

# Neeraja Kirtane

(+1) 217-305-2081

[kirtane3@illinois.edu](mailto:kirtane3@illinois.edu)

[Website](#) [◇ LinkedIn](#) [◇ GitHub](#) [◇ Google Scholar](#)

---

## RESEARCH INTERESTS

**Research Interests:** Deep Learning, Natural Language Processing, Graph Deep Learning.

**Focus Areas:** Trustworthiness in NLP, NLP for low-resource languages, Class imbalance handling in Graph Learning.

---

## EDUCATION

<b>University of Illinois Urbana-Champaign</b>	2023 – 2025
<i>Masters of Science in Computer Science (Thesis Track: Advisor- Prof. Hao Peng) (MSCS)</i>	CGPA: 4.0/4.0
<b>Manipal Institute of Technology, Manipal, India</b>	2018 – 2022
<i>B.Tech in Computer Science and Engineering (Minor: Computational Intelligence)</i>	CGPA: 9.14/10

---

## EXPERIENCE

<b>University of Illinois, Urbana-Champaign</b>	Aug 2023- Present
<i>Graduate Student Researcher</i>	Advisor – <a href="#">Prof. Hao Peng</a>
<ul style="list-style-type: none"><li>Working on various problems to make large language models (LLMs) more efficient and trustworthy and to mitigate hallucinations in these models.</li><li>Investigating how hidden layers play a role in detecting hallucinations in factual data.</li></ul>	
<b>University of Illinois, Urbana-Champaign</b>	May 2024- Present
<i>Graduate Student Researcher</i>	Advisors – <a href="#">Prof. Hao Peng</a> and <a href="#">Prof. Dilek Hakkani-Tur</a>
<ul style="list-style-type: none"><li>Working on jailbreaking LLMs for bias</li><li>Investigating how jailbreaking works for multiple turns.</li></ul>	
<b>Indian Institute of Technology Madras, Chennai, India</b>	Jul 2022- Aug 2023
<i>Post Baccalaureate Fellow</i>	Advisors – <a href="#">Prof. Balaraman Ravindran</a> & <a href="#">Dr. Rajashree Baskaran</a>
<ul style="list-style-type: none"><li>Worked on the <a href="#">Project Hidden Voices</a> at the <b>Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI)</b>.</li><li>Building intelligent tools to aid in adding 10,000 notable women’s biography drafts to Wikipedia.</li><li>This aims to <b>reduce the gender gap</b> in wikipedia data.</li><li>Worked on building knowledge graphs and doing graph to text generation using large language models. Finetuned models like GPT-J and GPT-Neo for this process.</li></ul>	
<b>Indian Institute of Technology Madras, Chennai, India</b>	Jan 2022- Jun 2022
<i>Research Intern</i>	Advisors – <a href="#">Prof. Balaraman Ravindran</a> & <a href="#">Dr. Ashish Tendulkar</a>
<ul style="list-style-type: none"><li>Worked on handling class imbalance in Graph neural networks at <b>RBC-DSAI</b>.</li><li>Used <b>implicit ways</b> at the algorithmic level to handle this imbalance. Used a custom loss function and tuned the attention weights to focus more on minority nodes.</li><li>Additional Links: <a href="#">Report</a>   <a href="#">Slides</a>   <a href="#">Github</a></li></ul>	
<b>Centre for development of advanced computing, CDAC Pune</b>	Jun 2020 – Aug 2020
<i>ML Intern</i>	Advisor – <a href="#">Rahul Dangi</a>
<ul style="list-style-type: none"><li>Extracted keywords and named entities from a document for <b>better comprehension</b>.</li><li>Used word embeddings of the GLoVe dataset for the predictions. Major libraries used in Python were NLTK (for text processing), Gensim (to use the LDA algorithm), Flask (to create the front end of the project).</li><li>Additional Links: <a href="#">Github</a>   <a href="#">Report</a></li></ul>	

## PUBLICATIONS

---

### 1. FactCheckmate: Preemptively Detecting and Mitigating Hallucinations in LMs

Under review [Paper](#)

Oct 2024

- Authors: *Neeraja Kirtane, Deema Alnuhait, Muhammad Khalifa, Hao Peng*

### 2. Hidden Voices: Reducing gender data gap, one Wikipedia article at a time

Wikiworkshop 2023 [Paper](#)

May 2023

- Authors: *Neeraja Kirtane, Anuraag Shankar, Chelsi Jain, Ganesh Katrapati, Raji Baskaran, Balaraman Ravindran*

### 3. ReGrAt: Regularization in graphs using attention mechanism to handle class imbalance

GCLR workshop at AAAI 2023 [Paper](#)

Sep 2022

- Authors: *Neeraja Kirtane, Jeshuren Chelladurai, Balaraman Ravindran, Ashish Tendulkar*

### 4. Efficient Gender Debiasing of Pre-trained Indic Language Models

Deployable-AI workshop at AAAI 2023 [Paper](#)

Aug 2022

- Authors: *Neeraja Kirtane, V Manushree, Aditya Kane*

### 5. Mitigating gender stereotypes in Hindi and Marathi

Gender bias in NLP workshop at NAACL 2022 [Paper](#)

May 2022

- Authors: *Neeraja Kirtane, Tanvi Anand*

### 6. Transformer based ensemble for emotion detection

WASSA workshop at ACL 2022 [GitHub](#) | [Paper](#)

Mar 2022

- Authors: *Aditya Kane, Shantanu Patankar, Sahil Khose, Neeraja Kirtane*

### 7. Occupational Gender Stereotypes in Indian Languages

Widening NLP workshop at EMNLP 2021 [Paper](#) | [Video](#) | [Poster](#)

Nov 2021

- Authors: *Neeraja Kirtane, Tanvi Anand*

## PROJECTS

---

### Evaluating Mathematical Reasoning Chains [Github](#)

- Developed a pretrained metric to evaluate the chains generated by LLMs for Math reasoning tasks.
- The metric evaluated nine different characteristics of the chain.

### Labelling privacy policies using LLMs

- Labeled privacy policies and trained a Llama model to categorize privacy policies.
- Our results were comparable to GPT models

### Smart Document Explorer [GitHub](#)

Summer 2021

- Created a program to make a document more **accessible and understandable**. Used history and geography textbooks as the data to help children benefit from this.
- Extracted named entities, keywords. Summarized the text, found similar sentences given a sentence.
- Used relationship extraction to map dates with events in history textbooks.

## TECHNICAL SKILLS AND RELEVANT COURSEWORK

---

**Languages:** Python, C++, Java, C, SQL

**Tools and Libraries:** PyTorch, NumPy, TensorFlow

**Courses:** CS 412: Introduction to Data mining, CS 546: Advanced NLP, CS 562: Advanced Topics in Security, Privacy, and Machine Learning, CS 568: User-centered ML, CS 598: LLMs post Pretraining.

## TEACHING EXPERIENCE AND EXTRACURRICULAR

---

- **TA for CS 105:** Introduction to Computing: (Non-Tech) for Fall 2023, Spring 2024.
- Volunteer at **EMNLP 2021, NAACL 2022**
- **Regional Mathematics Olympiad (RMO)** Finalist.
- Outstanding TA award for Spring 2024 semester.
- College level Finalist at **Smart India Hackathon** among forty plus teams