



AI-Assisted ISIC Classification System

Integrating BERT, RAG, and LLMs for Accurate & Transparent Industrial Classification



The Classification Challenge

Manual Process Issues

- Slow processing times
- Human error prone
- Inconsistent results

Critical Need

Accurate ISIC classification is crucial for reliable statistics and informed policy decisions across industries.

Why AI is the Solution

BERT Advantages

- Fast automated classification
- Highly accurate predictions
- Handles large volumes efficiently
- Reduces manual errors

BERT Limitations

- Cannot explain predictions
- Lacks transparency
- No human-readable justifications



RAG complements BERT by retrieving relevant ISIC manual sections and generating transparent, context-aware explanations.

The Complete AI Stack



BERT MODEL

Core classifier providing high-speed, accurate predictions for business activity descriptions.



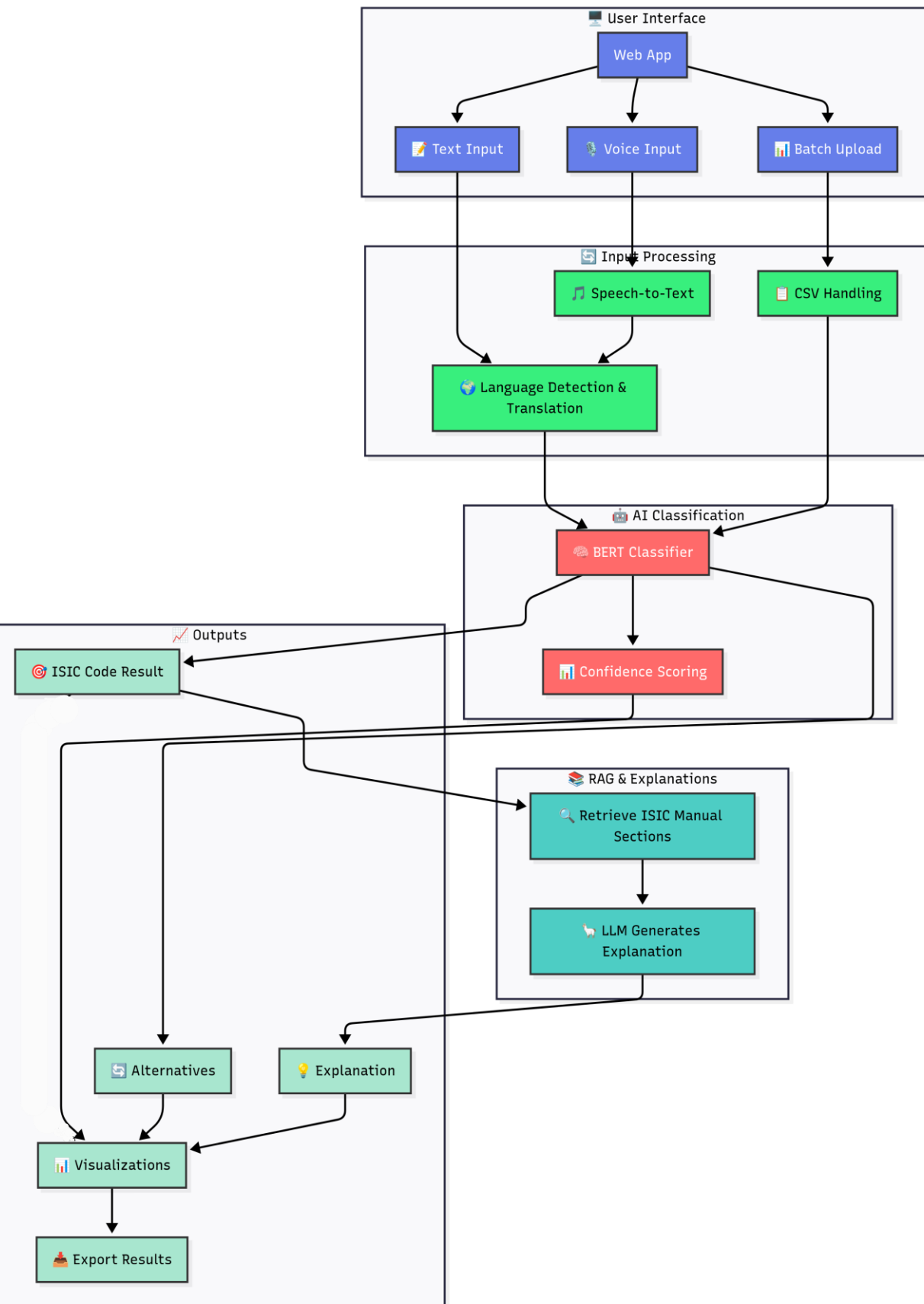
RAG System

Retrieves relevant ISIC manual passages and supports version-aware explanations.



LLM (LLaMA)

Converts retrieved sections into human-readable justifications and explanations.



System Workflow



Input Processing

Business description in any language with auto-detection and translation to English.



BERT Prediction

Fast, confident ISIC code prediction with confidence scoring.



RAG + LLM

Retrieves manual sections and generates textual explanations.

Output includes predicted ISIC code, confidence score, and reference-based explanation for complete transparency.

Multilingual Capabilities



Global Language Support

Processes business descriptions in English, French, Spanish, and other languages with automatic detection.



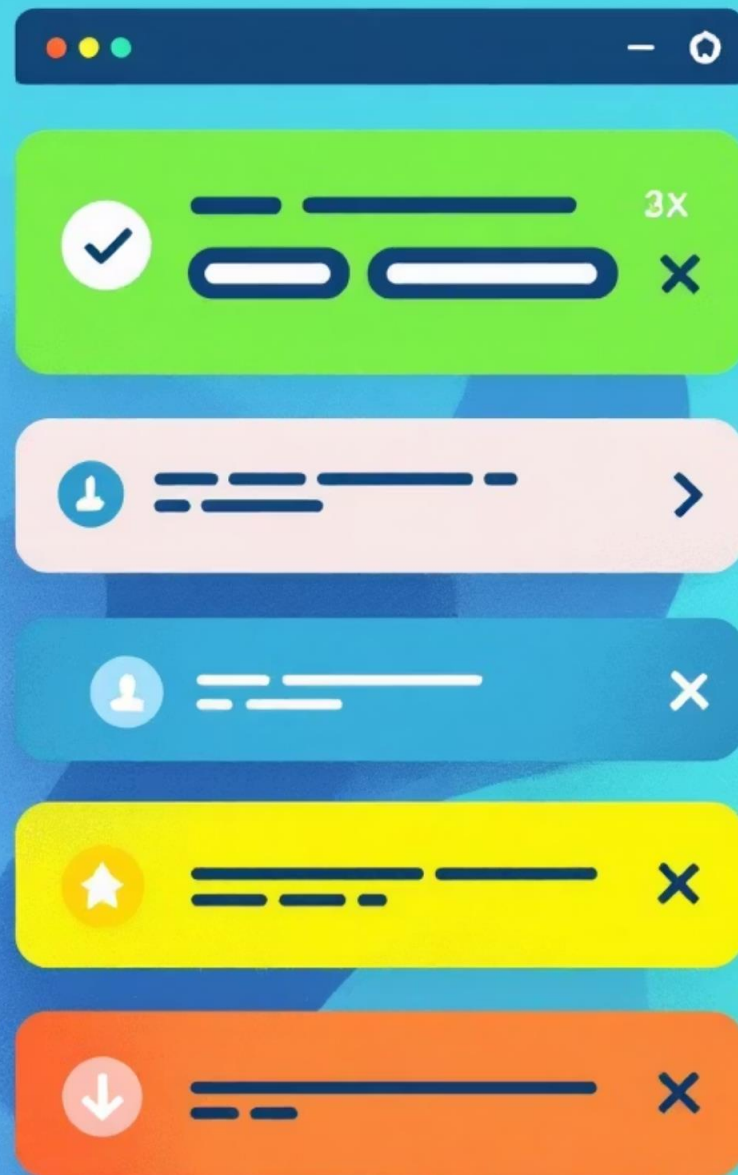
Seamless Translation

Auto-translates input to English for classification while maintaining original context and meaning.



Consistent Results

Delivers uniform classification accuracy regardless of input language or regional variations.



Handling Manual Updates

1

Current Version

BERT predicts code 3100 based on Rev4 manual training data.

2

Manual Update

ISIC Rev5 released with activity reclassified under code 3110.

3

RAG Adaptation

System indexes new sections and suggests updated code 3110 with justification.

BERT remains effective; retraining may be needed with manual changes, but not immediately thanks to RAG + LLM.

System Benefits



Speed & Scale

Fast, accurate classification handling large volumes of business descriptions efficiently.



Transparency

Explainable predictions with manual references and confidence scores for auditability.



Future-Proof

Version-aware system adapts to manual updates without requiring model retraining.



Addressing Common Concerns

Retraining Requirements

Manual updates are handled efficiently with RAG indexing, reducing the immediate need for BERT retraining

RAG Limitations

RAG alone insufficient for fast, large-scale classification - BERT provides the speed and accuracy foundation.

Implementation Complexity

Combined solution balances complexity with functionality, delivering best of both approaches.

The Future of Classification

Accurate • Explainable • Future-Proof

3x

Faster Processing

Compared to manual classification
methods

95%

Accuracy Rate

Consistent, reliable ISIC code predictions

100%

Transparency

Every prediction backed by manual
references

Transforming industrial classification for statistical offices worldwide