

Education

- 2023–2025 : **M.Sc., Electrical Engineering**, Washington University in St. Louis, STL, US.
Generative AI, Posterior Inference, Probabilistic Sampling (GPA: 4.0/4.0)
- 2024–2025 : **M.Sc., Computer Engineering**, Washington University in St. Louis, STL, US.
Deep Learning, Machine Learning, Parallel Computing (GPA: 3.83/4.0)
- 2018–2022 : **B. E., Electronic & Information Engineering**, Xidian University, Xi'an, CN.
Deep Learning, Computer Vision, Video Processing (GPA: 3.4/4.0), 2nd Class Scholarship (top 10%)

Skills

ML	Machine Learning & Deep Learning, Diffusion/Flow Model , Adversarial AI, Computer Vision
Mathematics	Monte Carlo Method, Mathematical Optimization, Stochastic Process
Programming	Python, PyTorch, Tensorflow, PySpark, C, OpenMP
System	System Architecture, Parallel Computing , Docker Container

Research Experience

Washington University in St. Louis

- Sep.2025 – **Probabilistic Inference in AI.**
present Sampling and variational inference for SDEs/ODEs provide a principled framework for solving measure transport problems. Our research leverages these techniques on probabilistic generative models (Diffusion/Flow/Consistency Models), including: Sampling, Optimal Control, MCTS, and Applied Stochastic Processes. See our papers for more details. ([Optimal Control, VI and SMC](#))([Tree Search Scheme](#))
- Advisor 1: **Dr. Yevgeniy Vorobeychik**, Professor, Department of Computer Science & Engineering, Washington University in St. Louis ([Personal Web-page](#))
- Advisor 2: **Dr. Jens Sjölund**, Assistant Professor of Machine Learning · WASP Fellow · ELLIS Member, Uppsala University ([Personal Web-page](#))

Washington University in St. Louis

- Sep.2024 – **Diffusion Model Based Sequential Decision Making.**
Aug.2025 Strategic sampling within a limited sampling budget from unobserved regions is essential in various scientific and engineering domains. We model this problem as Active Target Discovery(ATD) and introduce novel frameworks that leverage Diffusion Models with Tweedie's estimate and online reward learning to solve ATD problems. Additionally, when the prior Diffusion Model is not good enough, we use Doob's h-transform to efficiently fine-tune our diffusion process, enabling expectation maximization nature. See our NeurIPS 2025 papers for more details.([White Box Decision Making](#))([EM via Doob's h-transform](#))
- Advisor : **Dr. Yevgeniy Vorobeychik**, Professor, Department of Computer Science & Engineering, Washington University in St. Louis ([Personal Web-page](#))

Feb.2024 – **Accelerating GNN Inference on Multi-Core Systems.**

present Developing GNN inference computing kernel on CPUs. Our method achieves up to $3.35\times$ and $3.20\times$ speedups over DGL for GCN and GraphSAGE respectively, and even greater improvements over PyG, across a range of large-scale, high-dimensional graph datasets with diverse structural properties. See our IPDPS 2025 Workshop paper for more details on the first version of our GCN model implementation on multi-core CPU: ([Fast GNN Inference](#)). A following work optimizing both sparse and dense (GEMM) kernels of GCN, GraphSage, Graph Isomorphism Networks (GIN) and Simple Graph Convolution (SGC) is in submission.

Advisor : **Dr. Roger Chamberlain**, Professor, Department of Computer Science & Engineering, Washington University in St. Louis ([Personal Web-page](#))

Publications

In Conference Proceedings

* Equal contribution

2025 Anindya Sarkar*, **Binglin Ji***, and Yevgeniy Vorobeychik. "Active target discovery under uninformative priors: The power of permanent and transient memory." *In Advances in Neural Information Processing Systems (NeurIPS), 2025.*

2025 Anindya Sarkar*, **Binglin Ji***, and Yevgeniy Vorobeychik. "Online feedback efficient active target discovery in partially observable environments." *In Advances in Neural Information Processing Systems (NeurIPS), 2025.*

2025 **Binglin Ji**, Chenfeng Zhao, and Roger D. Chamberlain. "FGI: Fast GNN Inference on Multi-Core Systems." *In IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). 2025.*

In Submission

2026 **Binglin Ji***, Anindya Sarkar*, Hengchang Lu, and Jens Sjölund, Yevgeniy Vorobeychik "Sequentially-Controlled Interactive Multi-Particle Flow-Maps for Online Feedback-Driven Search." *arXiv preprint arXiv:2607.01144*

2026 **Binglin Ji***, Anindya Sarkar*, Hengchang Lu, and Jens Sjölund, Yevgeniy Vorobeychik "Bootstrap Flow-Map Tree Sampling Enables Online Feedback Driven Search." *arXiv preprint arXiv:2607.02915*

Work Experience

Lenovo Rresearch

Jun.2022 – **Computer Vision & MLOps.**

Jan.2023 Developing Computer Vision models to solve industrial problems. Developing a Docker-based resource management system to manage datasets, algorithm & code, model parameter versions, runtime environments, and historical training/testing records.
(Shanghai, CN)

Director : **Dr. Haoyu Xu**, Director, Industry Solution, Lenovo Research ([Personal Web-page](#))

National Laboratory of Pattern Recognition, Chinese Academy of Sciences

Oct.2020 – **Deep-Learning Based Video Segmentation.**

Oct.2021 Training and developing efficient algorithm pipelines (Computer Vision models and unsupervised learning methods) to segment videos.
(Beijing, CN)

Advisor : **Dr. Guan Luo**, Associate Professor, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences ([Personal Web-page](#))

Position of Responsibility

2026 **Reviewer of NeurIPS 2026.**

