Marianne Arriola

New York, NY | marriola@cs.cornell.edu | m-arriola.com

Interests: Deep Learning, Generative AI for text, Discrete Diffusion

Education

Ph.D. student in Computer Science, Cornell University — Sep 2023–Jun 2027
Advisor: Volodymyr Kuleshov

B.S., Computer Science, University of California, Santa Barbara — Aug 2019–Jun 2023

Current Research

Diffusion Language Models
Committee: Volodymyr Kuleshov (Chair), Mohamed Abdelfattah, Fei Wang

- I explore diffusion LMs which enable faster & more controllable generation with better benchmark performance for math & planning compared to traditional LLMs
- I design novel parameterizations and architectures to improve model quality [1,2,3,4], training efficiency [1,3,4], and inference speed [1,3]

Publications

Selected Papers

[1] Marianne Arriola, Aaron Gokalsan, Justin Chiu, Zhihan Yang, Zhixuan Qi, Jiaqi Han, Subham Sahoo, Volodymyr Kuleshov. "Block Diffusion: Interpolating between Autoregressive and Diffusion Language Models." *ICLR 2025*. **Oral presentation (Top 1.77%)**.

[2] Subham Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin Chiu, Alexander Rush, Volodymyr Kuleshov "Simple and Effective Masked Diffusion Language Models." *NeurIPS 2024*.

---

[3] Marianne Arriola*, Yair Schiff*, Hao Phung, Aaron Gokaslan, Volodymyr Kuleshov "Encoder-Decoder Block Diffusion Language Models for Efficient Training and Inference." *NeurIPS 2025*.

[4] Marianne Arriola, Naveen Venkat, Jon Granskog, Anastasis Germanidis. "Adapting Autoregressive Vision Language Models for Parallel Diffusion Decoding." *Runway Research Blog.*

[5] Marianne Arriola, Weishen Pan, Manqi Zhou, Qiannan Zhang, Chang Su, Fei Wang. "Joint Analysis of Single-Cell Data across Cohorts with Missing Modalities." *arXiv* (Feb 2024).

[6] Marianne Arriola & Kadina Johnston "Identifying Optimal Proteins by Their Structure Using Graph Neural Networks." *Caltech URJ* (Jun 2022).

Employment

Runway, Research Intern. New York, NY — Jun 2025–Sept 2025
MIT CSAIL, Research Intern. Cambridge, MA — Jun 2022–Nov 2022
Caltech, Research Intern. Caltech, Pasadena, CA — Jun 2021–Aug 2021

Open-Source Contributions

Block Diffusion Language Models (GitHub)

- Led development of a diffusion LM with arbitrary-length generation and KV caching
- 850 stars as of Oct 2025

Projects

MIT CSAIL, Cambridge, MA — Jun 2022 – Nov 2022
PI: Justin Solomon

- Designed a memory-efficient representation for complex point clouds (e.g., from LiDAR) which summarizes spatial data using geometric primitives
- Developed a graph neural network for hybrid point clouds that matches state-of-the-art segmentation performance while using 50% less memory

Caltech, Pasadena, CA — Jun 2021 – Aug 2021
PI: Frances Arnold

- Proposed a data-driven method to iteratively refine existing protein structures toward desired properties (e.g., substrate specificity)
- Built a graph neural network to predict protein functional capacity using graph-based structural representations

Patents

Xiangru Huang, Marianne Arriola, Yue Wang, Vitor Campagnolo Guizilini, Rares Andrei Ambrus, Justin Solomon. "Hybrid Geometric Primitive Representations for Point Clouds." U.S. patent pending.

Selected Talks

International Conference on Learning Representations (ICLR) — Apr 2025
Amazon Artificial General Intelligence (AGI) — Apr 2025

Awards

Top Reviewer, NeurIPS 2025 — Oct 2025
NSF Graduate Research Fellowship — Mar 2023–Jun 2028
Bowers CIS Dean's Excellence Fellowship, Cornell University — Mar 2023–Jun 2029

Activities

MIT Summer Research Program Application Reviewer — Jan 2024–Feb 2024

Programming Skills

Python, PyTorch, C++, MATLAB, Bash