Research Interests and Background: My research interests span Machine Learning (ML) and Computer Vision (CV). Building robust systems that model visual perception and understanding find important applications in medicine, surveillance, photography, autonomous vehicles, and much more. I am particularly interested in building reliable systems that model visual perception with limited supervision. My prior research contributions have tackled key tasks in computer vision including (1) image compositing (background replacement), (2) unsupervised representation learning, (3) image understanding with limited data, and (4) AI-based vision tools for healthcare. These have resulted in publications at top conferences like: **IJCAI, WACV, SIGIR, EMBC**, and **IMWUT/Ubicomp**. I wish to earn a Ph.D. in Computer Science to tackle core problems in computer vision, computer graphics, and machine learning.

Computer Vision and Machine Learning: I delved into research in my sophomore year, when I started working with **Prof. P. J. Narayanan** at the Center for Visual Information Technology (CVIT) at IIIT-Hyderabad. I began with exploring problems in the space of Computational Photography and Structure from Motion. For my first major research project, I worked on the task of **color-consistent background replacement**. The aim was to automatically replace a dull or overexposed sky in outdoor images with an aesthetically pleasing one. I proposed a hypothesis — "the skies of two images with similar illumination and foregrounds (in terms of color, layout, and semantic structure), can be swapped to produce realistic composites without requiring foreground color correction." We used an ensemble of hand-crafted and deep-learned features to encode the illumination and foreground. These features are then used to retrieve realistic and aesthetically appealing candidates for sky replacement. The candidates are re-ranked by optimizing a max-sum diversification objective to make the selection more diverse. The proposed method was validated with quantitative experiments and a comprehensive user study. I received the **Dean's Research Award** for this project, and it led to a publication at **MMM 2018**.

This motivated me to work on deeper problems and I got interested in using **representation learning** to model the style and content in images. Style recognition and similarity are important measures for understanding abstract concepts like art, fashion, design, etc. However, the definition of style is contextual and vague. Deep neural networks (DNNs) perform well on image understanding tasks but require densely annotated large-scale datasets. Obtaining such annotations is expensive (in terms of time and money) and often infeasible. I developed a DNN based framework for learning a neural embedding that captures the 'look-and-feel' of an image. Unlike previous supervised learning methods, the proposed framework is **unsupervised** – it does not use categorical labels but uses a gram matrix feature-based proxy measure for forming triplets of anchor, similar, and dissimilar images. These triplets are used to train a siamese network with a triplet loss to learn an embedding useful for style-based search and retrieval. The embeddings achieved state-of-the-art results for style-based image retrieval and recognition on six datasets that capture different notions of style. This work led to a publication at **IEEE WACV 2020**.

During my research at CVIT, a few of my lab mates and I had been exploring **multimodal machine learning**. We decided to enroll in The Clickbait Detection Challenge co-organized by Google and the Webis Research Group. The challenge required us to build a system that automatically identifies clickbait posts (given its content and images) used to lure users into clicking on articles that fail to fulfill the post description. We proposed a siamese network that fuses visual and textual data to learn robust neural embeddings that are used to classify an online post as clickbait or not. I took the lead on developing the visual feature extractor and the siamese network module. We secured 3^{rd} place in the competition, and an extension of this work was published at **ACM SIGIR 2018** .

My research journey at IIIT-Hyderabad culminated in the form of my Master's thesis. Working at CVIT introduced me to the fascinating world of research, where I developed a solid foundation in computer vision and machine learning, and skills to collaborate effectively and conduct principled research. The brainstorming sessions with my advisor and senior graduate students taught me to communicate well, ask questions, and develop methods for answering them. I left CVIT feeling optimistic about research and wished to explore similar problems further.

After completing my thesis, I joined the **Media and Data Science Research** lab at **Adobe** as an intern. Here I worked on building an efficient solution to detect and segment unseen object classes in images. This is a common use-case for softwares like Adobe Experience Manager – a digital asset management solution for brands like Audi, Nike, etc., to maintain their product images. These are large repositories often with images of very similar objects. An automated solution to segment 'novel' object classes can save a lot of time and effort. During my internship, I carried out an extensive literature survey of the domain and narrowed our focus to the **Few-Shot Image Segmentation** problem, *i.e.,* learning to segment images given only a few samples (1-5) for each unseen object class. I proposed a novel method that utilizes image similarity in terms of both foreground and background information to perform few-shot image segmentation. The method achieved state-of-the-art performance on both the one-shot and five-shot segmentation benchmarks. The work was published at **IJCAI 2020**, and a **US patent** has been filed for the same. My internship experience made me a better engineer and researcher. I learned to write well-documented and optimized code, and build applications to integrate my research project as a software tool.

Applying Vision, ML, HCI to Healthcare: To further my research experience, I joined the **Microsoft Research (MSR)** Lab as a **Pre-doctoral Research Fellow** (residency program). I am currently working with **Dr. Mohit Jain** and **Dr. Nipun Kwatra** on developing AI-based low-cost diagnostic solutions. Over the past year, I have worked on building a smartphone-based corneal topographer to diagnose keratoconus – a severe eye disease. Such a device can be highly useful in rural and remote locations where modern medical facilities are inaccessible.

The work at MSR has challenged me in new and unique ways. A commercial corneal topographer is an expensive (∼$10,000) device with proprietary hardware and software. There is limited relevant literature and no open-source implementation available. To add to the challenge of building such a device, there are no standard datasets available to test our pipeline. Despite these challenges, we remained persistent. We set up a collaboration with the Sankara Eye Hospital and worked with three senior ophthalmologists and an optician to understand the principle behind the working of a commercial topographer. I explored related technical papers and patent documents that gave us key insights in designing the system. At a high level, a corneal topographer projects a concentric ring pattern on the eye and reconstructs the corneal surface using principles of ray-optics and 3D geometry. The generated topography maps highlight any deformities in the cornea. After months of consistent efforts, iterating over the system design, and running pilot studies, I was able to build a working prototype. The final system consists of a 3D printed conical head to project a concentric ring pattern, an AI-assisted smartphone app to capture eye images, and an analysis pipeline to reconstruct the corneal surface. We conducted an extensive clinical evaluation of our system and found that it performed at par with the gold-standard medical device "Optikon Keratron". This project led to a publication at **ACM IMWUT/Ubicomp 2021**. We are now working towards deploying the device at Sankara clinics across India for mass screening of keratoconus. The proposed attachment costs **$33** (**303x** less than Keratron) and weighs just **140 grams**, making it ideal for use by community health workers.

In the context of healthcare diagnostics, I also worked on the task of **abnormality detection** from auscultation signals. I developed a "ResNet34" based lung sound classification model to identify abnormalities like crackling and wheezing. Interestingly, a simple CNN model worked best for this task compared to Hybrid CNN-LSTM networks and Transformer networks, which had been the focus of prior work. We proposed and combined 'novel' data augmentations and transfer learning techniques to achieve top scores on a small-sized dataset. The work was accepted at **EMBC 2021**. This reinforced my motivation to study machine learning as a continuous process, that builds effectively on past knowledge and adapts to new task settings with minimal supervision.

My experience at MSR broadened my perspective on how technology can have an impact beyond the realm of computer science as a field. I also noticed that there exists a huge gap between research and deployment. Once the system is deployed, many unforeseen scenarios can arise – system crashes, unexpected data (due to changing environment), and edge cases. These unexpected scenarios are not accounted for when working in a controlled lab setting. This emphasized the importance of building robust and generalizable systems. It made me realize that research is not a destination but a continuous process of experimenting, problem solving and making improvements.

Future Goals and a Ph.D.: After completing my Ph.D., I wish to continue in academia. I have found learning new concepts, conducting research, collaborating with fellow researchers, and teaching – fulfilling and enriching. A Ph.D. would help me build a strong foundation and depth to achieve this goal. This decision is informed by my research background and positive experience working as a teaching assistant for graduate and undergraduate level courses. My experience as a TA improved my understanding of multiple core subjects, and it was gratifying to see students indulge in critical thinking and ask challenging questions. During my Ph.D. I wish to continue working on fundamental problems in Computer Vision and Machine Learning. My experience with deploying deep learning (DL) based systems brought to light an important practical issue that arises when moving from a research setting to deployment. Although the DL model achieves state-of-the-art performance on standard datasets, it fails to generalize on real-world datasets, where the distribution of images and the number of classes may be very different. This is a common problem and it is impractical to have densely annotated large-scale data for every use case. I wish to work towards solving this by investigating techniques like transfer learning, domain adaptation, weakly-supervised and/or self-supervised learning to make such systems robust and easily usable in practice.

I am particularly interested in working with **Prof. Bernt Schiele** and **Prof. Andrea Vedaldi** as my research interests strongly align with their work on image and video understanding. I also am interested in working with **Professors Matthias Bethge, Matthias Hein and Zeynep Akata,** as my research interests are highly intertwined with the spirit of their past and current work around unsupervised learning, representation learning and explainable ML. Similarly, I am also interested in **Prof. Gerard Pons-Moll's** work on image synthesis and 3D modeling. Building an accurate virtual model of the real-world can help alleviate the need for human annotated datasets, and can propel significant advancements in deep learning.

The **ELLIS** program offers a unique opportunity to work at the best labs across Europe, and be supervised by leading researchers in the area of machine learning and its application in computer vision. Hence, it would be ideal for conducting fundamental and inter-disciplinary research. I believe my strong background in research, software engineering experience, and alignment of interests with faculty makes me a good fit for the ELLIS Ph.D. program.