

Xavier Thomas

xthomas@bu.edu — [LinkedIn](#) — [Website](#) — [Google Scholar](#) — [Github](#)

RESEARCH INTERESTS

Computer Vision, Generative Models, Vision-Language Models, Multimodal Learning, Interpretability

EDUCATION

Boston University, Boston, United States

Sep 2025 — Expected May 2030

Ph.D. in Computer Science

Advisor: [Prof. Deepti Ghadiyaram](#)

Boston University, Boston, United States

Sep 2023 — May 2025

Master of Science in Artificial Intelligence (Thesis Track)

Thesis: Diffusion Models as Representation Learners: Interpreting and Leveraging their Latent Features

Advisor: [Prof. Deepti Ghadiyaram](#)

Manipal Institute of Technology, Manipal, India

Jul 2018 — Aug 2022

Bachelor of Technology in Electronics and Instrumentation

Minor in Computational Intelligence

SELECTED PUBLICATIONS

Semantic Richness or Geometric Reasoning? The Fragility of VLM’s Visual Invariance

Jason Qiu*, Zachary Meurer*, **Xavier Thomas***, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

COLM 2026

Generative Action Tell-Tales: Assessing Human Motion in Synthesized Videos

Xavier Thomas, Youngsun Lim, Ananya Srinivasan, Audrey Zheng, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

ECCV 2026; VGBE & PhysHuman, CVPRW 2026 (Oral)

Some Modalities are More Equal Than Others: Decoding and Architecting Multimodal Integration in MLLMs

Tianle Chen*, Chaitanya Chakka*, Arjun Reddy Akula, **Xavier Thomas**, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

CVPR Findings 2026; Sight and Sound, CVPRW 2026 (Oral)

What’s in a Latent? Leveraging Diffusion Latent Space for Domain Generalization

Xavier Thomas, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

ICCV 2025

Revelio: Interpreting and leveraging semantic information in diffusion models

Dahye Kim*, **Xavier Thomas***, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

ICCV 2025; MIV, CVPRW 2025 (Oral)

Progressive Prompt Detailing for Improved Alignment in Text-to-Image Generative Models

Ketan Suhaas Saichandran*, **Xavier Thomas***, Prakhar Kaushik, Deepti Ghadiyaram | [Webpage](#) | [Paper](#)

AIACC, CVPRW 2025 (Oral)

Diversity vs. Recognizability: Human-like Generalization in One-Shot Generative Models

Victor Boutin, Lakshya Singhal, **Xavier Thomas**, Thomas Serre | [Paper](#)

NeurIPS 2022

Adaptive Methods for Aggregated Domain Generalization

Xavier Thomas, Dhruv Mahajan, Alex Pentland, Abhimanyu Dubey | [Paper](#)

Preprint

* denotes equal contribution

RESEARCH & WORK EXPERIENCE

Boston University

Boston, United States

Graduate Researcher

Jun 2024 — Present

- **Improving Temporal Perception and Reasoning in MLLMs:** Diagnosing bottlenecks in MLLMs’ temporal representations and developing methods to strengthen understanding of events, motion, and order in video.
- **Visual Reasoning in MLLMs:** Evaluated MLLMs across visual domains (e.g., photos, sketches) under basic geometric transformations, showing that performance collapses as semantic cues diminish and models rely on recognition shortcuts.
- **Evaluation of Video Generation Models:** Developed an evaluation framework grounded in human action dynamics to measure motion fidelity and temporal coherence in generated videos – missed by existing video-generation metrics.

- **Diffusion Models as Representation Learners:** Probed intermediate diffusion features to uncover the semantic concepts and domain-specific information they encode, and leveraged them for classification and domain generalization.
- Advisors: [Deepti Ghadiyaram](#) | [Makarand Tapaswi](#), [Shraman Pramanick](#), [Subhashini Venugopalan](#)

ShareChat | Content and User Understanding Team

Machine Learning Engineer Intern

Bangalore, India

Jul 2022 — Jun 2023

- Designed **MAViC, a Multimodal Active Learning algorithm for Video Captioning** that reduces annotation effort by integrating semantic similarity and uncertainty from visual and language modalities. Achieved **96%** of CIDEr on MSR-VTT and **79%** of CIDEr on MSVD **using only 25% of the training data**, compared to full-data performance.
- Integrated an **advanced computer vision pipeline** into **production**, improving both content classification and moderation capabilities on ShareChat (180M+ MAUs) and Moj (160M+ MAUs).
- Developed an **AI-powered content creation tool** for seeding content, and built an **Automated News Dashboard** to compile daily news into genre-specific video snippets, improving user engagement and retention.
- Integrated **Natasha, an AI assistant**, into ShareChat’s messaging service.
- Manager: [Vikram Gupta](#) | Research Advisors: [Anant Raj](#), [Hisham Cholakkal](#)

Serre Lab | Brown University

Research Intern

Providence, United States

Sep 2021 — May 2022

- Developed a **novel evaluation framework for one-shot image generative models**, introducing metrics for sample recognizability and diversity to enable systematic comparisons between models and human one-shot generalization.
- Benchmarked 4 representative generative architectures against human performance on the Omniglot dataset.
- Advisors: [Victor Boutin](#), [Thomas Serre](#)

Massachusetts Institute of Technology

Research Assistant

Cambridge, United States

Jan 2021 — Nov 2021

- Created a **novel algorithm for privacy-preserving domain generalization** that recovers domain information by removing class-specific noise from latent features, enabling the training of robust, domain-adaptive classifiers.
- **Outperformed state-of-the-art methods** that require domain supervision on multiple benchmarks, demonstrating that privacy and adaptability can be achieved simultaneously without sacrificing accuracy.
- Advisor: [Abhimanyu Dubey](#)

École de technologie supérieure (ÉTS), Montréal

Mitacs Globalink Research Intern

Montreal, Canada

Jul 2021 — Sep 2021

- Designed a **margin-based separation loss** that sharpens the model’s confidence in correct sub-category predictions relative to incorrect ones, improving pseudo-label quality for Weakly Supervised Semantic Segmentation.
- Introduced a **multi-crop augmentation strategy**, training on multiple randomly-scaled crops of each image so pseudo-labels stay reliable even when objects appear at different scales or positions.
- Advisor: [Jose Dolz](#)

FOR.ai

Researcher

Oct 2020 — Aug 2021

- Contributed to a **large-scale benchmarking study of Out-of-Distribution (OOD) detection** in computer vision models, establishing baselines for evaluating robustness under distribution shifts.
- Collaborated with researchers from Google Brain, University of Oxford, and Vector Institute as part of the FOR.ai (now Cohere For AI) research collective.
- Advisor: [Sheldon Huang](#)

SELECTED TALKS & PRESENTATIONS

- **Generative Action Tell-Tales: Assessing Human Motion in Synthesized Videos;** Oral presentations at the VGBE and PhysHuman Workshops, CVPR 2026; ML Collective Deep Learning: Classics and Trends Reading Group, Apr 2026; New England Computer Vision Workshop, Nov 2025.

SERVICE & MENTORSHIP

- **Research Mentor:** Mentored Atul Das (master’s); Jason Qiu and Zachary Meurer (undergraduate); and Audrey Zheng, Ananya Srinivasan, Preksha Shyamsukha, and Siwoo Song (high school). Audrey, Preksha, and Siwoo participated through [BU RISE](#). Mentored projects led to coauthored papers at ECCV 2026 and COLM 2026.
- **Teaching Assistant:** DS 542: Deep Learning for Data Science (Spring 2024, Fall 2024, Spring 2025); CS 541: Applied Machine Learning (Spring 2026).
- **Reviewer:** IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026; Neural Information Processing Systems (NeurIPS), 2026; International Journal of Computer Vision (IJCV); Actionable Interpretability Workshop, ICML 2025; Mechanistic Interpretability Workshop, NeurIPS 2025.