haidark@gmail.com
(845) 633-0903

HAIDAR KHAN

linkedin.com/in/haidark
haidark.github.io

EXPERIENCE

National Center for AI

08/2023 - Present Research Scientist **Riyadh, Saudi Arabia**

Building a national foundation model (ALLaM) for the Kingdom of Saudi Arabia serving the government and private sectors with capabilities in Arabic and English.

- Created a training stack capable of training models at 49% MFU based on MegatronLM. Lead a team of 3 scientists and 6 engineers.
- Pretrained and aligned our models on trillions of tokens of English and Arabic data.
- ALLaM outperformed all open and closed models on automated and human evaluations in Arabic

Amazon Alexa

10/2021 – 08/2023 Senior Applied Scientist **New York, NY**

Large scale training of large (1B – 100B parameter) language models on web-scale datasets as part of the Alexa Teacher Model (AlexaTM) program.

- Developed infrastructure to scale training of large language models using DeepSpeed.
- Compressed large language models for natural language understanding, automatic speech recognition rescoring, and semantic parsing.
- Combined AlexaTM with visual understanding and image generation models to create new multimodal Alexa experiences.

Amazon Alexa

08/2019 – 10/2021 Applied Scientist **New York, NY**

Natural language understanding (NLU) research for virtual assistants including language modeling, semantic parsing, and intent/entity recognition.

- Deployed efficient transformer-based models for Alexa NLU that satisfy production latency constraints (<10ms inference).
- Lead a team of 4 scientists and engineers to speed up sequence-to-sequence semantic parsing systems by 3x with parallel decoding.

Rensselaer Polytechnic Institute

09/2014 – 05/2019 Research Assistant **Troy, NY**

Epileptic seizure prediction using machine learning (collab with The Mount Sinai Epilepsy Center).

- Discovered a novel patient-independent pre-seizure state by applying change point detection combined with deep learning to 500+ hours of electroencephalogram (EEG) containing 200+ focal onset seizures.

Siemens Corporate Technology

05/2018 – 08/2018 Research Intern **Princeton, NJ**

Modeling agents and adversaries in a power plant network with reinforcement learning.

- Increased the possible number of modeled agents by a factor of 2 with available hardware.

IBM T.J. Watson Research Center

05/2016 – 08/2017 Research Intern **Yorktown Heights, NY**

Empirically studied the minibatch size/convergence rate tradeoff for deep neural network training.

- Designed a variant of parallel SGD and analyzed performance on benchmark datasets and networks
- The algorithm implemented on an IBM HPC cluster reduced total training time from 14 to 4 days.

EDUCATION

09/2014 – 05/2019 Rensselaer Polytechnic Institute Troy, NY

- Ph. D Computer Science

Dissertation: *Predicting Change Points in Multivariate Time Series Data*

- M.S. Computer Science

09/2009 – 05/2013 SUNY New Paltz New Paltz, NY

- B.S. in Computer Engineering

SELECTED PUBLICATIONS

N Alzahrani, et al. "When Benchmarks are Targets: Revealing the Sensitivity of Large Language Model Leaderboards" *ACL 2024*

M Ozdayi, et al. "Controlling the Extraction of Memorized Data from Large Language Models via Prompt-Tuning" *ACL 2023*

S Rongali, M Harakere, H Khan, K Arkoudas, W Hamza, A McCallum "Low-Resource Compositional Semantic Parsing with Concept Pretraining." *EACL 2023*

S Soltan, et al. "AlexaTM 20B: Few-shot learning using a large-scale multilingual seq2seq model." *ArXiv 2022*

M Kumar, Y Merhav, H Khan, R Gupta, A Rumshisky, W Hamza "Controlled data generation via insertion operations for NLU." *NAACL HLT Industry Track 2022*

L Xu, Y Gu, J Kolehmainen, H Khan, A Gandhe, A Rastrow, A Stolcke, I Bulyko "RescoreBERT: Discriminative speech recognition rescoring with BERT." *ICASSP 2022*

W Sun, H Khan, N Mesnards, M Rubino, K Arkoudas "Unfreeze with Care: Space-efficient fine-tuning of semantic parsing models." *The Web Conference 2022*

B Kleiner, J FitzGerald, H Khan, G Tur "Mixture of domain experts for language understanding: An analysis of modularity, task performance, and memory tradeoffs." *SLT 2022*

N Ström, H Khan, W Hamza "Squashed weight distribution for low bit quantization of deep models." *INTERSPEECH 2022*

J FitzGerald, et al. "Alexa Teacher Model: Pretraining and distilling multi-billion-parameter encoders for natural language understanding systems." *KDD 2022*

S Soltan, H Khan, W Hamza "Limitations of knowledge distillation for zero-shot transfer learning." *EMNLP 2021 Workshop on Simple and Efficient Natural Language Processing (SustaiNLP)*

X Liu, M Li, L Chen, P Wanigasekara, W Ruan, H Khan, W Hamza, C Su "ASR n-best fusion nets." *ICASSP 2021*

Q Zhu, H Khan, S Soltan, S Rawls, W Hamza, "Don't parse, insert: multilingual semantic parsing with insertion based decoding." *CoNLL 2020*

P Prakash, SK Shashidhar, W Zhao, S Rongali, H Khan**,** M Kayser "Compressing transformer-based semantic parsing models using compositional code embeddings." *EMNLP Findings 2020*

C Peris, G Oz, K Abboud, P Wanigasekara, H Khan "Using multiple ASR hypotheses to boost i18n NLU performance." *ICON 2020*

D Park, H Khan**,** B Yener "Generation & evaluation of adversarial examples for malware obfuscation." *ICMLA 2019*

M Perrone, H Khan, C Kim, A Kyrillidis, J Quinn, V Salapura "Optimal mini-batch size selection for fast gradient descent." *Arxiv 2019*

H Khan, B Yener "Learning filter widths of spectral decompositions with wavelets." *NeurIPS 2018.*

H Khan, L Marcuse, M Fields, K Swann, B Yener "Focal onset seizure prediction using convolutional networks." *IEEE Transactions on Biomedical Engineering 2017*

AWARDS

- **(2017)** Best Poster Award - Rensselaer Polytechnic Institute Computer Science Department
- **(2013)** Outstanding Graduate - SUNY New Paltz Engineering Department
- **(2012)** Steve Bogart Engineering Scholarship
- **(2011)** Rensselaer Polytechnic Institute Joseph H. Smith Jr. '45 Award

Languages and Technologies

- Python; C++; C; Java
- Packages: PyTorch; MxNet; Deepspeed; Keras; Tensorflow
- https://github.com/haidark

Hobbies & Extracurriculars

- Avid reader of science fiction novels (Herbert, Asimov, Orwell, LeGuin)
- Learning and speaking human languages (4 so far…)
- Horse riding (Western and Hunter)