BHAWNA PALIWAL

Research Engineer, Microsoft Research

@ bhawna7374@gmail.com — bhawnapaliwal.github.io — paliwal-bhawna — Google Scholar

EDUCATION

Indian Institute of Technology (IIT), Ropar *2017 - 2021*
Bachelor of Technology in Computer Science
Awarded with **Director's Gold Medal**
Grade (CGPA): 9.4/10

PUBLICATIONS

CROSS-JEM: Accurate and Efficient Cross-encoders for Ranking
Bhawna Paliwal*, Deepak Saini*, Mudit Dhawan, Siddarth Asokan, Nagarajan Natarajan, Surbhi Aggarwal, Pankaj Malhotra, Jian Jiao, Manik Varma
ENLSP@Neural Information and Processing Systems (NeurIPS), 2024
To be submitted at ICML 2025 [PDF]

OAK: Enriching Document Representations using Auxiliary Knowledge for Extreme Classification
Shikhar Mohan*, Deepak Saini*, Anshul Mittal, Sayak Ray Chowdhury, **Bhawna Paliwal**, Jian Jiao, Manish Gupta, Manik Varma
International Conference on Machine Learning (ICML), 2024 [PDF]

Enhancing Tail Performance in Extreme Classifiers by Label Variance Reduction
Anirudh Buvanesh*, Rahul Chand*, Jatin Prakash, **Bhawna Paliwal**, Mudit Dhawan, Neelabh Madan, Deepesh Hada, Vidit Jain, Sonu Mehta, Yashoteja Prabhu, Manish Gupta, Ramachandran Ramjee, Manik Varma
International Conference on Learning Representations (ICLR), 2024 [PDF]

Extreme Meta-Classification for Large-Scale Zero-Shot Retrieval
Sachin Yadav*, Deepak Saini*, Anirudh Buvanesh*, **Bhawna Paliwal**, Kunal Dahiya, Jian Jiao, Manik Varma
Knowledge Discovery and Data Mining (KDD), 2024 [PDF]

NGAME: Negative Mining-aware Mini-batching for Extreme Classification
Kunal Dahiya et al. including **Bhawna Paliwal**
International Conference on Web Search and Data Mining (WSDM), 2023 [PDF]

This explains That: Congruent Image–Report Generation for Explainable Medical Image Analysis with Cyclic Generative Adversarial Networks
Bhawna Paliwal*, Abhineet Pandey*, Abhinav Dhall, Ramanathan Subramanian, Dwarikanath Mahapatra
IMIMIC Workshop, MICCAI, 2021 [PDF]

EXPERIENCE

Microsoft Research Sep 2021 - Present
Research Engineer *Bangalore*

- *Advisor: Dr. Manik Varma*
- Working on efficient language modeling architectures and training algorithm for large-scale systems in search and recommendation.
- My work has resulted in publications at ICLR, ICML, and KDD in addition to major tech transfers to Microsoft's production systems.

Monash University Aug 2020 - Jan 2021
Research Intern *Remote*

· *Advisor: Dr. Abhinav Dhall*
· Developed an interpretable deep learning architecture for generating medical diagnosis reports, resulting in a publication at the IMIMIC workshop at MICCAI.
· Led the design and implementation of a multi-modal deepfake detection model. The developed approach incorporating physiological signals alongside audio and video modalities from counterfeit videos.

Microsoft May 2020 - July 2020
Software Engineering Intern *Remote*

· Developed a task aware chatbot and channel for android applications, leveraging efficient intent classification models for seamless user interaction

Indian Institute of Science (IISc) May 2019 - July 2019
Research Intern *Bangalore*

· *Advisor: Dr. Vijay Natarajan*
· Worked on matrix visualization tools for visualizing large-scale topological data structures.

SELECTED RESEARCH PROJECTS

Dataset Selection for Fine-tuning LLMs 2024
Advisors: Dr. Sainyam Galhotra

- Motivated by the concept that the optimal training data should align with a model's current state, I worked on a data selection algorithm based on perplexity scores.
- My findings show that training on high perplexity data points contributes minimally to task performance enhancement, and identification of high yield data points to train on can be approximated by much smaller language models.
- This approach showed that we could save around 66% on total training time and achieve accuracy within 1% range of full training. This work is currently under preparation to be submitted at ACL'25.

Efficient Ranking for Large-scale Recommendation Systems 2024
Advisors: Dr. Nagarajan Natarajan, Dr. Manik Varma

- In large-scale recommendation systems, the ranking task involves reordering retrieved items based on relevance scores derived from a computationally intensive transformer-based language model.
- By showcasing the token position-invariance of the standard transformer architecture in handling short text inputs, I introduced a joint architecture capable of scoring multiple items per query in a single encoder pass.
- This approach achieved a 4× speed improvement compared to vanilla cross-encoders, while being upto 2% more accurate. This work has been accepted at ENLSP@NeurIPS'24 and is to be submitted at ICML'25.

Real-Time Query Completion using Extreme Classification 2023
Advisors: Dr. Yashoteja Prabhu, Dr. Manish Gupta, Dr. Manik Varma

- Query Completion is a real-time service that provides suggestions for incomplete queries entered by users on search engines.
- Developed a new approach where Query Completion is reformulated as a classification task with a large label set consisting of all possible completions of a query. Different from conventional autoregressive methods, the devised architecture enables the generation of a set of completions in a single pass of the language model.

- Outperformed autoregressive models in both accuracy and latency, providing suggestions four times more accurately with a two-fold reduction in CPU latency. A/B tests on Bing demonstrated a 7% decrease in user typing effort. This work was included in our ICLR'24 paper.

Physiological Signals for Improving Deepfake Detection 2022

Advisor: Dr. Abhinav Dhall

- The proliferation of realistic fake videos created by deep learning models poses a significant risk for the dissemination of harmful misinformation.
- In contrast to traditional methods that rely on audio and visual artifacts for video classification, my work demonstrated substantial improvement in deepfake detection accuracy by incorporating physiological signals in addition to audio and visual modalities.
- Devised graph convolutional network (GCN) based approach for incorporating physiological signals to the standard classification architecture. This work is out on arxiv.

Explainable Medical Report Generation 2021

Advisors: Dr. Abhinav Dhall, Dr. Ramanathan Subramanian

- Addressed the limitation of deep learning-based black-box models in medical diagnosis by providing prototypical explanations alongside model predictions.
- Recognized the common reliance on post-hoc modules for explanation generation, which often results in explanations inconsistent with model decisions.
- Proposed and implemented a cycle-GAN based architecture to produce coherent image-report pairs for chest X-Ray image diagnosis, ensuring that the generated report effectively clarifies the image, and vice versa. This research was published at IMIMIC, MICCAI, 2021.

HONOURS AND AWARDS

- Awarded patent for *CROSS-JEM: Accurate and Efficient Cross-encoders for Ranking* and *Extreme Meta-Classification for Large-Scale Zero-Shot Retrieval*
- Class of 2021 Director's Gold Medal recipient for best all-round performance in academics, research, and leadership (one awardee out of 250 students)
- Global Rank 2 in CVPR 2021 workshop challenge on eye gaze detection
- Awarded with Merit scholarship during all 8 semesters at IIT Ropar
- Academic Rank 2 among 250 students at IIT Ropar Class of 2021
- Winner of Smart India Hackathon 2020 organized by MHRD, Govt. of India
- Top 0.1% in Joint Engineering Entrance (JEE) 2017 Exam among 1.3M candidates.

SKILLS

Programming Languages: Python, C, C++, Java, SQL
Libraries and Frameworks: PyTorch, Numpy, HuggingFace, OpenCV, Weights and Biases

PROFESSIONAL OUTREACH

Presented our work on **Extreme Memory Based Knowledge Models** to Bill Gates and Microsoft Research India Advisory Board 2024.

Volunteer: WiCV (CVPR'21), Microsoft Booth at NeurIPS'22, COLT'23

Reviewer: KDD'24, NeurIPS'24

Mentor: Coding Club and Debating Society at IIT Ropar