NIRAJ MODI

AI ENGINEER

Bringing 5.5 years of expertise in AI conversational designs, deploying LLMs, implementing RAGs and data-driven solutions implementation. Seeking to leverage skills in computer vision, image processing, NLP, and chatbot development to drive innovation and operational efficiency in a dynamic organization

CONTACT DETAILS

+ 91-9053118020
+971-545454438
nirajmodi51@gmail.com
Dubai
https://www.linkedin.com/in/niraj-modi-495436119/

EDUCATION

Completed a Bachelor of Technology in Computer Science from Guru Teg Bahadur Academy, Bengaluru, with a commendable academic performance of 76%.

CORE COMPETENCIES

- Large Language Models
- Intelligent Document Processing (IDP)
- OCR
- Document Classification
- Entity Extraction
- Named Entity Relation (NER) Model
- Data Analysis
- Data Visualization
- Data Modeling/Predictive Modeling
- Supervised and Unsupervised Learning
- Model Deployment

PROFILE SUMMARY

- **Open-Source Large Language Models (LLMs):** Proficient in deploying open-source LLMs, such as Llama 3, Mistral-7b to develop advanced natural language processing solutions.
- **Retrieval-Augmented Generation (RAG):** Experienced in implementing RAG techniques to enhance AI models with domain-specific knowledge, improving accuracy and reducing hallucinations.
- **Azure Suite:** Developed and integrated AI-driven solutions using **Azure AI Foundry, Azure Document Intelligence, Logic Apps, Function Apps, Event Grid, and Storage Containers** to support scalable, event-based architectures
- **AWS:** Deployed and optimized LLMs using **OLLAMA and vLLM on AWS EC2**, with hands-on experience in model training and inference workflows using **SageMaker**.
- **On-Premises LLM Deployment:** Skilled in deploying LLM solutions within on-premises environments, ensuring compliance with data security protocols and industry regulations.
- **Deep Learning (DL):** Proficient in leveraging Deep Learning techniques to solve complex problems and enhance model performance for various applications
- **Advanced IDP Solutions Implementation:** Successfully deployed advanced IDP solutions to automate the extraction, classification, and processing of unstructured data from documents, resulting in 70% increase in operational efficiency.
- **Customized Document Workflows:** Designed and implemented customized document processing workflows, incorporating OCR (Optical Character Recognition) and NLP (Natural Language Processing) techniques to streamline document handling processes.
- **Conversational UX Design:** Designed and developed highly intuitive and user-friendly chatbot interfaces, incorporating natural language understanding (NLU) and intent recognition for seamless user interactions.
- **Machine Learning (ML):** Strong foundation in Machine Learning algorithms and methodologies, with a track record of building high-performance models for diverse applications.
- **ANN / RNN / CNN:** Proficient in designing and implementing Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) for specialized tasks in image recognition, sequence analysis, and more.
- **Image Processing:** Adept at applying advanced Image Processing techniques to enhance, analyze, and manipulate images for diverse applications, ensuring high-quality visual data output.
- **Data Analysis and Visualization:** Skilled in performing in-depth data analysis to uncover meaningful insights and trends. Proficient in creating visually compelling data visualizations, enabling stakeholders to grasp complex concepts easily.
- **Data Preprocessing:** Skilled in data cleaning, transformation, and preparation to ensure optimal quality and accuracy, setting the foundation for robust analysis and modeling.

WORK EXPERIENCE

AL ML ENGINEER, RAKBANK (Contract), Dubai
MAY 2025 – Present
KEY RESULT AREAS:

• Clearview :

- Designed and implemented 'Clearview', an AI-driven cheque processing solution using **OCR, entity recognition, and signature verification** to automate manual workflows.
- Achieved automation of over 350K cheques per month, significantly reducing processing time and improving accuracy
- Delivered operational efficiency gains equivalent to **15+** FTEs, enhancing scalability and reducing manual intervention

• Ops Tool :

TECHNICAL SKILLS

- Retrieval Augment Generation (RAG)
- Deep Learning (DL)
- Computer Vision (CV)
- Image Processing
- Chatbot Development
- Data Preprocessing
- Natural Language Processing (NLP)
- Machine Learning (ML)
- Model Training
- Python
- MongoDB
- RASA Framework
- Google Dialogflow
- OpenCV
- RDBMS & NoSQL
- TensorFlow/Keras/ PyTorch
- Numpy/Pandas/Scikit-Learn
- PySpark
- AWS/Azure/GCP

KNOWLEDGE PURVIEW

- Competent in conducting **Bug Bounty Hunting** across diverse platforms including Hackerone, BugCrowd, Intigriti, and Pentabug.

.

- Developed a multi-functional OPS Tool integrating document translation, layout extraction, OCR, checklist generation, and document comparison using **Azure Document Intelligence and Azure Translator.**
- Architected the solution using **Azure Logic Apps, Function Apps, Storage Containers, Event Grid, and Power Automa**te to enable seamless, scalable document processing workflows.
- Enabled data governance and compliance while achieving operational efficiency gains equivalent to **5+ FTEs** through intelligent automation

SR. SOFTWARE ENGINEER, Ampcus Inc., Remote (US)

JULY 2024 – MAY 2025
KEY RESULT AREAS:

• **Deployed open-source Large Language Models (LLMs)**, optimizing and fine-tuning them to meet specific project requirements, resulting in improved model performance and adaptability. **Quantized high-parameter LLMs** to efficiently utilize available GPU resources, achieving a 30% reduction in computational load without compromising model accuracy.

• **Developed and implemented Retrieval-Augmented Generation (RAG) pipelines**, integrating external data sources to enhance the contextual relevance and accuracy of generated content. Implemented **keyword-based metadata filtering mechanisms**, improving data categorization and retrieval efficiency, leading to a 20% increase in processing speed.

• **Implemented on-premises Large Language Model (LLM) solutions**, ensuring strict adherence to data security protocols and compliance with industry regulations, thereby safeguarding sensitive information and maintaining organizational integrity.

• **Established multi-level indexing systems** to optimize data retrieval processes, reducing query response times by 25% and enhancing system scalability.

• **Designed and deployed conversational AI pipelines**, utilizing Natural Language Processing (NLP) techniques to improve user engagement and interaction quality.

SR. SOFTWARE ENGINEER, SMECORNER, BENGALURU

FEB 2022 – JULY 2024
KEY RESULT AREAS:

- Led the design and implementation of a sophisticated Intelligent Document Processing (IDP) use case, encompassing the cross-validation of information across 27 distinct document types.
  - This multifaceted initiative leverages advanced technologies including NLP, OCR, Computer Vision, PDF Scrapping, Document Classification, entity extraction, and Named Entity Relation (NER) models.
- Actively involved in the end-to-end development of guided and unguided chatbots, focusing on optimizing their functionality and user experience.
  - Conducting ongoing data analysis to refine chatbot performance and enhance overall user acceptance.
  - Additionally, overseeing the seamless deployment of RASA chatbots across diverse channels, including MS Teams, Slack, and Web APP.
- Collaborated with esteemed US-based pharmaceutical clients to conceptualize and develop Proof of Concepts (PoCs) and custom chatbot solutions tailored to their specific use cases.
  - This involves a deep understanding of client requirements and a proactive approach to deliver innovative and effective chatbot solutions.

COMPUTER VISION SCIENTIST
SMECORNER

JULY 2019 – FEBRUARY' 2021
KEY HIGHLIGHTS:

- Played a pivotal role in contributing to the entire product development lifecycle, encompassing requirement generation, annotation, modeling, and back-end development for the Auto-KYC Verification Process.
  - Successfully addressed four different document types, achieving an exceptional accuracy rate of 93%.
  - Employed a Deep Learning Object Detection model for precise localization, coupled with the customization of OCR to extract data from localized entities.
- Designed and implemented an innovative Discrepancy Resolution System that effectively handled discrepancies throughout the entire Loan Application to Loan Disbursement process.
  - This multifaceted project featured Computer Vision, Deep Learning, NLP-based data extraction, and PDF scraping solutions.
- Devised a unified computer vision solution aimed at enhancing automatic data extraction from bank statements of more than 80 different banks, resulting in an impressive 80% reduction in turnaround time.

- This strategic approach made nearly 85 older modules redundant and harnessed Image Processing techniques for working with coordinate systems and PDF scraping.
- Developed a Personal Discussion Chatbot equipped with voice recognition capabilities and support for vernacular languages.
  - Implemented an AI-based solution for generating comprehensive summary reports, pivotal in the loan approval process.
  - The chatbot was constructed using the RASA framework and seamlessly deployed on MS Teams. Leveraged NLP techniques for summary generation.

INTERNSHIP EXPERIENCE

DEEP LEARNING INTERN, JAZARI AI

APRIL 2019 - JUNE 2019

KEY RESULT AREAS:

- Successfully deployed a lightweight deep learning model on the Nvidia Jetson Nano board to enable edge computing for smart cameras.
  - This initiative aimed to optimize operations at Haryana National Highway Toll Booths.
  - Conducted a successful pilot for a duration of seven days to validate the effectiveness of the implementation.
- Developed specialized computer vision modules dedicated to error detection within the packaging industry.
  - These modules were integrated with smart cameras and robotic arms, enhancing quality control processes and ensuring a higher level of precision and accuracy.