

# VIGNESH KOTHAPALLI

+1-212-961-7356 | vk2115@nyu.edu | github.com/kvignesh1420 | in/kvignesh1420

## EDUCATION

- **New York University, Courant Institute of Mathematical Sciences** Sep 2021 - May 2023  
Master of Science in Computer Science, Advisor: Prof. Joan Bruna  
CGPA: 3.97/4.0
- **Indian Institute of Technology Guwahati** Jul 2014 – May 2018  
B.Tech in Electronics and Communication Engineering  
CGPA: 8.22/10.0

## EXPERIENCE

- **LinkedIn**, Mountain View, CA July 2023 – Present  
*Software Engineer - Machine Learning*
  - Core member of the *360Brew* foundation model team, working on post-training of the 140 billion parameter Large Language Model (LLM) for recommendation and ranking tasks across LinkedIn.
  - Core member of the *Liger-Kernel* project, developing triton kernels to reduce LLM training latencies and GPU memory consumption by over 60%. The library (3K+ stars on github) has been integrated with HuggingFace, PyTorch Lightning and is being adopted across industry and academia for training LLMs.
  - Analyzing the spectral properties of feature covariance matrices and weights for semi-structured N:M pruning of LLMs.
- **Department of CS, Dartmouth**, Remote: Mountain View, CA Nov 2023 – Present  
*Research Collaborator. Advisor: Prof. Yaoqing Yang*
  - Analyzing *feature-learning* in shallow neural networks and its implications on the weight matrix spectrum during training.
  - Our work presents the precise scale of learning rates for Adam to learn single-index models in high-dimension settings.
- **Math and Data Group, NYU**, New York, NY Sep 2022 – May 2023  
*Graduate Researcher. Advisor(s): Prof. Joan Bruna, Prof. Jonathan Weare*
  - Extended the unconstrained features model for analyzing the “Neural Collapse” phenomenon in Graph Neural Networks.
  - Developed randomized Schur-complement-based graph augmentation techniques for contrastive learning.
- **LinkedIn**, Mountain View, CA May 2022 – Aug 2022  
*Summer Intern*
  - Developed low-rank gradient compression techniques for reducing the pre-training duration of BERT models by 20%.
  - Implemented batched memory copy for tensor fusion in allgather and added support for int8 allreduce in Horovod.
- **IBM CIO Labs**, Bangalore, India Jul 2018 – Aug 2021  
*Software Developer*
  - Developed a dependency graph framework to facilitate root-cause analysis of events in distributed data platforms.
  - Employed MLOps techniques to train and serve auto-encoder models in production for detecting anomalies in Kafka, Solr, and HDFS telemetry data. The framework reduced the Mean Time to Respond (MTTR) by 80%.

## PUBLICATIONS

- **Liger Kernel: Efficient Triton Kernels for LLM Training**  
Pin-Lun Hsu, Yun Dai, Vignesh Kothapalli, Qingquan Song, Shao Tang, Siyu Zhu, Steven Shimizu, Shivam Sahni, Haowen Ning, Yanning Chen  
<https://arxiv.org/abs/2410.10989> Under Review
- **Crafting Heavy-Tails in Weight Matrix Spectrum without Gradient Noise**  
Vignesh Kothapalli, Tianyu Pang, Shenyang Deng, Zongmin Liu, Yaoqing Yang  
<https://arxiv.org/abs/2406.04657> Under Review
- **Kernel vs. Kernel: Exploring How the Data Structure Affects Neural Collapse**  
Vignesh Kothapalli, Tom Tirer  
Workshop on Symmetry and Geometry in Neural Representations (NeurIPS) Dec 2024
- **A Neural Collapse Perspective on Feature Evolution in Graph Neural Networks**  
Vignesh Kothapalli, Tom Tirer, Joan Bruna  
Advances in Neural Information Processing Systems (NeurIPS) Dec 2023
- **Randomized Schur Complement Views for Graph Contrastive Learning**  
Vignesh Kothapalli  
International Conference on Machine Learning (ICML) Jul 2023
- **Neural Collapse: A Review on Modelling Principles and Generalization**  
Vignesh Kothapalli  
Transactions on Machine Learning Research (TMLR) Apr 2023
- **Edge detection using fractional derivatives and information sets**  
Vignesh Kothapalli, Shaveta Arora, Madasu Hanmandlu  
Journal of Electronic Imaging Jun 2018

- **Binary Document Image Super Resolution for Improved Readability and OCR Performance**  
Ram Krishna Pandey, Vignesh Kothapalli, AG Ramakrishnan, B Chandrasasa  
<https://arxiv.org/abs/1812.02475> Dec 2018
- **Robust Recognition of Tone Specified Mizo Digits Using CNN-LSTM and Nonlinear Spectral Resolution**  
Vignesh Kothapalli, Biswajit Dev Sarma, Abhishek Dey, Parismita Gogoi, Wendy Lalhminghlui, Priyankoo Sarmah, SR Mahadeva Prasanna, SR Nirmala, Rohit Sinha  
IEEE INDICON Dec 2018
- **Abnormal Event Detection on BMTT-PETS 2017 Surveillance Challenge**  
Vignesh Kothapalli, Gaurav Yadav, Amit Sethi  
IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops Jul 2017

---

#### TALKS

- **Learning on Graphs (LoG) Meetup, Stanford, 2023:** *A Neural Collapse Perspective on Feature Evolution in GNNs.*

---

#### TECHNICAL SKILLS

- **Languages:** C, C++, Python.
- **Machine Learning Technologies:** Horovod, Tensorflow, Keras, PyTorch, PyTorch geometric, DGL, Scikit-learn, Triton.
- **Tools/Frameworks:** Docker, Flask, MySQL, MongoDB, Git, Kafka, Spark, Impala, Airflow, Streamsets MLFlow, Travis.

---

#### PROFESSIONAL SERVICE

- **Open Source Contributions:** Horovod, Tensorflow, Tensorflow-IO (*maintainer*), Liger-Kernel (*committer*).
- **Reviewer:** IEEE Transactions on Cybernetics, IEEE Access, IEEE Transactions on Industrial Informatics, Transactions on Machine Learning Research (TMLR), Neural Information Processing Systems (NeurIPS).

---

#### HONORS AND AWARDS

- NeurIPS Top Reviewer - 2024.
- Google Open Source Peer Bonus Award (TensorFlow) - 2021.
- IBM Managers Choice Award 2018-19.
- Merit-based scholarship from Govt of Telangana, India from 2015-2017.