

IrishTaylor®

Australia | India | USA

CASE STUDY

AI Systems for Contextual Intelligence and Predictive Insights

OCTOBER 2025



Abstract

This paper presents a suite of artificial intelligence systems developed for contextual understanding, predictive analytics, and intelligent text processing. The focus areas include context-based retrieval through Retrieval-Augmented Generation (RAG) models, predictive modeling using supervised learning algorithms, multilingual text extraction via Tesseract OCR, and adaptive question-answering systems powered by LLaMA-based large language models. Together, these technologies enable smarter automation, cross-domain learning, and future trend forecasting, enhancing business and educational intelligence.

Introduction

The convergence of language models, predictive analytics, and computer vision has accelerated the development of AI systems capable of learning, reasoning, and adapting to contextual data. This paper outlines our AI development journey, covering:

- Contextual retrieval frameworks (RAG)
- Prediction systems (supervised regression/classification)
- Text recognition (OCR with Tesseract)
- Question–Answer correction systems (LLaMA)
- Trend forecasting (time-series and pattern recognition)

These modules have been integrated to form a robust technical stack for automated reasoning, data-driven predictions, and language understanding.

Context-Based Retrieval with RAG Models

Overview:

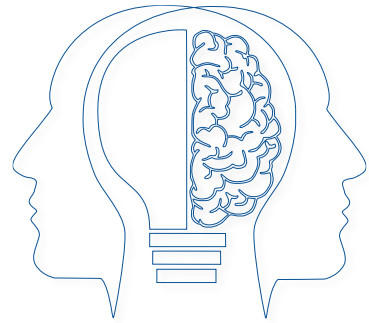
Retrieval-Augmented Generation (RAG) models combine information retrieval with generative AI to produce context-aware responses.

Our system uses:

- **Vector databases (Qdrant / FAISS)** for embedding-based retrieval.
- **Embedding models** (e.g., sentence-transformers, OpenAI embeddings) for semantic context matching.
- **LLM generation (LLaMA or similar)** for synthesizing final outputs based on retrieved chunks.

Algorithmic Approach:

1. Input query → embedded into a vector space.
2. Retrieve top-k relevant documents using cosine similarity.
3. Feed retrieved context + query into LLM for contextualized response.
4. Post-process output for accuracy and relevance.



Use Cases:

- Chatbots for educational and enterprise knowledge bases.
- Document summarization and Q&A from technical manuals.
- Intelligent search within organizational knowledge repositories.

Predictive Models using Supervised Learning

Overview:

We have designed and trained prediction systems using supervised algorithms for both **regression** and **classification** tasks.

Key Algorithms Implemented:

- **Regression:** Linear Regression, Random Forest Regressor, XGBoost, LSTM (for time series).
- **Classification:** Logistic Regression, Decision Trees, Support Vector Machines, and Neural Networks.

Applications:

- Employee lifecycle and performance predictions.
- Real estate price forecasting.
- Trend analysis for business KPIs.
- Medical or training progress classification models.

Pipeline Example:

Data preprocessing → Feature engineering → Model training → Validation (k-fold) → Deployment → Continuous model updates.

Multilingual Text Handling using Tesseract OCR

Overview:

Tesseract OCR has been integrated into our pipeline to extract text from PDFs and scanned documents in multiple languages (English, Hindi, Tamil, etc.).

Technical Workflow:

1. Image/PDF preprocessing (binarization, noise removal).
2. Language model selection (lang eng+tam+hin).
3. Text extraction and segmentation.
4. Post-processing using NLP (for structure, heading/chapter detection).

Applications:

- Automated document indexing and content extraction.
- OCR for education materials and multilingual datasets.
- Enhancing RAG-based systems with real-world text ingestion.

Question and Passage Answer Correction using LLaMA Models

Overview:

Using fine-tuned **LLaMA-based language models**, we developed systems that correct and evaluate question-answer pairs. The models leverage semantic similarity and contextual correctness to assess accuracy.

Algorithmic Design:

- Input question and student answer are embedded using transformer encoders.
- Model evaluates semantic overlap and contextual correctness.
- Grading and feedback are generated automatically using a scoring layer.

Applications:

- Automated evaluation for e-learning platforms.
- AI tutors for comprehension and answer improvement.
- Contextual error detection in human-generated responses.

Future Trend Prediction

Overview:

We have implemented forecasting models for identifying future trends using regression-based and hybrid deep learning methods (Prophet, ARIMA, LSTM).

Use Cases:

- Market and economic trend forecasting.
- Organizational resource planning.
- Predictive maintenance for operational systems.

Pipeline Example:

Historical data → Feature extraction → Model training → Forecast generation → Visualization dashboard.

Conclusion

The AI systems presented demonstrate how hybrid intelligence—combining contextual retrieval, predictive modeling, and text understanding can transform business, education, and operational intelligence. By integrating RAG models, OCR, and LLaMA-based feedback systems, we've created an ecosystem that learns from data, reasons with context, and predicts with accuracy. Future work includes enhancing multimodal inputs (image + text + speech) and deploying scalable inference pipelines for enterprise environments.