

Hadi Askari

Haskari@ucdavis.edu | (408) 333 5895 | <https://github.com/HadiAskari> | linkedin.com/in/HadiAskari67 | <https://shorturl.at/GyA4R>

EDUCATION

University of California Davis

PhD in Computer Science

CGPA: 3.90

Davis, California

Sept 2021-June 2026

University of California Davis

Masters in Computer Science

CGPA: 4.00

Davis, California

Sept 2021-Dec 2024

National University of Sciences and Technology Main Campus

Major: Bachelors in Electrical Engineering

Minor: Bachelors in Computer Science

CGPA: 3.67

Islamabad, Pakistan

Sept 2017-June 2021

WORK EXPERIENCE

Pinterest - Content Understanding Team

Machine Learning Research Intern

Palo Alto, California

June 2024-Sept 2024

- Improved upon the teams' classification models by fine-tuning **LLM models** such as **Llama3-8B-chat** and **Llama3.1-8B-chat**. Performance was better than **GPT-4o** and production model by **4.3%**.
- Fine-tuned **Vision Language Models (VLMs)** such as **InternVL2-8b** and **MiniCPM-Llama3-V-2_5** to compare.

Comcast - Media Analytics Team

Machine Learning Research Intern

San Francisco, California

June 2023-Sept 2023

- Worked on improving the teams' **Knowledge Graph embeddings** via **Graph Neural Network algorithms** (for e.g **GraphSage**) for improving their content recommendation.
- Worked on **entity alignment** of heterogenous public Knowledge Graphs (for e.g DBpedia, conceptNet) to enhance the teams' Knowledge Graph.
- Leveraged **Unsupervised Clustering Algorithms (tSNE)** to optimize recommendations via **text embedding models (text-embedding-ada-002, bge-large-en)**.
- Created a **LangChain** tool using **open-source LLMs** and querying online databases as context to improve recommendations

Comcast - Media Analytics Team

Machine Learning Research Intern

San Francisco, California

June 2022-Sept 2022

- Improved the **semantic relevance** between closed caption data from TV Shows and **IAB Taxonomy** ad-words by **103.7% (mAP)** by using **Sentence Transformer** models in a **Zero Shot** setting and contributed to a deployment ready product.
- Conducted research on **Unsupervised Learning Methods** such as **TSDAE**, **SimCSE**, **GPL** and **Few Shot Learning** Methods such as **SetFit** to improve the classification.
- Code Presentation**

LUKA Research Lab

Graduate Research Assistant

Davis, California

Sept 2021-Present

- Created a multi-input **LSTM** based model to minimize the KL Divergence of YouTube's video category distributions for a published Research project in **Reinforcement Learning** and online profile **obfuscation**.
- Fine tuned **GPT-2** Language Model to create conversational bots to be deployed real-time on Twitter.
- Investigated position bias and paraphrasing bias by fine-tuning summarization models like **Pegasus** and **BART** and comparing their results with the **quantized** versions of **LLMs** like **GPT-3.5-Turbo**, **Dolly-v2** and **LlaMa-2**.
- Working on ways to improve the **In-context Learning (ICL)** capabilities of **LLMs** and **VLMs**.
- Working on **Named Entity Recognition (NER)** and **Relation Extraction (RE)** for a **Software Security Knowledge Graph**.

SKILLS

Softwares: Python, C, C++, Pytorch, Matlab, R, Docker, MLFlow

Interests: Machine Learning, Deep Learning, NLP, Reinforcement Learning, Recommender Systems, Web Crawling, Data Mining, Computational Social Science, Knowledge Graphs, Graph Neural Networks, Large Language Models (LLMs).

UC Davis Research Projects

Improving In-Context Learning for Low Resource Settings, NLP, (Under submission at ACL 2025)

Feb 2024-ongoing

- Introduced the **Indirect ICL** domain which is a low-resource domain for **In-context Learning**. Investigated two main settings, **Mixture of Tasks** in the candidate pool and **Noisy Labels** in the candidate pool.
- Leveraged **Influence Functions** for **Demonstration Selection** in **In-context Learning** for **Indirect ICL** domain. Our method outperformed BertScore and Cosine Similarity by up to **3%**.

Named Entity Recognition and Relation Extraction for Software Security Knowledge Graph, NLP, (NSF-Grant)

Jan 2024-ongoing

- Performed NER by combining **BERT** based NER and **LLM** based NER (**UniNER**) from free-form articles (Hacker News).
- Used prompting techniques like In-Context Learning (**ICL**) and Chain-of-Thought (**CoT**) with **Llama3-8b** to perform **Relation Extraction (RE)** between the extracted libraries and vulnerabilities to add to the Knowledge Graph.

Analyzing Paraphrasing Robustness in LLMs, NLP, (Accepted at NAACL 2025)

Nov 2023-June 2024

- Analyzed the robustness capabilities of **LLMs** like **GPT-3.5-Turbo**, **Dolly-v2**, **Llama-2-13b-chat** and **Mistral-7b-v1** under a process called Relevance Paraphrasing for the end task of **abstractive summarization**.

Investigating Position Bias in LLMs, NLP, (Accepted at NAACL 2024)

Sept 2023-Oct 2023

- Investigated position bias by fine-tuning summarization models like **Pegasus** and **BART** and comparing their results with the **quantized** versions of **LLMs** like **GPT-3.5-Turbo**, **Dolly-v2** and **Llama-2-13b-chat**.
- Performed **Prompt Engineering** with **LangChain** and **Distributed Inferencing** in **PyTorch** and **HuggingFace** to generate efficient results.
- ***Link to publication***

Twitter Bots for News Incentivization, NLP, Web Crawling, Computational Social Science (Accepted at PNAS Nexus 2024)

Jan 2022-April 2024

- Using NLP techniques such as **Language Modeling**, **Topic Modeling**, **Transfer Learning** and **Named Entity Recognition** to train Twitter bots for automated responses.
- Designed the entire pipeline from data collection for training, user collection for the study, selecting which Tweets to respond to and evaluation methods of the experiment. Included database management in **SQLite**
- ***Code***

Obfuscating and De-obfuscating User Recommendations on YouTube, Reinforcement Learning, Deep Learning, Data Mining and Web Crawling (Accepted at PETS 2023)

Sept 2021-April 2023

- Built a De-obfuscation model on **TensorFlow** using **LSTMs** to minimize the **KL Divergence** of RL driven, obfuscated YouTube recommendations.
- Scraped Reddit data from 2017 to 2021 to collect 100k users to test our model on real world data. Performed several validation experiments.
- Conducted **system overhead** experiments on the web extension.
- ***Link to publication.***

TikTok/YouTube Shorts/Insta Reels, NLP, Systems, Web Crawling, Deep Learning.

Jan 2023-May 2023

- Implemented **Selenium** based sock-puppets for an audit of **TikTok**, **YouTube-Shorts** and **Instagram Reels**.
- Implemented Python scripts using **Android Debug Bridge (ADB)** to control the **android** versions of the apps.
- Developed **multi-modal ML classifiers** for analysis. Combining **textual**, **audio** and **video** classifiers to answer key research questions.
- ***Code-1, Code-2, Code-3, Code-4***

Teaching

UC Davis

Teaching Assistant

Davis, California

Jan 2023 - Mar 2023

- Taught **Data Structures and Algorithms** where I held discussion sections, office hours and graded student's assignments.

National University of Sciences and Technology Main Campus

Islamabad, Pakistan

Teaching Assistant

Aug 2020 - June 2021

- Taught **Advanced Machine Learning** where I was responsible for creating and grading assignments in **Tensorflow** and **PyTorch**.

Professional Services

Program Committee Member

- COLING 2025

Research Reviewer

- ACL ARR October 2024
- ACL ARR June 2024
- ACL ARR February 2024
- LREC-COLING 2024

Professional Memberships

- Association for Computational Linguistics.