

Sai Chandra Pandraju

saichandrapandraju@gmail.com | +91-9063506674 | LinkedIn: [linkedin.com/in/saichandra-pandraju/](https://www.linkedin.com/in/saichandra-pandraju/) | Portfolio: www.saichandra.dev

| | | |
|-----------------|--|------------------------------|
| EDUCATION | QIS College of Engineering and Technology , Ongole, Andhra Pradesh, India, affiliated to JNTUK Bachelor of Technology in Electronics and Communications Engineering, CGPA: 9.138/10 | April 2019 |
| | Sri Vani Junior College , Chirala, Andhra Pradesh, India Senior Secondary (XII) in Mathematics, Physics, and Chemistry, Percentage: 97.9% | March 2015 |
| | Vedamatha English Medium School , Chirala, Andhra Pradesh, India Secondary School (X), CGPA: 9.8/10 | March 2013 |
| SKILLS | Programming Languages: Python, Java, C, TypeScript, HTML/CSS, SQL, MATLAB Libraries: TensorFlow, PyTorch, Hugging Face Transformers, DeepSpeed, spaCy, scikit-learn, NumPy, Pandas Frameworks: Spring Boot, Angular, Django, Flask | |
| CERTIFICATIONS | TensorFlow Certified Developer , Mathematics for Machine Learning Specialization , Deep Learning Specialization , Machine Learning A-Z | |
| WORK EXPERIENCE | Senior Systems Engineer, INFOSYS R&D (iCETS) | Oct. 2021 – Present |
| | <ul style="list-style-type: none">Developing API design for building Infosys AI Platform.Developing a Visual Studio Code extension for assisting Infosys developers with code suggestions, code translation, and code summarization. | |
| | Systems Engineer, INFOSYS R&D (iCETS) | Dec. 2019 – Oct. 2021 |
| | <ul style="list-style-type: none">Worked on Full-Stack web development using Angular, Spring Boot, and MySQL, among others, for building Live Enterprise Application Management Platform (LEAP).Implemented end-to-end ML pipelines for classification, clustering, and forecasting using LEAP's MLStudio.Researched Deep Learning Natural Language Understanding (NLU) and made Infosys among the top 10 in the SuperGLUE Benchmark.Developed NLP/NLU solutions for Question Generation, Policy Chatbot, Bio-medical Relation Extraction, and CodeBot. | |
| | Systems Engineer Trainee, INFOSYS Ltd. | Aug. 2019 – Dec. 2019 |
| | <ul style="list-style-type: none">Trained in JAVA EE, Angular, Python, MySQL, Data Structures & Algorithms.Built a 'Travel Booking Site' from scratch and led my team to be in the top 3% for this final project.Completed the training as a 'Top Performer' and was selected to elite R&D department in Infosys(iCETS). | |
| PROJECTS | CodeBot, Infosys R&D (iCETS) | Aug. 2021 – Present |
| | <ul style="list-style-type: none">Built datasets by creating data pipelines to extract, clean, and pre-process internal code repositories.Pretrained PLBART and T5 models on CodeSearchNet and Infosys Internal GitHub repositories using MegatronLM+DeepSpeed (Model and Data Parallelism) on an NVIDIA DGX A100 GPU Cluster.Deployed all of our models as APIs and created a User-Interface (UI) to try these models for a variety of code tasks like 'Translation', 'Summarization', 'Generation', 'Refinement', 'Defect Detection' and 'Clone Detection'.Working on developing a Visual Studio Code extension for assisting Infosys developers with code suggestions, code translation, and code summarization. | |
| | Biomedical Relation Extraction, Infosys R&D (iCETS) | June 2021 – Aug. 2021 |
| | <ul style="list-style-type: none">Created an end-to-end pipeline that downloads and extracts PubMed abstracts, performs NER, creates a dataset, and extracts the relations by converting the problem into one of the NLU problems.After collating and filtering the relations, created a User-Interface (UI) with the generated knowledge base. For the bio-medical entity that the user inputs, the application will return corresponding entities of 'caused by' and/or 'treated by' relations. | |
| | Policy Chatbot, Infosys R&D (iCETS) | Apr. 2021 – June 2021 |
| | <ul style="list-style-type: none">Trained and deployed a conversational chatbot using the T5 model for answering user queries related to legal policies by extracting data from Infosys Internal Policy Repository.Performed data cleansing, pre-processing, and anonymization to prevent biases in the model using statistical methods such as Parity Difference, Equal Opportunity Difference, Average Odds Difference, Disparate Impact, and Theil Index. | |

- Generated the sentence embeddings for the corpus and used embedding-based content retrieval, cosine similarity to extract the closest context for the question. The answer is then generated based on the question and the context.

End-to-End Question and Answer Generation System, Infosys R&D (iCETS)

Feb. 2021 – Apr. 2021

- Developed an end-to-end pipeline to generate questions and answers from structured and unstructured datasets (PDF, Word, Web URLs, Spreadsheets, Images) without human intervention.
- Extracted text along with its structure from PDFs and Images using fine-tuned LayoutLMv2 and Detectron models.
- For textual data, a fine-tuned T5 model is used to generate questions whose answers can be Boolean, one-word, sentence-length, or summary.
- For tabular data, a modified version of the Table-to-Text(ToTTo) dataset is used to fine-tune the T5 model to generate questions based on highlighted cells. In addition, TAPAS is used for Sequential and Conversational style answers.

SuperGLUE Benchmark, Infosys R&D (iCETS)

Dec. 2020 – Feb. 2021

- Ranked top 10 in SuperGLUE, a rigorous benchmark for Natural Language Understanding Tasks with a score of 86.0.
- Rather than taking a model-centric approach like trying out big models, we chose a data-centric approach. We used Snorkel AI's weak supervision to improve the model's performance.
- We took a relatively small model(RoBERTa-large) compared to top positions on the leaderboard like T5-11B, TuringNLG, etc., and created snorkel's data functions and were able to reach 6th (in Feb' 21) position on SuperGLUE benchmark.
- Integrated DeepSpeed to efficiently train large language models with minimal infra.

Real-Time Ticket Clustering, Infosys R&D (iCETS)

Oct. 2020 – Nov. 2020

- Performed Exploratory Data Analysis (EDA) and implemented Latent Dirichlet Allocation (LDA) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to cluster the ticket database and also implemented it in real-time ticket allocation application.

Multivariate Forecasting for SAP HANA database, Infosys R&D (iCETS)

Sep. 2020 – Oct. 2020

- Performed Exploratory Data Analysis (EDA) and implemented Vector Auto Regressive Moving Average (VARMA) model to forecast the total record count of HANA tables to reach 2 million (can be any arbitrary count).
- Implemented a mail triggering module using Python's 'smtpplib' to send mail to respective teams on the forecasted date.

Live Enterprise Application Management Platform (LEAP), Infosys R&D (iCETS)

Dec. 2019 – Sep. 2020

- Developed a variety of interactive widgets and dashboards that help enterprises to better visualize their applications' data.
- Enhanced existing widgets to let users customize the aesthetic aspects like color, font size, font family, font-weight, shape, etc.
- Redesigned dashboard module such a way that lets client's support teams build and deploy custom widgets without disturbing the core modules.

Travel Booking Site, Infosys Ltd.

Nov. 2019 – Dec. 2019

- Built a single-page web application (SPA) for travel booking with several features like - New user registration, Login & Logout, Filter the available trips, View details such as itinerary, highlights & day-wise plan, Booking the trip, Viewing the booked trips, and Cancel booking.
- Led my team to be in the top 3% for this final project. Completed the training as a Top Performer and was selected to elite R&D department in Infosys(iCETS).

Human Activity Recognition using Sensors, QIS College of Engineering & Technology

Nov. 2018 – Mar. 2019

- Created a Deep Convolutional and LSTM network for Human Activity Recognition with wearable sensor data.
- Used MATLAB to code the model along with training and testing functions from scratch without using any NN libraries like TensorFlow.

RESEARCH

Answer-Aware Question Generation from Tabular and Textual Data using T5 – [Paper](#), [GitHub](#)
Unsupervised Convolutional Filter Learning For COVID-19 Classification – [Paper](#), [GitHub](#)

AWARDS

[INSTA Award for ML Research](#)
[INSTA Award for Full Stack Development](#)