

AI/ML Architect specializing in **NLP, LLMs**, and scalable deep learning systems. Expertise in **Generative AI, model fine-tuning**, and deploying large-scale production-ready AI solutions

PROJECTS

❖ **CURIE – Intelligent Assistant** for Actionable Insights - (*A **flagship model** from DRAUP*)

- Understands user queries & intent using custom models, performing deep searches across multiple **RAG applications (30M+ embeddings)** with Milvus and several **Text-to-SQL systems** to extract insights (**2M+ table entries**).
- **Combines RAG and Text-to-SQL outputs** using **Chain-of-Thought** reasoning with custom models and LLMs for context-aware insights. Optimizes real-time retrieval and response generation for accurate, intelligent assistance.

❖ **LLM-Powered News Engine**: Built a scalable news processing system similar to Google News by developing in-house **transformer-based language models**

- Continual pre-trained a **RoBERTa**-based transformer language model on (10M+ news) 550M tokens, reducing loss by 60% & improving various downstream NLP tasks like **classification, NER, semantic embedding, recommendations**, etc.
- Built **an in-house scalable clustering algorithm** to generate news trend recommendations at scale.
- Led a team in training **Mistral-7b** with **LORA** techniques for “**relationship extraction**” from diverse news articles.

❖ **Scaling Profile Recommendations (~400M profiles)**:

- Implemented a **semantic similarity search system** using multiple language models and the Milvus vector database. Embedded 400M+ profiles and **fine-tuned a BERT-based sentence transformer** for efficient matching.

- ❖ **Automated Job Responsibility Extraction System:** Led the development of an automated system to extract and summarise job responsibilities for professionals.
 - Trained and implemented multiple machine learning models to analyze job postings, and developed a **Seq2Seq BART model** for large-scale responsibility **summarisation**, removing the need for manual work.
 - Utilized data-driven ranking and summarisation to enhance profile context, significantly improving the accuracy of work experience representation for millions of profiles.
- ❖ **Information Retrieval (IR) System for Domain-Specific URLs**
 - Designed and implemented a machine learning system utilizing **XGBoost and Random Forest** algorithms
 - Engineered multiple Machine Learning tools for **effective annotation** of training data, including the creation of an [HTML Tag Annotator](#)

EMPLOYMENT

DRAUP Inc. (Oct 2018- present), **Data Science Architect**

- ❖ Lead a **12-member Data Science team**, driving the adoption of ML and deep learning technologies for real-world applications.
- ❖ Collaborate with **product, big data, and research teams** to develop AI-driven solutions for business and client needs.
- ❖ **Designed, deployed, and optimized** multiple ML/DL/LLM models in production on AWS ECS, leveraging GPUs and **Apache Spark** (AWS EMR) for large-scale processing..
- ❖ Built and scaled Retrieval-Augmented Generation (**RAG**) applications and Text-to-SQL systems, integrating **vector databases**, trained embedding models, and LLMs to enhance NLP **search, retrieval, and query** generation.

CALIENT Technologies Inc.,(Oct 2017 - Sep 2018), **Software Engineer - Machine Learning**

- ❖ Built Machine Learning models to categorize network device error logs into multiple categories (**word2vec + logistic regression**) and implemented an alarm system using **Java** for quick response.

COGOPORT, (Feb 2017 - Sep 2017), **Software Developer**

- ❖ Designed and Developed Bid Request System including **data analytics pipeline** and API development using **Ruby on Rails**.

TECHNICHE e-commerce solutions, (Apr 2015 - Feb 2017), **Full Stack Developer**

- ❖ Proficiently designed and developed multiple **e-commerce websites** using **Node JS (MEAN stack)**, encompassing essential features such as user authentication, authorization, shopping cart functionality, and comprehensive order management

Hewlett-Packard(HP), (Aug 2013 - Dec 2014), **Technical Consultant**

- Created **Java** applications for internal use, utilizing extensive SQL expertise for banking project

Skills, Languages & Technologies

- **Python, Apache Spark, PyTorch, TensorFlow, HuggingFace** 🙌, **Scikit-Learn, Keras, Spacy, BetterTransformer,**
- Traditional ML, DL & NLP algorithms like **Neural Networks, LSTMs, Transformers (BERT, BART, RoBERTa, GPT),** etc
- AWS ecosystem: EMR, Sagemaker, ECS, S3, EC2.

EDUCATION

Bachelor of Engineering (B.E.) in Computer Science (2009 - 2013) - VTU