Milleno Pan

millenopan@gmail.com · (516) 331-0640 · millenopan.com · Google Scholar · github.com/millenopan

EDUCATION

2023 – 2025 **Cornell University** New York, NY
M.S. in Computational Biology, Weill Cornell Medicine · GPA: 4.0/4.0
Thesis: *Causal Inference on Gene Expression Analysis* · Advisor: Wesley Tansey

2019 – 2022 **University of California, Berkeley** Berkeley, CA
B.A. in Computer Science and Mathematics

PUBLICATIONS

2026 **Empirically Calibrated Conditional Independence Tests** arXiv · Code
Milleno Pan, Antoine de Mathelin, Wesley Tansey
Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 300:5194–5202, 2026

RESEARCH EXPERIENCE

2024 – Present **Memorial Sloan Kettering Cancer Center** New York, NY
Research Associate, Tansey Lab · Advisor: Wesley Tansey

2026 – Present **Hierarchical Discrete Diffusion** · Mentor: Christopher Tosh

- Developing continuous diffusion models for discrete data that generate coarse-to-fine through tree- and DAG-structured hierarchies, building on Dirichlet flow matching and Fisher flow.
- Evaluating on image (CIFAR-10) and text data; manuscript targeting ICML 2027.

2025 – 2026 **Empirically Calibrated Conditional Independence Tests (ECCIT)**

- Identified two ways conditional independence tests lose error control in practice: noise misestimation in small samples, and skewed null p-values under model misspecification.
- Proposed ECCIT, a test-agnostic wrapper that trains an adversary (a Gumbel-softmax feature mask and response function) on bootstrap resamples to maximally miscalibrate a base test, then fits a monotone map to recalibrate its p-values.
- Proved that when the true response lies in the adversary class, calibrating against the worst-case FDP adversary controls FDR under Benjamini–Hochberg.
- Benchmarked GCM and HRT on synthetic and semi-synthetic GDSC gene expression data, achieving valid FDR with higher power than prior calibration methods like CONTRA.

2024 – 2025 **Causal Inference on Gene Expression Analysis**

- Applied conditional independence testing to estimate the causal effect of gene expression on drug response across cancer types.
- Implemented permutation- and residual-based tests, balancing Type I error against power, and built tests over graph-based subsets of genes and samples.

2023 **Stanford University** Stanford, CA

Research Associate, Altman Lab · Advisor: Russ Altman

- Studied whether social media language can forecast emerging drug epidemics at fine geographic resolution (NIH R21-funded).
- Ran topic modeling and sentiment analysis on Reddit and Twitter posts with DLATK and built reporting pipelines for risk tracking.

2021 – 2022 **University of California, Berkeley** Berkeley, CA

Undergraduate Researcher · Advisor: Marti Hearst · Mentor: Peitong Duan

- Applied reinforcement learning with Monte Carlo Tree Search to group mobile UI elements into semantic hierarchies.
- Combined MCTS with HCI layout-tree methods, outperforming earlier LSTM/RNN models.

INDUSTRY EXPERIENCE

2022 – 2023 **Google** Mountain View, CA

Software Engineer, YouTube Ads

- Built TensorFlow/Keras models to detect viewer presence on YouTube TV.
- Designed metrics and experiments to tune ad-break length in extended ad pods.

2021 **Amazon** Seattle, WA

SDE Intern, Supply Chain Optimization Technologies

- Trained a weighted KNN/XGBoost ensemble to predict order cutoff times (89% within one hour).
- Built a precomputation library that cut expected latency by about 20% on 30% of requests.