FANGZHU SHEN

fangzhu.shen@duke.edu https://fangzhushen.github.io
github.com/fangzhushen linkedin.com/in/fangzhu-shen-shelly Google Scholar

EDUCATION

Duke University Durham, NC, USA
Ph.D. in Computer Science *January 2023 - Present*
- Advisor: Prof. Sudeepa Roy
- GPA: 3.98/4.0
- Expected graduation: Spring 2027

Duke University Durham, NC, USA
M.S. in Economics and Computation *August 2021 - December 2022*
- Transitioned to the Ph.D. program

Central University of Finance and Economics Beijing, China
B.A. in Finance *September 2016 - July 2020*

PUBLICATIONS AND PREPRINTS

[1] **Whiel: A Neurosymbolic Platform for Certified Relational Program Verification**
Jesse Comer*, **Fangzhu Shen***, Shuo Zhang*, Mayur Naik, Sudeepa Roy, and Val Tannen.
Preprint, 2026. [pdf]

[2] **Selectivity Estimation for Linear Queries via Online Learning**
Fangzhu Shen, Debmalya Panigrahi, and Sudeepa Roy.
Preprint, 2026. [pdf]

[3] **Efficient Shapley-Based Influence Attribution in Social Networks**
Fangzhu Shen, Amir Gilad, and Sudeepa Roy.
KDD 2026. [pdf] [code]

[4] **The Cost of Representation by Subset Repairs**
Yuxi Liu*, **Fangzhu Shen***, Kushagra Ghosh, Amir Gilad, Benny Kimelfeld, and Sudeepa Roy.
PVLDB 2024; presented at VLDB 2025. [pdf] [code]

[5] **Causal What-If and How-To Analysis Using HypeR**
Fangzhu Shen*, Kayvon Heravi*, Oscar Gomez, Sainyam Galhotra, Amir Gilad, Sudeepa Roy, and Babak Salimi.
ICDE 2023, Demonstration Track. [pdf]

RESEARCH EXPERIENCE

Whiel: LLM Agents for Relational Program Verification *(Preprint)*
- Developed a neurosymbolic platform that combines an LLM agent with formal verification to reason about imperative relational programs, encoding typically undecidable database problems, such as equivalence and containment of recursive queries and query rewriting using views, as Hoare validity problems.
- Built the agent harness around an iterative propose–verify–refine loop using Lean 4/Vampire verifier feedback and reusable skills; the platform solved 90% of 86 tasks versus 23% for direct Lean proving with the same model.

Online Learning for Selectivity Estimation *(Preprint)*
- Recast selectivity estimation as an online-learning problem in which both the database and query workload may change dynamically, using historical query feedback to support selectivity prediction.
- Proposed optimal algorithms and established tight regret bounds for point, range, subset, and linear queries over static and dynamic databases, including a polynomial separation between squared and absolute loss.

Shapley-Based Influence Attribution *(KDD 2026)*

- Introduced a Shapley-value framework for ex-ante influence attribution that assigns credit to seed users before diffusion occurs, accounting for overlapping influence to support influencer valuation and campaign budget planning.
- Identified a sharp tractability boundary: polynomial-time exact algorithms for single-step activation but #P-hardness beyond one step; developed approximation algorithms with provable guarantees and validated them on large-scale networks.

Representation-Aware Data Repair *(PVLDB 2024; VLDB 2025)*

- Developed a framework for data cleaning that resolves data inconsistencies while preserving specified subgroup representation, supporting data preparation for downstream tasks.
- Developed polynomial-time algorithms for tractable cases and efficient heuristics for general instances to maximize data retention under functional dependencies and representation constraints; demonstrated effectiveness on real-world datasets.

Causal What-If and How-To Analysis with HypeR *(ICDE 2023 Demo)*

- Built and presented an interactive system that lets users express causal what-if and how-to queries through a SQL-like interface and explore results through visual explanations.

INDUSTRY EXPERIENCE

PhD Software Engineer Intern, Uber *Summers 2025 and 2026*

Mobility Simulation & Planning, Pricing Model Research

Summer 2025 Project: Dynamic Structural Estimation *New York, NY, USA*

- Developed the end-to-end framework combining machine learning forecasting and economics-based structural pricing model; integrated marketplace prediction with pricing simulation; the prototype evolved into Uber's production model.
- Designed residual multi-task neural architectures for high-variance marketplace parameters, reducing one-week-ahead prediction error by 11% across 106 cities.
- Implemented structural equilibrium equations as differentiable PyTorch components, enabling end-to-end learning through the pricing model, and built a reproducible data and training pipeline over 42,000+ records.

Summer 2026 Project: Temporal Dependency Modeling *Sunnyvale, CA, USA*

- Extended the joint-modeling framework developed in Summer 2025 to capture temporal dependencies in marketplace forecasts; diagnosed why standard temporal attention amplified co-movement and distorted the observed dependency structure.
- Designed a baseline-conditioned neural decoder that separates market-wide level shifts from hour-specific reallocations, reducing temporal correlation error by 27% and weighted forecast error by 21% in a 12-city, 26-week backtest.

Software Engineering Intern, Carl Zeiss Meditec *Summer 2022*

AI Solution & Service Team *Raleigh, NC, USA*

- Built feature-engineering pipelines over Azure Cloud SQL data and evaluated tree-based, Bayesian, and neural models for file-conversion-time prediction.
- Deployed the selected model as a real-time FastAPI service for integration with production systems.

SKILLS

- Programming: Python, SQL, C, Rust
- ML/AI: PyTorch, Scikit-learn, agent harness, vLLM
- Formal Methods and Tools: Lean 4, Vampire
- Data and Development: PostgreSQL, MySQL, MongoDB, Docker, Git, Linux

TEACHING EXPERIENCE

- **Teaching Assistant**, Introduction to Database Systems, Duke University *Fall 2025*
- **Teaching Assistant**, Causal Inference in Data Analysis *Spring 2025*
- **Teaching Assistant**, Introduction to Databases, Duke University *Fall 2024*
- **Teaching Assistant**, Introduction to Database Systems, Duke University *Fall 2023*

SERVICE

- Reviewer, VLDB 2026 Demonstration Track
- Shadow PC Member, VLDB 2026

AWARDS

- VLDB 2025 Travel Award
- ICDE 2023 Travel Award
- Duke Scholar Award (25% Tuition Waiver) *August 2021 - December 2022*
- Academic Scholarships, Central University of Finance and Economics *2017, 2018, 2019*