VISHNU SASHANK DORBALA

◇ **Website**: https://vdorbala.github.io ◇ **Phone**: (+1) 202-322-2098 ◇ **Email**: vdorbala@umd.edu

EDUCATION

University of Maryland, College Park *July 2021 - Present*
Doctor of Philosophy (PhD), Computer Science

University of Maryland, College Park *July 2019 - May 2021*
Masters of Engineering, Robotics

Symbiosis International University, Pune, India *June 2013 - July 2017*
Bachelor's of Engineering, Electronics and Telecommunication

RESEARCH INTERESTS

Foundation Models (LLMs & VLMs), Embodied AI, Robotics, Human-AI Interaction

WORK EXPERIENCE

GAMMA Lab, University of Maryland January 2020 - Present
Advised by ***Prof. Dinesh Manocha***

Generalizing Embodied AI with Foundation Models

- Conducting research on generalizing tasks and approaches in *Embodied AI*, using foundation models (VLMs, LLMs and Diffusion) for **data generation and inference**.
- Developed **LGX**, a novel *LLM-based exploration scheme* that queries LLMs for commonsense knowledge to achieve SOTA zero-shot robot navigation with no training (RA-L 2023).
- Developed an LLM-based pipeline for *synthesizing human-like instructions*, mitigating reliance on user annotated data and empowering researchers to create embodied navigation datasets (NAACL 2023).
- Studied the influence of **VLM-driven conversation** between overhead and ground mobile robots to improve language-guided navigation (ICRA 2025).
- Established the generalizability of various embodied navigation schemes on *dynamic* topological graph environments with ***portable*** **household targets** such as watches or phones (Paper Under Review).

Sony AI May 2024 - Sept 2024
Applied Scientist Intern at ***Core AI Research Team (CAIR)***

How good are Vision-Language Models (VLMs) at Question Answering?

- Developed a scalable data generation approach to evaluate *Question Answering* (QA) performance on VLMs. This was used to generate a dataset of *spatial*, existential and compositional queries for images.
- Used LoRA and vLLM to finetune and evaluate this dataset with 15 different VLMs;inferred that **larger VLM models are more sensitive to visual inputs over smaller ones performing QA**.

Amazon Alexa AI May 2023 - Sept 2023
Applied Scientist Intern under ***Dr. Reza Ghanadan***

Large Language Models for Generating Subjective Queries

- Proposed a novel and challenging variant to the *Embodied Question Answering* problem that tackles *situational* queries over simple ones proposed in prior art. Our work highlights the **limitations** of LLM-generation, showing them to be **good at producing queries, but poor at answering them** (Under Rev. at Robotics Conf.).

Amazon Alexa AI May 2022 - Sept 2022
Applied Scientist Intern under ***Prof. Gaurav Sukhatme***

CLIP-Nav: Using CLIP for Zero-Shot Vision-and-Language Navigation

- Developed a novel zero-shot navigation scheme that uses VLM grounding scores to solve **Vision-And-Language Navigation (VLN)** on REVERIE. Our approach showed improved generalizability over dominant supervised learning methods, and established a **zero-shot baseline** for this task. (Published at CoRL, 2022.)

Nokia Bell Labs June 2020 - December 2020

Robotics Research Intern/Co-op

Robust Navigation on an Indoor Robot

· Implemented indoor **Visual SLAM** on a Loomo robot, in a complex office environment.

· Developed a novel SLAM pipeline that utilized landmark information to improve navigation performance in the presence of glass windows and reflections.

Center for Visual Information Technology, IIIT, Hyderabad Sept 2017 - May 2019

Research Fellow under ***Prof. C.V. Jawahar***

Visual Navigation for an Assistive Robot

· Worked on developing SARA, an assistive ***robot wheelchair*** platform, capable of interacting with people.

· Performed ***autonomous corridor following*** on SARA via **Visual Servoing** (Published at IROS, 2019).

SELECTED PUBLICATIONS

· ***V.S. Dorbala***, et. al, "Improving Zero-Shot ObjectNav with Generative Communication", *Accepted* at IEEE Int. Conf. on Robotics and Automation (**ICRA**), 2025 *(Link to Website)*

· ***V.S. Dorbala***, et. al, "Right Place, Right Time! Dynamizing Topological Graphs for Embodied Navigation", *Under Review* at Double Blind Computer Vision Conference, 2024 *(Link to Paper)*

· ***V.S. Dorbala***, et al., "Can LLM's Generate Human-Like Wayfinding Instructions? Towards Platform-Agnostic Embodied Instruction Synthesis", *Published* at the Annual Conf. of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2024 *(Link to Paper)*

· ***V.S. Dorbala***, et al., "Can an Embodied Agent Find Your "Cat-shaped Mug"? LLM-based Zero-Shot Object Navigation.", IEEE Robots and Automation Letters (**RA-L**), 2023 *(Link to Paper)*

· ***V.S. Dorbala***, et al., "*CLIP-Nav: Using CLIP for Zero-Shot Visual-Language Navigation*" Conference on Robot Learning (**CoRL**), 2022 *(Link to Paper)*

· ***V.S. Dorbala***, et al., "*Can a Robot Trust You? A DRL-Based Approach to Trust-Driven Human-Guided Navigation*" IEEE Int. Conf. on Robotics and Automation (**ICRA**), 2021 *(Link to Paper, Video)*

· ***V.S. Dorbala***, et al., "*A Deep Learning Approach for Robust Corridor Following*" IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (**IROS**), 2019 *(Link to Paper, Video)*

SELECTED PROJECTS

• ***Do Face Generation Models Generate Good Faces?*** Oct 2023

- Used GPT-4 to generate queries that *indirectly* ask a person about their facial appearance.
- **Inference *(Report)***: Observed unusual annotations, a beard bias and exaggerated facial markings.

• ***Did Subtitles Ruin Your Movie?*** Nov 2022

- Used Neural Machine Translation (NMT) to translate movie subtitles of 58 movies from French to English and vice versa. Correlated translation quality with movie ratings.
- **Inference *(Report)*** - In some crucial cases, the NMT subtitles convey scene sentiments far better than human-written ones.

• ***Classifying Glucose Levels with EEG Data*** Dec 2021

- Used a Muse headband to capture EEG brain signals non-intrusively for approximating glucose levels.
- **Inference (*Report*)**: As the data was incredibly noisy given the sheer temporal volume (collected in milliseconds), training a CNN based classifier did not give favourable results. Classical hand-crafted feature extraction proved superior.

ACHIEVEMENTS

· Research conducted on LLM-based exploration (LGX) is being used by a YC funded robotics startup.

· Co-Chaired at a conference session on Human-Robot Interaction: Robot Navigation at ICRA 2021.

· Received the Dean's fellowship for PhD students awarded by the University of Maryland, College Park.