# ARPANDEEP KHATUA

202-710-8808 901 W Western Ave, Urbana IL, 61801
akhatua2@illinois.edu arpandeepkhatua.com github.com/akhatua2

## EDUCATION

**University of Illinois at Urbana-Champaign** August 2019 - May 2023
**Bachelor of Science, Computer Engineering (Minor, Mathematics)** **Cumulative GPA – 4.00/4.00**
Edmund J. James Honors Scholar, Dean's List (2019, 2020, 2021)
**Relevant Coursework: ECE:** Analog Signal Processing, Computer Systems & Programming, Digital Signal Processing, Computer Systems Engrg, Digital Systems Laboratory. **CS:** Discrete Structures, Data Structures Honors, Artificial Intelligence, Deep Learning, Algs & Models of Comp, Machine Learning, Natural Language Processing. **Math:** Differential Eq, Fundamental Math, Probability with Engrg, Applied Linear Algebra, Number Theory.
**Senior Thesis:** *Generating Large Real World and Synthetic Graph Datasets for GNN Applications.*

## INTERNSHIP EXPERIENCE

**Meta Inc.** **May 2022 - Aug 2022**
Software Engineering Intern (Live Video Community Experiences) - Mr. Gabe Ochoa New York, NY

- Created infinite scroll comments on WWW using React, PHP, and Javascript and live polls and video overviews on iOS and Android using Native Templates and Bloks. Generated in a sig-stat increase in overall FB Watch time based on A/B testing over 200+ million users.
- Simplified E2E testing framework and documentation for internal languages used by 50+ teams.

**Facebook Inc.** May 2021 - August 2021
Software Engineering Intern (Facebook App Commerce Shop Rankings Team) Menlo Park, CA

- Built a product retrieval transformer model personalized to users based on their history and viewing habits. This model beats the current production retrieval model by 35% based on initial offline tests and also performs better than the online ranking model by 30%.
- Optimized the existing collection pipelines to operate 600x faster using Hack, PHP, and Python. Created a new API to help engineers query PyTorch models which is currently used as an internal tool.

## RESEARCH EXPERIENCE

**Illinois Geometry Lab Fellow Spring 2022** Social Computing Lab - Prof. Jana Diesner Aug 2022 - Present

- Using Twitter API to collect $10M$ tweets for detecting and prioritizing needs during crisis events (like the Russia-Ukraine conflict) in order to (1) extract a list of needed resources, (2) how they are fulfilled and (3) detect *who-needs-what* phrases.

**Undergraduate Researcher** FORWARD LAB - Prof. Kevin C.C. Chang Jan 2022 - Present

- Designed a pipeline to generate structure-aware articles given a topic using information retrieval, ranking, text generation, and summarization for creating Wikipedia-like articles and richer search results.
- Created a 2-stage attention-based seq2seq model to generate subtopics for a given title. Given a topic and intro, stage one generates a list of categories for the topic and stage two creates a list of subtopics, which is then used for structured text generation.
- Filtered out noisy web retrieved data using text-rank and trained a few-shot classifier model to classify them into the dynamically generated sub-topics.

**Illinois Geometry Lab Fellow Fall 2022** Prof. A.J. Hildebrand Jan 2022 - Present

- Performed large-scale statistical analysis of the first 30 billion continued fraction (CF) digits of $\pi$ using chi-square tests, $p-tests$, and Kolmogorov–Smirnov tests to provide compelling evidence that these digits behave like those of a random real number.
- Implemented massive multi-dimensional space walks to compare random and $\pi$ CF digits and used extreme-digit, single-digit, and $z-scores$ for further evidence.

**C3SR Scholar** IMPACT Lab - Prof. Wen-Mei Hwu, Dr. Mailthody Aug 2021 - Present

- Generated the largest publicly available homogeneous graph dataset - Illinois Graph Benchmark (IGB) with 260M nodes and 3.9B edges collaborating with researchers from Amazon AWS, Deep Graph Library (DGL), NVIDIA and IBM for multi-class node classification and efficient system designing.
- Combined Microsoft Academic Graph (MAG) and Semantic Scholar databases to annotate 162× more data for supervised learning tasks to test emerging Graph Neural Network (GNN) models at scale.

**NCSA SPIN Intern** National Center for Supercomputing Applications Sep 2021 - May 2022

- Applied transfer learning on a Mask-RCNN based model detectron2 using the Pubmed dataset to detect 5 non-text classes in documents. Coupled with a large image classification model to detect and classify over 1000 classes of non-text objects and images on scanned documents with over 97% accuracy.
- Used openCV for preprocessing pipeline to process over 16M volumes (5B pages) in the HathiTrust Digital Library to improve run-time by 3× on a $V100$ GPU.

**Student Researcher** Advisor - Prof. Sanmi Koyejo Oct 2020 - May 2021

- Worked on a cross-department project to predict phenotype combinations in maize/sorghum crop genes to maximize heritability using reinforcement learning (RL). Wrote classic local and global search algorithms like Particle Swarm Optimization (PSO) and Simulated Annealing (SA) to set baselines.
- Developed a Multi-Layer Perceptron (MLP) to serve as a mapping function between wavelength and experimental ground-truth data which improved the RL search over multi-dimensional wavelength space.

**Software Engineering Intern** Health Care Engineering Systems Center Jun 2020 - May 2021

- Developed a novel operation training procedure in virtual reality using Unity with anatomically accurate physics scripts, by conducting literature reviews, capable of real-time rendering, optimized to run without a GPU.
- Implementing a Reinforcement Learning model to automate operation procedure and create a predictive model to assign probability of success using 6-dimensional c-space A*, RRT, and RRT* search algorithms.

**Undergraduate Researcher** Advisor - Prof. Matthew Caesar Sep 2020 - May 2021

- Experimenting combinations of different neural net architectures for object detection (ResNets, MobileNets) and object tracking using deepSORT with an SSD trained on NCSA's HAL cluster.
- Implemented the Hungarian Algorithm and Kalman Filter to track and predict object position during large periods of obstruction.

**Undergraduate Research Assistant** EarthSense (FRESH Labs) Oct 2019 - Mar 2020

- Soldered and worked on the power source and circuit design for cameras and sensors of an agricultural bot.
- Wrote Python scripts for autonomous path planning using OpenCV. Labeled over 5000 pictures and helped train a Mask R-CNN to detect the path and its surrounding with less than 5% error.

## TEACHING EXPERIENCE

**Computer Systems Engineering (ECE 391) Course Staff** Jan 2022 - Present

- Creating new course material and internal grading scripts and conducting Office Hours to help students debug codes and provide machine problem overviews for intensive upper-level OS kernel-building class.

**Probability with Engrg Applic (ECE 313) Course Staff** Jan 2022 - Present

- Graded and provided feedback for students on weekly home-works and exams in upper-level probability and statistic class for signal processing and control systems.

**Analog Signal Processing (ECE 210) Course Staff** Aug 2020 - Dec 2021

- Graded and provided feedback for students on weekly home-works and exams for a sophomore-level class covering circuit analysis, Fourier, and Laplace transform.

**Intro to Electronics (ECE 110/120 + Honors Lab) Course Staff** Jan 2020 - Dec 2020

- Mentored underclassmen to work on projects involving robotics path planning, computer vision, NLP, FPGAs, and power systems in the honors lab. Conducted office hours and discussion sections for intro-level electrical and computer classes.

## PROJECTS

**XerSIZE - AI Personal Trainer** Sep 2020

- Used Google Tensorflow's PoseNet and trained a neural net classifier by collecting data of different poses to create an AI personal trainer and built a full-scale web app using HTML, CSS, JS, ml5.js, Flask, and SQLite.

**Kaizen - Duke University HackDuke Winner** Dec 2020

- Built a custom NLP model to classify text based on mental health conditions and a web page for easier access by patients and health-care professionals with an OCR and voice-to-text functionality.

**CourseLoop - Illinois HackThis Winner** Aug 2020

- Auto-grading on text extracted from PDF assignments with an OCR pipeline, using NLP in python with 98%+ accuracy in 1 week. Reduced grading time by 50% utilizing better algorithms and libraries.

## TECHNICAL TOOLS, LEADERSHIP & HONORS

**Tools:** ***Very Knowledgeable:*** React, Hack/PHP, GraphQL, Python, C++, C, SQL • ***Knowledgeable:*** HTML, CSS, JavaScript, Django, Flask, Python Data/ML Libraries.

**Leadership:** ECE PULSE Corporate Committee • HKN Honor Society Secretary & Corporate Director • Promoting Undergraduate Research in Engineering (PURE) Vice President & Corporate Director.

**Honors:** O. Thomas and Martha S. Purl Scholarship (awarded to 2 students out of 2000 every year) • Grainger Engineering Scholar • NCSA Student Pushing Innovation Fellow • C3SR-URAI Scholar • Illinois Geometry Lab Scholar (×2) • Edmund J. James Scholar • Dean's List • PURE Scholar • HCESC Jump ARCHES • COVID-19 Wall of Recognition in Engineering.

## REFERENCES

Prof. Wen-mei Hwu - **AMD Jerry Sanders Chair Emeritus at UIUC, Senior Distinguished Research Scientist at NVIDIA** - w-hwu@illinois.edu
Dr. Vikram Sharma Mailthody - **Research Scientist at NVIDIA** - vsm2@illinois.edu