Yuanqi Du

Curriculum vitæ (5th January 2026)

Ph.D. Student
Department of Computer Science
Cornell University

yuanqidu@cs.cornell.edu
yuanqidu.github.io
Yuanqi Du
github.com/yuanqidu

Education

08/2022–05/2026 (Expected) **Cornell University**, Ph.D. in Computer Science

GPA: 3.96/4.00, Advisor: Prof. Carla P. Gomes

08/2017–05/2021 **George Mason University**, B.S. in Computer Science

GPA: 4.00/4.00

Work experience

08/2025–12/2025 **NVIDIA Research**, Santa Clara, CA

Research Intern, Supervisor: Dr. Tomas Geffner, Dr. Julius Berner, Dr. Arash Vahdat

05/2025–08/2025 **Microsoft Research New England**, Boston, MA

Research Intern, Supervisor: Dr. Carles Domingo-Enrich

10/2021–08/2022 **DP Technology**, Beijing, China

Researcher, Supervisor: Dr. Linfeng Zhang, Mr. Guolin Ke

06/2021–08/2022 **University of Amsterdam**, Remote

Visiting Student, Supervisor: Prof. Max Welling, Prof. Bernd Ensing (AMLab)

11/2020–05/2021 **Microsoft Research Asia**, Beijing, China

Research Intern, Supervisor: Dr. Jianwei Zhu (Machine Learning Group)

Selected publications (full list in the end)

(* denotes co-first authorship and † denotes advising role)

2025

1. **Y. Du**[*,†], B. Yu*, T. Liu*, et al.: *Accelerating Scientific Discovery with Autonomous Goal-evolving Agents*. arXiv preprint arXiv:2512.21782, 2025.
2. Z. Song*, J. Lu*, **Y. Du**[*], et al.: *Evaluating Large Language Models in Scientific Discovery*. arXiv preprint arXiv:2512.15567, 2025.
3. **Y. Du***, J. He*, F. Vargas, Y. Wang, C. P. Gomes, J. M. Hernández-Lobato, and E. Vanden-Eijnden: *FEAT: Free Energy Estimators with Adaptive Transport*. Advances in Neural Information Processing Systems, 2025.
4. **Y. Du**, J. He, E. Vanden-Eijnden, and C. Domingo-Enrich: *REACT: Rare Event Analysis via Stochastic Optimal Control*. arXiv soon, 2025.
5. C. Duan*, G.-H. Liu*, **Y. Du***, T. Chen, Q. Zhao, H. Jia, C. P. Gomes, E. A. Theodorou, and H. J. Kulik: *React-OT: Optimal Transport for Generating Transition State in Chemical Reactions*. Nature Machine Intelligence 7:4, pp. 615–626, 2025. **(Cover Article)**
6. J. He, ..., **Y.Du** [†], and F. Vargas[†]: *RNE: A Plug-and-play Framework for Diffusion Density Estimation and Inference-time Control*. arXiv preprint arXiv:2506.05668, 2025.
7. H. Wang, M. Skreta, C.-T. Ser, ..., **Y. Du**[†], A. Aspuru-Guzik[†], K. Neklyudov[†], and C. Zhang[†]: *Efficient Evolutionary Search over Chemical Space with Large Language Models*. International Conference on Learning Representations, 2025.
8. J. He, ..., **Y. Du**, S. Syed, and F. Vargas: *CREPE: Controlling diffusion with REPlica Exchange*. arXiv:2509.23265, 2025.

9. H. Wang, J. Guo, L. Kong, R. Rampi, P. Schwaller, **Y. Du**†, and C. Zhang†: *LLM-Augmented Chemical Synthesis and Design Decision Programs.* Forty-second International Conference on Machine Learning, 2025.

2024

10. **Y. Du***, M. Plainer*, R. Brekelmans*, C. Duan, F. Noe, C. P. Gomes, A. Aspuru-Guzik, and K. Neklyudov: *Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling.* Advances in Neural Information Processing Systems, 2024. **(Spotlight)**
11. A. Schneuing*, C. Harris*, **Y. Du***, A. Jamasb, I. Igashov, W. Du, T. Blundell, P. Lió, C. Gomes, M. Welling, et al.: *Structure-based Drug Design with Equivariant Diffusion Models.* Nature Computational Science 4:12, pp. 899–909, 2024.
12. **Y. Du***, A. R. Jamasb*, J. Guo*, T. Fu, C. Harris, Y. Wang, C. Duan, P. Liò, P. Schwaller, and T. L. Blundell: *Machine Learning-aided Generative Molecular Design.* Nature Machine Intelligence 6:6, pp. 589–604, 2024.

2023

13. W. Du*, **Y. Du***, L. Wang*, D. Feng, G. Wang, S. Ji, C. P. Gomes, Z.-M. Ma, et al.: *A New Perspective on Building Efficient and Expressive 3D Equivariant Graph Neural Networks.* Advances in Neural Information Processing Systems, 2023.
14. L. Holdijk*, **Y. Du***, F. Hooft, P. Jaini, B. Ensing, and M. Welling: *Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths.* Advances in Neural Information Processing Systems, 2023.
15. C. Duan, **Y. Du**, H. Jia, and H. J. Kulik: *Accurate Transition State Generation with an Object-aware Equivariant Elementary Reaction Diffusion Model.* Nature Computational Science 3:12, pp. 1045–1055, 2023. **(Cover Article)**
16. H. Wang*, T. Fu*, **Y. Du***, W. Gao, et al.: *Scientific Discovery in the Age of Artificial Intelligence.* Nature 620:7972, pp. 47–60, 2023.

Selected invited talks & guest lectures

1. AI-Driven Scientific Discovery
   a) UC Berkeley BIDMaP Institute *Jan 2026*
   b) Flatiron Institute Center for Computational Mathematics *Jan 2026*
   c) Microsoft Research New England *Nov 2025*
   d) Harvard Kempner Institute *Nov 2025*
2. Towards Generalist Agents for Accelerating Scientific Discovery
   a) ACS Generative and Agentic Systems in Chemistry Symposium *March 2026*
   b) alphaXiv Seminar *March 2026*
3. Scaling Probabilistic Models at Inference time for Scientific Discovery
   a) OpenEye CUP XXV Meeting on Generative Design *March 2026*
   b) ACS Machine Learning in Chemistry Symposium *March 2026*
4. Bridging Non-equilibrium Simulation and Probabilistic Machine Learning
   a) ACS Integrating Deep Learning and Dynamics Symposium *March 2026*
   b) Google DeepMind *Nov 2025*
   c) UC Berkeley Pitzer Center for Theoretical Chemistry Seminar *Oct 2025*
   d) University of Maryland Statistical Physics Seminar *Oct 2025*
   e) SIAM PNW Biennial Meeting Scientific ML Symposium *Oct 2025*
   f) Caltech Yisong Yue's Group (CMS) *July 2025*
   g) Caltech AI + Science Lab *July 2025*
   h) UCLA IPAM Workshop on Sampling, Inference, and Data-Driven Physical Modeling in Scientific Machine Learning *July 2025*

i) MIT Zhang Group (Chemistry) *July 2025*

j) UC Berkeley Krishnapriyan and Mandadapu's Group (EECS & CBE) *June 2025*

5. Scientific Knowledge Exists in Large Language Models and You Can Extract It
   a) UW Data Science Seminar *Nov 2025*
   b) UMass Amherst Machine Learning and Friends Lunch *Nov 2025*
   c) UC Berkeley Teresa Head-Gordon Lab (Chemistry) *Oct 2025*
   d) AI Alliance AI for Materials & Chemistry Webinar *August 2025*
   e) USC Melady Lab *July 2025*
   f) EPFL AI for Chemistry and Beyond Seminar *May 2025*
   g) ACS Machine Learning in Chemistry Symposium *March 2025*
6. Accelerating Molecular Discovery with Machine Learning: A Geometric, Sampling and Optimization Perspective
   a) UCSD Rose Yu's Group (Computer Science) *March 2025*
   b) ACS Generative Modeling for Chem, Bio, & Material Symposium *March 2025*
   c) Princeton AI for Accelerating Invention Research Talk Series *Jan 2025*
   d) UCLA Quantum Physics and AI Guest Lecture *Oct 2024*
   e) UT Austin Henkelman Group (Chemistry) *Oct 2024*
   f) MIT Kulik Research Group (Chemical Engineering) *August 2024*
   g) Microsoft Research AI4Science Colloquium *June 2024*
   h) Stanford Theoretical Chemistry Group *June 2024*
   i) Gatech Applied and Computational Mathematics Seminar *April 2024*
   j) Cornell Deep Learning (SYSEN 6888) Guest Lecture *Nov 2023*
7. Accelerating Transition Dynamics Simulation with Machine Learning
   a) Cornell SCAM Seminar (Center of Applied Mathematics) *March 2025*
   b) ACS Inferring Kinetics, Thermodynamics, and Mechanisms from Enhanced Sampling Simulation Symposium (also in PHYS Sci-Mix) *March 2025*
   c) ByteDance AI Lab *March 2025*
   d) NYU Courant Institute Generative Model/Sampling Seminar *Nov 2024*
   e) UC Berkeley BIDMaP Young Scholar Seminar *Nov 2024*
   f) SIAM New York-New Jersey-Pennsylvania Section Annual Conference *Nov 2024*
   g) MIT IAIFI Summer Workshop *August 2024*

Selected community leaderships

1. Founder and Lead Organizer @AI for Science workshop series, with *NeurIPS 2021*, *ICML 2022*, *NeurIPS 2022*, *NeurIPS 2023*, *ICML 2024* and *NeurIPS 2025*.
2. Co-founder and Lead Organizer @Learning on Graphs conferences 2022, 2023 and Logistic Chair 2024 and 2025. Lead Organizer for NYC workshops (2024 spring, 2024 fall, 2025 spring) and Seattle meetup (2024 fall).
3. Co-founder and Lead Organizer for workshops related to probabilistic machine learning @SPIGM with *ICML 2023*, *ICML 2024*, *NeurIPS 2025*, @DGM4HSD with *ICLR 2022*, and @FPI with *ICLR 2025*,
4. Co-founder and Lead Organizer for MSR New England Generative Modeling and Sampling Seminar (2025 summer) and Co-founder and Lead Organizer for Cornell AI for Science seminar (2024 spring).

Community services

1. Journal Reviewer: Nature, Science Advances, Nature Computational Science, Nature Machine Intelligence, Nature Communications, JACS, JCTC, TPAMI, TMLR, npj Digital Medicine, Bioinformatics
2. Conference Area Chair: NeurIPS (2025-)
3. Conference Reviewer: NeurIPS (2022-2024), ICML (2022-), ICLR (2022-), AAAI (2023-2024), ECCV (2024), CVPR (2023), KDD (2023), ICCV (2023), AISTATS (2022)
4. Proposal Reviewer: ICML, NeurIPS workshops
5. Panel Moderator: "Is probabilistic inference relevant in the era of foundation models?" @SPIGM workshop NeurIPS 2025
6. Admission Committee: Cornell CS PhD (2024-2025)

Selected awards & achievements

1. NeurIPS Top Reviewer Award (both main track and dataset track, top 8%) *2024*
2. Microsoft Research Asia Star of Tomorrow Award (top 10%) *2021*
3. Distinguished Academic Achievement Award *2021*
4. NSF REU Fellowship *2019-2020*
5. Distinguished Undergraduate Research Award *2019-2020*
6. GMU OSCAR Fellowship *Summer 2019*
7. Outstanding Undergraduate Teaching Award *2018-2020*

Teaching experience

08/2024–12/2024 **Practicum in Artificial Intelligence**, Teaching Assistant

08/2018–12/2019 **Object-oriented Programming, Data Structures, Data Mining**, Teaching Assistant

02/2019–05/2019 **Research & Career & Course Student Advisor**, Peer Mentor

Full publication list

2025

1. **Y. Du**[*,†], B. Yu*, T. Liu*, et al.: *Accelerating Scientific Discovery with Autonomous Goal-evolving Agents.* arXiv preprint arXiv:2512.21782, 2025.
2. Z. Song*, J. Lu*, **Y. Du**[*], et al.: *Evaluating Large Language Models in Scientific Discovery.* arXiv preprint arXiv:2512.15567, 2025.
3. **Y. Du***, J. He*, F. Vargas, Y. Wang, C. P. Gomes, J. M. Hernández-Lobato, and E. Vanden-Eijnden: *FEAT: Free Energy Estimators with Adaptive Transport.* Advances in Neural Information Processing Systems, 2025.
4. **Y. Du**, J. He, E. Vanden-Eijnden, and C. Domingo-Enrich: *REACT: Rare Event Analysis via Stochastic Optimal Control.* arXiv soon, 2025.
5. C. Duan*, G.-H. Liu*, **Y. Du***, T. Chen, Q. Zhao, H. Jia, C. P. Gomes, E. A. Theodorou, and H. J. Kulik: *React-OT: Optimal Transport for Generating Transition State in Chemical Reactions.* Nature Machine Intelligence 7:4, pp. 615–626, 2025. **(Cover Article)**
6. J. He, ..., **Y.Du** [†], and F. Vargas[†]: *RNE: A Plug-and-play Framework for Diffusion Density Estimation and Inference-time Control.* arXiv preprint arXiv:2506.05668, 2025.
7. J. He, ..., **Y. Du**, S. Syed, and F. Vargas: *CREPE: Controlling diffusion with REPlica Exchange.* arXiv:2509.23265, 2025.
8. D. Blessing, ..., **Y. Du**, A. Vahdat, and G. Neumann: *Trust Region Constrained Measure Transport in Path Space for Stochastic Optimal Control and Inference.* NeurIPS, 2025. **(Spotlight)**

9. H. Wang, M. Skreta, C.-T. Ser, ..., **Y. Du**†, A. Aspuru-Guzik†, K. Neklyudov†, and C. Zhang†: *Efficient Evolutionary Search over Chemical Space with Large Language Models*. International Conference on Learning Representations, 2025.
10. H. Wang, J. Guo, L. Kong, R. Rampi, P. Schwaller, **Y. Du**†, and C. Zhang†: *LLM-Augmented Chemical Synthesis and Design Decision Programs*. Forty-second International Conference on Machine Learning, 2025.
11. J. He*, **Y. Du***, F. Vargas, D. Zhang, S. Padhy, R. OuYang, C. Gomes, and J. M. Hernández-Lobato: *No Trick, No Treat: Pursuits and Challenges Towards Simulation-free Training of Neural Samplers*. arXiv preprint arXiv:2502.06685, 2025.
12. L. Zhang, P. Potaptchik, J. He, **Y. Du**, and et al: *Accelerated Parallel Tempering via Neural Transports*. arXiv preprint arXiv:2502.06685, 2025.
13. **Y. Du***, L. Kong*, W. Mu*, K. Neklyudov, V. De Bortol, H. Wang, D. Wu, A. Ferber, Y.-A. Ma, C. P. Gomes, et al.: *Diffusion Models as Constrained Samplers for Optimization with Unknown Constraints*. Artificial Intelligence and Statistics, 2025.
14. J. Lu, Z. Song, Q. Zhao, **Y. Du**, Y. Cao, H. Jia, and C. Duan: *Generative Design of Functional Metal Complexes Utilizing the Internal Knowledge and Reasoning Capability of Large Language Models*. Journal of the American Chemical Society, 2025. DOI: 10.1021/jacs.5c02097 **(Cover Article)**
15. J. Gan, P. Zhong, **Y. Du**, Y. Zhu, C. Duan, H. Wang, C. P. Gomes, K. A. Persson, D. Schwalbe-Koda, and W. Wang: *Large Language Models Are Innate Crystal Structure Generators*. arXiv preprint arXiv:2502.20933, 2025.
16. B. Yin, J. Wang†, W. Du, P. Wang, P. Ying, H. Jia, Z. Zhang, **Y. Du**†, C. P. Gomes, C. Duan†, and H. Xiao†: *AlphaNet: Scaling Up Local Frame-based Atomistic Foundation Model*. npj Computational Materials, 2025.
17. X. Zhang, L. Wang, J. Helwig, Y. Luo, C. Fu, Y. Xie, M. Liu, Y. Lin, Z. Xu, K. Yan, et al.: *Artificial Intelligence for Science in Quantum, Atomistic, and Continuum Systems*. Foundations and Trends® in Machine Learning 18:4, pp. 385–912, 2025.
18. X. Chen, Y. Wang, J. He, **Y. Du**, S. Hassoun, X. Xu, and L. Liu: *Graph Generative Pre-trained Transformer*. Forty-second International Conference on Machine Learning, 2025.

2024 & earlier

19. **Y. Du***, M. Plainer*, R. Brekelmans*, C. Duan, F. Noe, C. P. Gomes, A. Aspuru-Guzik, and K. Neklyudov: *Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling*. Advances in Neural Information Processing Systems, 2024. **(Spotlight)**
20. A. Schneuing*, C. Harris*, **Y. Du***, A. Jamasb, I. Igashov, W. Du, T. Blundell, P. Lió, C. Gomes, M. Welling, et al.: *Structure-based Drug Design with Equivariant Diffusion Models*. Nature Computational Science 4:12, pp. 899–909, 2024.
21. **Y. Du***, A. R. Jamasb*, J. Guo*, T. Fu, C. Harris, Y. Wang, C. Duan, P. Liò, P. Schwaller, and T. L. Blundell: *Machine Learning-aided Generative Molecular Design*. Nature Machine Intelligence 6:6, pp. 589–604, 2024.
22. L. Kong*, H. Wang*, W. Mu*, **Y. Du**, Y. Zhuang, Y. Zhou, Y. Song, R. Zhang, K. Wang, and C. Zhang: *Aligning Large Language Models with Representation Editing: A Control Perspective*. Advances in Neural Information Processing Systems, 2024.
23. G. Wei, Y. Huang, C. Duan, Y. Song†, and **Y. Du**†: *Traversing Chemical Space with Latent Potential Flows*. Advances in Neural Information Processing Systems, 2024.
24. Y. Li, L. Kong, **Y. Du**, Y. Yu, Y. Zhuang, W. Mu, and C. Zhang: *MUBen: Benchmarking the Uncertainty of Molecular Representation Models*. Transactions on Machine Learning Research, 2024.
25. H. Wang*, T. Fu*, **Y. Du***, W. Gao, et al.: *Scientific Discovery in the Age of Artificial Intelligence*. Nature 620:7972, pp. 47–60, 2023.
26. W. Du*, **Y. Du***, L. Wang*, D. Feng, G. Wang, S. Ji, C. P. Gomes, Z.-M. Ma, et al.: *A New Perspective on Building Efficient and Expressive 3D Equivariant Graph Neural Networks*. Advances in Neural Information Processing Systems, 2023.

27. L. Holdijk*, **Y. Du***, F. Hooft, P. Jaini, B. Ensing, and M. Welling: *Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths.* Advances in Neural Information Processing Systems, 2023.

28. C. Duan, **Y. Du**, H. Jia, and H. J. Kulik: *Accurate Transition State Generation with an Object-aware Equivariant Elementary Reaction Diffusion Model.* Nature Computational Science 3:12, pp. 1045–1055, 2023. **(Cover Article)**

29. **Y. Du***, Y. Wang*, Y. Huang, J. C. Li, Y. Zhu, T. Xie, C. Duan, J. Gregoire, and C. P. Gomes: *M^2 Hub: Unlocking the Potential of Machine Learning for Materials Discovery.* Advances in Neural Information Processing Systems 36, pp. 77359–77378, 2023.

30. **Y. Du**, X. Liu, N. M. Shah, S. Liu, J. Zhang, and B. Zhou: *ChemSpacE: Interpretable and Interactive Chemical Space Exploration.* Transactions on Machine Learning Research, 2023.

31. W. Du*, H. Zhang*, T. Yang*, and **Y. Du***: *A Flexible Diffusion Model.* International Conference on Machine Learning, 2023.

32. D. Chen, Y. Zhu, J. Zhang, **Y. Du**, Z. Li, Q. Liu, S. Wu, and L. Wang: *Uncovering neural scaling laws in molecular representation learning.* NeurIPS, 2023.

33. W. Du, H. Zhang, **Y. Du**, Q. Meng, W. Chen, N. Zheng, B. Shao, and T.-Y. Liu: *SE (3) Equivariant Graph Neural Networks with Complete Local Frames.* International Conference on Machine Learning, 2022.

34. **Y. Du***, X. Guo*, H. Cao, Y. Ye, and L. Zhao: *Disentangled Spatiotemporal Graph Generative Models.* Proceedings of the AAAI Conference on Artificial Intelligence, 2022.

35. X. Guo*, **Y. Du***, and L. Zhao: *Deep Generative Models for Spatial Networks.* Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.