dhruba.j.paul@gmail.com
linkedin.com/in/dj-paul

Dhruba Jyoti Paul

(425) 766-7356
djpaul.dev

Education

Madison, WI **University of Wisconsin-Madison** **Sep 2020 - Aug 2023**

- **Bachelor of Science** in Computer Science and Data Science with a CGPA of 3.94/4.00 (Dean's List, Distinction in Major)
- **Coursework**: Artificial Intelligence; Matrix Methods for ML; Algorithms; Machine Organization; Programming III; Discrete Mathematics; Differential Equations; Linear Algebra; Multivariable Calculus; Data Science II; Statistics and Data Modeling II; Operating Systems; Virtual Reality; Theory and Design of Programming Languages
- Inducted into **PBK Honor Society**; invited by Dean of College of Letters and Science, recognizing top 0.4% of undergraduates.

Publication

Pister, K., **Paul, D.J.**, Brophy, P., Joshi, I. (2024). *PromptSet: A Programmer's Prompting Dataset*. ICSE '24
Rzig, D.E., **Paul, D.J.**, Pister, K., Henkel, J., Hassan, F. (2024). *PromptDoctor: Toward Automated Prompt Linting and Repair.* WIP

Work Experience

Machine Learning Engineer **ALL3D** **Sept 2024 - Present**

- Improved model inference speed by 3x while reducing resource utilization by 2/3, resulting in performance gains and cost savings
- Reverse-engineered a relevant product in record time, providing a competitive advantage and informing business decisions
- Designed and set up geometries to enhance high-quality synthetic data generation for training stable video diffusion models
- Integrated ML models into scalable production pipelines and designed evaluation pipelines for 3D model reconstruction

NLU Research Assistant **University of Wisconsin-Madison** **Oct 2023 - Sept 2024**

- Utilized various NLP techniques to generalize a process to optimize prompts, improving results by 10% on synthetic datasets
- Extracted and conducted static analysis on about 60.7% of the open-source API-based LLM usage on GitHub
- Categorized clusters of developer prompts in PromptSet, generated through t-SNE and K-Means of the prompt embeddings
- Explored prompt detection strategies like testing heuristics with Tree-sitter and fine-tuning flair NLP framework for text classification
- Proposed static analysis methods to improve prompt quality and reliability within software development pipelines

Co-Instructor **Microsoft TEALS** **Dec 2023 - May 2024**

- Supported an initiative to extend comprehensive educational support and actively foster student engagement outside the traditional classroom environment through the development of a RAG-based Discord bot hosted on GCP
- Developed and delivered engaging lectures to enhance student participation and learning outcomes
- Rapid Issue Resolution. Consistently addressed at least 96% of student problems within 6 hours, ensuring timely support.

Software Architect **4P Marketing Consultancy** **Jan 2024 - Apr 2024**

- Designed a data pipeline using geofencing and FaceNet facilitated facial detection with a Flask server and an Android app
- Enhanced data security via Fourier transformations for pixel pattern detection in digital data and AES encryption
- Automated data entry using Google Vision OCR and NER models, enhanced by Levenshtein distance-based heuristics

Projects

- **Tagore GPT** (2024). A simple language model based on the paper "Attention is All You Need" and OpenAI's GPT-2, trained on a custom dataset of literary pieces by Bengali poet and writer, Rabindranath Tagore. ***Python, PyTorch***
- **Face Emotion Classifier** (2023). Classifies faces by emotion with a 3-layer neural network. Trained using stochastic gradient descent. Accuracy estimated through 8-fold cross-validation. ***Python, Keras, Tensorflow***
- **Runscan** (2023). Recover image files from ext2 disk images by analyzing inodes and data blocks to identify file type and content by checking file signatures for known JPG header patterns. ***C, debugfs, mkfs***

Technical Skills and Frameworks

- **Languages**: Python, R, Java, C/C++, HTML, CSS, JavaScript, SQL, x86 assembly, PHP
- **Frameworks and Tools:** Numpy, Pandas, Playwright, Flask, Git, MySQL, SQLite, Scikit-Learn, PyTorch, SciPy, Keras, LangChain

Additional

1st Place, CheeseHacks Hackathon (2022). Built Facial Detection Attendance Tracker using cosine similarity (ResNet)
Florence Waste Pulver Scholarship (2022). Merit Scholarship awarded for academic excellence.
UW-Madison Undergraduate Scholarship for Summer Study (2022-2023). Merit Scholarship. Awarded 2 consecutive years.