

Kanishk Jain

📱 mobile | @ email | 🔗 linkedin | 🐙 github | 🌐 webpage

EDUCATION

Université de Montréal | Mila

Montréal, Canada

PhD in Computer Science and Engineering; GPA: 4.3

Sep 2023 – Present

- Research Focus: Multi-Agent Reinforcement Learning, Video Understanding, Generative Modeling
- Advisor: Dr. Aishwarya Agrawal

International Institute of Information Technology

Hyderabad, India

MS by Research in Computer Science and Engineering; GPA: 9.33/10

Aug 2021 – Dec 2022

- Research Focus: Multimodal Learning, Visual Grounding, Language-Guided Autonomous Navigation, Semi-Supervised Learning
- Advisor: Dr. Vineet Gandhi

International Institute of Information Technology

Hyderabad, India

B.Tech (Honors) in Electronics and Communication Engineering; GPA: 6.73/10

Aug 2013 – Jun 2017

RESEARCH PUBLICATIONS

Benchmarking Vision Language Models for Cultural Understanding

Paper

Shravan Nayak, Kanishk Jain, Rabiul Awal, Siva Reddy, et al.

EMNLP 2024 (Oral)

- Introduced CulturalVQA, a benchmark for evaluating Vision Language Models' (VLMs) comprehension of geo-diverse cultural elements, revealing performance disparities across regions and cultural facets.

Test-Time Amendment with a Coarse Classifier for Fine-Grained Classification

Paper

Kanishk Jain, Shyamgopal Karthik, Vineet Gandhi

NeurIPS 2023

- Developed a post-hoc correction strategy using coarse-grained predictions at test-time to reduce mistake severity and enhance fine-grained classification accuracy.

Instance-Level Semantic Maps for Vision Language Navigation

Paper

Laksh Nanwani, Anmol Agarwal, Kanishk Jain, et al.

ROMAN 2023

- Designed instance-level semantic maps to enable robust, language-directed navigation across diverse environments, improving navigation accuracy and adaptability.

Ground then Navigate: Language-guided Navigation in Dynamic Scenes

Paper

Kanishk Jain, Varun Chhangani*, Amogh Tiwari, K. Madhava Krishna, Vineet Gandhi*

ICRA 2023

- Developed a visual-grounding approach for language-guided navigation in dynamic outdoor environments, enhancing autonomous navigation systems' responsiveness to linguistic commands.

Bringing Generalization to Deep Multi-View Pedestrian Detection

Paper

Jeet Vora, Swetanjali Dutta, Kanishk Jain, Shyamgopal Karthik, Vineet Gandhi

WACV-W 2023

- Formulated an evaluation framework and dataset to benchmark generalization in multi-view pedestrian detection systems, facilitating improved model robustness across diverse environments.

Comprehensive Multi-Modal Interactions for Referring Image Segmentation

Paper

Kanishk Jain, Vineet Gandhi

ACL Findings 2022

- Proposed a novel architecture for Referring Image Segmentation, leveraging synchronous multi-modal interactions and hierarchical aggregation to enhance segmentation performance.

Grounding Linguistic Commands to Navigable Regions

Paper

Kanishk Jain, Nivedita Rufus*, Unni Krishnan*, Vineet Gandhi, K. Madhava Krishna*

IROS 2021

- Introduced Referring Navigable Regions (RNR), a task for grounding navigable road regions based on linguistic commands, advancing language-guided autonomous navigation capabilities.

WORK EXPERIENCE

Noah's Ark Lab | Huawei

Intern Researcher

Sep 2025 – Present

- Building modality-invariant foundation models and unified fusion layers to better encode and aggregate multi-modal signals
- Generalizing contrastive learning beyond two modalities with a variation of traditional InfoNCE loss

Université de Montréal | IFT 6765

Teaching Assistant

Jan 2025 – Apr 2025

- Led weekly tutorials and office hours, answering student queries
- Designed and graded assignments and presentations; mentored teams on experiment design and ideas.

Université de Montréal | Mila

PhD Student

Sep 2023 – Present

- Leveraging multi-agent reinforcement learning for emergent communication protocol and lifelong learning.
- Exploring early-exit methods for streaming video understanding, for real-time decision making with reliability guarantees.

CVIT, IIIT Hyderabad

Research Engineer

Mar 2023 – Aug 2023

- Developed Hierarchical Ensembles to reduce mistake severity in fine-grained classification, achieving improved performance in fully supervised and semi-supervised settings.
- Designed and deployed a real-time player tracking solution in Bird's Eye View for live sports analysis during the 2022 Asia Cup.

CVIT, IIIT Hyderabad

Research Assistant

Sep 2019 – Dec 2022

- Led research on language-guided autonomous navigation, introducing explicit grounding techniques for navigable regions with real-time visual feedback integration.
- Enhanced multi-modal interaction models for referring image segmentation, increasing the efficiency and precision of visual-linguistic model performance.
- Developed an analytics tool for Counter-Strike: Global Offensive (CSGO) games, providing actionable insights and improving gameplay strategies for competitive teams.

Turvo

Software Engineer

Jul 2017 – Aug 2019

- Integrated Xero Accounting Platform with Turvo's platform, using a Pub-Sub messaging pattern to handle diverse accounting scenarios efficiently.
- Implemented batch payment processing, enabling users to schedule and manage multiple payments simultaneously, improving workflow efficiency.
- Developed an Optical Character Recognition (OCR) system for document images, utilizing active learning and template detection to extract data from unstructured documents.

SELECTED PROJECTS

T2I Generation using SSMs: Analyzing the effectiveness of Selective State Space Models (SSM) in reducing computational complexity for text-to-image generation.

Top-View Player Tracking: Developed and deployed a real-time player tracking solution in Bird's Eye View, used live during the 2022 Asia Cup Cricket Tournament for performance analysis.

Stereo SLAM: Created 3D point clouds from stereo images and estimated motion/pose by applying the iterative Perspective-n-Point (PnP) algorithm for 2D-3D correspondences.

Pose Graph Optimization: Implemented the Levenberg–Marquardt algorithm to optimize robot poses using Odometry and Loop Closure constraints for 1D and 2D Simultaneous Localization and Mapping (SLAM).

Unity Game for Amblyopia: Developed a Unity-based game for diagnosing Amblyopia, controlled through eye gaze movements tracked by an eye-tracking device.

Neuro Rehab Systems: Designed a rehabilitation tool aimed at assisting recovery from nervous system injuries and minimizing functional impairments.

TECHNICAL STRENGTHS

Languages: Python, Java, C++, C#, Node JS, Javascript

Frameworks: PyTorch, Keras, Tensorflow, OpenCV, scikit-learn, scikit-image

Tools: CARLA, TensorRT, Open3D, Unity 3D, Matlab, Spring Boot, Maven, REST

Databases: MySQL, Mongo DB, Elasticsearch, Apache Solr, Redis

RELEVANT COURSES

ML/AI Courses: Statistical Methods in AI, Computer Vision, Mobile Robotics, Topics in Optimization Methods, Topics in ML, Cognitive Science and AI

Core Science: Computer Programming, Operating Systems and Algorithms, Data Structures

Other Courses: Digital Image Processing, Digital Signal Processing, Linear Algebra, Probability and Random Processes, Discrete Mathematics

ACADEMIC SERVICE

Reviewer: NeurIPS (2025); CVPR (2024); WACV (2024, 2025); ACL (2024); ECCV (2024); ICRA (2024).

ACHIEVEMENTS

Morocco Solidarity Hackathon: Member of the winning team; developed a solution to predict human trafficking behavior from social media posts.

Qualcomm Innovation Fellowship: Led the winning team of the Qualcomm Innovation Fellowship (QIF) 2020 in India.

JEE Mains: Ranked in the top 0.2% nationally among 1,200,000 candidates in JEE Mains.

JEE Advanced: Secured a rank of 4539 among 150,000 candidates in JEE Advanced.

R&D Showcase: Presented the Amblyopia Game at the college's annual R&D showcase.

REFERENCES

Dr. Vineet Gandhi, Dr. K. Madhava Krishna