesh Ramakrishnan*
2021–2022 **Undergraduate Researcher, IIT Bombay** , *Advisor: Prof. Ganesh Ramakrishnan*
2022 **AI Research Intern, AWL Inc., Japan**, *Advisor: Alok Bishoyi*
2021 **Research Intern, TCS Research**, *Advisor: Arijit Ukil*

Publications

Published/Accepted Peer-reviewed Papers

[1] Raghav Singhal*, **Kaustubh Ponkshe***, Praneeth Vepakomma. FedEx-LoRA: Exact Aggregation for Federated and Efficient Fine-Tuning of Foundation Models. In *Proceedings of ACL*, 2025.

[2] **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. In *SCOPE Workshop, ICLR*, 2025.

[3] **Kaustubh Ponkshe***, Praneeth Vepakomma. Power Learning: Differentially Private Embeddings for Collaborative Learning with Tabular Data. In *Trustworthy ML Workshop, CIKM*, 2024. (Oral)

[4] **Kaustubh Ponkshe**, Venkatapathy Subramanian, Ganesh Ramakrishnan, Natwar Modani. StructFormer: Document Structure-Based Masked Attention and Its Impact on LLM Pre-Training. In *DocUI Workshop, AAAI*, 2025.

[5] Raghav Singhal*, **Kaustubh Ponkshe***, Rohit Vartak, Praneeth Vepakomma. ABBA: Highly Expressive Hadamard-Product Adaptation for Large Language Models. In *ES-FoMo Workshop, ICML*, 2025.

[6] Raghav Singhal*, **Kaustubh Ponkshe***, Rohit Vartak, Praneeth Vepakomma. Fed-SB: Extreme Communication Efficiency for Private Federated LoRA Fine-Tuning. In *ES-FoMo Workshop, ICML*, 2025.

[7] Priya Mishra*, Suraj Racha*, **Kaustubh Ponkshe**, Adit Akarsh, Ganesh Ramakrishnan. GUIDEQ: Guided Questioning for Progressive Information Collection and Classification. In *Findings of NAACL*, 2025.

[8] Parjanya Prashant, **Kaustubh Ponkshe**, Chirag Garg, Ishan Pendse, Prathamesh Muley. Crop Yield Prediction of Indian Districts Using Deep Learning. In *ICIIP*, 2021.

Preprint/Under Review

[1] **Kaustubh Ponkshe***, Shaan Shah*, Raghav Singhal*, Praneeth Vepakomma. Safety Subspaces Are Not Distinct: A Fine-Tuning Case Study. arXiv preprint, 2025.

[2] Parjanya Prashant*, **Kaustubh Ponkshe***, Babak Salimi. TokenSwap: A Lightweight Method to Disrupt Memorized Sequences in LLMs. arXiv preprint, 2025.

Selected Projects

Efficient ML and Fine-Tuning

- **LoRA-SB: Simulating Full Finetuning via Low-Rank Approximation** *[Pdf][Page][Code]*
  - Designed a low-rank initialization that bridges the gap to full finetuning with 27-$90\times$ fewer parameters.
- **ABBA: Hadamard-Product Reparameterization for Expressive PEFT** *[Pdf][Code]*
  - Reparameterized high-rank updates as Hadamard products of two learnable low-rank matrices

AI Safety and Security

- **Safety Subspaces: Geometric Limitations of Alignment Preservation** *[Pdf][Code]*
  - Showed empirically that safe and unsafe behaviours co-activate overlapping subspaces, challenging subspace-isolation assumptions in alignment.
- **Power-Learning: Differentially Private Embeddings for Distributed Learning** *[Pdf][Page][Code]*
  - Generated privacy-preserving embeddings via normalizing flows, enabling model-agnostic training with formal privacy guarantees.
- **TokenSwap: Mitigating Verbatim Memorization in Deployed LLMs** *[Pdf]*
  - Swapped token probabilities at inference, reducing verbatim leakage up to $10\times$ while preserving fluency.

Distributed Learning

- **FedEx-LoRA: Exact Aggregation for Federated LoRA Finetuning** *[Pdf][Page][Code]*
  - Proposed a provably exact aggregation scheme, achieving full baseline fidelity with negligible overhead.
- **Fed-SB: Efficient & Private Federated Finetuning via Subspace Alignment** *[Pdf][Page][Code]*
  - Averaged square LoRA-SB matrices, yielding $230\times$ communication savings under fixed noise budgets.

Robotics and Control

- **IITB Racing: Autonomous Stack for Electric Racecar**
  - Led perception, planning, and control stack for FSSIM; deployed full autonomous system at FSUK 2023.

Scholastic Achievements

- **Undergraduate Research Award (URA 02)** for one of the best bachelors theses g at IIT Bombay.
- Graduated with **3rd rank** in the M.Tech. Artificial Intelligence & Data Science programme at IIT Bombay.
- Qualified for National Mathematics Olympiad (**State rank 3**); precursor to the International Math Olympiad.
- Secured All-India Rank **846** in JEE (Advanced) and Rank **993** in JEE (Main) among **1.5 million** candidates.
- **PRISM 2020 Finalist** KAIST competition on Crop-Yield Prediction (Top 20 international projects).