PHILLIP SI

408-480-7009 • psi6@gatech.edu • Atlanta, GA • https://ps789.github.io/

PROFESSIONAL SUMMARY

Ph.D. candidate in Computer Science at Georgia Tech researching probabilistic machine learning, generative data assimilation, and differentiable surrogate models for physical systems. Develops efficient, uncertainty-aware digital twins that combine sparse observations, dynamical forecasting, and optimal control. Publications at NeurIPS, ICLR, and ICML; research experience at Oak Ridge and Argonne National Laboratories.

EDUCATION

Georgia Institute of Technology Anticipated 05/28
Ph.D in Computer Science, GPA 4.00
Advised by Peng Chen and Chao Zhang

Carnegie Mellon University 05/23
M.S. in Machine Learning. QPA 4.00 (Coursework, transitioned to Ph.D at Georgia Tech)

Cornell University 05/22
B.A. in Computer Science (Magna Cum Laude)
& Mathematics GPA: 3.91/4.00

SELECTED PUBLICATIONS

- **Si, Phillip**; Chen, P. "LEVDA: Latent Ensemble Variational Data Assimilation via Differentiable Dynamics". **NeurIPS 2026**.
- **Si, Phillip**; Qiu, Y; Sallam, O.; Feinstein, J.; He, Z.; Yan, E.; Chen, P. **"**Toward AI-driven digital twins for metropolitan floods: A conditional latent dynamics network surrogate of the shallow water equations". **Journal of Hydrology, November 2026.**
- Tian, Z.; Shen, K; Hao, W; **Si, Phillip**; Chen, P; Zhu, W. "LD-EnFF: Latent-Dynamics Ensemble Flow Filtering for Data Assimilation with Sparse Observations". **arXiv** Preprint, 2026.
- Tong, Y.; **Si, Phillip**; Qiu, Y; Chen, P. "C-STRIDE: An Observation-Driven AI Digital Twin for Predicting Basin-Wide Flood Fields from Sparse Stream-Gauge Histories". **arXiv** Preprint, 2026.

- Xiao, P.; **Si, Phillip**; Chen, P. "LD-EnSF: Latent Dynamics with Ensemble Score Filters for Fast Data Assimilation with Sparse Observations". **ICLR 2026.**
- **Si, Phillip**; Chen, P. "Latent-EnSF: A Latent Ensemble Score Filter for High-Dimensional Data Assimilation with Sparse Observation Data". **ICLR 2025**.
- **Si, Phillip**; Chen, Z.; Sahoo, S.; Schiff, Y.; Kuleshov, V. "Semi-Autoregressive Energy Flows: Exploring Likelihood-Free Training of Normalizing Flows ". **ICML 2023**.
- Choi, S.; Mousavi, S.; **Si, Phillip**; Yhdego, H. G.; Khadem, F.; Afghah, F. "ECGBERT: Understanding Hidden Language of ECGs with Self-Supervised Representation Learning". **arXiv** Preprint, 2023.
- **Si, Phillip**; Bishop, A.; Kuleshov, V. "Autoregressive Quantile Flows for Predictive Uncertainty Estimation". **ICLR 2022 (Spotlight).**

RESEARCH EXPERIENCE

Graduate Researcher, Oak Ridge National Laboratory 05/26 – Present

- Developing an AI digital twin for unsteady Navier–Stokes flow that integrates sparse-sensor state and unknown-gust estimation with latent-dynamics forecasting and model predictive control (MPC), in collaboration with Guannan Zhang.

- Built a derivative-supervised neural surrogate and LD-EnSF assimilation–control loop; achieved 27.8× lower wake-fluctuation cost than no control and ~0.59 s per online update in simulated closed-loop tests (ongoing manuscript).

Research Aide, Argonne National Laboratory 06/25 – 05/26

- Led development of CLDNet, a rainfall-driven latent neural ODE with terrain-conditioned, coordinate-based decoding for high-resolution flood prediction over the 4.2-million-cell Des Plaines River basin.

- Enabled 96-hour basin-wide forecasts ~115× faster than the numerical solver; validated the flood simulator against USGS measurements and integrated LD-EnSF updates from sparse stream gauges for a flood digital twin.

Graduate Researcher, Georgia Institute of Technology 08/23 – Present

- Developed Latent-EnSF, combining coupled variational autoencoders and ensemble score filtering for 225x spatially reduced observations; achieved 0.198-second assimilation steps on ERA5, versus 0.501 seconds for full-state EnSF.

- Helped develop LD-EnSF, combining latent dynamics and history-aware observation encoding; reduced tsunami forecast propagation time by ~4,000x versus the full physics solver while retaining accurate sparse-observation filtering.

- Developed LEVDA, a differentiable latent ensemble-variational smoother for joint state–parameter estimation from sparse, irregular observations, without physical-model adjoints or observation-to-latent encoders.

- Contributed to LD-EnFF, achieving 5.74%, 6.53%, and 4.42% final-state relative RMSE on Kolmogorov, tsunami, and atmospheric benchmarks, respectively; also contributed to C-STRIDE, reducing flood-depth errors ~40% versus gauge-history-only predictions using six gauges and terrain/rainfall inputs.

Undergraduate Researcher, Cornell University 09/20 – 05/23

- Studied calibrated predictive uncertainty with Professor Volodymyr Kuleshov, developing probabilistic models for aleatoric and epistemic uncertainty.
- Developed Autoregressive Quantile Flows, using quantile-based objectives to train expressive conditional distributions without likelihood Jacobian evaluation.
- Extended generative flow models to semi-autoregressive energy flows trained with sample-based objectives, improving flexible distribution modeling.

Natural Language Processing Research Intern at Cornell Tech 05/19 – 08/19

- Conducted NLP research under Professor Yoav Artzi at Language in Context Lab
- Developed system for using collaborative interactions to study second language acquisition in goal-specified environment with Unity and Python
- Used NewsRoom dataset to generate articles with multiple summary labels for parameter based summary generation with various Python architectures

SERVICE

- **Annual Conference on Neural Information Processing Systems (NeurIPS)**
  ***Top Reviewer (8%)* — 2026**
  Reviewer — 2023, 2025
- **International Conference on Learning Representations (ICLR)**
  Reviewer — 2024, 2025, 2026
- **International Conference on Machine Learning (ICML)**
  Reviewer — 2025, 2026

INDUSTRY EXPERIENCE

Machine Learning Research Intern, CardioPhi — Milpitas, CA 03/23 – 08/23

- Trained Transformer/ResNet models on high-frequency ECG streams for arrhythmia detection and prediction.
- Contributed to ECGBERT representation learning research.
- Partnered with product to translate model outputs into actionable UI signals.

Software Engineering Intern, Adobe — San Jose, CA 05/21 – 08/21

- Optimized a generative search algorithm by incorporating native speedups using C++ and lower-level code as part of an innovation team developing cutting-edge and currently realizable technology.

Software Engineering Intern, Adobe — San Jose, CA 05/20 – 08/20

- Built a Python ML routing pipeline; improved accuracy from ~30% to ~90% on curated validation.
- Deployed internal tool through an online docker container and Amazon Lambda(Serverless); then integrated with slack bot as a slash command.

HONORS AND AWARDS

- **Top Reviewer,** NeurIPS 2026
- **President's Fellow**, Georgia Institute of Technology
- **Tanner Dean's Scholar**, Cornell University
- **Dean's List**, Cornell University

LEADERSHIP

Intelligent Systems Team Lead at Cornell Data Science - Ithaca, NY 09/19 – 05/22

- Led a 22-member Intelligent Systems team at Cornell Data Science.
- Work on combining Bayesian Networks with Kernel Density Estimators as a nonparametric statistical modeling method to reduce the effects of high- dimensional data
- Using Reinforcement Learning to develop a Pokerbot agent to achieve strategy optimization under uncertainty

TECHNICAL SKILLS

Programming & Tools: Python, Java, R, PyTorch, NumPy, SciPy, Git, Docker, Linux, AWS Lambda
Machine Learning: Generative modeling, normalizing flows, diffusion/score-based models, Bayesian inference, uncertainty quantification, reinforcement learning
Scientific Computing: Data assimilation, neural surrogates, neural ODEs, PDE simulation, differentiable optimization, model predictive control, computational fluid dynamics