Kumar Tanmay

Graduate Research Assistant (Harvard DtAK Lab)
Harvard University, Cambridge, MA

kr.tanmay147@gmail.com | kumartanmay@fas.harvard.edu
in | | Google Scholar | Website
+1-617-955-9468

Education

Harvard University, MA, USA 2024 - Present
Masters in Computer Science & Statistics (Data Science)

Indian Institute of Technology, Kharagpur, India 2018 - 2022
Bachelor of Technology (Hons.) in Electrical Eng. & Minor in Computer Science & Eng.
CGPA: 8.97/10.0

Publications (Google Scholar)

* = equal contribution

Sanchit Ahuja*, **Kumar Tanmay***, Hardik Hansrajbhai Chauhan, Barun Patra, Kriti Aggarwal, Luciano Del Corro, Arindam Mitra, Tejas Indulal Dhamecha, Ahmed Awadallah, Monojit Choudhary, Vishrav Chaudhary, Sunayana Sitaram "sPhinX: Sample Efficient Multilingual Instruction Fine-Tuning Through N-shot Guided Prompting". *NAACL*. 2025. (Under Review, pdf)

Utkarsh Agarwal*, **Kumar Tanmay***, Aditi Khandelwal*, Monojit Choudhury. "Ethical Reasoning and Moral Value Alignment of LLMs Depend on the Language we Prompt them in". *LREC-COLING*. 2024. (Accepted, pdf)

Aditi Khandelwal*, Utkarsh Agarwal*, **Kumar Tanmay***, Monojit Choudhury. "Do Moral Judgment and Reasoning Capability of LLMs Change with Language? A Study using the Multilingual Defining Issues Test". *EACL*. 2024. (Published, Oral, pdf)

Kumar Tanmay*, Aditi Khandelwal*, Utkarsh Agarwal*, Monojit Choudhury. "Probing the Moral Development of Large Language Models through Defining Issues Test". *Workshop on AI meets Moral Philosophy and Moral Psychology (MP^2) - Neurips*. 2023. (Published, Workshop, pdf)

Abhinav Rao*, Aditi Khandelwal*, **Kumar Tanmay***, Utkarsh Agarwal*, Monojit Choudhury. "Ethical Reasoning over Moral Alignment: A Case and Framework for In-Context Ethical Policies in LLMs". *Findings of EMNLP*. 2023. (Published, Poster, pdf)

Kriti Aggarwal*, Aditi Khandelwal*, **Kumar Tanmay***, Owais Mohammed Khan, Qiang Liu, Monojit Choudhury, Hardik Hansrajbhai Chauhan, Subhojit Som, Vishrav Chaudhary, Saurabh Tiwary. "DUBLIN: Visual Document Understanding By Language-Image Network". *EMNLP Industry Track*. 2023.(Published, Poster, pdf)

Kumar Tanmay*, Kumar Ayush*. "Augmented Reality Based Recommendations based on Perceptual Shape Style Compatibility with Objects in the Viewpoint and Color Compatibility with the Background". *Advances in Image Manipulation Workshop (AIM) - ICCV*. 2019. (Published, Workshop, pdf)

Kumar Ayush*, Burak Uzkent*, **Kumar Tanmay**, Marshall Burke, David Lobell, Stefano Ermon. "Efficient Poverty Mapping from High Resolution Remote Sensing Images". *AAAI*. 2021. (Published, Oral, pdf)

Shuvam Chakraborty, Kumar Ayush*, Burak Uzkent*, **Kumar Tanmay**, Evan Sheehan, Stefano Ermon. "Efficient Conditional Pre-training for Transfer Learning". *Workshop on Learning with Limited Labelled Data for Image and Video Understanding (L3D-IVU) - CVPR*. 2022. (Published, Workshop, pdf)

Kumar Ayush*, Burak Uzkent*, Chenlin Meng*, **Kumar Tanmay**, Marshall Burke, David Lobell, Stefano Ermon. "Geography-Aware Self-Supervised Learning". *ICCV*. 2021. (Published, Poster, Webpage, pdf,)

Work Experience

Research Engineer, Turing Team, Microsoft Research and Development, India
Visual Document Understanding, Image QA, Table QA Jul 2022 - Aug 2023
Advisors: Dr. Subhojit Som, Vishrav Chaudhary, Prof. Monojit Choudhury, and Dr. Saurabh Tiwary

- Co-led the development of **DUBLIN**, a state-of-the-art large-scale transformer-based model for visual document understanding, achieving superior performance across tasks like QA, information extraction, summarization, and classification. *Accepted in EMNLP Industry Track 2023*.
- Introduced a novel method for handling variable input resolution images, improving performance on long documents by 7.4% (on average) in benchmarks like InfographicsVQA and DocVQA.
- Enhanced table comprehension with 80K synthetic data and curriculum learning, achieving an average 8% performance gain across 20 multimodal benchmarks, including +21% in AI2D, +7.5% in InfographicsVQA, and

+5.6% in DocVQA.
- Contributed to the deployment of advanced Table QA and Image QA models in Microsoft Bing, enhancing user interaction and data retrieval capabilities.

Ethical Alignment and Value Pluralism in Large Language Models (LLMs) Dec 2022 - Oct 2023
Advisor: Prof. Monojit Choudhury

- Developed a psychometric assessment tool inspired by the Kohlbergian Moral Development and Defining Issues Test frameworks to evaluate the moral judgment and reasoning capabilities of LLMs, including GPT-4, LLaMA, PaLM, and GPT3.x series. *Accepted in MP² Neurips 2023.*
- Established a comprehensive framework to facilitate the infusion of ethical policies for moral alignment in LLMs by leveraging in-context learning to address complex social dilemmas characterized by conflicting values. *Accepted in Findings of EMNLP 2023.*
- Conducted an in-depth examination of the "Foreign Language Effect" in LLMs, with a specific focus on understanding the moral reasoning abilities of these models within multilingual contexts. *One paper accepted in EACL 2024 and another in LREC-COLING 2024.*

Multilingual Instruction Finetuning Sep 2023 - Jul 2024
Advisors: Vishrav Chaudhary, Dr. Sunayana Sitaraman and Prof. Monojit Choudhury

- Co-developed **sPhinX**, a 6.7B parameter LLM showcasing advanced reasoning across 51 languages. *Work submitted to NAACL 2025.*
- Established a 1.8M multilingual instruction dataset using GPT-4, significantly enhancing multilingual performance across models in Microsoft Copilot and M365.
- Devised a novel and efficient instruction tuning technique achieving a 10% performance improvement over the state of the art, without even pre-training in 51 languages.

Microsoft Bing Copilot for Telegram Sep 2023 - Jul 2024
Advisors: Dr. Tejas Dhamecha

- Developed a deployment pipeline for integrating Bing Chat with Telegram using C# and Azure Bot Framework, enabling multilingual suggestion chips (across 10 languages) and interactive emoji/GIF creation for enhanced user engagement.
- Achieved **100K+** daily active users (DAU) post-deployment, with sustained growth in user engagement over several weeks.

Internship/Academic Research Experience

Data to Actionable Knowledge Lab (DtAK), Harvard University, MA Fall 2024
Intrepretable Machine Learning and Safe AI for Identifying and Mitigating Biases in LLMs
Advisors: Prof. Finale Doshi-Velez and Prof. Weiwei Pan

- Working on mechanistic interpretibilty using TransformerLens for identifying where biases are located in transformer based models and devising ways to mitigate them through model editing and activation steering.
- Researching task-specific representation geometries in high-dimensional space to facilitate controlled shifts in language model behavior between tasks through representation steering and generalized fine-tuning.

qure.ai, Mumbai
Classification and Localization of Tubes and Catheters in Chest X-Ray Images Summer 2021
Advisors: Tarun Raj and Dr. Pooja Rao [qXR-BT]

- Developed an end-to-end pipeline for classification and segmentation of medical entities (tubes & catheters) in Chest X-Rays, addressing class imbalance through a weighted-binary cross-entropy loss function.
- Developed a two-step strategy for precise tube tip localization using deep learning-based segmentation and advanced image processing techniques.
- Conducted extensive experiments with various CNN architectures, surpassing baseline methods and contributing to the successful integration of models into qure.ai's deep learning stack for the qXR-BT product.

Sustainability and AI Lab, Stanford University, CA
Machine Learning for Socioeconomic, Sustainability and Computer Vision Tasks Fall 2020
Advisor: Prof. Marshall Burke and Prof. Stefano Ermon *Remote Collaborator*

- **Geography-Aware Self-Supervised Learning (ICCV 2021)**: Helped develop and validate a novel contrastive learning method for remote sensing data, leveraging spatio-temporal structures and geo-location showing improvements in image classification, object detection, and semantic segmentation.
- **Efficient Poverty Mapping from High Resolution Remote Sensing Images (AAAI 2021)**: Helped

develop and validate a novel reinforcement learning method to optimize acquisition of high-resolution satellite images based on analysis of free low-resolution imagery, enhancing poverty prediction in Uganda.

- **Efficient Conditional Pre-training for Transfer Learning (CVPRW 2022)**: Helped develop and validate efficient and adaptable methods for pre-training dataset filtering to reduce high training costs, emphasizing performance, adaptability, and flexibility.

Graduate and Undergraduate Research Projects

Harvard Academic Atlas Fall 2024
Advisor: Prof. Pavlos Protopapas

Developing a course planner leveraging the Harvard Course Database to recommend courses for students based on their short- and long-term goals. Implemented a Retrieval Augmented Generation (RAG) System using ChromaDB's vector database and fine-tuned the LLama 3.1 8B model to deliver responses in a conversational tone. This system simplifies course selection for Harvard students, helping them align their academic schedules with career objectives and efficiently plan their weekly schedules.

Medical Tube Abnormality Detection in Chest-X Rays using Deep Learning Sep 2021 - Mar 2022
Advisor: Prof. Pabitra Mitra [Bachelor's Thesis]

Developed DL algorithms for abnormal tube position detection in Chest X-Rays, achieving a 24% improvement via knowledge distillation. Created a proprietary dataset, incorporating segmentation masks for anatomical regions and ideal tip positions, leading to a 35% avg. improvement in abnormalities detection across tube types. Developed CNN-based segmentation models and applied image processing techniques for optimal detection.

Typologically Diverse QA - Zero-Shot and Few Shot Language Jackknifing Fall 2021
Self-Project [PDF]

Investigated cross-lingual generalization of multilingual BERT (mBERT) using TYDI QA dataset, exploring its performance on QA tasks across various languages. Identified linguistic feature impacts on zero-shot transfer, finding that fine-tuning mBERT as a language model on QA questions significantly improved zero-shot cross-lingual understanding, addressing question interpretation challenges.

Augmented Reality Based Context Aware Recommendations Summer 2019
Self-Project [PDF]

Developed a novel consumer targeting system by modeling AR-based data. Created persuasive recommendations by using statistical modeling, an exemplar part-based 2D-3D alignment method to find the best matching 3D models of furniture present in the user's preferred purchase viewpoint in the AR app and a combination of 3D style compatibility and color compatibility algorithms. *Work published in ICCV Workshop 2019.*

Talks/Presentations

DUBLIN: Visual Document Understanding By Language-Image Network [Slides, Video] Dec 2023
EMNLP 2023

A Case and Framework for In-Context Ethical Policies in LLMs [Slides, Video] Dec 2023
EMNLP 2023

Technical Skills

Languages: Python, C++
Libraries: PyTorch, PyTorch Lightning, Tensorflow, Keras, NumPy, Scikit-learn, Selenium, Pandas, OpenCV, Beautiful Soup, Scipy, NLTK, Flask, Tesseract, EasyOCR

Awards and Achievements

OpenAnalytics Intra-collegiate Tech Competition (Captain) - Second Prize - IIT Kharagpur 2022
Led a team, developing a neural network based model to predict popularity of music tracks based on the features associated with the music. Competition featured 50 teams.

Kishore Vaigyanik Protsahan Yojana (KVPY) Scholarship 2017
Awarded to 1323 students from around 100,000 applicants by Dept. of Science and Technology, Govt. of India for exceptional aptitude in basic sciences.

Abhay Seva Sansthan Gold Medal (Class X & Class XII) 2016-2018
Awarded for scoring 100% in Computer Science in ICSE (10th grade) and ISC (12th grade) board examination.

C. D. Sinha Award 2018
Awarded for outstanding performance (95.6%) in ISC (12th grade) board examination.

S. P. Sinha Scholarship 2016
Awarded full scholarship for 11th & 12th grade due to outstanding performance (96.5%) in ICSE (10th grade) board examination.

Relevant Coursework + Online MOOCs

Programming & Data Structures (+Lab)	Signals & Networks (+Lab)	Data Analytics
Mathematics I & II	Advanced Practical Data Science (MLOps)	Digital Signal Processing
Algorithms I (+Lab)	Probability and Stochastic Processes	Social Computing
Marketing & Market Research	Big Data Processing	Artificial Intelligence
NLP with Probabilistic Models	Principles Of Programming Languages	Linear Models
Machine Learning	Convolutional Neural Networks	Sequence Models

Extracurricular Activities

Reviewer for NAACL 2025, EMNLP 2024, NAACL 2024, TACL 2024, ACL 2023, AACL-IJCNLP 2023.

Teaching Assistant (2023) for a workshop course in Practical NLP and Large Language Models (LLMs) for undergraduate students at the Indian Institute of Science (IISc) as part of Kotak-IISc AI-ML Centre.

Volunteer Teacher at eVidyaloka (2023): Taught Maths and Science to underprivileged children in remote areas of Jharkhand via Skype.

Technovation Mentor for Youth Coding Initiative (2022): Mentored 5 high-school girls to build a business plan and mobile app for autistic children that we piloted with 10 families.

Captain (2022) of the OpenAnalytics (intra-collegiate tech competition) team of my hall of residence, IIT Kharagpur.