Brandon Y. Yang
brandonyifanyang.com branyang@virginia.edu

In the modern era of embodied Artificial Intelligence (AI) systems, the integration of Natural Language Processing (NLP) and Computer Vision (CV) foundational models trained on extensive data has propelled efforts to develop generalist robotic policies. Despite these efforts, current models often fall short in **reasoning**, **planning**, and **executing** *long-horizon* tasks. My research focuses on addressing these challenges at the intersection of NLP and robotics, where I aim to:

(1) **Develop embodied agents that follow and interact with natural language *collaboratively*.**
(2) **Leverage language for reasoning and planning to navigate complex environments and solve long-horizon tasks.**

Reinforcement Learning and the Need for Reasoning Robotics is inherently collaborative, yet enabling true *autonomous* cooperation in *multi-agent* systems remains challenging. I joined the Collaborative Robotics Lab at the **University of Virginia** (UVA) with Prof. Tariq Iqbal, focusing on multi-agent RL to develop collaborative policies. I developed simulation environments for complex assembly tasks and designed offline centralized RL policies. However, I realized our robots' successes depended on reward functions requiring labor-intensive design processes. This highlighted a significant limitation: *the inability of robots to adapt to new environments without extensive human input*. I became curious about whether robots could mimic how humans navigate complex environments through innate reasoning rather than external supervision.

Language and Vision Guided Robotic Manipulation Motivated by the limitations of reward supervision, I explored *language* and *vision*-driven approaches to enable autonomous reasoning and action. I developed *GLOMA: Grounded Location for Object Manipulation*, a framework using large language models (LLM) and image diffusion models to generate goal images from language instructions. *GLOMA* enables robots to execute goal-conditioned policies without the need for manually crafted reward functions, allowing them to autonomously imagine subgoals for long-horizon tasks. I led the development of *GLOMA*, including dataset creation and model fine-tuning. However, I realized that 2D image-based methods fail to capture the 3D semantics of natural environments, motivating me to explore 3D-based methods for robotic perception.

Advancing Robotic Perception with 3D Gaussian Splatting To continue goal synthesis motivated by autonomous policies, I expanded to 3D by collaborating with Prof. Jia-Bin Huang at the **University of Maryland** and colleagues from **MIT**. Our research addresses the limitations of 2D synthesis in environments needing 3D understanding, such as vertical displacement. We leverage 3D Gaussian Splatting (3DGS) for highly accurate 3D field representation, enabling more effective robotic perception. To enhance semantic understanding, we inject embeddings from large 2D foundational models into 3DGS, allowing robots to comprehend scene semantics and perform object-level edits. Our preliminary results demonstrate that 3D goal synthesis enables more robust and precise robotic manipulation. As project lead, I developed the codebase and explored various embedding injection techniques.

Enhancing Robotic Planning with Skill-Conditioned Architectures While goal-synthesis methods in 2D and 3D can be robust and interpretable, they incur significant computational overhead due to their multi-step processes. Recent research has focused on end-to-end Vision-Language-Action (VLA) models to streamline this process. However, these models lack interpretability and generalization, especially with out-of-distribution data. To overcome these limitations, I am collaborating with Prof. Yen-Ling Kuo at **UVA** to develop *SkillVLA*, a novel VLA model that enhances long-horizon, language-guided robotic policies by introducing a *skill-conditioned* action output space. In *SkillVLA*, each action is grounded to a specific

Brandon Y. Yang
brandonyifanyang.com branyang@virginia.edu

skill—such as *grasp* or *lift*—which improves the interpretability of the policy. This enables robots to perform complex tasks more efficiently and adapt to diverse environments. I plan to submit this work to **RSS 2025** and believe *SkillVLA* can advance skill-based learning in modern robotic manipulations.

Future Plans I plan to extend my research in skill-conditioned reasoning and planning for embodied agents by developing systems that utilize **complex semantic concepts** (e.g., *object affordances, spatial relations, grasping strategies*) to enhance their understanding of the physical world. Acquiring these concepts alongside skill manipulation priors (e.g., *picking, placing*) enables embodied agents to **reason**, **plan**, and **execute** actions based on a deeper understanding of objects and scenes. This approach could facilitate **robust generalist robotic policies** for language-guided manipulation in complex environments. I will leverage insights from NLP and multimodal communities, which have made significant progress in modeling semantic structures across perceptual inputs. Furthermore, I believe richer 3D environment representations can enhance concept learning, allowing agents more informed reasoning about the physical world.

I also aim to develop embodied agents that work ***collaboratively***. While modern LLMs can converse with humans, current robot systems have yet to fully leverage this capability for collaboration. By integrating NLP methods for collaborative dialogue into embodied AI, I aim to enable robots to perform *collaborative* tasks, such as assembling furniture or cleaning, in partnership with humans or other agents.

Why UPenn The Robotics Master's program at UPenn is the ideal environment to pursue my academic and research goals. I am drawn to the specialized coursework involving robotics control and design that I have not previously studied. Courses such as *Design of Mechatronic Systems* and *Control and Optimization* will help me bridge the critical gaps between high-level AI decision-making, computer vision perception, and low-level control. These classes will not only fill my knowledge gaps in robotics but also provide a robust foundation for more advanced research.

In addition, GRASP offers the perfect environment to extend classroom learning into research settings. I am particularly excited about the opportunity to work with faculty whose research aligns closely with my interests. Prof. Dinesh Jayaraman's work in Object-Centric Spatial Attention resonates with my goal of building foundational agents capable of performing specific tasks with specific skills. Prof. Nadia Figueroa's research in Human-Robot Interaction and robot safety directly supports my research objectives. The collaborative environment at GRASP, including student seminars and reading groups, promises a rich, engaging academic experience, and I am very excited about the opportunity to contribute to this community. GRASP's collaborative setting is particularly appealing to me, as I believe great research emerges from environments that foster open dialogue and interdisciplinary exchange.

Drawing from my three years of experience as an undergraduate teaching assistant, I hope to continue fostering educational growth at Penn. I believe teaching is just as critical as learning, and I am committed to creating supportive, collaborative learning environments that empower students to reach their full potential. I am excited about the opportunity to contribute to the educational mission at UPenn and help students develop the skills they need.

Ultimately, I believe the Robotics Master's at UPenn GRASP will help me strengthen my technical skill set, research skills, communication skills, and leadership abilities to better help me prepare for a future PhD in robotics and/or ML. I am excited about the opportunity to contribute to the vibrant academic community at UPenn and look forward to the possibility of joining the Robotics Master's program.

Brandon Y. Yang
brandonyifanyang.com branyang@virginia.edu

Living in China, Germany, and the United States has shaped my deep appreciation for multiculturalism and diverse perspectives. I've learned that the most meaningful connections come from genuine curiosity about others' cultures and ways of thinking. At Penn, I plan to explore community by joining reading groups and discussion forums that bring together students from different fields. I'm especially excited to engage with peers who have distinct academic backgrounds, as their viewpoints often spark new ideas for my work in robotics and AI. For me, community is built through collaboration and the exchange of ideas across disciplines, and I hope to find that at Penn.

Teaching has also been a central part of how I connect with others. As an undergraduate teaching assistant for computer science courses, I loved working with students who approached problems in unique ways. Whether explaining gradient descent or discussing systems architecture, I found that breaking down complex ideas in a way that resonated with each individual not only helped them learn but also deepened my own understanding. At Penn, I want to continue building community through mentorship and by creating inclusive study groups where students can learn from one another.

By contributing to Penn's culture of collaboration, I hope to both shape and be shaped by the community. I'm eager to learn from the diverse perspectives around me, which will push me to think more creatively and fuel my passion for research and teaching. In turn, I aim to bring my experiences and commitment to interdisciplinary collaboration to help strengthen Penn's vibrant academic environment.