

# Jatin Prakash

## Education

- 2016–2018 **Ramjas School Pusa Road**, Delhi, India  
Central Board of Secondary Education (CBSE) Class XII – CGPA: 95.8%/100
- 2018–2022 **Indian Institute of Technology Delhi (IIT Delhi)**, *Hauz Khas*, Delhi, India  
B.Tech (Bachelors) in Computer Science and Engineering – CGPA: 8.88/10  
Advisor: [Prof. Chetan Arora](#)

## Publications

*C = Conference, P = Preprint/Submitted — \* implies equal contribution, † implies core contribution*

- C.1 **A Stitch in Time Saves Nine: A Train-Time Regularizing Loss for Improved Neural Network Calibration**  
Ramya Hebbalaguppe\*, **Jatin Prakash\***, Neelabh Madan\*, Chetan Arora  
CVPR 2022 **Oral (top 4.2% papers)** [[Paper](#)] [[Code](#)]
- C.2 **A Novel Data Augmentation Technique for OOD Detection using Compounded Corruptions**  
Ramya Hebbalaguppe, Soumya Suvra Ghosal, **Jatin Prakash**, Harshad Khadilkar, Chetan Arora  
ECML-PKDD 2022 [[Paper](#)] [[Code](#)]
- C.3 **Renee: end-to-end training of extreme classification models**  
Vidit Jain, **Jatin Prakash**, Deepak Saini, Jian Jiao, Ramachandran Ramjee, Manik Varma  
MLSys 2023 [[Paper](#)] [[Code](#)]
- P.1 **Enhancing Tail Performance in Extreme Classifiers by Label Variance Reduction**  
Anirudh Buvanesh\*, Rahul Chand\*, **Jatin Prakash**<sup>†</sup>, Bhawna Paliwal, ... Manik Varma  
*Under review*
- P.2 **Understanding Calibration Transfer in Knowledge Distillation**  
Ramya Hebbalaguppe\*, Mayank Baranwal\*, **Jatin Prakash\***, Neelabh Madan, Kartik Anand, Chetan Arora  
*Under review*

## Research Experience

- July'22- Present **Microsoft Research**, *eXtreme Classification (XC) group*, Bengaluru, India  
*Research Fellow*  
Advisors: [Dr. Manik Varma](#), [Dr. Amit Sharma](#), [Dr. Ramchandran Ramjee](#)
- Enhanced extreme classifiers on under-represented or tail labels [ICLR 2024]
  - Increased training efficiency of large-scale models [MLSys 2023]
  - Deployed large-scale extreme classifiers on Microsoft's Bing search engine
  - Using pseudo-labels from LLMs to tackle missing labels/improve extreme classifiers
- November'21- January'22 **Tata Consultancy Services (TCS) Research**, Noida, India  
*Research Intern*  
Advisor: [Ramya Hebbalaguppe](#)
- Out-Of-Distribution (OOD) detection [ECML-PKDD 2022]

## Scholastic Achievements

- Oral presentation (top 4.2% papers) at CVPR 2022 for undergraduate thesis at IIT Delhi.
- Qualified for the ACM-ICPC Regionals 2021 programming competition.
- Secured **99th** percentile in JEE Advanced and JEE Mains 2018 examinations among a million contesting candidates.

## Real World Deployments

- June'23 **Extreme classifiers on Microsoft Bing for sponsored search**
- Used regularization framework [ICLR 2024] and efficient training [MLSys 2023] to obtain large-scale extreme classifiers to predict a subset of 80 million labels given a short search query
  - Employed a heuristic strategy to supervise classifier using cross-encoder as a teacher in the regularization framework [ICLR 2024], depending on the nature of label (popular or under-represented) for maximum accuracy
  - Model was deployed on Bing search engine for ad recommendation
  - Resulted in gain of 25% on offline metrics and increase of 3.6% in online metrics
  - Consequently, resulted in significant increase in clickthrough-rates (CTR) and revenue

---

## Software Engineering Experience

- April'22- July'22 **Ivy**, *Graph Compiler group*, London, UK  
*ML Research Engineer Intern*
- Significant contributions towards the graph compiler that transpiles code in one ML framework to another. [\[Website\]](#) [\[GitHub\]](#)
- May'21- July'21 **Microsoft**, *Security and Compliance team*, IDC, Hyderabad, India  
*Software Engineering Intern*
- Worked on improving the Document Fingerprinting algorithm in M365 services for sensitive document classification.
- April'20- November'20 **OpenMined**, Open Source Contributions  
Repository maintainer for SyferText, a privacy preserving NLP library. [\[Website\]](#) [\[GitHub\]](#)

---

## Teaching

- Jan-Jul'22 **Introduction to Computer Science**, *Teaching Assistant*

---

## Extra Curricular

- Competitive programming: [Expert \(1854\)](#) on [\[Codeforces\]](#), [4 stars](#) on [\[Codechef\]](#).
- Core team member of the software development club of IITD, DevClub. [\[GitHub\]](#)

---

## Selected Research Projects

- May'23- Present **Using pseudo-labels to tackle missing labels/improve extreme classifiers**  
Advisors: [Dr. Amit Sharma](#), [Dr. Manik Varma](#) — [In Progress]
- Used LLM (GPT-4) to align (distill) a small LM towards the task of generating diverse pseudo-labels for a data point that are representative of missing labels.
  - Used the associated unlabeled or uncurated data available with the data point to help smaller LM generate pseudo-labels – without relying on it's limited parametric knowledge.
  - High scalability of a smaller LM enables the proposed approach to generate pseudo-labels for hundreds of millions of data points.
  - Pseudo-labels can be used to supervise extreme classifiers for maximum efficiency in deployment.
- Dec'22- May'23 **Enhancing extreme classifiers on tail labels**  
Advisors: [Dr. Manik Varma](#) — [under review ICLR 2024]
- Used siamese-style encoder as a teacher in a knowledge distillation inspired framework to regularize over-parameterized extreme classifiers specifically on the tail labels.
  - Proposed framework enhances the tail performance of the one-vs-all linear classifiers by upto 5% absolute points without compromising their already excellent head accuracies.
  - This framework can be used in a plug-and-play manner on any XC training algorithm / architecture.
- July'22- Nov'22 **Efficient end-to-end training of large-scale XC models**  
Advisors: [Dr. Ramchandran Ramjee](#), [Dr. Manik Varma](#) — [MLSys 2023]
- Identified core bottlenecks in end-to-end training of large-scale XC models that consist of a transformer encoder followed by a massive one-vs-all linear layer (also called extreme classifiers)
  - Proposed "shortcut loss" technique that bypasses PyTorch's AutoGrad to manually calculate gradients of the massive one-vs-all linear layer.
  - Increased training efficiency by 15x, making end-to-end XC model training practical
  - Simple end-to-end training established SOTA in both efficiency and accuracy on benchmark XC datasets.
  - Using the proposed approach, XC model comprising of 8 billion parameters could be trained on a dataset containing 120 million labels and half a billion training points in under a day as compared to the baseline that takes 10 days on a Nvidia DGX-2 node.
- March'22- May'23 **Calibration through knowledge distillation**  
Advisor: [Prof. Chetan Arora](#) — [under review CVPR 2024]
- Proposed a simple framework to distill calibrated and accurate students using knowledge distillation.
  - Empirically showed that only calibrated teachers can potentially distill calibrated students. Further, showed that we can distill calibrated students from even smaller but calibrated teachers too.
  - This simple method outperforms or is on par to the contemporary and complicated methods for calibration.
  - Observed better trade-offs in calibration/OOD accuracy/refinement v.s. corruption severity of datasets.
- April'20- March'22 **Calibration of deep neural networks**  
Advisor: [Prof. Chetan Arora](#) — [CVPR 2022]
- Proposed a simple regularization loss that penalizes miscalibration at a mini-batch level to obtain calibrated deep neural networks.
  - Proposed loss is inspired from a multi-class calibration metric. Deep nets trained using the proposed loss are observed to have both top-class and multi-class calibration.
  - Seamlessly integrates into other classification tasks in NLP domain, as well as image segmentation.