

Akshay Parekh, Ph.D.

Consultant Data Scientist (NLP Applied Research)

email: sendtoakshayparekh@gmail.com

mobile: +91 8462000540

website: akshayparakh.in

github: akshayparakh25

I am working as an NLP Data Scientist at Advanced Analytics and Data Science (AADS), Eli Lilly & Co., Bangalore, since June 2022. I have worked on several projects involving Generative AI, Knowledge Graphs, and Information Extraction. My current work focus on building and designing a GenAI authoring framework that can become base layer for all the authoring projects.

I completed my Ph.D. from IIT Guwahati in 2023. My PhD research work centers around Natural Language Processing (NLP), Information Extraction (IE), and Deep Learning (DL), with a primary focus on Large-scale Relation Classification (RC) from noisy datasets.

I enjoy playing Table Tennis & Football. I also like to travel, trek, and explore hidden places and local food. I have keen interest on Finance and Startup Industry.

EDUCATION

- **Indian Institute of Technology Guwahati** Guwahati, India
Ph.D., Advisors: Dr. Ashish Anand & Dr. Amit Awekar Jul 2016 – 2023
- **Center for Development in Advanced Computing (CDAC)** Hyderabad, India
Diploma in System Software Development(DSSD); CGPA: 73.25% (Batch Topper) Feb 2014 – Aug 2014
- **G. H. Raisoni College of Engineering** Nagpur, India
Bachelor of Engineering in Computer Science & Engineering; CGPA: 72.40% Jul 2009 – May 2013

INDUSTRY EXPERIENCE

- **Advanced Analytics & Data Science (AADS), Eli Lilly & Co. - LCCI** Bangalore, India
NLP Data Scientist Feb 2022 - Present
 - **Consultant Data Scientist:**
 - * **AI Authoring:** Using generative AI, Large Language Models (LLMs), and domain-specific fine-tuning for drafting first-draft of manuscripts, reports, presentations, and other documents. Using state-of-the-art research work and paid/open-source tool for building intelligent document parser.
 - * **Knowledge Graphs:** Building Knowledge graph for enterprise data; Leveraging NLG for Information Extraction; Fine-tuning LLMs for Relation Extraction and generating GraphDB queries
 - * **Single cell transformer models:** Exploring single cell transformer models (such as, scGPT) for Drug Repurposing.
 - **NLP Research Intern:** Biomedical Information Retrieval, Unsupervised Key-Phrase Extraction, Transformers, BioClinicalBERT, Biomedical Named Entity Recognition
- **IBM Research Labs** Bangalore, India
Research Intern May 2019 - Aug 2019
 - **Relation Extraction for Ontology Learning:** Relation Extraction, NLP, DL & Ontology Graphs

STARTUP EXPERIENCE

- **Clearfeed.ai** Bangalore, India
Consultant Data Scientist Oct 2021 - Jan 2022
 - **Dataset Preprocessing & Preparation:** Preparing dataset from large open source slack dumps.
 - **Text Classification:** Urgency Detection, Bug/Error Report Identification, Thread Identification
 - **Summarization:** Generating summaries of threaded conversation; Generating title for a thread.
- **Digitales Communications Pvt. Ltd.** Mumbai, India
Software Engineer / Data Engineer Jan 2015 - Jun 2015
 - **Shopista:** Web Scraping, Information Extraction & Android UI Designer

PUBLICATIONS

PhD Thesis

Understanding and Mitigation of Noise in Crowd Sourced Relation Classification Dataset.

- **Akshay Parekh**, Ashish Anand, Amit Awekar

link

Patent Application

Neural-Based Ontology Generation and Refinement

- Balaji Ganesan, Riddhiman Dasgupta, **Akshay Parekh**, Hima Patel, Berthold Reinwald

Application No.: 16/937641

US Patent

Conference

Taxonomical hierarchy of canonicalized relations from multiple Knowledge Bases.

- **Akshay Parekh**, Ashish Anand, Amit Awekar

CoDS-COMAD 2020

Improving Relation Classification using Relation Hierarchy.

- **Akshay Parekh**, Ashish Anand, Amit Awekar

NLDB 2022

Document Structure aware Relational Graph Convolutional Networks for Ontology Population.

- Abhay M Shalghar, Ayush Kumar, Shobha G, Balaji Ganesan, Aswin Kannan, **Akshay Parekh**

DLG4NLP Workshop at ICLR 2022

Relation Classification Model Performance as a Function of Dataset Label Noise.

- **Akshay Parekh**, Ashish Anand, Amit Awekar

CoDS-COMAD 2023

arXiv

A Neural Architecture for Person Ontology population.

- Balaji Ganesan, Riddhiman Dasgupta, **Akshay Parekh**, Hima Patel, Berthold Reinwald

arXiv:2001.08013

Budget Sensitive Reannotation of Noisy Relation Classification Data Using Label Hierarchy.

- **Akshay Parekh**, Ashish Anand, Amit Awekar

arxiv:2112.13320

Noise in Relation Classification Dataset TACRED: Characterization and Reduction.

- **Akshay Parekh**, Mayank Baranwal, Shivang Dalal, Ashish Anand, Amit Awekar

arxiv:2311.12298

PROJECTS

- **Relation Extraction for Ontology Creation:** Internship project at IBM Research Labs, Bangalore. In this project, we developed an end-to-end model to extract relations between real-world entities from texts for ontology creation. We based our model on On2vec and used TACRED and other industry datasets for evaluation.
- **Generating Knowledge Graph from Wikipedia Infobox:** This is part of my initial work for my Ph.D. thesis work. In this project, I extract all the infobox from the Wikipedia XML dump. Categorize infobox template types in Person, Organization, and Location. Parse all the infobox that falls into one of the categories. Also, generate a list of relations for each of the 3 types. Finally, extract information in the form of triple (e1, r, e2)
- **Packet Filtering Firewall for Linux Kernel:** Course project for DSSD, developed in C/C++ using iptables, netfilter, and Linux Kernel module.

TALKS

- **IEEE Computing Society & Coding Club, GHRIET, Nagpur. (October 2023):** Webinar: Generative AI and it's Significance
- **Workshop: AI in Healthcare 2019, IIT Guwahati:** Hands-on session on Deep Learning
- **Management Development Programme on AI in its possible implementation in the Indian context, IIT Guwahati (20 November 2019):** Hands-on Session on Introduction to Python for Data Analysis and Text Processing.
- **TEQIP: DL for NLP, IIT Guwahati (19 November 2019):** Hands-on Session on Introduction to Tensorflow==1.x.
- **RSF, EEE, IIT Guwahati (16 November 2019), IIT Guwahati:** Hands-on Session on Python for Machine Learning.

TECHNICAL SKILLS

- **Languages:** Python, C
- **Web-Technologies:** HTML, CSS
- **Data Analysis and Visualization:** Numpy, Pandas, Matplotlib, Seaborn
- **ML-DL:** Scikit-learn, PyTorch
- **NLP & Knowledge Graphs:** NLTK, spacy, gensim, transformers, langchain, neo4j, & cypher
- **Operating System:** Windows, Linux(Ubuntu)
- **Tools:** Ipython, Jupyter, Latex

CERTIFICATIONS

- **Deep Learning Specialization:** DeepLearning.AI, Coursera
- **Natural Language Processing Specialization:** DeepLearning.AI, Coursera
- **Introduction to Generative AI:** Google Cloud, Coursera
- **Generative AI with Large Language Models:** DeepLearning.AI, Coursera

ACHIEVEMENTS

- **Eli Lilly & Co., Lilly Innovation Award 2023:** Winner
- **Persistent Systems Pvt. Ltd., SemiColon 2012, Hackathon:** Winner