

# DEEPTHI KARKADA

Email: karkada.deepthi@gmail.com

Mobile: (+1)980-267-0085

## EXPERIENCE

---

### Intel Corporation

*Deep Learning Software Engineer*

*Hillsboro, OR, USA*

*Feb 2016 - Oct 2022*

### Intel Corporation, Big Data pathfinding group

*Pathfinding Intern*

*Folsom, CA, USA*

*June 2015 - Aug 2015*

### Infosys

*Systems Engineer*

*Hyderabad, TS, India*

*Sep 2011 - Dec 2013*

## SELECT PROJECTS

---

### Natural Language processing (NLP) and Conversational AI

- Designed and developed low precision optimization techniques for machine learning modalities such as computer vision and natural language processing (NLP). This includes optimizations for popular neural network models such as ResUnet, Transformer, and BERT-based models for semantic segmentation, semantic similarity analysis, language modeling, question answering, text and token classification, and speech recognition tasks using the Intel Neural Network Compression Framework.
- Designed and developed several components for a conversational tutoring system's natural language understanding module. Developed tools for annotations and analysis of the collected corpus and knowledge graph-based natural language understanding (NLU) module.
- Integrated the NLU for developing a reinforcement learning-based dialogue policy.
- Used self-supervised learning and uncertainty estimation techniques to improve NLU performance.
- Implemented an efficient 8-bit post training quantization methodology for the Transformer model for low precision inference. The quantized model showed a 50% improvement in inference time with less than 0.5% relative drop in accuracy.
- Integrated Horovod distributed training framework bindings into the Mozilla DeepSpeech Speech-To-Text engine. Implemented a novel data partitioning scheme to speedup recurrent neural networks (RNN) for automatic speech recognition (ASR) training.

### Deep Learning Framework Optimizations

- Optimized Deep learning frameworks TensorFlow and MXNet frameworks on Intel<sup>®</sup> Architecture. These optimizations include integrating Intel<sup>®</sup> Math Kernel Library bindings into core kernels.

## AWARDS AND PATENTS

---

- **Intel Achievement Award (IAA)**, the highest achievement award at Intel corporation.
- Offered **travel grant** for attendance at the Workshop for Young Female Researchers in Speech Science & Technology at Interspeech, 2018.
- **3 Divisional Recognition awards** for performance at Intel corporation.
- **Joint inventor on four patents and patent applications** filed with the US Patent Office.
  - Synchronization scheduler of distributed neural network training.
  - Automated resource usage configurations for deep learning neural network workloads on multi-generational computing architectures.
  - Methods and apparatus for distributed training of a neural network (Under review).
  - Workload Behavior Modeling and Prediction for Data Center Adaptation (Under review).

- S.Pati, U.Baid, B.Edwards, M.Sheller, S.Wang, G. A.Reina, P.Foley, A.Gruzdev, **D.Karkada**, et al. “Federated Learning Enables Big Data for Rare Cancer Boundary Detection”. **Nature communications**, 2022.
- **D.Karkada**, R.Manuvinakurike, M.Paetzel-Prussman, K.Georgila “Strategy-level Entrainment in Dialogue Systems for a Creative Visual Reference Resolution Task”. Language Resources and Evaluation Conference (**LREC**), 2022.
- B. Baheti, S.Thakur, S.Pati, **D.Karkada**, et.al “Optimization of Deep Learning based Glioblastoma detection for Low Resource Environments”. Society for Neuro Oncology (**SNO**), 2022.
- S.Thakur, S.Pati, R.Panchumarthy, **D.Karkada**, J.Wu, D.Kurtaev, C.Sako, P.Shah, S.Bakas, “Optimization of Deep Learning based Brain Extraction in MRI for Low Resource Environments”. BrainLes workshop at MICCAI (**MICCAI workshop**), 2021.
- V.Tassopoulou, N.Kasmanoff, V.Soni, F.Khoo, K.Ward, **D.Karkada**, M.Ramasubramanian, R.Ramachandran, F.Soboczenski, P.Biliski, “Generating informative and accurate descriptions of natural hazards and phenomena using large transformer-based models”. American Geophysical Union (**AGU**) Fall Meeting, 2021. (Abstract)
- A.Bergqvist, R.Manuvinakurike, **D.Karkada**, M.Paetzel, “Nontrivial Lexical Convergence in a Geography Themed Game”. Special Interest Group on Discourse and Dialogue (**SIGDIAL**), 2020.
- M.Paetzel, **D.Karkada**, R.Manuvinakurike, “RDG-Map: A Multimodal Corpus of Pedagogical Human-Wizard Spoken Interactions”. Language Resources and Evaluation Conference (**LREC**), 2020.
- **D.Karkada**, R.Manuvinakurike, M.Paetzel, K.Georgila “Knowledge Graph Based Natural Language Understanding in a Rapid Spoken Dialogue Map-Game”. Natural Language, Dialog and Speech Symposium, New York academy of Sciences (**NYAS, NDS**), 2020. (Abstract)
- A.Bhandare, V.Sripathi, **D.Karkada**, V.Menon, S.Choi, K.Datta and V.A.Saletore, “Efficient 8-bit quantization of Transformer model for Language Translation”. Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations, International Conference on Machine Learning (**ICML workshop**), 2019.
- **D.Karkada**, “Visual Reference Resolution for a Novel Task”. Workshop for Young Female Researchers in Speech Science & Technology, (**Interspeech workshop**), 2018. (Abstract)
- **D.Karkada**, R.Manuvinakurike, and K.Georgila, “Towards Understanding End-of-trip Instructions in a Taxi Ride Scenario”. Fourteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation, International Conference on Computational Linguistics (**COLING workshop**), 2018.
- **D.Karkada**, V.A.Saletore, “Training speech recognition models on HPC Infrastructure”. The 4th Workshop on Machine Learning in HPC Environments, The International Conference for High Performance Computing, Networking, Storage, and Analysis (**IHPC workshop**), 2018.
- E.Ould-Ahmed-Vall, M.Abuzaina, M.Amin, J.Bobba, R.Dubtsov, E.Fomenko, M.Gangadhar, N.Hasabnis, J.Huang, **D.Karkada**, Y.J.Kim, S.Makineni, D.Mishura, K.Raman, AG.Ramesh, V.Rane, M.Riera, D.Sergeev, V.Sripathi, B.Subramanian, L.Tokas, A.Valles, “Accelerating TensorFlow on Modern Intel<sup>®</sup> Architectures”. The First International Workshop on Architectures for Intelligent Machines (**AIM**), 2017.

## MENTORSHIP

---

### NASA Frontier development Lab project on Automated reporting of natural events

- Mentored a group of researchers for a summer project on [automated reporting of natural events using natural language processing \(NLP\) techniques](#).
- This project used natural language generation techniques using large transformer-based language models such as GPT-2 for narratives generation using metadata on natural events. [Project link](#)

## TECHNICAL REPORTS

---

- **Knowledge graphs inference optimizations using the Intel OpenVINO toolkit:** ([medium blog](#))
- **Image classification for large high-content microscopy images:** (<https://www.intel.ai/teaching-machines-image-classification-health-life-sciences-intel-xeon-scalable-processors-lab-coats/>)
- **TensorFlow\* Optimizations on Modern Intel Architecture:** (<https://software.intel.com/en-us/articles/tensorflow-optimizations-on-modern-intel-architecture>)
- **Boosting Deep Learning Training Inference Performance on Intel Xeon and Intel Xeon Phi Processors:** ([boosting-deep-learning-training-inference-performance-on-xeon-and-xeon-phi](#))
- **AI parallel processing using MPI and TensorFlow:** ()

## EDUCATION

---

<b>University of North Carolina at Charlotte</b>		<i>Charlotte, NC, USA</i>
MS in Electrical Engineering	<i>Jan 2014 - Dec 2015</i>	Overall GPA: 3.54
<b>Visvesvaraya Technological University</b>		<i>Bangalore, KA, India</i>
BE in Electronics and Computer Engineering	<i>Sep 2007 - Jul 2011</i>	Overall Percentage: 73.54

## TECHNICAL STRENGTHS

---

<b>CS Skills</b>	Data Structures, Algorithms, System Design
<b>Languages</b>	Python, HTML, CSS, C, PHP, $\LaTeX$
<b>Toolkits and libraries</b>	Huggingface transformers, OpenVINO, OpenMP, MPI, Apache Spark MLlib, Hadoop MapReduce, Tensorboard
<b>Deep Learning frameworks</b>	TensorFlow, PyTorch