CIS 6270: Discrete Generative Models

Fall 2026

Instructor

Professor Pranam Chatterjee, PhD
pranam@engineering.upenn.edu
Lab Website: chatterjeelab.com
Office Hours: Wednesdays 5 - 6 pm

TAs

Tong Chen (CIS PhD Student, Chatterjee Lab)
chentong@engineering.upenn.edu
Office Hours: Thursdays 5 - 6 pm
Rosie Zhang (CIS PhD Student, Chatterjee Lab)
roxiez@engineering.upenn.edu
Office Hours: Tuesdays 5 - 6 pm
Sophie Vincoff (BE PhD Student, Chatterjee Lab)
svincoff@engineering.upenn.edu
Office Hours: Mondays 5 - 6 pm

Class Times and Location

Tuesdays and Thursdays, 8:30 - 10 am in Towne 337

Office Hours Location

Scheduled per session at either: One uCity Square Lobby or Amy Guttman Hall

Course Description

Generative models operate on continuous and discrete state spaces. This course develops their mathematical, algorithmic, and empirical foundations, with discrete state modeling as the main focus. We begin with probability, linear algebra, differential equations, information theory, maps, pushforward distributions, changes of variables, and conservation of probability. We then develop continuous generative models through ordinary and stochastic differential equations, flow matching, diffusion, score matching, denoising objectives, probability-flow ODEs, and guidance. These constructions lead to discrete diffusion, including masked and uniform corruption, categorical posteriors, continuous-time jump processes, block diffusion, and guided sequence generation. We next derive discrete flow matching through probability paths, target rate matrices, conditional and marginal rates, simplex-valued interpolants, Fisher-Rao geometry, Gumbel-softmax flows, rectification, and multi-objective guidance. The course then studies one-step and few-step generation through flow maps, consistency objectives, MeanFlow, discrete and latent flow maps, posterior guidance, expanding state spaces, and stochastic flow maps. The final lecture connects these methods to optimal transport and Schrödinger bridges through couplings, the Monge and Kantorovich problems, Wasserstein distances, entropic regularization, Sinkhorn iterations, path-space relative entropy, stochastic control, Doob transforms, iterative proportional fitting, and discrete bridge matching. Lectures emphasize derivations of objectives, transition kernels, rate matrices, probability-evolution equations, training losses, and sampling rules, together with worked numerical examples and analyses of computational behavior. Integrated code examples give students experience implementing architectures, training objectives, samplers, guidance procedures, and evaluations. Applications and projects span biological sequences and structures, images, audio, natural language, code, graphs, point clouds, 3D molecular data, probability-simplex representations, and continuous latent representations. By the end of the semester, students will be able to derive and analyze continuous and discrete generative processes, translate mathematical specifications into working algorithms, evaluate models across data modalities, and read and contribute to current research in generative modeling.

Learning Objectives

1. Build a rigorous understanding of the mathematical foundations of generative models, including probability, the probability simplex, continuous-time Markov chains, the master equation, and generative transport frameworks.

2. Derive, analyze, and implement the core generative frameworks: classical generative models, discrete diffusion, discrete flow matching, optimal transport, stochastic control, and discrete Schrödinger bridges.

3. Apply these frameworks across biological sequences and structures, images, audio, natural language, code, graphs, point clouds, 3D molecular data, probability-simplex representations, and continuous latent representations.

4. Connect derivations to implementations by building algorithms from scratch and studying their behavior empirically through benchmarks and ablations.

5. Read current research and design, communicate, and defend novel discrete generative modeling strategies for sequence design and analysis.

Prerequisites

Students must have completed one of the following courses: CIS 5190, CIS 5200, or ESE 5460.

Textbooks

The course follows original lecture notes written by Professor Chatterjee, which covers the full material in this class. Deep Generative Modeling by Jakub M. Tomczak (Springer, 2024) is the suggested introductory reference.

Communications

Course communications will be via Canvas: `https://canvas.upenn.edu/`. Students are also encouraged to email the instructor and TAs.

Grading

- **Participation (10%):** Active participation in lecture and discussion is expected. This includes contributing to proofs and derivations in class, engaging in algorithm design sessions, and working through implementation examples. Students are encouraged to arrive to class on time each day, but must arrive by 8:40AM to be marked present. *Starting at 8:41AM, new arrivals will be considered absent.* Unexcused absences will reduce the final grade (1 point per absence, 0.5 if reported at least a week in advance). The instructor and TAs reserve the right to deduct points based on level of contribution to the class.

- **Exams (30%):** Two written exams (15% each), closed-book with a single reference sheet allowed. Exams emphasize rigorous mathematical reasoning and algorithm design. Students will construct formal proofs of convergence, invariance, or optimality; derive equations central to discrete generative models (for example forward and reverse kernels in diffusion, the master equation and rate matrices, the generating velocity and continuity equations in flow matching, and variational bounds such as the ELBO); and analyze the behavior of stochastic processes using tools such as eigenvalue decompositions and coupling arguments. Some questions will ask students to design or adapt algorithms in pseudocode, bridging theory and practice. The exams assess depth of understanding of the theorems, derivations, and algorithmic frameworks introduced in lecture and the textbook. **Exam 1 covers material taught from August 25 through September 29:** probability and linear algebra foundations, maps and changes of variables, ODEs and probability conservation, continuous flow matching, SDEs, Fokker-Planck, continuous diffusion, score matching, DDPMs, guidance, masked diffusion, and MDLM foundations. **Exam 2 is cumulative and covers all material taught from August 25 through November 12:** all Exam 1 topics together with MDLM implementation, categorical corruption, continuous-time discrete diffusion, block diffusion, discrete flow matching, simplex and Gumbel-softmax flows, rectification and multi-objective guidance, flow map matching, latent and discrete flow maps, posterior guidance, expanding state spaces, and strong stochastic flow maps.

- **Group Projects (60%):** Students will be assigned to groups at the beginning of the semester based on their level of expertise, with students at similar levels grouped together so that each team works at a comparable pace. There are three projects, and each one ends with a defense.

- **Project 1, Continuous Generative Models (15%):** Teams conduct a two-modality empirical study of continuous generative modeling. They implement and compare continuous flow matching and diffusion under a matched evaluation protocol, demonstrate guided generation, and introduce and ablate a methodological innovation. The innovation may modify the probability path, coupling, objective, guidance rule, sampling procedure, or another justified component. Teams first evaluate the method on one continuous modality and then transfer the same innovation to a second modality to test whether the improvement generalizes.
- **Project 2, Discrete Generative Models Workshop (20%):** Teams build a base generative model for a discrete state space, such as discrete diffusion or discrete flow matching, and demonstrate guidance toward a specified condition or property. The study should include appropriate baselines, quantitative evaluation, guidance-strength analyses, and controlled ablations. The code and written report follow an ICLR workshop format. Top papers will be invited to submit to an appropriate ICLR 2027 workshop.
- **Final Project, ICML-Ready Paper (25%):** Teams extend Project 1, Project 2, or both into a complete ICML-ready paper. The final work should develop a coherent methodological contribution and support it with the theory, baselines, ablations, multi-seed experiments, guidance analyses, and cross-setting validation needed for a main-conference story. Deliverables include a complete code repository and a written report prepared in the ICML style. Top papers may be submitted to ICML 2027.

Defense format. Every project ends with a 30-minute defense per group: a 10-minute presentation of the work followed by 20 minutes of questions from the teaching team. The code and writeup are due before class (8:30 am) on the Tuesday or Thursday before the defense, which leaves that day for the team to prepare its slides and talk. Defenses are then held the following day, on a Wednesday for Projects 1 and 2 and on Friday, December 4 for the final project, so that the full day can be scheduled for talks.

Practice Problems

There are no graded problem sets. We will release practice problems with full solutions throughout the semester. These are ungraded and are meant to help you prepare for the exams and build fluency with the derivations and implementations from lecture.

Final Grade

No curve; traditional grading scheme (90-93 = A-, 94-97 = A, 97-100 = A+). Grades posted on Canvas and directly scaled to percentage.

Tentative Schedule

Date	Class Topic
Tue, Aug 25	**Lecture 1: Probability Foundations** Course logistics and roadmap; continuous and discrete state spaces; probability distributions; expectation and Monte Carlo estimation; joint, marginal, and conditional probability; Bayes' rule
Thu, Aug 27	**Lecture 1: The Probability Simplex and Linear Algebra** Probability vectors and simplex geometry; logits and softmax; vectors and matrices; linear transformations; eigenvalues, eigenvectors, diagonalization, and matrix exponentials
Tue, Sep 1	**Lecture 1: Maps, Pushforwards, and Changes of Variables** Maps, inverse maps, and composition; pushforward distributions; derivatives and Jacobians; determinants and local volume change; finite change of variables
Thu, Sep 3	**Lecture 1: Vector Fields, Probability Conservation, and Optimization** Gradients, directional derivatives, vector fields, divergence, and the continuity equation; entropy, cross-entropy, KL divergence, constrained optimization, and the complete generative-model framework
Tue, Sep 8	**Lecture 2: ODEs and Continuous Probability Transport** Ordinary differential equations, velocity fields, flow maps, Euler's method, numerical integration, conservation of probability, the continuity equation, instantaneous change of variables, and continuous normalizing flows
Thu, Sep 10	**Lecture 2: Flow Matching** Endpoint couplings, probability paths and interpolants, conditional and marginal velocities, the flow-matching regression objective, conditional-expectation derivation, sampling, neural parametrization, and implementation *Project 1 (Continuous Generative Models) Assigned*
Tue, Sep 15	**Lecture 3: SDEs, Fokker-Planck, and Score Matching** Flow-matching recap, Brownian motion, stochastic differential equations, drift and diffusion, Itô's lemma, the Fokker-Planck equation, variance-preserving diffusion, Gaussian scores, and denoising score matching

Thu, Sep 17	**Lecture 3: Reverse-Time Diffusion and DDPMs** Noise prediction, score-driven sampling, reverse-time SDEs, DDPM forward transitions and posteriors, the learned reverse mean, time-conditioned networks, U-Nets, and DDPM training and sampling
Tue, Sep 22	**Lecture 3: Diffusion Paths and Probability-Flow ODEs** Latent diffusion, AMP-Diffusion, alternative noise paths, probability-flow ODEs, deterministic sampling, and the connections among score matching, diffusion, and flow matching
Thu, Sep 24	**Lecture 3: Guidance for Continuous Generative Models** Conditional scores, classifier guidance, classifier-free guidance, reward guidance, conditional flow matching, multi-objective guidance, time-dependent property reliability, and guided latent flow and diffusion models
Tue, Sep 29	**Lecture 4: Masked Diffusion and MDLM** Discrete sequence corruption, absorbing masks, exact reverse reveal probabilities, clean-token prediction, the masked diffusion ELBO, continuous-time weighting, and MDLM training and sampling *Project 1 (Continuous Generative Models) Code and Writeup due at 8:30 am*
Wed, Sep 30	*Project 1 defenses (Continuous Generative Models)* 30-minute defense per group, 10-minute talk, 20-minute Q&A with the teaching team
Thu, Oct 1	**Fall Term Break (Oct 1-4)**
Tue, Oct 6	**Lecture 4: MDLM Implementation and Categorical Corruption** MDLM implementation for DNA, language and genomic results, token-dependent masking, discrete image generation, D3PM, absorbing and uniform corruption, cumulative categorical kernels, and the discrete diffusion ELBO
Thu, Oct 8	*Exam 1* Material taught from August 25 through September 29: mathematical foundations, maps and changes of variables, ODEs and probability conservation, flow matching, SDEs, continuous diffusion, score matching, DDPMs, guidance, masked diffusion, and MDLM foundations

Tue, Oct 13	**Lecture 4: Continuous-Time Discrete Diffusion** The continuous-time limit of categorical corruption, rate and generator matrices, the master equation, reverse-time rates, probability ratios, discrete scores, score entropy, and continuous-time training losses *Project 2 (Discrete Generative Models Workshop) Assigned*
Thu, Oct 15	**Lecture 4: Block Diffusion and Guided Sequence Generation** Blockwise factorization and training, the connection to autoregressive generation, classifier and classifier-free rate guidance, completion-based value estimates, PepTune, and Pareto-guided peptide generation
Tue, Oct 20	**Lecture 5: Discrete Flow Matching with Jump Rates** Chosen probability paths, probability velocities, the master equation, conditional and marginal jump rates, endpoint posteriors, target rate matrices, the discrete flow-matching loss, and DNA sampling
Thu, Oct 22	**Lecture 5: Dirichlet and Fisher-Rao Flows** Simplex-valued states, Dirichlet conditional paths and scores, endpoint-conditioned velocities, Fisher information geometry, square-root coordinates, spherical geodesics, and simplex flow matching
Tue, Oct 27	**Lecture 5: Gumbel-Softmax Flows and Multi-Objective Guidance** Gumbel-max sampling, differentiable Gumbel-softmax paths, temperature schedules, pathwise velocities, straight-through estimation, MOG-DFM, cone-constrained edit directions, and rectified flow *Project 2 (Discrete Generative Models Workshop) Code and Writeup due at 8:30 am*
Wed, Oct 28	*Project 2 defenses (Discrete Generative Models Workshop)* 30-minute defense per group, 10-minute talk, 20-minute Q&A with the teaching team
Thu, Oct 29	**Lecture 5: Rectification and Multi-Objective Discrete Refinement** Reflow and endpoint recoupling, dependence preservation, ReDi, reward-tilted target distributions, guided local proposals, Metropolis-Hastings and monotone acceptance, AReUReDi, Pareto preferences, and finite-neighborhood convergence

Tue, Nov 3	**Lecture 6: Flow Map Matching and Consistency** Finite-interval flow maps, two-time conditioning, composition and tangent identities, Lagrangian and Eulerian losses, path consistency, Shortcut Models, and MeanFlow *Final Project (ICML-Ready Paper) Assigned*
Thu, Nov 5	**Lecture 6: Latent and Discrete Flow Maps** One-step and few-step generation, latent flow maps and decoder sensitivity, categorical representations, posterior means, cross-entropy and KL objectives, probability-valued teachers, and endpoint-error bounds
Tue, Nov 10	**Lecture 6: Posterior and Reward-Guided Flow Maps** Reward-tilted endpoint distributions, posterior sampling, Diamond Maps, nested conditional velocities, value-gradient guidance, Meta Flow Maps, and stochastic guidance through Doob transforms
Thu, Nov 12	**Lecture 6: Expanding-State and Strong Stochastic Flow Maps** Expansion and transport, generators with birth events, token insertion and variable-length generation, Brownian-path conditioning, shared-noise composition, pathwise refinement, interval drift and diffusion targets, sampling, and molecular applications
Tue, Nov 17	**Exam 2 Review** Cumulative review of all material taught from August 25 through November 12, with emphasis on discrete diffusion, discrete flow matching, guidance, and flow maps
Thu, Nov 19	*Exam 2* Cumulative coverage of all material taught from August 25 through November 12: mathematical foundations, continuous flow and diffusion models, discrete diffusion, discrete flow matching, guidance, and flow maps
Tue, Nov 24	**Thanksgiving Break (Nov 24-29) - no class**
Thu, Nov 26	**Thanksgiving Break (Nov 26-29) - no class**
Tue, Dec 1	**Lecture 7: Optimal Transport and Continuous Schrödinger Bridges** Monge maps, Kantorovich couplings, primal and dual transport, Wasserstein distances, dynamic transport, entropic regularization, Sinkhorn iterations, path-space relative entropy, stochastic control, Girsanov, Doob transforms, and iterative proportional fitting

Thu, Dec 3	**Lecture 7: Discrete Schrödinger Bridges and Course Synthesis** Controlled jump processes, discrete Doob transforms, DDSBM, categorical bridge matching, reciprocal and Markov projections, iterative Markovian fitting, masked bridge models, BranchSBM, EntangledSBM, and course synthesis *Final Project (ICML-Ready Paper) Code and Writeup due at 8:30 am*
Fri, Dec 4	*Final Project defenses (ICML-Ready Paper)* 30-minute defense per group, 10-minute talk, 20-minute Q&A with the teaching team

Student Success

If you are having trouble contributing to or completing assignments or projects, please reach out to me so we can discuss appropriate strategies. You may email me or come to office hours to describe any personal or academic difficulties you are facing. I may also direct you to other resources on campus that can help.

The Weingarten Center for Student Accessibility provides free academic support services for all Penn students, including learning consultations, academic coaching, tutoring, and study skills resources. The Weingarten Center also coordinates academic accommodations for students with disabilities (see below).

If you are concerned about your physical or mental health, Counseling and Psychological Services (CAPS) offers individual and group counseling, workshops, and crisis support. For urgent concerns, CAPS provides 24/7 phone support at 215-898-7021. Additional wellness resources can be found through Student Health and Counseling.

Academic Accommodations

If you need to request accommodations for a disability or medical condition, contact the Weingarten Center for Student Accessibility as early as possible. I will work with their office to ensure equal access to course materials and assessments.
Absences will be accommodated with proper documentation:

- Incapacitating illness or injury (Incapacitation Form required)
- Personal emergencies (dean's excuse required)
- Religious observances (must clear in advance)
- University athletic obligations (NOVAP required)

- Required court or legal appearances (must clear in advance)

Academic Integrity

As members of the Penn community, students are expected to uphold the principles outlined in the Code of Academic Integrity. This code reflects Penn's commitment to scholarship, honesty, fairness, respect, and responsibility. Students must:

- Not engage in cheating, plagiarism, or other forms of academic dishonesty
- Conduct themselves honorably in all academic and nonacademic endeavors
- Seek clarification when unsure about the expectations for an assignment, exam, or group work

Please speak with me to get any clarification about this course.