Ashima Suvarna

Los Angeles, California - 90024

asuvarna31@g.ucla.edu
https://asuvarna31.github.io/

Education

- **University of California, Los Angeles (UCLA)** — Sept 2021 - Sept 2023
  Masters in Science - Computer Science; Deepmind Scholar — *GPA: 3.8/4.0*

- **Delhi Technological University (Formerly DCE)** — Aug 2016 - June 2020
  Bachelor of Technology - Computer Engineering — *GPA : 8.9/10*

Work Experience

- **PLUSLAB, UCLA** — Los Angeles, California
  Research Associate — *Jun 2023 - Present*
  - Researched **Questions Under Discussion** to model discourse relationships within documents.
  - Proposed supervised models for **automatic evaluation of QUD Parsers** using theoretical properties of QUDs.
  - Analysed the **phonological knowledge** of LLMs through prompt-based probing in English.

- **Walmart Inc.** — Sunnyvale, California
  Data Scientist Intern — *Jun 2022 - Sep 2022*
  - Implemented **tree-based Click-Through-Rate** prediction models for re-ranking products in Walmart Search.
  - Designed a **dynamic slot allocation** strategy to combine multiple ranking objectives like revenue and relevance.
  - Extracted relevance scores in product titles and image using **CLIP as an important feature** in ranking tree.

- **Dimagi Inc.** — New Delhi, India
  Technical Project Analyst — *June 2020 - August 2021*
  - **Designed requirements** for mobile and web apps for Aanganwadi workers and Lady Supervisors on Commcare.
  - **Aggregated maternal health data** from Low Income Countries and extracted important features and insights.
  - Identified high risk pregnancies using **SVMs and Random Forest Classifiers** using vital health indicators.
  - Implemented **outlier detection strategies** in maternal and child health data collected via the Commcare.

- **University of New Brunswick** — Fredericton, Canada
  Mitacs Research Intern — *June 2019 - August 2019*
  - Studied the **syntactic properties** of Native American languages from the eastern Maritime Province in Canada.
  - Optimized **RNN-based language models** with **sub-word modeling** for inflective languages like Miqmaq.
  - Designed a **novel metric based on keystroke savings** in next word prediction to evaluate sub-word models.

Publications

- **(Workshop)** Ashima Suvarna*, Xiao Liu*, Tanmay Parekh, Kai-Wei Chang, Nanyun Peng, Automatic Evaluation of Question Under Discussion (QUD) Discourse Parsers [Poster] **SoCal NLP 2023**
- **(Journal)** Minni Jain, Ashima Suvarna, Amita Jain, An evolutionary game theory based approach for query expansion **Multimedia Tools 2022**
- **(Conference)** Jeremie Boudreau, Akankshya Patra, Ashima Suvarna, Paul Cook, Evaluating the Impact of Sub-word Information and Cross-lingual Word Embeddings on Mi'kmaq Language Modelling [Paper] **LREC 2020**
- **(Workshop)** Ashima Suvarna*, Grusha Bhalla*, NotAWhore : A Computational Linguistic Perspective of Rape Culture and Victimization on Social Media [Paper] **ACL-SRW 2020**

Selected Projects

Should Large Language Models go to School for Phonology ? *Advisor: Prof. Nanyun Peng*
Designed tasks to probe LLMs for **phonological understanding** by curating frequent and rare words.
Benchmarked four popular LLMs (Alpaca-7B, GPT3.5, Vicuna-7B) in zero-shot and few-shot settings.
Created a **research report** to fulfill MS in CS requirements at UCLA. [**Report**]

Examining Gender Bias in Languages with Grammatical Gender for Hindi *Advisor: Prof. Kai-Wei Chang*
Literature survey of the **gender bias prevalent in languages with grammatical gender** such as Spanish.
Created a blog report by exploring the **grammatical and semantic gender** axes in **Hindi**. [**BLOG**]

Automated Prompt Engineering for Stereotype Detection in language models *Advisor: Prof. K.W. Chang*
Created an **automatic prompt generation** benchmark to evaluate the social biases in language generation.
Evaluated LLMs across four social bias categories - **disability, religion, race, and gender**. [**CODE**

Service

Reviewer: ACL-SRW 2023, NAACL-SRW 2022, ICLR Tiny Papers 2023, ICLR Tiny Papers 2024
Program Committee: SoCal NLP Symposium 2023
Volunteer: Enactus DTU 2016-2018, Open Office-Hours

Skills Summary

Machine Learning: Pytorch, Apache Spark, AWS | **Programming:** Python, C/C++, SQL, HTML, CSS.

Other Projects

Demo on Zero-shot Sonnet Generation with Discourse Level Planning *Advisor: Prof. Nanyun Peng*
Scaled and refined a pipeline for **sonnet generation** for a web-based demo using **AWS and Streamlit**.**[CODE]**
Optimised the **decoding strategy** for improved time by 50% using pruning methods and diverse generations.

Creative Language Generation in Hindi *Advisor: Prof. Nanyun Peng*
Assessed the efficacy of state-of-the art **pun generation frameworks** in generating puns in Hindi language.
Proposed a **novel algorithm for generating short structured poems** called shayari in Hindi using IndicBart.

Mixture-of-Denoisers for Protein Sequence Modeling *Advisor: Prof. Sriram Sankararaman*
Implemented a **mixture-of-denoisers** for **protein sequences** of Pfam dataset using a BERT-based model.
Proposed that an **xtreme mask-filling objective** is highly effective than vanilla masking objective.

Talks

- Interpretable Machine Learning (CS 224) [SLIDES]
- A Distributional Approach to Controlled Text Generation (CS 269) [SLIDES]

Awards and Scholarships

- **Deepmind Scholarship** to fully fund Masters at UCLA CS; awarded to 2 students in Fall 2021 in the MS cohort.
- **Mitacs Globalink Research Scholarship** to conduct research and visit a Canadian University for Summer 2019.
- General Secretary of **Enactus DTU** in 2018; won the **KPMG Grant** to rehabilitate acid-attack survivors in Delhi.

Relevant Coursework

Natural Language Generation, Fairness, Transparency and Robustness in Natural Language Processing, Automated Reasoning Theory and Applications, Deep Learning, Artificial Intelligence, Machine Learning Foundations, Big Data Analytics, Discrete Structure, Algorithm Design and Analysis, Theory of Computation.