Leo Villani

lv259@cornell.edu +1 914 414 9381 leov974.github.io leo-villani LeoV974

Research Interests

I am interested in statistical theory and machine learning, with a focus on uncertainty, structure, and generalization in representations and graphs.

Education

Cornell University *Aug 2025 – Present*
Doctor of Philosophy in Statistics
GPA: 4.180 / 4.000

University of California, Berkeley *Aug 2021 – May 2025*
BA Applied Mathematics and Computer Science
Highest Honors (Mathematics), High Distinction (Computer Science)
GPA: 3.96 / 4.00

Experience

Researcher *Berkeley, CA*
BLISS *Mar 2025 – Present*

- Designed controlled game-to-text benchmarks (Nim variants) and fine-tuned transformer language models to study shortcut learning, modular reasoning, and generalization under distribution shift.
- Developed probing + adversarial training workflows (discriminator probes, DANN-style invariance) to reduce reliance on spurious prompt features, including automated dataset generation, checkpoint evaluation, and analysis on NCSA Delta.

Undergraduate GSI *Berkeley, CA*
UC Berkeley, Department of Mathematics *Jul 2023 – Jan 2025*

- Facilitated discussion sections, designed and graded exams, and maintained regular office hours to support student learning.
- Consistently earned a student rating of 6.7/7.0, significantly above the department average of 5.8/7.0.

Projects

Controlled Creativity in Generative Image Models *Ithaca, NY*
Cornell University *Oct 2025 – Jan 2026*

- Developed a theory-driven framework for *controlled creativity* in generative image models by designing geometry-based latent-space perturbations that increase novelty while preserving semantic identity.
- Implemented and benchmarked multiple generative pipelines (VAE- and GAN-based), including post-hoc sampling methods that enforce semantic constraints via PCA subspaces and Jacobian null-space projections.
- Built end-to-end experimental tooling for reproducible creative sweeps (step-size/constraint grids), quantitative semantic-stability evaluation using pretrained vision classifiers, and publication-quality figure generation for an ICML-style manuscript.

Selected Class Projects

- Secure File Sharing System: Developed an end-to-end encrypted file sharing system featuring secure file storage, user authentication, and revocable sharing.
- BYOW: Implemented an engine for generating explorable worlds. Extra Credit Presentation

Technologies

Languages: Python, Java, C, C++, MATLAB, R, SQL, HTML, Go, Bash, Squiggle, LaTeX

Packages: PyTorch, TensorFlow, Pandas, Statsmodels, Scikit-learn, Matplotlib, Dask

Awards

Outstanding Graduate Student Instructor (OGSI) Award *UC Berkeley Mar 2024*

Bronze Medal *MathCON Apr 2019*