Pankti Fadia

Data Scientist

pankti.fadia.99@gmail.com 8401976784 linkedin.com/in/pankti-f-97687b144 github.com/Pankti99

Experience

DXFactor Solutions Pvt. Ltd. — *2022-present*

AI based Medical Prior Authorization Letter Generation

Technologies and libraries: Gen AI, Qdrant

Developed an AI-powered solution to automate the generation of prior authorization letters for medical procedures based on patient clinical notes and medical history. Using prompt engineering a draft letter is generated, ensuring it complies with medical guidelines. Additionally, the letter is enhanced through integration with external medical guidelines and reference searches. By using embedding techniques, the system pulls information from trusted medical sources, improving the quality and accuracy of the letter. The final output includes properly formatted citations to support the medical necessity of the procedure.

Fitness AI chat bot

Technologies and libraries: Flask, sqlite3, chroma db, Streamlit, JS,Gen AI

Developed a chat bot solution for a fitness gym client using OpenAI API and prompt engineering techniques. Integrated Chroma DB for efficient querying of client information. Developed a Streamlit dashboard for visualizing data and analyzing chats. Incorporated an SQLite3 database for efficient storage of chat logs and associated information. Utilized a Node.js API to retrieve class schedules.

Features:

- Tracked user journeys to analyze conversions for gym membership.
- The context aware agent analyzes the user's interaction with website to offer personalized and relevant information, enhancing the user experience.
- Managed persistent chat and sessionIDs across all pages of the website.
- Implemented privacy-preserving mechanisms to keep personally identifiable information (PII) safe.
- Dynamically generated new ad pages, with an intelligent agent providing information based on user search query and dynamic UI.

Oyster Cage detection

Technologies and libraries: ArcGIS pro, QGIS, Azure Blob Storage Container, S3 browser

Detected ocean gears mainly - oyster cages in Chesapeake Bay using airplane and drone imagery. Used YOLOv5 for accurate cage detection. Calculated total count and area of cages using geo- processing and image processing techniques. Generated output in shape file format for seamless integration with GIS systems. Created Streamlit app for drone imagery. Optimized entire pipeline by implementing multi-threading to download large scale images.

Challenges and Solutions:

1. Small and blurry cages:Improved results by using SAHI during inference.
2. Number of images: Reduced unnecessary land images from 800,000 to 200,000 using a shoreline layer, which significantly reduced costs and processing time.
3. Image extraction: Leveraged ArcGIS Pro for image extraction, generating grid files and cropping based on tile orientation.
4. Image downloading time: Extracting a single image from ArcGIS took nearly 50 seconds due to TIFF file format. Reduced time to 2 seconds by generating dynamic URLs with cropped image extent values.

Deployment:
Used HPC cluster for image extraction and prediction. Implemented TORQUE job scheduler.

GPT bot:
Created "Text to Chart" wrapper around GPT using Open AI API for fetching and displaying population data for blue catfish. Did prompt engineering to improve the results.

LinkedIn and Instagram Automation and Targeted Lead Generation

Technologies and libraries: Gen AI, sqlite3, Apollo, Selenium, Rapid API

Developed a LinkedIn automation script using Apollo and Selenium to automate profile following, connection building, comment generation, automatic comment posting, messaging for outreach, and lead generation.

Additionally Implemented a lead generation project aimed at identifying potential clients for product expansion. The goal was to leverage LinkedIn to identify companies actively hiring for positions relevant to our product and identifying Ideal Customer Profile (ICP) brands within this pool and initiating outreach efforts. This innovative solution optimized outreach efforts, boosting engagement and lead acquisition.

Document Tampering Detection

Technologies and libraries: Streamlit, GenAI, Adobe, Computer Vision

Conducted a series of experiments focused on identifying fraudulent activities within bank statements through innovative document analysis techniques. Experimented with computer vision methods and image processing for detecting inconsistencies and alterations in financial documents. Utilized the Adobe API to extract detailed metadata for analysis. Analyzed patterns within fraudulent documents to uncover common characteristics and indicators of tampering.

GPT based Chat bots

Technologies and libraries: Looker studio, Flask, OpenAI, Lang chain, AWS, Streamlit, JS

1.Created GPT based Chatbot for fitness industry- PT slot booking using OpenAI API, prompt engineering and flask APIs. Created analytics dashboard for the Chatbot in looker studio.

2.Created a Task Management Bot - that retrieves project updates from CSV files and inserts new updates using Lang chain and OpenAI API. The bot has been deployed on AWS.

3.Created a Status Update Bot - that retrieves data from client meetings and PDFs to provide real-time project updates and project information. The bot has been deployed on AWS.

4.Created Voice enabled Bot - Developed a voice-enabled healthcare bot leveraging OpenAI API. Integrated speech recognition technology to enable natural language processing and seamless communication. Designed to assist users with medical inquiries and general healthcare guidance through voice commands.

QA Tool

Technologies and libraries: Flask, JS, pdftotext

Developed a QA tool leveraging OpenAI API for generating test cases based on project requirements, streamlining QA automation processes. Implemented features for uploading PDF files and downloading outputs using JavaScript. Incorporated functionality to extract text from PDF files using the 'pdftotext' utility. The tool enhances efficiency in test case creation and facilitates seamless integration into QA workflows.

Healthcare Chat bot

Technologies and libraries: Hugging Face, flask

Fine-tuned Llama 2 7B chat model for the health care use case. Deployed it using hugging face inference endpoint.Utilized text generation client to generatethe response.

Information Extraction from tabular data

Technologies and libraries: Pytesseract

Designed and implemented a computer vision solution for extracting textual information from images. Utilized OCR engine to recognize and extract text from images.Utilized image preprocessing techniques to enhance the accuracy of text extraction.

Inferenz Tech Pvt. Ltd. (as a ML Engineer) *2020-2022 (sept)*

Virtual Makeup and Jewelry

Technologies and libraries: Pytorch, Dlib,Mediapipe, Matplotlib, NumPy,scikit-learn, OpenCV, javascript, THREE.js, blender,unity, Meshroom

Given the data set applied virtual makeup and virtual hair color. Preprocessed data to convert 3D image into 2D mask, extracted the features using ROI. Performed segmentation of hair and facial features using deep learning models. Apply virtual makeup and hair color based on the OpenCV techniques on real time using Mediapipe landmark. Implemented the python code into JavaScript to optimize the process.

Integrated 2D and 3D virtual jewelry. Using THREE.js applied virtual jewelry in real time. Generated 3D models using Meshroom software. Edited 3D models into blender. Experimented applying virtual jewelry in unity.

Information Extraction usingOCR

Technologies and libraries: Pytesseract, pdf2image, pdfplumber, numpy, pandas, jTessBoxeditor

Created the dataset of various type of the documents in different formats. Extracted key information from various kind of documents using Tesseract-OCR. Trained the tesseract to improve the OCR result using jTessBoxEditor and achieved **96% accuracy**. Fine-tuned the LSTM model to add newcharacters to be detected by OCR. Applied various image preprocessing techniques to remove noise from the documents.

RASA Chatbot

Technologies and libraries: RASA

Created stories and trained the NLU model basedon domain. Performed testing of bot using RASA X.

Skills

Languages	Python, Javascript
ML Libraries	TensorFlow, Keras, PyTorch, OpenCV, scikit-learn, NumPy, SciPy, Pandas, Seaborn, Matplotlib, PIL, Spacy, NLTK, flask, Tesseract engine.
GIS Software	ArcGIS pro, QGIS
Cloud services & storage	Apache server, Azure Blob Storage Container, S3 browser (Amazon S3 client), AWS (EC2)
IDEs/Development Tools	Azure DevOps server, Docker,confluence, VS code, Jupyter Notebooks, Colab, GitHub, Tensor Board
Frameworks	Streamlit, Azure Bot Framework , RASA
Vector Databases	ChromaDB, Qdrant
Web Scrapping Library	Selenium, Beatuifulsoup

Education

M.Sc. Artificial Intelligence & Machine Learning *2019-2021*

Department of Computer Science- Gujarat University, Ahmadabad

B.Sc. (CA & IT) *2016 – 2019*

K.S. School of business management

Research & Experiments

- Created game in the construct3D platform as a part of research group at Inferenz (under progress)
- Participated in hackathon and Build AR web app for travelling at Inferenz
- Created smart contracts in Blockchain (Beginner level)

Activities & Achievements

- Research Paper ImageColorization published in International Association of Biologicals and computational Digest:
  - **https://www.researchgate.net/publication/349302947_Image_Colorization**
- Track winner in Hack2skill - Snap chat new year lensathon : **2021 The year in rewind**
- Written blog on Free Online Meeting And Video Calling Apps:
  - **https://technodossier.com/online-video-calling-apps/**
- Written blog on Image Colorization:
  - **https://technodossier.com/image-colorization/**
- Volunteered as campus ambassador for the TCS.

Certificates & Courses

- Neural Networks and Deep learning(2020) | **deeplearning.ai, Coursera**
- ChatGPT prompt engineering for developers
- Participation in innovation hackathon at DXFactor

Languages

English | हिंदी | ગુજરાતી