Vivek V Kashyap

vivekvkashyap10@gmail.com | Website | Linkedin | Github

EDUCATION

Bangalore Institute of Technology   **August 2019 - August 2023**
Bachelor of Engineering in Electrical & Electronics   **CGPA** : 8.42/10.0

PUBLICATIONS

- **"Constructing Bayesian Pseudo-Coresets using Contrastive Divergence"** Piyush Tiwary, **Vivek Kashyap,** Shubham Kumar, Prathosh AP. **Accepted UAI 2024**   **Paper**
- **"GPT Neo for Common Sense Reasoning - A Theoretical and Practical Lens"** **Vivek Kashyap,** Rohan Kashyap   **Paper**

EXPERIENCE

Indian Institute of Science   **January 2024 – Present**
Research Assistant, Advisor : **Dr. Pawan Goyal, Dr. Prathosh AP**
- Collaborating with **IIT Kharagpur** and **IIT Delhi** on **LLMs** using **constraint decoding** to translate **English-Sanskrit poems** in the **Anushtup meter** from a **low-resource dataset**, achieving a **BLEU score of 30**.
- Working on an architecture for **automatic symmetry groups** to handle **invariant symmetries permutation subgroups** into **neural networks**.

Indian Institute of Technology Bombay   **June 2023 – Nov 2023**
Research ML Intern, Advisor : **Dr. Ganesh Ramakrishnan**
- Implemented and analyzed various models to build a **Llama chatbot model on health,** model training was done on 8 **million sentences** to provide **reliable information to users**.
- Optimized **Retrieval-Augmented Generation** methods for **faster inference** without **compromising accuracy**, incorporating **re-rankers** and **filtering** techniques to achieve optimal results.

Indian Institute of Science   **June 2022 – May 2023**
Research Intern, Advisor : **Dr. Prathosh AP**
- Developed a **novel loss function** combining **Contrastive Divergence** and **Energy-Based Models**, achieving **93% accuracy** using **10 training images.** Accepted at **Uncertainty in Artificial Intelligence (UAI) 2024**.
- Designed a **compute and memory-efficient model, reducing GPU usage by 50%** while delivering SOTA results on out-of-distribution datasets and cross-architecture performance for various image datasets.

Omdena   **March 2021 – July 2021**
ML Engineer, Optimiz Claims
- Worked directly with the **CEO** to design a **machine learning model** for **identifying similar damaged claims**, leading to a **70% reduction** in **company costs**.
- Applied **classification decision trees** to estimate cost retrieval for damaged cargo claims with **78% accuracy**, leading to **15% increase in customer traffic** and a **rise in revenue**.

RESEARCH PROJECTS

Neural Machine Translation Using Memory Efficient Transformers   **January 2021**
- Worked on memory efficient Transformer models to perform machine translation on Portuguese-English sentences, achieving an accuracy of 93% and performed assessment of computational and memory factors.
- Analyzed the impact of positional embedding and rotary embedding on transformer architecture, leveraging the PennTreeBank Dataset to evaluate key factors such as FLOPs, speed, and memory usage, yielding valuable findings.

Diffusion Transformer: Advancing Image Synthesis with Self-Supervised Learning   **March 2023**
- Built a Diffusion Transformer from scratch by combining diffusion probabilistic models with transformers for high-fidelity image generation, achieved a 15% improvement in FID scores over GANs.
- Leveraged self-supervised learning to minimize labeled data needs and applied contrastive learning for improved robustness and scalability. Reduced training instability and mode collapse common in other generative models.

TECHNICAL SKILLS & COURSEWORK

Mathematics & CS   Probability & Statistics, Linear Algebra, Calculus, Differential Equation, Optimization, Introduction to Programming, Data Structures & Algorithms, Python, Pytorch, Tensorflow.

Machine Learning   ML Algorithms, Transformers, Large Language Models, Diffusions, Computer Vision, Advanced Representation Learning, Advanced NLP, MLOps, Feature Engineering.