

Amogh Inamdar

Data analysis underpins intelligent decision making, and data representation begets analysis.

Representation learning shows promising results in distilling information from big data, but its top performing algorithms are hard to understand and deploy as resilient systems. Studying these challenges was central to my experience as a Master's student at Columbia University, where I discovered the joys of diving deep into machine learning through academic research. Through a career in research via the Ph.D. program at Columbia, I aim to engineer algorithms that represent data better, leverage them in end-to-end applications, and work towards provable AI explainability.

Explainability has long been a key aspect of my research - in my very first graduate research project (supervised by Dr. Ansaf Salieb-Aouissi), I worked with doctors from New York-Presbyterian to create a dataset of spinal deformity patients from HIPAA-protected demographic and medical information. Supervised by Dr. Ansaf Salieb-Aouissi, I used a battery of random forest algorithms with gradient boosting to classify post-operative Proximal Junctional Kyphosis (PJK), a complication that often necessitates re-surgery in patients. By aggregating decision trees and validating calculated risk factors for PJK, I presented potentially actionable insights to medical experts at a regular cadence. Through this, I learned that explainability in AI is non-negotiable in critical fields like medicine, and to design systems that balance explainability with data privacy. Working in medicine was also extremely rewarding.

Habituated to task-specific deep learning, I was fascinated by self-supervised learning when I studied the SimCLR algorithm in a class taught by Dr. Richard Zemel. To explore further, I began working with Dr. Zemel on contrastive attribute learning. I adapted the SimCLR algorithm to learn vectors composable from subspaces via multi-head instance discrimination, to constrain variation in image data to perceivable 'attributes' like shape and color. Experimenting on the dSprites and CelebA datasets showed that weakly supervised models outperformed self-supervised counterparts on attribute-level tasks, corroborating theoretical findings from Locatello et. al. 2019. Gratifyingly, visualizing data in representation space with t-SNE also qualitatively linked disentanglement and performance. Based on these results, I shifted focus to studying whether large text-image models such as CLIP and Flamingo can leverage language association for zero-shot learning through experiments on attributes datasets like MIT-states. This project

Amogh Inamdar

sparked my interest in disentangled representations, few-shot learning, and domain adaptation, and I see these as potential focal points of my Ph.D. research. I continue to contribute to it as a working professional and enjoy it greatly.

I also have a strong record of engineering praxis. Guided by Dr. Nakul Verma and Dr. Salleb-Aouissi, I built `logiclearner.ctl.columbia.edu`, a propositional logic practice app, for the Center for Teaching and Learning at Columbia. I developed a system to parse and generate a search frontier for Boolean expressions, optimized heuristics with a genetic algorithm, and adapted A\* search to solve logic proofs. Through my research, LogicLearner provides hints on-the-fly from wherever a student may be stuck, which cannot be achieved by precomputing solutions. Recognizing a unique opportunity for self-supervision, I developed a large question bank by backtracking from solution states and used a Siamese GRU to embed expressions in vector space, preserving transition similarity according to logic rules. Used as a heuristic, this network solved many human-generated questions in under a second. Logiclearner has been used in Dr. Salleb-Aouissi's discrete math class at Columbia, and has motivated me to further study representation learning for non-perceptual tasks. Currently, as a software engineer at Salesforce, I build and manage highly optimized systems that control millions of dollars of cloud infrastructure for the Genie customer data platform.

Having worked closely with Dr. Richard Zemel at Columbia, I am keen on continuing my exploration of representation learning and generalizability as a Ph.D. student in his group. Working full-time as a software engineer on a data-driven stack has only reinforced my belief in the need for agile machine learning models. With Dr. Zemel, I hope to explore the frontiers of such models and build robust, explainable models that generalize from small amounts of data. As a returning student, I would be able to make meaningful contributions from day one. Dr. Carl Vondrick's excellent computer vision course served as my introduction to representation learning, and I was fortunate to take Dr. Zemel's course alongside his accomplished students. Being especially interested in computer vision, I hope to work on research under his guidance. I am also excited by the breadth of interdisciplinary research at Columbia. While I wasn't able to work with Dr. Liam Paninski during my MS due to visa work limits, I was privileged to have

Amogh Inamdar

interacted with him on research in pose estimation. My experience with medical and cognitive science make me well-suited to tackle the fascinating research in computational neuroscience that Dr. Paninski's group focuses on.

Overall, Columbia University presents an ideal environment for my growth as a researcher, with an incredible peer group and faculty with diverse research interests. Through my Master's degree, I have demonstrated excellence in intense graduate coursework and both core and interdisciplinary research at a high level. My data engineering profession has also made me a strong coder and system designer. My time at Columbia was immensely rewarding but all too brief - as a Ph.D. student, I would mature as an academic and gain a breadth of perspective. After my degree, I plan on continuing research in machine learning and its applications as a research scientist. In the long term, I hope to have a career in academia, combining my love for research and pedagogy as a university professor.