Kumar Abhishek

Senior Undergraduate, Mechanical Engineering

in kumar-abhishek | kabhishek20
+91-8797090479 | kabhishek20@iitk.ac.in

Academic Qualifications

Year	Degree/Certificate	Institute	Performance
2020 - 2024	B. Tech.	Indian Institute of Technology Kanpur	7.5/10
2020	Class XII (CBSE)	Delhi Public School, Ranchi	95.60%
2018	Class X (ICSE)	Bishop's School, Ranchi	96.00%

Internship / Work Experience

Dr. Reddy's Laboratories, Hyderabad | Data Science and Analytics Intern | *Received **PPO** for valuable contribution* | *May'23 - July'23*

Competitor's Warning Alert, using Google's PaLM-2 Large Language Model

Objective	• **Scrap articles** related to FDA inspection from various **media sources** to predict **warning letter** before it is officially published • Extract all **tokens** relevant to USFDA inspection from these articles and return that in the form of a daily **alert message**
Approach	• Scraped websites like **FiercePharma, MedicalDialogues, EconomicTimes** to make **realtime database** with 100+ FDA articles • **Filtered and Preprocessed** the scrapped contents, passed it as context to **Pathways Language Model** with 540B parameter • Employed **Prompt Engineering** to structure the query, implemented **few-shot learning** to effectively develop the prompt
Outcomes	• Articles from all listed sources were correctly **filtered**, and **Alert Message** of any new article in the database was set up • **PaLM-2 model** was implemented successfully, it extracted **tokens** relevant to USFDA and returned in an **Alert Message**

Drug Shortage Classifier, using Google's BERT AI-Language Model

Objective	• **Categorize the drugs** currently experiencing shortage in the market by utilizing the official **USFDA drug shortage database** • Provide **real-time updates** about these drug by **mailing CSV file** of the complete drug shortage list with all details **weekly**
Approach	• Scraped out the USFDA website, saved the drugs information with **3000+ values** and performed **data labelling manually** • Imported **BERT masking and encoding layers** from **tensorflow-hub**, **fine-tuned** it on the manually labelled classification data • Added **dropout layer** to prevent **overfitting**, **saved the fine-tuned model** and used it for **real-time drug shortage classification**
Outcomes	• **Fine-tuned BERT Large Language model** was able to perform real-time shortage classification with an **accuracy of 97.40%** • Precision and recall value were **0.97 and 0.98** respectively, it also made a CSV file with details of all drugs in shortage

Student Undergraduate Research Graduate Excellence(SURGE) | Research Intern | *May'22 - July'22*

- Worked on developing a system for **labelling point cloud data** taken from laser lights over a **moving rover** for **deep learning** purposes
- Labelled laser data into 10+ classes with the use of **CloudCompare**, an open source software for data labelling and found its flaws
- Contributed in making a **plugin** to make data labelling in CloudCompare easier, by using a **polygon** instead of cube for classification.

Kalvi Career Education Private Limited | Front-end Web Development Intern *Jun'22 - Aug'22*

- Worked on developing **cutting-edge labs** and practice exercises by utilizing tools like **HTML, CSS, JavaScript, React** for learning purpose
- Coordinated with a team comprising of **5 members** to create daily projects based on **industry inputs** to foster outcome based learning
- Involved in building engaging lectures and practice lessons to help other grasp concepts using **Markdown** and **Stackblitz** platform

Self Projects

Chess Pieces Detection and Labelling using YoloV5 | Computer Vision Project | *(July'23)*

- Collected **30 snapshots** from Lichess, **annotated** them using **CVAT** tool and made dataset in **YAML** format for YoloV5 to understand
- Trained the model for **6 classes**, **1381 epochs** with **1100, 180, 200, 170, 220, 250** instances of **Pawn, King, Knight, Queen, Bishop and Rook**
- Tested and **Stored the weight** of the model which had the **mAP(mean Average Precision) of 99.30%, recall and precision of 0.971 and 0.933**

Document Question-Answering using davinci model of GPT3 | LLM Project | *(Jun'23)*

- Utilized **Streamlit** as **GUI interface** to upload PDFs with limit of **200MB**, text was extracted from PDF using **PyPDF2** library of python
- Exploited **OpenAI embeddings** to create text embeddings, **FAISS package** from facebook was used for **semantic search** and get response
- Implemented **davinci engine** of **GPT3** for getting answers using semantic search, built question-answer chain using **langchain package**

Realtime Facial Emotion Detection | Deep Learning Project | *(July'22)*

- Trained a **Convolutional Neural Network (CNN)** model with **28,709 images** to recognize **5+ different** real-time facial emotions correctly
- Utilized **OpenCV** packages to automatically **detect faces** in **images or live webcam feed** and draw **bounding boxes** of 2 px around them
- **Tested the trained model** on live feed through system's webcam with 7 different facial expressions and achieved an **accuracy of 84%**

Technical Skills

Languages: C, C++, C#, Python, HTML, CSS, JS, SQL	**Libraries**: Pandas, Numpy, Matplotlib, Streamlit, Tensorflow, Sklearn
AI models: GPT3, Llama-2, GPT3.5, Flan-T5, BERT, PaLM-2	**Utilities**: Microsoft Office, PowerBI, Anaconda, Github, GCP, Bigquery

Position Of Responsibility

Student Guide, Counselling Service Team	**Senior Executive**, Udghosh Events Team
Programming Secretary, Game Development Society	**Web Secretary**, Society Of Civil Engineers

Relevant Courses

[i] Exceptional Performance [*] Online [o] Ongoing

Python for everybody[*]	**Data Science, Codebasics**[*]	Data Structure and Algorithm[o]
Fundamentals of Computing[i]	Linear Algebra and differential equation[i]	Real Analysis and Multivariable Calculus
Applied Probability and Statistics	**Natural Language Processing(NLP)**[*]	**Traditional Machine Learning(ML)**[*]