Viraj Bagal

+91-7219652039 | virajbagal12@gmail.com | LinkedIn | GitHub |

About Me

Experienced NLP and Computer Vision professional with **3+ years** of expertise in both research and engineering. Proficient in developing and productionizing models for diverse problems such as **summarization, text classification**, object detection, and segmentation. **3 Research papers** on multimodal data understanding and generative NLP published in international conferences like WACV, ISBI, JCIM. Demonstrates exceptional skills as both an independent contributor and a valuable team player. Excellent problem-solving and communication skills, with the ability to work effectively in a cross-functional team.

Publications

- Viraj Bagal et al., 'MMBERT: Multimodal BERT Pretraining for Improved Medical VQA', ISBI 2021.
- Minesh Mathew, Viraj Bagal et al., 'InfographicVQA', WACV 2022.
- Viraj Bagal et al., 'MolGPT: Molecular Generation using Transformer-Decoder Model', JCIM 2021 (AAAI workshop 2021)

Experience

Machine Learning Engineer April 2021 – Present

Synapsica Healthcare — YC W20 *Bangalore, India*

- Manage end-to-end deep learning pipeline covering data cleaning, transformation, model development, experiment tracking, model optimizations, evaluations, deployments.
- In NLP, I developed and productionized models on problem statements like **entity tagging, summarization, language generation, text classification** using models like BERT, RoBERTa, Flan-T5.
- Implemented batch processing instead of sequential processing and **ONNX runtime accelerator** in production pipeline. This resulted in **30%** and **65%** reduction in production pipeline runtime respectively.
- Instruction finetuned open source commercial licensed **LLMs** like **Falcon** using **Huggingface Transformers** for tone changes and conversion of medical text in different formats. Used PeFT techniques like **LoRA** for faster and memory efficient training.
- Used **Huggingface Accelerate** for distributed training and, **Huggingface Optimum** for optimizing and quantizing the model for inference
- Developed and productionized around 10 models solving problem statements like image classification, object detection, segmentation, key point detection, ranking in medical images using models like **EfficientNets, YOLO, SAM, Transformers**
- Tech stack used: Pytorch, Python, HuggingFace, Spacy, Docker, Weights and Biases, AWS

Deep Learning Research Intern May 2020 – April 2021

Generative NLP, CCNSB Lab, IIIT *Hyderabad, India*

- Developed custom transformer decoder model similar to **GPT** for conditional molecule generation. It is **94%** smaller and achieved new state-of-the-art results (increase in performance) on conditional molecular generation.
- Implemented RNNs, LSTMs, Graph models for performance comparison against our model.
- Shorter version of research paper accepted at **AAAI-SDA 2021** workshop. Longer version accepted in **Journal of ChemInformatics (JCIM)**. Virtually presented my work at AAAI 2021 (Conference H5-index: **126**, Impact Score: **25.57**). Click here for the paper. Click here for repo.

Deep Learning Research Intern May 2020 – April 2021

Multimodal (CV + NLP) Understanding, CVIT Lab, IIIT *Hyderabad, India*

- Proposed and implemented a novel interpretable visual question answering (VQA) model on medical images, questions and answers.
- The model achieves new state-of-the-art performance with increase in accuracy and bleu score by **5%** while being **66%** more efficient than previous best models.
- Implemented **self-supervised training** with Masked Vision-Language Modeling and Image-Text Matching on multimodal **BERT** model using **multi-GPU DDP trainig, HuggingFace, Pytorch Lightning, and monitored results using wandb (W&B)**.

- Research paper accepted at **IEEE ISBI 2021** (Conference H5-index: **43**, Impact Score: **6.6**). Click here for the paper. Click here for repo.
- Modified and trained **LayoutLM** model to perform VQA on Infographics rather than Document. Click here for paper

Education

Indian Institute of Science Education and Research — Pune, India
MS/MSc in Physics, Minor in Mathematics. ***GPA: 9.3/10*** — *Aug. 2016 – June 2021*

Projects

InsightAI: AI based Content Insights — April 2023

- Developed **LLM** based content insight product that can summarize and allow Q&A on any CSV, PDF, Doc, Image, Youtube video.
- Used **OpenAI embeddings** for creating embeddings of the content.
- Used **Activeloop Deeplake** as the vector database and **retrieval augmented generation** for Q&A.
- Performed **prompt engineering** for **ChatGPT and GPT-4 APIs**.
- Used **langchain** to perform all this.
- APIs for file-transfer, processing, etc. using **FastAPI**. Frontend using **Streamlit**
- Deployed on **AWS EC2** instance using **docker compose** and **traefik** for reverse proxy. Click here for product
- Click here for the code

MLOps — October 2021

- Model monitoring using Weights and Biases, and Training configuration setup using Hydra.
- Data Version Control using DVC and Model Packaging using Fast API + ONNX + Docker.
- CI/CD using GitHub Actions, and created Container Registry using AWS ECR.
- Serverless Deployment using AWS Lambda and Prediction Monitoring using Elasticsearch Cluster + Kibana.

Achievements

- Secured **All India Rank 69** in KVPY 2016.
- Secured **All India Rank 2302** in JEE Advance 2016.
- **National Top 1%** in National Graduate Physics Examination 2019.
- 2× Kaggle Expert. Only **8%** of total Kaggle competitors are at this or above this rank
- Three publications. One in IEEE ISBI 2021, one in WACV 2022 and one in Journal of ChemInformatics (JCIM).

Skills

- **Technical**: Deep Learning, Natural Language Processing, Computer VIsion, Python, Pytorch, Docker, AWS EC2, AWS ECR, AWS Lambda, S3, MLOps, CI/CD, FastAPI, Prompt Engineering, MongoDB, LangChain, Transformers, Accelerate, Optimum
- **Non-Technical**: Excellent at verbal and written communication. Have experience in managing a team of 2 people at the current company. Excellent as a solo as well as a team player in terms of contributions
- **Open Source Contribution**: Made contributions to **Albumentations** library (widely used in Computer Vision) and **Pytorch Lightning**.