

# Aishwarya H. Balwani

aishwarya.balwani@stjude.org | aishwaryahb.github.io | (+1) 470-439-4942

## Education & Training

---

### St. Jude Children's Research Hospital

Postdoctoral Research Associate, Department of Developmental Neurobiology

Memphis, TN

2025 – Present

### Georgia Institute of Technology

PhD, Electrical & Computer Engineering

Atlanta, GA

2018 – 2025

- Minors: Mathematics, Computer Science
- Thesis: *Through the Recurrent Neural Network Looking Glass: Structure-Function Relationships in Cortical Circuits for Predictive Coding*

MS, Electrical & Computer Engineering

2016 – 2018

### University of Mumbai

BE, Electronics & Telecommunication (First Class with Distinction)

Mumbai, India

2012 – 2016

## Research & Work Experience

---

### Postdoctoral Research Associate

Fall 2025 – Present

St. Jude Children's Research Hospital

- Elucidating neuronal circuit mechanisms underlying hallucinations in mouse models of schizophrenia using RNN models and mechanistic interpretability. (Advisor: Dr. Stanislav Zakharenko)

### AI Strategy Consultant, Frontier Tech

Fall 2025 – Present

Scale AI

- Developing benchmarks and applying Chain-of-Thought interpretability to evaluate and improve reasoning and agentic capabilities of frontier LLMs. (Host: Chaithanya Bandi)

### Technical Advisor

February 2026 – Present

SmaLLM

- Advising on ML algorithms, fine-tuning strategies, model architectures, and statistical techniques.

### ML Researcher & Research Mentor

Summer 2025 – Present

Algoverse AI Research

- Mentoring high-achieving college students in applied ML and LLM research targeting workshops at top ML/AI conferences.

### Graduate Research Assistant

Summer 2018 – Spring 2025

Georgia Institute of Technology

- Studied predictive coding, incorporating biological constraints in RNNs, representational geometry, and architectural biases in the cortex. (Advisor: Dr. Hannah Choi)
- Studied multi-scale models of brain structure and organization, representation learning. (Advisor: Dr. Eva Dyer)

### Forecasting Mentee

Summer 2023

Epoch FRI AI Mentorship Program

- Developed Bayesian frameworks to quantify how antecedent questions predict beliefs on complex topics. (Mentor: Molly Hickman)

### Winter Project Intern

December 2022 – January 2023

Good Futures Initiative, EA Berkeley

- Developed a mathematical framework using representational geometry to detect and address AI misalignment.

### Summer Research Associate

Summer 2022

Center for Computational Neuroscience, Flatiron Institute, Simons Foundation

- Developed a three-factor Hebbian learning rule for non-negative recurrent networks and analyzed auditory cortex representations in mice. (Supervisor: Dr. SueYeon Chung)

### R&D Intern, Algorithms Team

Summer 2017

Intellifusion, China

- Worked on algorithms for image processing, data compression, and encryption.

## Publications, Preprints & Peer Reviewed Abstracts

---

### Publications

- Balwani A.**, Wang A. Y., Najafi F., Choi H. “Constructing Biologically Constrained RNNs via Dale’s Backpropagation and Topologically-Informed Pruning.” *Science Advances*, 2025.
- Sharma M., Zhang C., Bandi C., Wang C., ... **Balwani A.**, ... & Liu B. “ResearchRubrics: A Benchmark of Prompts and Rubrics For Evaluating Deep Research Agents.” *arXiv*, 2025 (Accepted and to be presented at *ICLR*, 2026).
- Balwani A.**, Cho S., & Choi H. “On the Architectural Biases of the Canonical Cortical Microcircuit.” *Neural Computation*, 2025.
- Ozan Bozdogan G., Zamani-Dahaj S.A., Kahn P., Day T., Tong K., **Balwani A.**, Dyer E., Yunker P., & Ratcliff W. “*De Novo* Evolution of Macroscopic Multicellularity.” *Nature*, 2023.
- Balwani A.**, & Krzyston J. “Zeroth-order Topological Insights into Magnitude-based Neural Network Pruning.” *PMLR Volume on Topology, Algebra, and Geometry in Learning*, 2022.
- Balwani A.\***, Miano J.\*, Liu R., Kitchell L., Prasad J., Johnson E., Gray-Roncal W., & Dyer E. “Multi-Scale Modeling of Neural Structure in X-ray Imagery.” *IEEE International Conference on Image Processing (ICIP)*, 2021.
- Prasad J., **Balwani A.**, Johnson E., Miano J., Sampathkumar V., De Andrade V., ... & Dyer E. “A three-dimensional thalamocortical dataset for characterizing brain heterogeneity.” *Nature Scientific Data*, 2020.
- Liu R., Subakan C., **Balwani A.**, Whitesell J., Harris J., Koyejo S., & Dyer E. “A generative modeling approach for interpreting population-level variability in brain structure.” *MICCAI*, 2020.
- Balwani A.**, & Dyer E. “Modeling variability in brain architecture with deep feature learning.” *2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE*, 2019.
- Milligan K., **Balwani A.**, & Dyer E. “Brain Mapping at High Resolutions: Challenges and Opportunities.” *Current Opinion in Biomedical Engineering*, 2019.
- Lee T., Kumar A., **Balwani A.**, Brittain D., Kinn S., Tovey C., Dyer E., da Costa N., Reid R., Forest C., & Bumbarger D. “Large-scale neuroanatomy using LASSO: Loop based Automated Serial Sectioning Operation.” *PLoS One*, 13.10, 2018.

### Workshop Papers & Peer Reviewed Abstracts

- Balwani A.** “Time-Resolved Circuit Discovery in RNNs via Windowed Causal Interventions and Local Linearization.” (Poster), *Symmetry and Geometry in Neural Representations (NeurReps)*, *NeurIPS*, 2025. Also accepted as “Mechanistic Interpretability for Time-Resolved Neural Circuit Discovery in RNNs.” (Poster), *COSYNE*, 2026.
- Rahman A., Gurugubelli A., Ankit O., Zhu K., & **Balwani A.** “Probing the Origins of Reasoning Performance: Representational Quality for Mathematical Problem-Solving in RL vs SFT Finetuned Models.” (Poster), *XAI4Science*, *AAAI*, 2026.
- Arturi D., Zhang E., Ansah A., Zhu K., Panda A., & **Balwani A.** “Shared Parameter Subspaces and Cross-Task Linearity in Emergently Misaligned Behavior.” (Honorable Mention – Top 4% and Oral – Top 7% of submissions), *UniReps Workshop*, *NeurIPS*, 2025. Also presented as a Spotlight Talk at *Mechanistic Interpretability Workshop*, *NeurIPS*, 2025.
- Durai A., Hu J., Buch K., Zhu K., Sharma V., & **Balwani A.** “LoRA-Guided PPO for Cost-Aware and Compute-Efficient Agent Orchestration.” *Efficient Reasoning Workshop*, *NeurIPS*, 2025.
- Balwani A.**, Wang A., Najafi F., & Choi H. “Constructing Biologically-Constrained RNNs via Dale’s Backprop and Topologically-Informed Pruning.” (Poster), *COSYNE*, 2025.
- Zhou W., **Balwani A.**, Chung S., & Schneider D. “Motor-sensory Experience Reshapes Neural Manifolds in Auditory Cortex to Reflect Acoustic Expectations.” *Advances and Perspectives in Auditory Neuroscience (APAN)*, 2023.
- Balwani A.**, & Choi H. “On the Architectural Biases of the Canonical Cortical Microcircuit.” (Talk, Top 3.2% of submissions), *COSYNE*, 2023.
- Cho S., **Balwani A.**, & Choi H. “Leveraging Predictive Coding to Improve Artificial Neural Network Performance.” (Poster), *Collaborative Research in Computational Neuroscience (CRCNS)*, 2022.
- Balwani A.**, & Krzyston J. “Zeroth-order Topological Insights into Magnitude-based Neural Network Pruning.” (Spotlight, Top 9.8% of submissions), *Topology, Algebra, and Geometry in Machine Learning*, *ICML*, 2022. Also presented as a poster at *Sparsity in Neural Networks*, 2022.
- Balwani A.**, & Dyer E. “Modeling Brain Microarchitecture with Deep Representation Learning.” (Poster), *ML Interpretability for Scientific Discovery*, *ICML*, 2020.
- Balwani A.**, Miano J., Prasad J., & Dyer E. “Learning to Segment at Multiple Scales.” (Poster), *BioImage Informatics*, 2019.
- Milligan K., **Balwani A.**, Maguire A., Margulies S., & Dyer E. “Deep Learning for Characterization of Neuroinflammation in Traumatic Brain Injury.” (Poster), *BioImage Informatics*, 2019.

## Preprints

- Amarnath C., **Balwani A.**, Ma K., & Chatterjee A. “TESDA: Transform Enabled Statistical Detection of Attacks in Deep Neural Networks.” *arXiv*, 2021.
- Balwani A.**, & Dyer E. “A Deep Feature Learning Approach for Mapping the Brain’s Microarchitecture and Organization.” *bioRxiv*, 2020.

## Honours & Awards

---

### Grants

- Open Philanthropy, Career Development and Transition Funding, 2025.

### Academic Awards & Fellowships

- ECE Coulter MS Fellowship, Georgia Institute of Technology, 2016–2017.

### Registration & Travel Awards

- Conference Registration Award, NeurIPS, 2025 (Sponsored by New Theory).
- COSYNE Presenters Travel Award, 2023.
- ICML Diversity and Inclusion Fellowship, 2020.

### Competitions & Hackathons

- Best Poster – Frontiers in Science Conference and Symposium (Intelligence), Georgia Tech, 2025.
- Winner (Technical Track) – Hacklytics, Data Science at Georgia Tech, 2019.
- Winner (Best Project) – AI/ML for Social Good Hackathon at Georgia Tech, 2018.
- Gold Award – IEEE UBTech-Education Robotics Design Challenge, 2017.

## Teaching & Mentoring

---

### Algoverse AI Research

- Mentored ~8 groups (>30 mentees) spanning high schoolers to post-PhD professionals, producing 5 accepted and 5 under-review workshop papers at NeurIPS, ICLR, ICML, and AAAI in mechanistic interpretability, representation learning, AI alignment/safety, and efficient learning.

### Teaching Assistant

- Linear Algebra, Georgia Tech (Spring 2024)
- AI Safety Fundamentals, Georgia Tech (Facilitator, AI Safety Institute) (2023)
- Professional and Technical Communications for ECE, Georgia Tech (Summer 2021)
- Data Analytics for Engineers, Georgia Tech (Fall 2019, 2018)
- Hands-On Tech Day Camp, Georgia Tech (June 2019)
- Deep Learning for Microscopy Image Analysis, Marine Biological Laboratory (May 2019)
- Mathematical Foundations for Data Science, Georgia Tech (Spring 2018)
- Embedded Systems & IoT, Eduvance (Summer 2016)

## Professional Service

---

### Reviewing

- Journals: Nature Scientific Reports, PLoS One, Distill
- Conferences: AISTATS, MIDL, CoLLAs
- Workshops: Workshop on Geometrical and Topological Representation Learning, Topological Data Analysis and Beyond, Lifelong Learning Workshop, Workshop on Continual Learning in Computer Vision, Workshop on Continual Semi-Supervised Learning
- Other: Neuromatch Academy 2020, President’s Undergraduate Research Awards (PURA), Georgia Tech

### Professional & Student Organizations

- Senator (ECE), Graduate Student Association, Georgia Institute of Technology, 2017–2018

## Workshops & Seminars

---

### Attendee

- Define, Design, and Align, AI Safety @ UCLA (January 2023)
- AI Safety Workshop, Berkeley (December 2022)
- London Geometry and Machine Learning Summer School (July 2021)
- Banach Center – Oberwolfach Graduate Seminar: Mathematics of Deep Learning, Institute of Mathematics, Polish Academy of Sciences (November 2019)
- Foundation of Data Science Summer School, Georgia Institute of Technology (August 2019)
- Spinning Up in RL Workshop, OpenAI (February 2019)