DIPRO RAY

202 S. 4th St., Champaign, IL 61820 • 760-216-3327 • dipror2@illinois.edu
diproray.github.io • github.com/diproray • linkedin.com/in/diproray

EDUCATION

University of Illinois at Urbana-Champaign August 2017 – May 2021
Bachelor of Science (Engineering), Computer Science **Cumulative GPA: 3.99/4.00**
Minor, Mathematics

- Taken: NLP, Algorithms, Data Structures, Systems Programming, Software Design Studio, Bioinformatics, Compilers, Number Theory, Combinatorics, Ethics.
- Senior Thesis (completed junior year): "Leveraging the Pattern-Relation Duality for Domain-specific Keyphrase Mining"
- Standardized Tests: GRE 331/340 (163 Verbal, 168 Quant, 5.0 AWA)
- Awards**:** IMC Trading Scholarship 2020-21, Grainger Engineering Scholarship 2020-21, Mrs. E. J. Hoover Scholar 2020, Tau Beta Pi Outstanding Junior 2020, NCSA Student Pushing Innovation Fellow 2018-19, Edmund J. James Scholar and Dean's List (all semesters).

WORK EXPERIENCE

Facebook, Inc. August 2020 – November 2020
Software Engineering Intern – Speech Infra Team, AI Applied Research Org Menlo Park, CA (remote)

- Built a real-time on-demand automatic speech recognition (ASR) service latency tracing tools for Live Videos. Worked on Facebook's complex low-latency ASR pipeline to record time taken at expensive stages in the service in C++. Created the frontend of the tool using Hack/PHP to (1) inject a user-entered video ID into a RocksDB (which is read from in the backend service to enable logging) (2) display latency graphs for the video's lifetime that pinpoints where delays occur.
- Improved several internal debugging tools concerning live video captioning by pulling logs from real-time database services. Created a new internal tool for video caption debugging by comparing ASR results with user-uploaded SRT caption files.

Airbnb Summer 2020
Software Engineering Intern Recipient (Deferred due to COVID-19) Seattle, WA

Facebook, Inc. May 2019 – August 2019
Software Engineering Intern – Typeahead Team, Search Org Menlo Park, CA

- Performed feature engineering on two people search and usecase ML models to reduce their original features capacity by 67%, ran their A/B tests on 25+ million people which recorded stat-sig online metric gains for app search engagement.
- Improved internal simulator tooling for Typeahead engineers using Hack, PHP and Javascript. Impact: improved rendering speed by 10x; added functionality to view NLP signals and backend request/response, and compare ranking features data.
- Onboarded the people search typeahead ML model to DAG workflow for easier ML model lifecycle management.

National Center for Supercomputing Applications August 2018 – May 2019
Software Engineering and Research (SPIN) Intern – High Performance Computing for Genomics group Champaign, IL

- Fully automated a developed statistical analysis and pipeline design through R code, optimizing it through fast matrix calculation, forking and multi-threading, and building a GUI for the pipeline.
- Upgraded the statistical pipeline in place for preprocessing research data by integrating Apache SparkR usage, containerizing for cloud deployment and open-sourced through Github.

RESEARCH EXPERIENCE

FORWARD Data Lab August 2019 – present
Undergraduate Researcher & Senior Thesis, Advisor: Prof. Kevin Chen-Chuan Chang Champaign, IL

- Developed an automated, domain-based, keyword detection system based on the Pattern-Relation-Duality (PRDualRank) framework. Implemented the system in a multithreaded manner in Python using spaCy, NLTK.
- Developed novel precision and recall evaluation metrics for the system. Achieved comparable performance against state-of-the-art AutoPhrase phrase detection system, with significantly lower complexity overhead.
- Exploring extensions to the PRDualRank model by exploring graph-based SSL techniques and arbitrary window size POS-tag based patterns.

Laboratory for Parallel Numerical Algorithms October 2019 – present
Undergraduate Researcher, Advisor: Prof. Edgar Solomonik Champaign, IL

- Conducted literature survey on computationally efficient methods for Canonical Polyadic (CP) Decomposition of sparse, low-rank tensors, and analyzed sparse decomposition performance on SPLATT and Cyclops Tensor Framework.

- Analyzed Tucker decomposition-based tensor completion techniques for higher-dimensional tensors, and developed code implementation with Numpy and CTF. Benchmarked performance against FTCom, P-Tucker and CTF CP-ALS tensor completion code for the CamVid dataset and synthetic tensors.
- Aiding in data preprocessing and preparing a COVID-19 data tensor of time, zipcode, ICD code for spatiotemporal analysis through sparse tensor decomposition on the Stampede2 supercomputer (to observe latent statistical features).

Blender Lab (CS REU Program) May 2020 – August 2020
Undergraduate Researcher, Advisor: Prof. Heng Ji and ChengXiang Zhai Champaign, IL
- Worked on building a joint end-to-end neural network for biomedical event extraction. Explored new embedding approaches for biomedical text using KG-based embeddings and BioBERT.
- Worked on KG and IR-based approaches to the TREC-COVID ad-hoc retrieval task.

Illinois Geometry Lab (Mrs. E. J. Hoover Program) June 2020 – July 2020
Undergraduate Researcher, Advisor: Prof. Sean English Champaign, IL
- Explored Turan-type problems in extremal graph theory.
- Studied *c*-loose path saturation problems for paths in *k*-uniform hypergraphs.

Illinois Geometry Lab August 2019 – December 2019
Undergraduate Researcher, Advisor: Prof. Richard Sowers Champaign, IL
- Detected broad patterns in traffic activity, parking behaviors in large metropolitan cities by applied unsupervised ML (low rank factorization) on relevant datasets collected from San Francisco and New York..
- Developed analytical software for this purpose and visualizing software to interpret obtained results working mainly in Python, scikit-learn and other data science libraries.

TECHNICAL TOOLS

Very Knowledgeable: C, C++, Python, Java.
Knowledgeable: Git, SQL, Hack/PHP, MongoDB, Haskell, JavaScript, LaTeX, Data/ML/NLP Python libraries.
Somewhat Knowledgeable: React, Neo4j, Racket, Verilog, MIPS.

TEACHING EXPERIENCE

Discrete Structures (CS 173) Summer 2020
- Graded assignments and assisted discussions for discrete math topics.

Algorithms & Models of Computation (CS 374) Fall 2019 and Spring 2020
- Graded assignments and conducted discussions sections in topics related to automata theory, algorithms, complexity theory.

Probability and Statistics for Computer Science (CS 361) Fall 2018
- Graded assignments and developed course content for a course dealing with probability, statistics and machine learning.

PROJECTS

Spotify Indexer September 2019 – December 2019
- Created a web application that analyzes Spotify music-listening trends of the user and assigns him/her a score based on how similar their tastes are to what is mainstream.
- Made the application highly scalable using AWS, DynamoDB. Reduced the score computation time by 50% by incorporating richer information within the database.

EXTRACURRICULAR INVOLVEMENT

Reflections | Projections Tech Conference 2018, 2019, 2020
Corporate Committee Champaign, IL
- Served on Corporate staff for 3 years for R|P, the largest student-run tech conference in the Midwest. Served as a Corporate member in 2018, Corporate chair in 2019 and Corporate advisor in 2020.
- Brought in a total of $150,000 in sponsorships and 60 companies to the conference over three years.

Other Involvement: Tau Beta Pi • ACM • Human Capital • WCS Bits N Bytes Mentoring • ACM Freshman Mentoring