Statement of Purpose

Shreyas Kalvankar *Master of Information Management and Systems applicant*

My research interests lie in uncovering the mathematics that drives Machine Learning and deep learning and their theoretical nature. What are the theoretical foundations explaining the success of deep networks in approximating complex functions, and are there inherent limits to their expressiveness? What insights can we gain into the dynamics of training deep networks? What are the potential applications of learned representations in various domains, such as computer vision, natural language processing, and speech recognition? How do we address the challenges of interpretability and explainability in learned representations? I aim to explore these problems and develop an understanding of the fundamental theoretical limits of Machine Learning. With this goal in mind, I want to pursue my Master's in Information Systems Management at UC Berkeley. I am particularly interested in effectively applying these algorithms to build intelligent systems to conduct research in physics and mathematics to solve problems such as approximating numerical computations using neural networks, exoplanet discovery, assisted theorem proving, etc.

My undergraduate research endeavors have been in applying deep learning techniques to astronomy and astrophysics, a domain that is undergoing significant transformations brought about by exponential growth in our abilities to collect and process data. I attempted to solve this problem using deep learning algorithms and, along with my colleagues, worked on a project to apply CNNs for Galaxy Morphology Classification, which subsequently turned into my first authored research paper. Galaxy morphologies are studied by astronomers to study dark matter distribution, galaxy evolution, and galaxy interactions. The problem involved correctly classifying galaxy images into various morphologies and also predicting the presence of certain features like rings, spirals, etc. We sought to create a 7-class system encompassing most galaxy morphologies. I set up the entire training pipeline and the model architectures, enabling robust training of 7 EfficientNet models. Our results were better than those at the second position on the Kaggle Public Leaderboard, effectively making it the second-best submission. This project made me realize the impact intelligent systems can have in streamlining scientific research.

I continued exploring more applications of such systems to astronomy and started working with Dr. Snehal Kamalapur in the Intelligent Systems Lab. I looked into developing an efficient process to colorize and up-scale unprocessed astronomical images that lie dormant and unseen in extensive space archives. I was fortunate to present this study at the *Informatik 2022* conference's Astro ML workshop in Hamburg. My role as the project lead was a significant learning experience that taught me various aspects of management, communication, and project planning. Through this project, I honed my research skills, nurtured my critical thinking abilities, and sharpened my decision-making capabilities.

My interest in this field expanded my research thinking beyond the confines of ML. I started contributing to an open-source project called EinsteinPy. EinsteinPy is an open-source pure Python package that studies problems arising in General Relativity and gravitational physics. My work specifically involved incorporating various symbolic computations, such as the Reissner–Nordström metric and calculations for the event horizon and ergosphere of a Kerr-Newmann black hole. Working on problems motivated by the challenges in science and engineering requires understanding other fields like numerical methods, computational geometry, etc. I am particularly interested in pursuing this area of research at the Berkeley AI Research Lab. Prof. Dr. Krishnapriyan's interests in developing physics-inspired machine learning methods seem to align perfectly with mine, and I would like to explore more such research at the BAIR Lab.

After completing my degree, I embarked on my professional journey, starting as a Machine Learning Engineer and later transitioning to a Machine Learning Consultant role at the startup Relfor Labs. My primary focus revolved around applying deep learning to audio data analysis and classification by devising novel architectures. I worked on designing a system for pre-processing

large amounts of raw audio data by converting it into mel spectrograms, refining the labels. I set up an end-to-end training pipeline to streamline experimentation. Throughout my work on this and various other ML and data-intensive projects, I frequently encountered the challenge of managing and comprehending extensive datasets. I have experienced firsthand the complexities of automating an end-to-end ML pipeline which can be sustainable in the long term, including data quality and consistency issues, model drift, version control, and debugging. My experience in developing ML models and designing ML pipelines, along with my challenges, has motivated me to tackle these issues by creating tools that streamline data management and ML processes. I have followed the work of Prof. Dr. Parameswaran and his student Ms. Shankar and I want to contribute to their efforts of simplifying data science and building tools to manage ML pipelines effectively.

I've been a full-time Software Developer at Dalton Maag, a type-design studio in London, for the past two years. Shortly after joining, I became involved in a project to design a proof-of-concept system that utilizes genetic algorithms to generate CJK glyphs automatically. Because of the sheer volume of these glyphs; often ranging in thousands, designing and drawing them is an expensive, time-consuming task. I was responsible for creating a framework that could learn from a few hundred glyphs and draw the remaining ones in the same style. This experience highlighted the incredible potential of harnessing data to create intelligent systems across various fields, profoundly influencing my interests.

My exposure to the industry has ignited a strong passion for crafting comprehensive systems tailored to meet the specific needs of businesses, particularly those reliant on effective data and information utilization. One such system that I worked on, Pricebot, stands out as a prime example. I was responsible for developing and integrating the pricing model which is used to create project quotes and plans by understanding the business needs. This tool completely automates a previously manual process and consistently delivers precise results which effectively prevents project overruns and unforeseen expenses, cutting down planning time significantly and saving valuable resources.

My goal is to leverage the mathematical foundations of machine learning to apply AI to scientific applications, developing tools that streamline data management and systems for automating business processes through information systems. The MIMS program at the School of Information is a unique blend of data science, social good and technology. The interdisciplinary nature of the program makes it possible to marry multiple domains together and use information to build technology that has real impact. Conversations with an I-School alumnus, Rajvardhan Oak, have convinced me that the MIMS program would prepare me well for research careers in the industry, or further doctoral studies. My determination to pursue a research career is unwavering, and I aspire to work in an academic setting and eventually pursue a Ph.D. Thank you for considering me as a prospective student at your university.