Niraj Modi

**Mobile**: +971-545454438 | **Location**: Dubai, UAE | **Email**: [nirajmodi51@gmail.com](mailto:nirajmodi51@gmail.com)

**LinkedIn** : <https://www.linkedin.com/in/niraj-modi-495436119/>

With 6 years of experience, I specialize in LLM model deployment (open source) & RAG implementation. I have hands-on expertise with Azure AI Foundry, AWS SageMaker, and Deep Learning frameworks, applying Computer Vision, OCR, and NLP techniques to streamline document handling and build advanced IDP solutions

**CAREER SUMMARY**

**RAKBANK | AI ML Engineer**

May 2025 – Present

* **Clear View**:
  + 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**
  + 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 Automate** to enable seamless, scalable document processing workflows.
  + Deployed Open Source LLM models to **AWS SageMaker** and EC2 instance with frameworks like **OLLAMA and vLLMs**
  + Enabled data governance and compliance while achieving operational efficiency gains equivalent to **5+ FTEs** through intelligent automation

**Ampcus Inc. | AI Engineer**

July 2024 – May 2025

* **Deployed open-source Large Language Models (into an on-prem server** ensuring strict adherence to data security protocols and compliance with industry regulations), optimizing 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 significant reduction in computational load without compromising the model’s 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.

**Marlabs | AI & Chatbot Engineer**

Feb 2022 – July 2024

* 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
* Led the end-to-end development **of guided and unguided chatbots**, focusing on optimizing their functionality and user experience.

**SMECorner | Computer Vision Scientist**

July 2019 – Feb 2022

* 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.
* 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**, harnessing Image Processing techniques for working with coordinate systems and PDF scraping.

**Education**

Rawal Institute of Engineering and Technology (MDU, Rohtak)

August 2015 – July 2019

* Bachelor of Technology in Computer Science with academic performance of 76 %
* Created real-time object detection model for final year project