Tarun Raheja

📱 +1-445-444-0118 | @ tarunraheja1234@gmail.com | 🔗 LinkedIn | ⚙ GitHub | ◉ Google Scholar

EDUCATION

University of Pennsylvania   Philadelphia, USA
Master of Science in Scientific Computing; **GPA: 4.0/4.0**   *Aug. 2022 – Aug. 2024*
Thesis Advisor *: Prof Andrea Liu,* **Topic** *: Learning in Physical Systems*

Birla Institute of Technology and Science, Pilani   Hyderabad, India
Bachelor of Engineering (Hons.) in Computer Science; **GPA: 8.57/10**   *Aug. 2015 – Aug. 2019*
Thesis advisor *: Prof. Sreejith GJ,* **Topic** *: DMRG for Potts' Model*   *Graduated with First Degree Honours*

PUBLICATIONS

1. Lambert, Neill* and **Raheja, Tarun*** and Cross, Simon and Menczel, Paul and Ahmed, Shahnawaz and Pitchford, Alexander and Burgarth, Daniel and Nori, Franco : QuTiP-BoFiN: A bosonic and fermionic numerical hierarchical-equations-of-motion library with applications in light-harvesting, quantum control, and single-molecule electronics. *Phys. Rev. Res.* 5, 013181 (2023).

2. **Raheja, Tarun*** and Pochhi, Nilay*: An initial exploration of using Persistent Homology for Noise-Resilient Tactile Object Recognition. In *NeurIPS 2024 Workshop on Touch Processing: From Data to Knowledge.* Oral Presentation (2024).

3. **Raheja, Tarun*** and Sinha, Raunak and Deepak, Advit and Healy, Will and Srinivasa, Jayanth and Lee, Myungjin and Kompella, Ramana : Enhancing Large Language Models through Transforming Reasoning Problems into Classification Tasks. In *Proceedings of the 2024 ACL Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 6007–6016. ELRA and ICCL, Torino, Italia (2024).

4. Haputhanthri, Udith and Storan, Liam and Jiang, Yiqi and **Raheja, Tarun** and Shai, Adam and Akengin, Orhun and Miolane, Nina and Schnitzer, Mark J and Dinc, Fatih and Tanaka, Hidenori : Understanding and controlling the geometry of memory organization in RNNs *(Under review)*

5. **Raheja, Tarun*** and Pochhi, Nilay* : The Alignment Trilemma: A Theoretical Perspective on Recursive Misalignment and Human-AI Adaptation Dynamics. In *ICLR 2025 Workshop on Bidirectional Human-AI Alignment.* (2025).

6. **Raheja, Tarun*** and Pochhi, Nilay* : On the Limitations of Neural Networks for Option Pricing: Analysis of Volatility Regime Sensitivity. In *ICLR 2025: I Can't Believe It's Not Better: Challenges in Applied Deep Learning.* (2024).

7. **Raheja, Tarun*** and Pochhi, Nilay* : DAGnosis: A Benchmark for Complex DAG-Based Multi-Step Reasoning in Large Language Models. *(Under review)*

8. **Raheja, Tarun*** and Pochhi, Nilay*: Foundation Models Meet Continual Learning: Recent Advances, Challenges, and Future Directions. In *NeurIPS 2024 Workshop on Scalable Continual Learning for Lifelong Foundation Models.* Accepted (Poster) (2024).

* Denotes first authorship

PATENTS

1. **Automatic retrieval augmented generation with expanding context.**
   A. Deepak, W. Healy, **T. Raheja**, J. Srinivasa, R. R. V. R. Kompella, R. Sinha.
   US Patent App. 18/442,996 (2025).

2. **Cache-Generated Frequently Asked Questions Page.**
   T. Raheja, R. Sinha, W. Healy, J. Srinivasa, A. Deepak, R. R. V. R. Kompella, et al.
   US Patent App. 18/439,888 (2025).

3. **Memory-based function calling for logical formulation and inference for large language models.**
   T. Raheja, A. Deepak, W. Healy, J. Srinivasa, G. Liu, R. R. V. R. Kompella, et al.
   US Patent App. 18/439,966 (2025).

4. **Performance monitoring, mitigation, and retraining of large language models.**
   A. Deepak, R. Sinha, W. Healy, **T. Raheja**, J. Srinivasa, R. R. V. R. Kompella.
   US Patent App. 18/440,710 (2025).

RELEVANT EXPERIENCE

Kipo AI   San Francisco, USA
Founding Machine Learning Engineer   *May. 2024 – Present*
- Built and owned ML and MLOps infra as the **first ML hire**, and Employee No. **2**.
- Architected and implemented a state-of-the-art scalable production **NLP** and **Vision** pipeline for understanding electronics datasheets. Outperforming competitors in **speed**, **cost**, and **data granularity**.
- Architected and implemented a state-of-the-art scalable production **agentic AI-search** pipeline for electronics components. This arbitrarily-fine-grained, domain-aware, **multi-modal**, **spec-accurate deep-research** is novel to the electronics industry .
- Architected and implemented tooling for **MLOps, DataOps**, and **training infra** for in-house vision and NLP models for several in-house tasks, specifically **YOLOv10** and **BERT** instances.
- Critically helped the company secure **2 major contracts**, directly worked on demos to clients, attended several sales **trade shows**. Currently mentoring **2 interns**. Helped grow the team to 10 people, directly onboarded **2 engineers**.

Cisco Research   San Jose, USA
*Machine Learning Research Intern (with **Jayanth Srinivasa**)*   *May. 2023 – Aug. 2023*
- Led project on a **novel reasoning algorithm** for LLMs and an **LLM Gateway**. Resulted in **four patent submissions**.

Linguistic Data Consortium   Philadelphia, USA
*NLP Research Projects Programmer (with **Prof. Mark Liberman**)*   *Sep. 2022 – Jan. 2024*
- Worked on **forced alignment** and **generalized information representation schemes** for robust NLP models.

UPenn Department of Physics   Philadelphia, USA
*Graduate Research Assistant (with **Prof. Andrea Liu**, **Dr. Nachi Stern**)*   *Jan. 2023 – May 2024*
- Worked on algorithms for **learning in physical systems**. Uncovered novel **physical symmetry breaking** phenomena.

J.P. Morgan & Chase   Mumbai, India
*Quantitative Researcher (**QR Rates team**)*   *Aug. 2021 – Aug. 2022*
- Worked on developing mathematical pricing and volatility models for linear and nonlinear **rates derivatives** products.
- Worked on an OPA framework for **hedging residual Greeks**. Published docs for the **IR Future risk model**.
- Migrated internal **model usage restrictions**. Worked on two **high impact internal projects** reporting directly to MDs.

humit.app   Bangalore, India
Founding Backend Engineer   *Feb. 2021 – Aug. 2021*
- Built the social music sharing service, Humit. Designed and deployed several **core features** end-to-end including a **recommendation system**, and **async growth reporting services**.

RIKEN, Theoretical Quantum Physics Laboratory   Wako-shi, Japan
*Research Assistant (with **Prof. Franco Nori, Dr. Neill Lambert**)* *Oct. 2019 – Nov. 2020*
- Developed an open source library to simulate **open quantum systems** using the **HEOM** formalism. Modelled the environment as a mixture of **fermionic and bosonic baths**. Studied **exceptional points** of large non-hermitian systems using a **novel block diagonalization** algorithm. Optimized **Noisy Mean Field Annealing** for NP hard problems using ideas from **quantum thermodynamics**.
- Worked on several research problems in **quantum information**, **open quantum systems**, and **combinatorial optimization**. Also worked on the **Coherent Ising Machine** project, and collaborated with **IBM Research Tokyo**.
- Worked on the admin team to **QuTiP**, an open-source package with **250k+ DAUs**.

tensoralpha GmbH   Berlin, Germany
*Quantitative Developer Intern (with **Dr. Thomas Schmelzer**)*   *Jul. 2020 – Aug. 2020*
- Worked on statistical analysis of proprietary Crypto derivatives trading strategies. Researched a novel **tracker strategy**.

Microsoft Research   Bangalore, India
*Research Intern (with **Dr. Karthik Ramachandra**)*   *Jul. 2019 – Sep. 2019*
- Worked on **Project Froid**, on scalar **UDF inlining** for MSSQL. Improved if-else parsing from **exponential to linear**.

HONORS & AWARDS

Exa AI Discovery Agents Hackathon 2025 : Won first place among 50+ teams.

Sentient Foundation Werewolf AGIThon (AGI House) 2024: Won first place in agent strategy/jailbreaking hackathon.

CS50 Puzzle Day 2020 (Harvard x Facebook) : Scored 8/8. Global winner.

WorldQuant Challenge : Achieved Gold ranking for high Sharpe, market-neutral, medium frequency alphas.

JEE Mains & Advanced : Ranked among top 0.3% nationwide, among 1.7 million candidates in the engineering entrance test.

Regional Mathematical Olympiad : Highest scorer in Bathinda region in India, for the first stage to the IMO.

Problem Solving Assessment Examination, CBSE : 99.8 percentilex in national aptitude test, out of 1.5 million candidates.

Academic Service/Reviewing

ACM WebConf HCRS 2026
ICLR 2026
ICLR 2026 I Can't Believe it's Not Better
ICLR 2025
NeurIPS 2025 Efficient Reasoning Workshop
NeurIPS 2025 Frontiers of Probabilistic Inference Workshop
NeurIPS 2024 RedTeaming GenAI Workshop
NeurIPS 2024 EvalEval Workshop
NeurIPS 2024 Workshop on Touch Processing

Skills

Programming : C, C++, Java, Python, MATLAB, R, SQL, LaTeX, AWS, CI/CD
Technologies : Git, BASH, Django, Pandas, NumPy, PyTorch, TensorFlow, Spark, sklearn, Docker, Airflow
ML/SWE: Deep Learning, Computer Vision, Infrastructure, Data Structures, Algorithms, Scalable Design, Data Science