ARPANDEEP KHATUA

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

RESEARCH INTERESTS

*My research interest focuses broadly on **graph neural networks (GNNs)** and **natural language processing (NLP)** with an emphasis on optimizing language models, long text generation, and information retrieval.*

EDUCATION

University of Illinois Urbana-Champaign Aug 2019 - Dec 2022
Bachelor of Science, Computer Engineering (Minor in Mathematics) **GPA – 4.00/4.00**
Edmund J. James Honors Scholar, Dean's List (2019, 2020, 2021)
Relevant Coursework: CS: Natural Language Processing, Artificial Intelligence, Machine Learning, Deep Learning, Algs & Models of Comp, Data Struct. Honors, Discrete Struct., Databases. **ECE:** Analog Signal Processing, Computer Systems & Programming, Digital Signal Processing, Computer Systems Engg, Digital Systems Laboratory. **Math:** Differential Eq, Fundamental Math, Probability with Engg, 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 video overviews on iOS and Android generating a sig-stat increase in 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 - Aug 2021
Software Engineering Intern (Commerce Shop Rankings Team) - Mr. Artem Zinchenko 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%.
- 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

C3SR Scholar IMPACT Lab - Prof. Wen-mei Hwu, Dr. V.S. Mailthody Aug 2021 - Present

- Generated the largest publicly available graph dataset - Illinois Graph Benchmark (IGB) with 600M nodes and 6B edges collaborating with researchers from Amazon AWS, Deep Graph Library (DGL), NVIDIA, and IBM for supervised 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.

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

- Created a 2-stage attention-based seq2seq model to generate subtopics for a given title that performs better than SOTA models for fine-grain attribute tagging.
- 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 π 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 π CF digits and used extreme-digit, single-digit, and $z - scores$ for further evidence.

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.

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 using NLP (like the Russia-Ukraine conflict) in order to (1) extract a list of needed resources, (2) how they are fulfilled.

Undergraduate 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 (VR) using Unity with anatomically accurate physics scripts, capable of real-time rendering optimized to run without a GPU.
- Implementing a reinforcement learning (RL) 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 model 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 Engg Applications (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 homeworks 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 freshmen and sophomores 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 </> (among 500 people) 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 </> (among 300 people) 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 auto-grading time by 50% utilizing better algorithms and libraries.

PUBLICATIONS, POSTERS AND PREPRINTS

- **Khatua, A.**, Mailthody, V., Taleka, B., Song, X., Ma, T., Bigaj, P. & Hwu W. IGB: An Immense Graph Dataset for Machine Learning Workloads. *International Conference On Machine Learning, 2023 (In Preparation).*
- **Khatua, A.**, Agarwal, A. & Chang, K. Generating High-Level Article Structure based on Topic using Two-stage seq2seq Model. *Annual Meeting Of The Association For Computational Linguistics, 2023 (In Preparation)*
- **Khatua, A.***, Li, X.*, Garitsis, E. & Hildebrand, A. Large Scale Statistical Analysis of the Randomness of π Continued Fraction Digits. *Experimental Mathematics 2022 (In Submission).*
- *(Poster Session)* Jin, A.*, **Khatua, A.***, Li, X.*, Singh, S.*, Zhang, Z.*, Garitsis, E. & Hildebrand, A. Analysis of the Continued Fraction Digits of π. *MAA MathFest 2022.*
- *(Poster Session)* **Khatua, A.**, Mailthody, V. & Hwu, W. Generating Large Synthetic and Real Graph Datasets. *Illinois Undergraduate Research Symposium 2022.*
- *(Poster Session)* **Khatua, A.**, Worthey, G., Capitanu, B. & Downie, J. Detection of Page Objects from the HathiTrust Digital Library. *Engineering Open House 2022*

TECHNICAL TOOLS, LEADERSHIP & HONORS

Tools: ***Very Knowledgeable:*** Python, React, Hack/PHP, GraphQL, 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 Eng. (PURE) Vice President&Corporate Director President Toastmasters Club.

Honors: O. Thomas and Martha S. Purl Scholarship ($3,400 awarded to 2 students out of 2000) • Illinois Engineering Achievement Scholarship ($1,000) • NCSA Student Pushing Innovation Fellow ($6,000 research sch.) • C3SR-URAI Scholar ($1,000) • Illinois Geometry Lab Scholar (×2) • Edmund J. James Scholar • Dean's List • PURE Scholar • HCESC Jump ARCHES ($6,000 research sch.) • COVID-19 Wall of Recognition in Engineering .

REFERENCES

Prof. Wen-mei Hwu - w-hwu@illinois.edu - Professor UIUC, Senior Distinguished Research Scientist NVIDIA
Prof. Kevin Chenchuan Chang - kcchang@illinois.edu - Professor UIUC
Dr. Vikram Sharma Mailthody - vsm2@illinois.edu - Research Scientist NVIDIA
Mr. Gabriel Ochoa - gabeochoa@meta.com - Team Lead, Live Videos, Facebook, Meta Inc.

Updated November 2022