Xavier Thomas

xthomas@bu.edu — LinkedIn — Website — Google Scholar — Github

RESEARCH INTERESTS

Computer Vision, Representation Learning, Vision-Language Models, Multi-modal Large Language Models, Interpretability

SKILLS

- **Programming Languages:** Python, MATLAB, C++, SQL, HTML/CSS, TypeScript
- **Tools & Technologies:** Linux, Bash Scripting, Git, Jira, Confluence, Docker, Kubernetes, Jenkins, Spinnaker, CI/CD, Google Cloud, SQLite, Redis, gRPC, Open MPI, RESTful APIs, Django, Streamlit, FastAPI, Chainlit, LangChain, LangGraph, NumPy, Pandas, PyTorch, Accelerate, Hugging Face Transformers, OpenCV

EDUCATION

Boston University, Boston, United States   Sep 2025 — current
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

PUBLICATIONS, PREPRINTS, AND WORKING PAPERS

Generative Action Tell-Tales: Assessing motion physics in synthesized videos
Xavier Thomas, Youngsun Lim, Ananya Srinivasan, Audrey Zheng, Deepti Ghadiyaram
Under review

What's in a Latent? Leveraging Diffusion Latent Space for Domain Generalization
Xavier Thomas, Deepti Ghadiyaram | ⊙ Code | 🖹 Paper
International Conference on Computer Vision (ICCV), 2025

Revelio: Interpreting and leveraging semantic information in diffusion models
Dahye Kim*, **Xavier Thomas***, Deepti Ghadiyaram | ⊙ Code | 🖹 Paper
International Conference on Computer Vision (ICCV), 2025

Progressive Prompt Detailing for Improved Alignment in Text-to-Image Generative Models
Ketan Suhaas Saichandran*, **Xavier Thomas***, Prakhar Kaushik, Deepti Ghadiyaram | ⊙ Code | 🖹 Paper
AI4CC Workshop, IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2025 (oral)

Diversity vs. Recognizability: Human-like generalization in one-shot generative models
Victor Boutin, Lakshya Singhal, **Xavier Thomas**, Thomas Serre | ⊙ Code | 🖹 Paper
Neural Information Processing Systems (NeurIPS), 2022

Adaptive Methods for Aggregated Domain Generalization
Xavier Thomas, Dhruv Mahajan, Alex Pentland, Abhimanyu Dubey | ⊙ Code | 🖹 Paper
Preprint

MAViC: Multimodal Active Learning for Video Captioning
Gyanendra Das, **Xavier Thomas**, Anant Raj, Vikram Gupta | 🖹 Paper
Preprint

RESEARCH & WORK EXPERIENCE

Boston University   Boston, United States
Graduate Researcher   Jun 2024 — Present

- **Vision in Multimodal Large Language Models (MLLMs):** Investigating limitations of visual understanding in MLLMs and developing methods to improve cross-modal alignment for robust multimodal reasoning.

- **Evaluation of Video Generation Models:** Designing and implementing novel evaluation metrics to assess human action fidelity, temporal consistency, and motion coherence in generative video models.
- **Internal Representations of Diffusion Models:** Analyzing diffusion models as representation learners by probing their intermediate states; demonstrating their effectiveness for downstream tasks such as classification, multi-modal reasoning, and domain generalization.

ShareChat | Content and User Understanding Team Bangalore, India
Machine Learning Engineer Intern 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 the 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 engagement and retention.
- Integrated **Natasha, an AI assistant**, into ShareChat's messaging service.
- Manager | Advisors: Vikram Gupta | Prof. Anant Raj, Prof. Hisham Cholakkal
- 📄 Paper

Serre Lab | Brown University Providence, United States
Research Intern Sep 2021 — May 2022

- Developed a **novel evaluation framework for one-shot generative models**, introducing new metrics for recognizability (human interpretability) and diversity (concept coverage) to enable systematic comparisons.
- Benchmarked 4 representative generative architectures against human performance on the Omniglot dataset.
- Advisors: Dr. Victor Boutin, Prof. Thomas Serre
- 🎧 Code | 📄 Paper

Massachusetts Institute of Technology Cambridge, United States
Research Assistant 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: Dr. Abhimanyu Dubey
- 🎧 Code | 📄 Paper

École de technologie supérieure (ÉTS), Montréal Montreal, Canada
Mitacs Globalink Research Intern Jul 2021 — Sep 2021

- Extended **sub-category exploration methods** for Weakly Supervised Semantic Segmentation by clustering image features to generate more accurate pseudo-labels.
- Designed **novel constraint-based refinements** to enhance object localization in Class Activation Maps (CAMs), and improved mean Intersection-over-Union (mIoU) scores on the PASCAL VOC 2012 benchmark.
- Advisor: Dr. 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

SERVICES

- **Reviewer:** Actionable Interpretability Workshop, ICML 2025; Mechanistic Interpretability Workshop, NeurIPS 2025
- **Mentoring:** Atul Das (Masters, BU), Audrey Zheng (RISE program, High School), Ananya Srinivasan (High School), Jason Qiu (undergrad, BU), Zachary Meurer (undergrad, BU)