# Harsh **Kohli**

Graduate Research Assistant

*1450 Worthington St., Unit B, Columbus, USA*

+1-614-554-9384 | harsh14791@gmail.com | HarshKohli | harsh-s-kohli | Harsh Kohli


## Education

**The Ohio State University** — *Columbus, USA*

PhD Computer Science & Engineering — *Aug. 2022 - Present*

- Natural Language Processing

**Georgia Institute of Technology** — *Atlanta, USA*

M.S Computer Science — *Jan. 2018 - May 2020*

- Specialization in Machine Learning

**Birla Institute of Technology & Science, Pilani - K.K Birla Goa Campus** — *Goa, India*

B.E (Hons.) Computer Science — *Aug. 2010 - May 2014*

## Research Publications

### Conference Publications [6]

Training Bi-Encoders for Word Sense Disambiguation
Harsh Kohli
*Document Analysis and Recognition – ICDAR 2021*
DOI: 10.1007/978-3-030-86331-9_53 — 2021

Transfer Learning and Augmentation for Word Sense Disambiguation
Harsh Kohli
*Advances in Information Retrieval - ECIR 2021*
DOI: 10.1007/978-3-030-72240-1_29 — 2021

Training Mixed-Objective Pointing Decoders for Block-Level Optimization in Search Recommendation
Harsh Kohli
*Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*
DOI: 10.1145/3397271.3401236 — 2020

AQuPR: Attention Based Query Passage Retrieval
Parth Pathak, Mithun Das Gupta, Niranjan Nayak, Harsh Kohli
*Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*
DOI: 10.1145/3269206.3269323 — 2018

Document categorization using semantic relatedness amp; Anaphora resolution: A discussion
Kaustubh Dhole, Harsh Kohli
*2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*
DOI: 10.1109/ICRCICN.2015.7434279 — 2015

Optimal Route Searching in Networks with Dynamic Weights Using Flow Algorithms
Sunit Singh, Ram Prasad Joshi, Harsh Kohli
*2015 International Conference on Computational Intelligence and Communication Networks (CICN)*
DOI: 10.1109/CICN.2015.37 — 2015

### arXiv Pre-print [1]

Learning Representations for Zero-Shot Retrieval over Structured Data
Harsh Kohli

URL: https://arxiv.org/abs/2111.00123 — 2021

## Work Experience

**Amazon** — *Cambridge (MA), USA*

Applied Scientist Intern — *May 2023 - Present*

- Part of Alexa team working on the task of Bilingual Lexicon Induction
- Developed a model for incorporating additional word and corpus-level features to improve state-of-the-art on standard BLI benchmarks
- The work is submitted and under review at NAACL 2024

**The Ohio State University** — *Columbus, USA*

Graduate Research Assistant — *Aug. 2022 - Present*

- Working under Prof. Huan Sun in analyzing & profiling limitations of Large Language Models in tasks such as Reasoning & Grounding
- Constructing a new dataset for benchmarking the performance of LLM's on these tasks

### Compass

*Bangalore, India*

Senior Machine Learning Scientist 2

*Mar. 2021 - Apr. 2022*

- Worked on the ranking & recommendation stack for real-estate search
- Implemented a low latency partial-query auto-suggest system incorporating both textual and non-textual features
- Worked on creating a unified representation using the multi-modal features on a real-estate listing page
- Developing a task-oriented chatbot to answer consumer queries in the absence of an agent - trained the dialogue state tracker to recognize the important entities and detect user intent as well as imbued the bot with a Reading Comprehension model to answer questions from text

### Salesken

*Bangalore, India*

Technical Architect - Data Science

*Jan. 2020 - Mar. 2021*

- Involved in the company's ideation, planning and strategy to leverage NLP techniques to improve sales conversations
- Led and mentored a team of Data Scientists to build several of the products key components including models for semantic matching and inference, emotion detection, dialogue state tracking, and fast semantic vector search
- Optimized the various Deep Learning models for latency and throughput, and set up the pipeline to drive them to production - Docker containerization, half-precision GPU inference, and auto-scaling using Kubernetes clusters

### Microsoft

*Hyderabad, India*

Data Scientist 2

*Feb. 2018 - Jan. 2020*

- Developed an architecture for Mixed Objective, Block level optimization for web search recommendation in Bing using a Pointing Decoder model and reinforcement learning objectives - the research was accepted at SIGIR 2020
- Worked on Bing's domain-specific passage ranking and Machine Comprehension abilities. As part of the work, we published a novel lightweight Attention-LSTM model for online ranking through a paper at CIKM 2018
- Session-context aware and Multi-Task DL systems for follow up query suggestion in Bing's Related Questions feature
- Entity disambiguation using Bi-Encoders in Web Queries using various negative-sampling strategies and optimization techniques
- Runner-up at the Microsoft Synapse Data Science Challenge 2019 - Microsoft's annual Data Science competition

### IPsoft

*Bangalore, India*

R&D Engineer

*Jun. 2014 - Feb. 2018*

- Worked on several modules including Machine Comprehension, Dialogue Management, Machine Translation, Anaphora Resolution, Intent Detection for Ipsoft's virtual assistant 'Amelia'
- Applied traditional logic and rule-based approaches to solve for these problems using taxonomies, ontologies and various other linguistic resources such as WordNet, FrameNet, PropBank etc.
- Led a team working on improving state of the art approaches on tasks like Question Answering, Natural Language Inference etc. using Attention-Matching sequence models under the guidance of Prof. Manning

## Select Academic Projects

#### Word Sense Disambiguation with Transformer Models and Gloss Information

- Implemented a cross-encoder based model for WSD with multi-task pretraining and data augmentation through chained back-translation
- Trained a Siamese network on context-gloss pairs with several negative-sampling and optimization techniques
- Infused the architecture with relational information from WordNet through a distinct pre-training step
- Presented the work through 2 single-author papers at ICDAR 2021 and ECIR 2021 both of which achieved SOTA results on standard benchmarks

#### Jack Watson Research Squad for Plagiarism Detection

- Working under Prof. Thad Starner (GeorgiaTech) towards developing a Chatbot to catch plagiarism on homework-for-hire websites
- Developed modules for plagiarism intent detection, homework similarity-matching algorithms, general conversation through Seq2Seq models, text auto-correct, and keyword extraction
- The project was accepted as a Work-in-Progress paper at the Learning@Scale conference 2019

#### Zero-Shot Retrieval on Tabular Data

- Proposed a method to encode a natural-language question and table information using features such as column-names and row content
- Trained a Neural Network model for Information Retrieval over tabular data

#### Document Categorization using Semantic Relatedness

- Implemented a Document Categorization system using Wordnet-based semantic similarity of most frequently occurring nouns
- Did a study of various Information Content and path-based semantic similarity measures and how well they performed on document categorization as well as the impact of Anaphora Resolution
- Paper was accepted at the 2015 IEEE International Conference ICRCICN

#### Optimal Route Searching in Networks with Dynamic Weights

- Worked with Prof. Ramprasad Joshi to propose a method for traffic-routing through a modification of the Edmunds-Karp algorithm using the Ford Fulkerson method for flow
- Accounted for dual nature of edges, representing both length and traffic densities at any time in a traffic network
- Simulated a time-centric and distance-centric approach to derive metrics such as average distance, speed and time, and proposed a method for optimizing the number of times the algorithm is invoked
- Paper was accepted at the 2015 IEEE International Conference CICN