

## Statement of Purpose

Kaustubh Sapru

## Statement of Purpose – PhD Admissions, Spring 2024

Throughout my formative years I've harbored a singular focus - impacting the lives of millions. I have nurtured a deep curiosity about how things work, taking an interest in a range of disciplines from the restaurant business to healthcare. An underlying thread across these interests is the shared applicability of problem definition, articulation, and problem solving. Focusing on the metrics surrounding seemingly disparate areas such as customer feedback to utilizing patient data to discern trends for blood cancer treatment led me into the realm of machine learning. Having applied machine learning techniques in different environments I began developing solutions to complex problems on a larger scale. My interests lie at the intersection of statistical decision making, generative AI, computer vision, and predictive modeling. I believe a multi-faceted approach involving advances in algorithms, probabilistic inference, and human-robot interactions has enormous potential to create large scale change and impact. I am motivated to address medical negligence, an issue that has been an underlying cause of personal loss and grief. Working toward the development of an equitable, accessible, and automated early disease diagnosis system is my best shot at achieving the goal of impacting millions of lives.

My journey into AI began rather unconventionally in 2016 while running a takeout business of authentic Kashmiri cuisine. Initially meant to support mine and my brother's education, it swiftly turned into a means of representing our lost cultural heritage. In 2019, I expanded operations by utilizing data analytics coupled with machine learning to target specific demographics and identify return customers. Simultaneously, I reached out to AI engineers from universities across India and the UK, pitching ideas for an AI powered early cancer detection system, which eventually led to a partnership with a local cancer treatment startup for the underprivileged. The startup maintained a database that housed HLA data used to match patients with potential donors for bone marrow transplants. To expand the database, I worked with a team of doctors and engineers to organize stem cell donation camps across the city where I educated people on the concepts and importance of transplants. I applied my newly acquired AI skills to identify probabilities of optimal patient-donor matches.

To further my understanding of AI's broader implications, I joined the Neural Engineering Lab as a graduate data scientist in June 2021 at the City College Center for Discovery and Innovation (CDI). At the Neural Engineering Lab, I found the AI community's strength lay in its openness to diverse ideas and collaborations. A pivotal relationship was fostered through mentorship from Prof. Lucas Parra, marked by daily brainstorming sessions exploring innovative concepts and strengthening my problem solving ability. I was able to enhance our team's analytical capabilities by developing a computer vision software analyzing human neural responses to facial stimuli. This addressed the larger goal of quantifying attention and engagement in humans. Widely adopted within the Parra Lab, this software supported independent research, overcoming challenges like class imbalance, improving network architecture, heightened accuracy despite scarce and uneven sampled training data, and identifying features that most accurately predict neural response.

## Statement of Purpose

Kaustubh Sapru

Since my foray into AI at City College, I have worked with leaders of financial trading companies and developers of AI powered applications. Here I had the opportunity to build upon my foundational work at CDI and I scaled my solutions. I have been inspired by the work done by Prof. Regina Barzilay at MIT, who is transforming automated early cancer detection. Her work, published in 2022 (EquiDock), raised important questions about protein flexibility for drug discovery. My most recent endeavor with Prof. Bruce Johnson, at the Advanced Science Research Center, New York, aimed at developing models that predict protein flexibility profiles using generative AI. I was able to integrate these models to NMRFX Analyst (NMR data processing software) after overcoming challenges such as saturating model accuracies, slow training times, and adding new variables to the training data.

My academic training and intensive research experience has equipped me with skills such as data analysis, statistical inference, hypothesis testing, development of generative AI models, improving prediction accuracies, and data visualization. I am interested in building upon the important work of leaders in the field, including Prof. Vondrick, Prof. Zemel, Prof. Smith, Prof. Azizi, Prof. Elhadad, and Prof. Alquraishi at the Computer Science department, by devising AI models that utilize images, text, and structured data to identify trends that affect early diagnosis, treatment, and disease prevention. Computer Science at Columbia offers a unique setting where teaming up with individuals from diverse backgrounds pushes the boundaries of what is achievable. Capitalizing on industry partnerships and the interdisciplinary atmosphere, coupled with my entrepreneurial spirit, I am eager to harness my strong work ethic in pursuit of my mission: addressing the deficiencies in our current societal landscape. I bring to the table a strong sense of organization, self-motivation, and resourcefulness, along with a creative mindset that thrives in team settings.

Columbia's commitment to nurturing historically underrepresented communities fosters an environment conducive to open expression and innovation. Having witnessed firsthand the profound impact of small initiatives, whether with the kitchen or the lab, I am committed to not only pioneering groundbreaking solutions but ensuring their ethical deployment for the greater good. Yet, amidst these achievements, I've learned that true impact lies not just in the solutions we create, but in the communities we empower and the compassion we impart. As I stand at this career juncture, poised to apply past knowledge together with knowledge gained at Columbia, I'm reminded of the responsibility that accompanies expertise. It's not just about the technical advancements, but the ethical and societal implications ingrained within them. In conclusion, my journey is not solely about conquering technical challenges, but about leveraging AI as a tool for profound societal change. I am driven by a fervent desire to create solutions that innovate, empathize, and empower. This convergence of expertise and empathy, I believe, is the catalyst for true, enduring transformation.

## Personal Statement – PhD Admissions, Spring 2024

Entrepreneurship is the best source of community development. After running a takeaway of authentic Kashmiri cuisine for five years, connecting with various people from different backgrounds and cultures has been impactful beyond measure. Initially this venture was meant to support my and my brother's education, however it quickly turned into a means of representing our lost language, culture, and identity. My homeland, Kashmir, had a long standing tradition of scholarship older than two millennia. A bloody genocide in the 1990s, due to terrorism in the region, not only uprooted that tradition but also displaced what was once a close knitted community and culture. With a family that was attacked and held captive by terrorists, the aftereffects of such a mass upheaval were not just economic but also emotional. Quintessentially, as resilient people, we rekindled a long lost camaraderie through connections made from our takeaway. In pursuit of providing our patrons with an authentic experience, I underwent an evolution in my personality wherein I learned my own language and the history of my homeland which originally spawned its scholarly tradition. This quest for knowledge prompted me to expand business operations by harnessing analytics over accumulated customer data to derive valuable consumer insights. Given my engineering background, I naturally gravitated towards machine learning backed data analytics. Subsequently, this fascination led me to pursue a diploma to focus on machine learning and data analytical approaches. Therein began my tryst with AI.

As I began to delve into the foundations of AI, in an unfortunate turn of events, the passing of my grandfather due to medical negligence had me pondering on 'Is there a way to predict indications of organ failure, expediting medical attention for its prevention?'. This question became the catalyst for my journey into AI driven healthcare, as I translated physiological medical records into machine readable features and applied the newly acquired fundamentals of machine learning to develop a universally applicable prediction system. To expand on the idea, I reached out to engineers and managers pitching ideas and receiving feedback that eventually led to my venture merging with a local startup also led by a fellow Kashmiri. This startup actively treated patients with blood disorders. I partnered with them to work on a remotely operated AI powered disease detection and prediction system. Owing to India's large rural population, the patient to doctor ratio is highly skewed. It was imperative that reliable diagnosis from urban and suburban centers could reach people who did not have access to sufficient healthcare. As I progressed into the deployment stage, it became apparent that I had a long way to go before I could bridge the gap between theoretical concepts and practical application, especially if it entailed utilizing the rapid innovations in AI. Therefore, I registered for a master's degree in biomedical engineering with a focus on data science at City College New York.

Since my journey into AI at City College, I have worked with leaders of financial trading companies and developers of AI powered applications, allowing exposure to the intricacies of articulating problems followed by testing and automating solutions, and then scaling them. My most recent development with Prof. Bruce Johnson at the Advanced Science Research Center was a model predicting protein movement profiles using GANs based on experimental NMR data. Additionally, I designed a multi-cascaded architecture of FNet Auto-Encoders, Time to Vector Transformers, and Residual ConvNets trained on multi-channel time series NMR data to predict properties of sparsely populated protein conformations crucial to drug discovery. Having integrated both models with the NMRFx Analyst, a software that primarily processes experimental data, I honed

## Personal Statement

Kaustubh Sapru

my skills to include data analysis, statistical inference, hypothesis testing, development of machine learning/deep learning models, data manipulation and visualization. With these skills, I believe I can make constructive contributions to machine learning, artificial intelligence, computer vision, NLP or the computational biology groups at the Computer Science department at Columbia.

One of the defining strengths of the AI community is its openness to diverse ideas and collaboration opportunities. One of the pivotal relationships that quickly propelled me towards becoming a proficient problem solver was the mentorship I received from my supervisor, Prof. Lucas Parra. Our interactions were marked by day to day brainstorming sessions, during which we consistently explored innovative concepts. Through these collaborative efforts, I enhanced the analytical capabilities of my research team by creating a computer vision software program designed to analyze the neural response of human subjects exposed to facial stimuli. This addressed the larger goal of quantifying attention and engagement in humans. My contribution ultimately culminated in the publication of our findings in Nature Communications and my thesis has had increasing viewership around the globe ever since. This accomplishment was further acknowledged and celebrated by both the Biomedical Engineering department and my advisor, who honored me with multiple prestigious awards for my dedicated work.

With a comprehensive background in biomedical engineering, which encompasses intensive research experience and a solid three-year tenure as a professional data scientist, I find myself well-positioned to explore the intricacies of computer science at a deeper level. This endeavor promises to provide my ideas with a more substantial foundation, enhancing their potential for impactful outcomes. Nestled in the vibrant city of New York, Columbia University offers a unique setting where collaboration with individuals from diverse backgrounds pushes the boundaries of what is achievable. Capitalizing on industry partnerships and the interdisciplinary atmosphere, coupled with my entrepreneurial spirit, I aspire to make a meaningful difference in the lives of millions, which constitutes my ultimate goal. I bring to the table a strong sense of organization, self-motivation, and resourcefulness, along with a creative mindset that thrives in team settings. My ability to identify problems and collaborate with teams to propose efficient solutions is a testament to my skill set. In addition to my formal education and professional experience, I have also demonstrated my commitment to becoming a proficient code developer through multiple online courses and successful completion of various coding tests. Since my initial encounter with artificial intelligence, I have diligently pursued every opportunity to expand my knowledge, and my enthusiasm for continued growth in this cutting-edge domain intensifies with each passing day. My interactions with researchers, including instructors and fellow students at Columbia University, have deepened my appreciation for Columbia's commitment to diversity and inclusion. I firmly believe that this dedication to historically underrepresented communities will afford me the freedom to express my ideas openly, brainstorm innovative solutions, and bring them to fruition. Having witnessed firsthand the profound impact even small initiatives can make, I am eager to harness my strong work ethic in pursuit of my mission: addressing the deficiencies in our current societal landscape to benefit the broader population. Consequently, I am at a juncture in my career where I am prepared to take the next significant stride and effectively apply the knowledge I acquire at Columbia University to positively shape our ever-evolving society.