

# Vidhi Jain

✉ [vidhij@andrew.cmu.edu](mailto:vidhij@andrew.cmu.edu)  
📁 [vidhijain.github.io](https://github.com/vidhijain)  
LinkedIn: [vidhijain96/](#)  
ORCID: [0009-0006-8170-4115](#)

## Education

- 2022– **PhD Candidate, Carnegie Mellon University.**  
2025 Robotics Institute, Advisor: [Yonatan Bisk](#)
- 2019– **Masters of Science in Robotics, Carnegie Mellon University.**  
2021 Robotics Institute, *GPA 4.0* Advisor: [Katia Sycara](#)
- 2014– **Bachelor of Engineering (Honors) in Computer Science, BITS Pilani.**  
2018 *CGPA 9.06/10.00, Distinction*, Advisor (off-campus thesis): [Aaron Courville](#)

## Experience

- Sep'24– **Stanford, Palo Alto, CA, USA.**  
Present Visting Research with [Dorsa Sadigh](#)
- Jun'23– **Google DeepMind, Mountain View, CA, USA.**  
Present Student Researcher with [Debidatta Dwibedi](#)
  - Trained video-conditioned robot policies with cross-attention transformers. [Vid2Robot \(RSS'24\)](#)
  - Extended models for visual grounding with conditional caption generation to question answering. [FlexCap](#)
  - Evaluated prompts for low-level motion control in 3d simulation [PromptBook \(ICRA'24\)](#)
  - Developed pipelines for zero-shot and fine-tuned task evals [Open X-Embodiment \(ICRA'24\)](#) **Best Paper**.
- Sep'21– **Meta (Facebook), Menlo Park, CA, USA.**  
Aug'22 AI resident with [Akshara Rai](#) and [Yixin Lin](#)
  - Developed synthetic data for robot demonstrations of loading the simulated dishwasher in AI Habitat. [Code](#)
  - Trained Encoder-decoder Transformer-based high-level policy for learning preferences in loading dishwasher from a single demonstration as prompt. [Paper](#) | [Website](#)
- Aug'18– **Microsoft Research, Bengaluru, India.**  
Aug'19 Research Fellow with [Amit Deshpande](#) and [Navin Goyal](#)
  - Divergence minimization in GANs** Investigated the properties of Jensen Shannon and Wasserstein divergences between the given and the generated distributions during training neural network generators, to discuss the learning dynamics like convergence and generalization.
- Jan'18– **Montreal Institute for Learning Algorithms (MILA), Montreal, QC, Canada.**  
Jun'18 Research Intern with [Aaron Courville](#)
  - Out-of-Distribution Detection in Generative Models** Experimented on autoregressive generative models and Variational Autoencoders, to detect any 'out-of-distribution (OOD)' sample at deployment. [Thesis](#)
- May'17– **Simon Fraser University (SFU), Burnaby, BC, Canada.**  
Jul'17 MITACS Globalink Research Intern with [Oliver Schulte](#)
  - Learning Bayesian Networks for Relational Databases** Remodelled existing codebase for project *FactorBase* to integrate cross-table sufficient statistics or population variables in the contingency tables for learning ground-level sufficient statistics. [Code](#) | [Invited Paper](#)
- May'16– **Forschungszentrum Informatik (FZI), Karlsruhe, Germany.**  
Jul'16 Research Intern with [York Sure-Vetter](#) and [Matthias Frank](#)
  - API annotation platform using Semantic Web** Designed and developed a customized Semantic MediaWiki for annotation of APIs for dynamic integration of provenance information in the decision support system of Project *BigGIS*; Presented Technical Poster at Grace Hopper Celebration India GHCI'16, Bengaluru, India; [Extended Abstract](#) | [Poster](#)

## Publications

- Aug 2024 **ANAVI: Audio Noise Awareness using Visuals of Indoors for NAVigation.**  
**Vidhi Jain**, Rishi Veerapanini, Yonatan Bisk. *Accepted at* Conference of Robot Learning (CoRL) 2024, Munich, Germany. [Preprint](#)
- Mar 2024 **FlexCap: Generating Rich, Localized, and Flexible Captions in Images.**  
Debidatta Dwibedi, **Vidhi Jain**, Jonathan Tompson, Andrew Zisserman, Yusuf Aytar. *Accepted at* Conference on Neural Information Processing Systems (NeurIPS) 2024, Vancouver, Canada. [Paper](#) | [Website](#)
- Jan 2024 **Vid2Robot: End-to-end Video-conditioned Policy Learning with Cross-Attention Transformers.**  
**Vidhi Jain**, Maria Attarian, Nikhil J Joshi, Ayzaan Wahid, Danny Driess, Quan Vuong, Pannag R Sanketi, Pierre Sermanet, Stefan Welker, Christine Chan, Igor Gilitschenski, Yonatan Bisk, Debidatta Dwibedi, *Presented at* Robotics: Science and Systems (RSS) 2024, Delft, Netherlands. [Paper](#) | [Website](#)
- Nov 2023 **How to Prompt Your Robot: A PromptBook for Manipulation Skills with Code as Policies.**  
Montserrat Gonzalez Arenas, Ted Xiao, Sumeet Singh, **Vidhi Jain**, Allen Z. Ren, Quan Vuong, Jake Varley, Alexander Herzog, Isabel Leal, Sean Kirmani, Dorsa Sadigh, Vikas Sindhwani, Kanishka Rao, Jacky Liang, Andy Zeng. *In Proceedings at* International Conference on Robotics and Automation (ICRA) 2024, Yokohama, Japan. [Paper](#)
- Nov 2023 **Open X-Embodiment: Robotic Learning Datasets and RT-X Models.**  
Open X-embodiment collaboration. *In Proceedings at* International Conference on Robotics and Automation (ICRA) 2024, Yokohama, Japan. ([Best Paper](#)) [Paper](#) | [Website](#)
- Jun 2023 **SLAP: Spatial-Language Attention Policies.**  
Priyam Parasher, **Vidhi Jain**, Xiaohan Zhang, Jay Vakil, Sam Powers, Yonatan Bisk, Chris Paxton. *In Proceedings of* Conference on Robot Learning (CoRL) 2023, Atlanta, USA., [Paper](#) | [Website](#)
- Jun 2023 **HomeRobot: Open-Vocabulary Mobile Manipulation.**  
Sriram Yenamandra, Arun Ramachandran, Karmesh Yadav, Austin S Wang, Mukul Khanna, Theophile Gervet, Tsung-Yen Yang, **Vidhi Jain**, Alexander Clegg, John M Turner, Zolt Kira, Manolis Savva, Angel X Chang, Devendra Singh Chaplot, Dhruv Batra, Roozbeh Mottaghi, Yonatan Bisk, Chris Paxton. *In Proceedings of* Conference on Robot Learning (CoRL) 2023, Atlanta, USA. [Paper](#) | [Website](#) | [Competition @ NeurIPS 2023](#)
- Dec 2022 **Transformers are Adaptable Task Planners.**  
**Vidhi Jain**, Yixin Lin, Eric Undersander, Yonatan Bisk, Akshara Rai. *In Proceedings of* Conference on Robot Learning (CoRL) 2022, Auckland, New Zealand. [Paper](#) | [Website](#) | [Video](#) | [Code](#)

## Workshop Presentations

- Dec 2022 **MAEA: Multimodal Attribution for Embodied AI.**  
**Vidhi Jain**, Jayant Sravan Tamarapalli, Sahiti Yerramilli, Yonatan Bisk. *Presented at* NeurIPS 2022, 5th Robot Learning Workshop: Trustworthy Robotics (RLW 2022), New Orleans, USA. [Short Paper](#) | [Website](#)
- Oct 2020 **Predicting Strategies in Simulated Search and Rescue.**  
**Vidhi Jain**, Rohit Jena, Huao Li, Tejus Gupta, Dana Hughes, Michael Lewis, Katia Sycara. *Presented at* NeurIPS AI+HADR 2020, Virtual. [Video](#) | [Preprint](#)
- Oct 2020 **Learning Embeddings that Capture Spatial Semantics for Indoor Navigation.**  
**Vidhi Jain**, Shishir Patil, Prakhar Agarwal, Katia Sycara. *Presented at* NeurIPS 2020 Object Representations for Learning and Reasoning (ORLR 2020), Virtual. [Video](#) | [Preprint](#) | [Code](#)
- Sep 2020 **Learning to Navigate in Unseen Cluttered Structured Environments.**  
**Vidhi Jain**, Ganesh Iyer, Katia Sycara. NeurIPS 2020 Women In Machine Learning (WiML 2020), Virtual. [Poster](#)
- Jul 2020 **Coping with Sample Inefficiency in Deep Reinforcement Learning (DRL) in embodied AI.**  
**Vidhi Jain**, Simin Liu, Ganesh Iyer. ICML'20 Women In Machine Learning (WiML) Un-Workshop, Virtual. [Discussion](#) | [Presentation](#)
- Aug 2017 **Model Selection Scores for Multi-Relational Bayesian Networks.**  
Sajjad Gholami, Oliver Schulte, **Vidhi Jain**, Qiang Zhao. IJCAI 2017 Declarative Learning Based Programming (DeLBP). [Invited Paper](#)

## Preprints

- Dec 2023 **Towards General-Purpose Robots via Foundation Models: A Survey and Meta-Analysis.**  
Yafei Hu\*, Quanting Xie\*, **Vidhi Jain\***, Jonathan Francis, Jay Patrikar, Nikhil Keetha, Seungchan Kim, Yaqi Xie, Tianyi Zhang, Shibo Zhao, Yu Quan Chong, Chen Wang, Katia Sycara, Matthew Johnson-Roberson, Dhruv Batra, Xiaolong Wang, Sebastian Scherer, Zsolt Kira, Fei Xia, Yonatan Bisk. [Paper](#) | [Website](#)
- Nov 2019 **Towards Zero-Shot Alignment and Retrieval for Forensic Detection.**  
**Vidhi Jain\***, Amrit Setlur\*, Aditya Pratapa\*. [Report](#)

## Tech Reports

- Aug 2021 **Towards Explainable Embodied AI.** Masters Thesis. [Report](#)
- Jun 2018 **Investigating the viability of Generative Models for Novelty Detection.** Bachelors Thesis. [Report](#)

## Selected Awards and Honors

- Aug 2020 **CIFAR Deep Learning Reinforcement Learning Summer School**
- Apr 2020 **Young Researcher at Heidelberg Laureate Forum (HLF '20)** among 224 researchers worldwide.
- Jul 2019 **J N Tata Endowment Scholarship for Higher Studies, India**
- Jun 2019 **K. C. Mahindra Scholarships for Post-Graduate Studies Abroad**
- May 2017 **ACM-W Scholarship for Attendance at Research Conferences**
- Mar 2017 **Citi Women Leader Award (CWLA), Mumbai, India**, Awarded one year of study scholarship covering upto INR 400,000 i.e. USD 6000, among the top 3 out of 1200 applicants
- Jun 2016 **GE Foundation Scholar-Leaders Program (GEFSLP)**, Awarded USD 2000 for two years of study, among top 5 scholars selected in India
- Mar 2016 **National Finalists at Microsoft's 'Build the Shield', Hyderabad, India**, Ranked 15<sup>th</sup> nationwide
- 2014–15 **Merit Scholarship** by Dean, BITS Pilani, *Top 1% among the batch of about 900 students*
- 2013 **Awarded Kishore Vaigyanik Protsahan Yojna (KVPY) Fellowship** by Dept of Science and Tech., Govt. of India
- 2010 **Awarded National Talent Search Exam (NTSE) Scholarship** by NCERT, India
- 2010 Ranked Top 1% in Regional State in **National Standard Examination in Junior Science (NSEJS)**

## Talks

- Mar 2024 **How to think of multimodality in Embodied AI and Robotic Manipulation?.**  
Guest Lecture in Multimodal Machine Learning course for graduate students at CMU, Pittsburgh. [Lecture Video](#)
- Jun 2023 **Transformers are Adaptable Task Planners.**  
CoRL Spotlight talk  
Google RL Reading Group
- Apr 2023 **Training vision models in PyTorch.**  
Hands-on lecture in Computer Vision course for graduate students at CMU, Pittsburgh. [Lecture Video](#)
- Dec 2022 **MAEA: Multimodal Attribution Framework for Embodied AI.**  
NeurIPS 2022, New Orleans. [Conference Video](#)
- Oct 2021 **3 brushes to paint your research canvas.**  
SAiDL (Virtual) [Video](#) | [Event Poster](#)
- Dec 2020 **Predicting Human Strategies in Simulated Search and Rescue .**  
NeurIPS 2020 Virtual [Video](#)

Dec 2020 **Learning Embeddings that Capture Spatial Semantics for Indoor Navigation.**

Neurips 2020 Virtual [Video](#)

May 2019 **Tutorial on Deep Learning with PyTorch.**

PyLadies Bangalore, India. [Presentation](#) | [Event Poster](#)

Jan 2019 **The One in Asankhya Project.**

by Sukriti Paul and Mansi Goyal (Virtual) [Video](#) | [Blog](#)

Aug 2018 **Invited speaker for 'Research in undergraduate studies'.**

IIIT-Bangalore ACM Student Chapter, Bangalore, India [Video](#)

Jul 2017 **Invited speaker for discussion on 'Innovation by Young India'.**

telecasted on national news channel NDTV India at New Delhi office. [Video](#)

## Professional Service and Leadership

Sep'23– **Reviewer** at IEEE International Conference on Robotics and Automation (ICRA), International Conference Present on Intelligent Robots and Systems (IROS)

Jan'20– **Reviewer** at International Conference on Machine Learning (ICML), Neural Information Processing and Present Signal (NeurIPS) and International Conference on Learning Representations (ICLR).

Jun'20– **Workshop Session Organizer** | Coping with Sample Inefficiency in Deep Reinforcement Learning (DRL) Jul'20 in embodied AI, ICML'20 WiML Un-Workshop, Virtual. [Report](#) | [Presentation](#)

Sep'17– **Co-Founder and Vice Chair** | BITS Pilani Association for Computing Machinery-Women (ACM-W) May'18 Chapter, *Started with 16 on-campus girls, organized project brainstorming and mentoring sessions*

May'15– **Technical Volunteer** | ARISE Impact, non-profit organization *Developed Android application for Aug'15 Audio-based learning for visually challenged*

## Teaching Experience

Jan'24– **Teaching Assistant** | 11777 Multimodal Machine Learning by Yonatan Bisk. Mentoring 5 teams with May'24 3-5 students each for semester-long research projects.

Jan'23– **Teaching Assistant** | 16720 Computer Vision by Deva Ramanan. for 106 students enrolled in the course, May'23 prepared homework, and graded.

Jan'17– **Teaching Assistant** | CS F111 Computer Programming by Vishal Gupta, *Mentored 35 students for 4 May'17 hours per week for Code programming lab sessions.*

Aug'14– **Teaching Volunteer** | Computer Literacy Programme, National Service Scheme, *Tutored 12-15 rural May'15 natives for 3 hours per week on computer applications.*

## References

**Yonatan Bisk**, Assistant Professor, Carnegie Mellon University, ybisk at andrew.cmu.edu

**Debidatta Dwibedi**, Research Scientist, Google DeepMind, Mountain View, debidatta at google.com

**Akshara Rai**, Research Scientist, Meta (previously Facebook), Menlo Park, akshararai at meta.com