

# Gourav Singh Bais

gouravbais08@gmail.com · +91-7869878010 · linkedin.com/in/gourav-singh-bais/ · github.com/gouravsinghbais/

## WORK EXPERIENCE

---

### Allianz Services India

Nov. 2021 – Present

Senior Data Scientist

Trivendram, IN

- **Lizzy - LLM Based Smart Assistant for Allianz Germany**
  - Worked on developing and deploying the RAG-based chatbot solution for answering different user queries regarding Allianz insurance products.
  - Effectively worked on evaluating the RAG-based solution with the help of LLM-as-a-judge and RAGAs metrics.
  - Effectively worked on improving the performance of the entire RAG pipeline and improved the accuracy by 40% while working on various parts of the pipeline like embedding model, vector database, and similarity metrics.
  - Worked on developing, deploying, and evaluating the text generation systems using open-source models and paid APIs.
- **A Master Template for end-to-end ML Solutions for Allianz Germany**
  - Designed a standardized project initiation process using the master template, enabling teams to quickly set up new ML projects with predefined directory structures, data ingestion pipelines, and model development frameworks.
  - Reduced project setup time by 40%, allowing teams to focus more on model development and less on infrastructure and boilerplate code. Also, reduced deployment cycles by 50%, enabling faster iterations and delivery of ML solutions to end-users by improving the existing CICD.
- **Product Recommender for Claims and Coverages**
  - Conducted in-depth research and analysis of customer preferences, historical claim data, and policy coverages to understand user behavior and identify opportunities for personalized recommendations.
  - Leveraged ensemble techniques to combine the strengths of multiple models, enhancing the overall F-1 Score to 90% and the robustness of the recommendation engine. Responsible for handling all the data science-related tasks under one product owner.
- **Business Classifier with Business Clustering**
  - Developed a classification system for small and medium commercials to classify them into different business areas. Responsible for improving the system's overall accuracy from 40% to 85%.
  - Developed data preprocessing, model training, and validation pipelines to ensure consistent and reliable results.
- Led a team of interns to develop the European Accidental Forms OCR project For Allianz France.
- Working on an internal LLM-based project for HR CV screening as a solution architect while managing a team of four developers.

### ValueMomentum

June. 2019 – Oct. 2021

Applied Machine Learning Engineer

Hyderabad, IN

- **Deep Learning-Based KYC Verification System**
  - Developed an end-to-end solution for verifying ID proofs like Aadhar Card and Pancard for a Fortune 500 banking Firm using Deep Learning-based object detection and OCR with an accuracy of 93%. Also developed an automated Aadhar number masking system that saved a lot of cost for customers.
  - Responsible for developing the deep learning algorithms for detection and OCR and deploying the model to production.
- **Intelligent Character Recognition (ICR) Based Systems**
  - Developed solutions for automatically extracting the data from different kinds of digital and scanned forms and documents using ML.

### REOFT Technologies

June. 2018 – Feb. 2019

Machine Learning Intern

Indore, IN

- Developed a solution for improving the sales of a Cafe using techniques like Market Basket Analysis, Customer Lifetime Value Analysis, and Churn Analysis.

## SKILLS

---

- Python • Pytorch • OCR • Deep Learning • NLP • LLMs • HuggingFace • Computer Vision • Tensorflow • Spacy
- OpenCV • Docker • Kubernetes • Helm • Flask • Azure • Amazon Web Services • REST APIs • Jenkins • GitHub Actions • FastAPI
- MLFlow • Technical Content Writing

## PROJECTS

---

- **European Accidental Forms OCR** - *March 2023*

An image-based OCR system to extract data from French and English forms.

- **Business Classifier and Product Recommender** - *November 2021*

A system to classify Small and Medium Commercials (SMC) into multiple domains and then recommend the best claims and coverage packages to them based on their profile.

- **Computer Vision-Based Vehicle Damage Assessment** - *June 2020*

Image-based vehicle damage detection for faster claims processing using object detection, image classification, and ML algorithms.

- **KYC Verification System** — *November 2019*

Automatic KYC verification using computer vision-based Face Match and ICR process.

- **Speech Recognition System** — *August 2019*

Finetunes Mozilla DeepSpeech model for speech-to-text using Python and Tensorflow.

## EDUCATION

---

**Rajiv Gandhi Prodyogiki Vishwavidyalaya, Bhopal**

**July 2019**

B.Tech., Computer Science

## CERTIFICATES

---

- **Generative AI with Large Language Models** - *Mar 2024*
- **Jenkins Essential Training** - *June 2023*
- **Data Scientist Nanodegree** - *June 2020*
- **Machine Learning Nanodegree** - *May 2020*
- **Deep Learning Nanodegree** - *May 2020*
- **Mathematics for Machine Learning** - *September 2019*

## AWARDS & OTHERS

---

- Working with O'Reilly publication for content and book reviews.
- Working as an evaluator in GreatLearning.
- Published more than 100 technical articles in the fields of data science, machine learning, and data engineering.
- Served as data master in ACM society for college.
- Dedicated to resolving issues and queries of data science practitioners over different platforms like LinkedIn, Medium, Twitter, etc.

## ARTICLES & BLOGS

---

- <https://neptune.ai/blog/llm-evaluation-text-summarization>
- <https://neptune.ai/blog/how-to-optimize-hyperparameter-search>
- <https://snyk.io/blog/secure-python-flask-applications/>
- <https://snyk.io/blog/ultimate-guide-creating-secure-python-package/>
- <https://www.redpanda.com/blog/real-time-machine-learning-processing-tensorflow-bigquery>
- <https://heartbeat.comet.ml/importance-of-machine-learning-model-retraining-in-production-a2c8d5e4c39e>