

# NEELESH GUPTA

 neelesh-gupta23 |  neeleshg23.github.io |  neeleshg@usc.edu |  +1 (832) 591 8299

## EDUCATION

---

**University of Southern California**

*Ph.D. Computer Engineering*

**Los Angeles, CA**

*August 2025 - December 2028*

**University of Southern California**

*B.S. Computer Science, Minor: Mathematics*

**Los Angeles, CA**

*August 2021 - May 2025*

## SKILLS

---

Programming Languages:

Python, C++, Verilog, Java

Machine Learning:

PyTorch, HuggingFace, Llama.cpp, Triton, NetworkX

Parallel Computing & Hardware:

CUDA, OpenMP, MPI, SYCL, Vivado, ModelSim

Tools & Platforms:

Linux, Bash, Git, LaTeX, Jupyter Notebooks

## WORK EXPERIENCE

---

**Research Assistant**

**Los Angeles, CA**

*USC Ming Hsieh Department of ECE - Data Science Lab*

*January 2023 - Present*

- Develop novel techniques for accelerating deep learning models on specialized hardware accelerators for compute-intensive applications under the supervision of Prof. Viktor Prasanna and Prof. Raj Kannan.
- Mentor exchange students on research projects related to high-performance computing and hardware acceleration, providing guidance on problem formulation, algorithm design, and experimental setup.

**Undergraduate Research Assistant**

**Marina del Rey, CA**

*USC Information Sciences Institute - STEEL: Security Research Lab*

*May 2022 - January 2023*

- Developed scripts to automate the PCA algorithm for detecting energy grid outages in synthetic data.
- Performed sentiment analysis on the Enron corpus to train a machine learning model to classify spam.

## PROJECTS

---

**Machine Learning for Data Prefetching**

*Jan. 2023 - Present*

- Develop novel computer architecture approaches to tackle post-Moore's Law memory scaling using ML to better L3 cache utilization compared to traditional rule-based prefetchers.
- Implement the first instance of a practical neural network-based prefetcher with inference latency comparable to rule-based prefetchers by distilling and approximating Transformers using tables.

**Efficient Table-Based Neural Network**

*Jan. 2023 - Present*

- Designed primitives for matrix multiplication, convolutions, and multi-headed self-attention that are approximated using a product quantization strategy and can be efficiently looked up in a table.
- Reduced arithmetic operations during inference by over 90% for compute-intensive operations.

## SERVICE

---

**Publicity Chair**

**Hyderabad, India**

*IEEE 32nd International Conference on High Performance Computing, Data, and Analytics (HiPC)*

## PUBLICATIONS

---

- [1] Pengmiao Zhang, **Neelesh Gupta**, Rajgopal Kannan, and Viktor K. Prasanna. “Net2Tab: Neural Network to Tables for Data Prefetching”. In: *Journal of Parallel and Distributed Computing (JPDC)* (2025). Manuscript under review.
- [2] Rakshith Jayanth, **Neelesh Gupta**, and Viktor K. Prasanna. “Benchmarking Edge AI Platforms for High-Performance ML Inference”. In: *IEEE High Performance Efficient Computing (HPEC)*. IEEE. Boston, MA, Sept. 2024.
- [3] **Neelesh Gupta**, Narayanan Kannan, Pengmiao Zhang, and Viktor Prasanna. “TabConv: Low-Computation CNN Inference via Table Lookups”. In: *Proceedings of the 21st ACM International Conference on Computing Frontiers*. CF ’24. Ischia, Italy: Association for Computing Machinery, 2024. ISBN: 9798400705977. DOI: [10.1145/3649153.3649212](https://doi.org/10.1145/3649153.3649212).
- [4] Pengmiao Zhang, **Neelesh Gupta**, Rajgopal Kannan, and Viktor K. Prasanna. “Attention, Distillation, and Tabularization: Towards Practical Neural Network-Based Prefetching”. In: *2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. San Francisco, CA, USA, May 2024, pp. 1–10. arXiv: [2401.06362](https://arxiv.org/abs/2401.06362) [[cs.NE](#)].
- [5] **Neelesh Gupta**, Pengmiao Zhang, Rajgopal Kannan, and Viktor K. Prasanna. “PaCKD: Pattern-Clustered Knowledge Distillation for Compressing Memory Access Prediction Models”. In: *IEEE High Performance Efficient Computing (HPEC)*. IEEE. Boston, MA, Sept. 2023.
- [6] Jeffrey Liu, Rajat Tandon, Uma Durairaj, Jiani Guo, Spencer Zahabizadeh, Sanjana Ilango, Jeremy Tang, **Neelesh Gupta**, Zoe Zhou, and Jelena Mirkovic. “Did your child get disturbed by an inappropriate advertisement on YouTube?” In: *Proceedings of KDD Undergraduate Consortium*. KDD-UC ’22. ACM. Washington, D.C., Oct. 2022.