# Boosting Quality Control in the Automotive Industry using LLMs and Contrastive Learning



- **MIT Faculty Advisor:** Prof. Retsef Levi
- **BMW Advisor:** Dr. Maka Karalashvili

**MIT Students: Dongming Shen & Tingying Yan**


## Problem Statement

- **The Problem Management Process:**

Defects Detected → Incidents → Problems → Cause → Measure → Lessons Learned

**Reporting Line Alignment (RLA)**

- **Matching an incident to a problem (RLA)** is a crucial part of managing product problems at BMW to **ensure vehicle quality** and **reduce customer complaints**.

## Objective

- **Improve RLA** and **make AI recommendations resemble technician's choices**

**Manual** identification of problems based on **individual experience and knowledge**

- Enhance **accuracy and efficiency**
- Database with **fewer duplicates** and **improved categorization**

## Datasets

- **Incident:** ***80K rows ✕ 171 features*** *from Jan. 2021 to Mar. 2024*
  An "incident" is a defect that has just surpassed a basic evaluation. Their texts contain basic information of defects and do not go deeper into further analysis.
- **Problem:** ***38K rows ✕ 307 features*** *from Jan. 2021 to Mar. 2024*
  A "problem" is a defect that is more mature. Their texts can be summaries of several similar incidents with more detailed information than the incidents.

## Challenges

**Multilingual free** text data with abbreviations & duplications

**Balanced** model selection

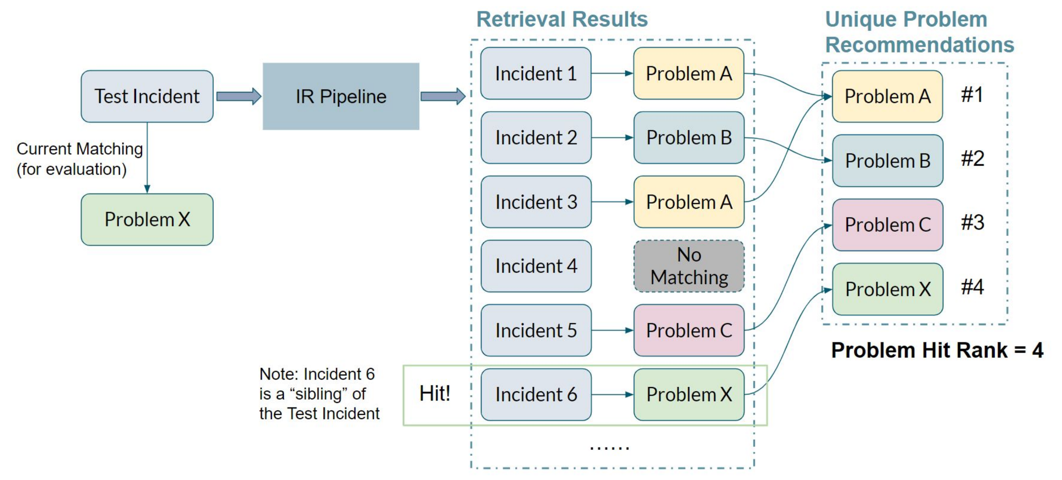
Need **systematic evaluation** method

**Domain knowledge** incorporation

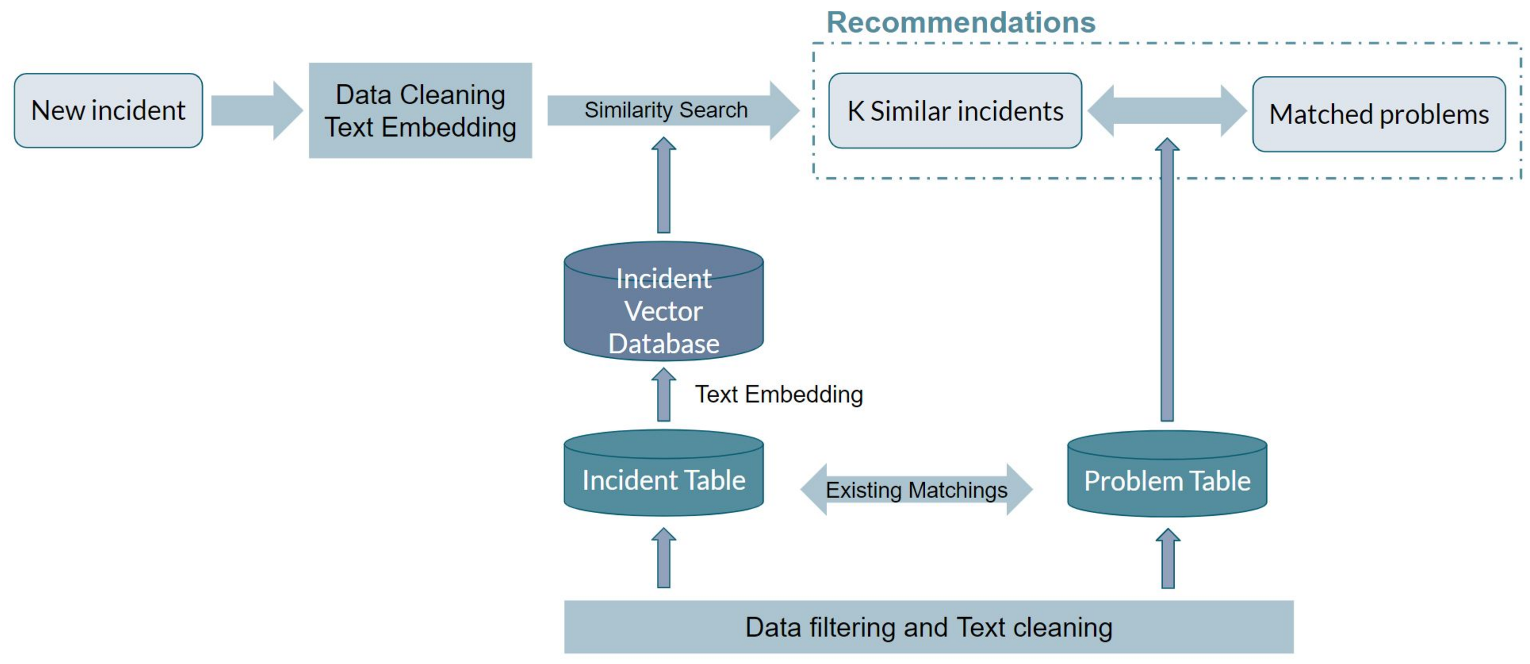
## Methodology

### Evaluation Methods:

- **Correct Labels:** previous incident-problem matchings in the database
- **Hit Rate:** the percentage of hits within a sample of 1,000 incidents with existing matchings.
- **Problem Hit Rank:** see the example on the right




### Retrieval Pipeline:

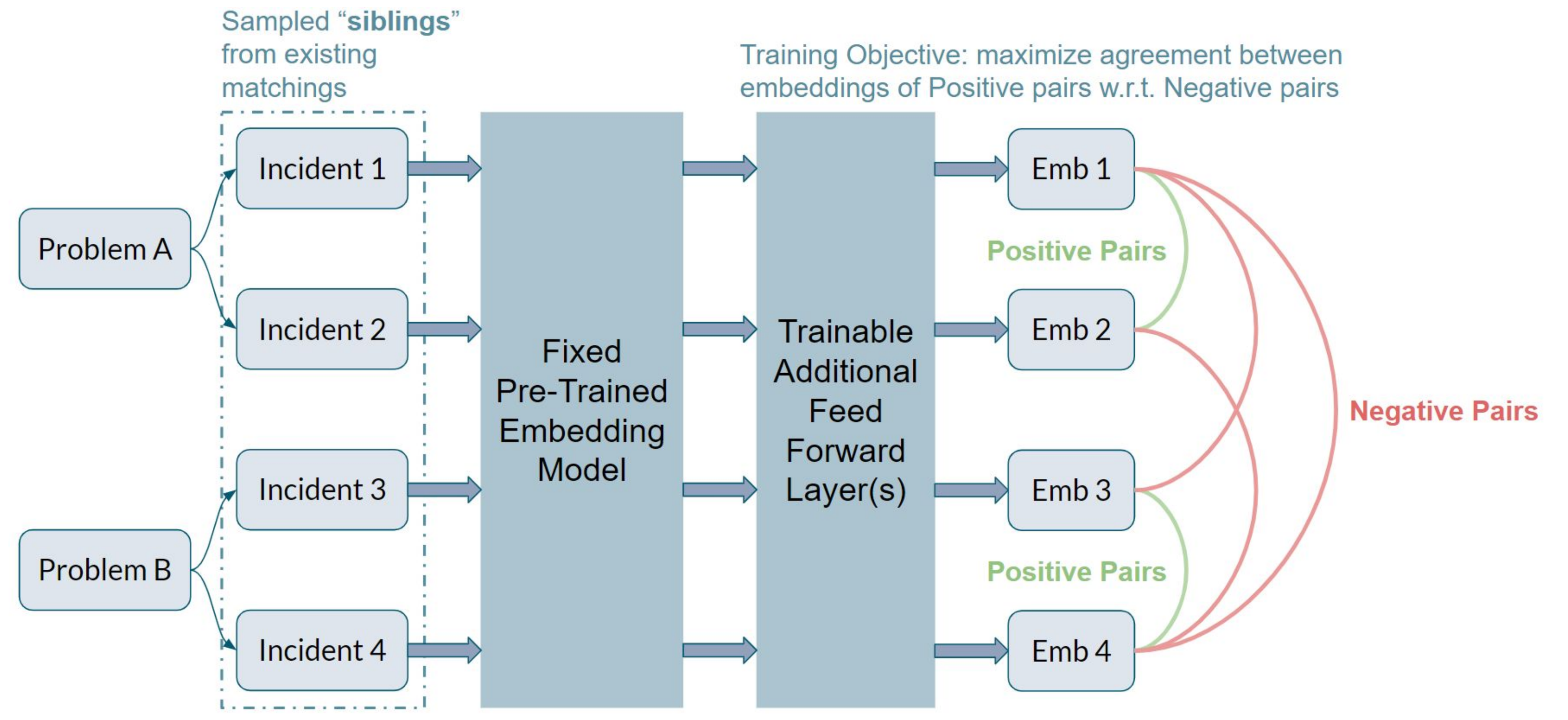



### Phase I: Embedding Model Selection

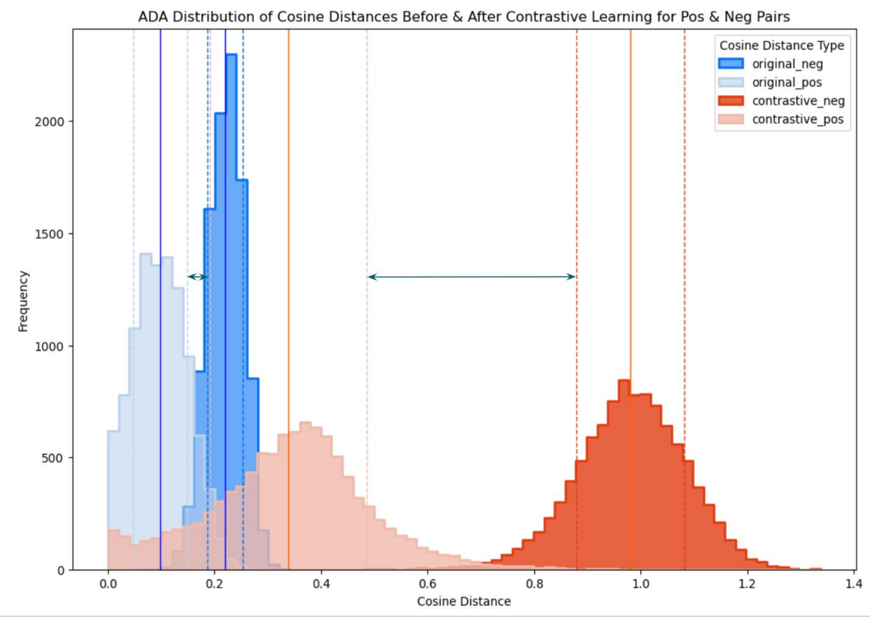
We tested text embedding of mBERT, RoBERTa, E5-large, and OpenAI ada with cosine distance to retrieve recommendations. We aim for a model that generates satisfactory results while acceptable in resources required.

### Phase II: Fine-tuning with Contrastive Learning:

We used the following pipeline to further finetune text embeddings:




### Cosine Distance Comparison



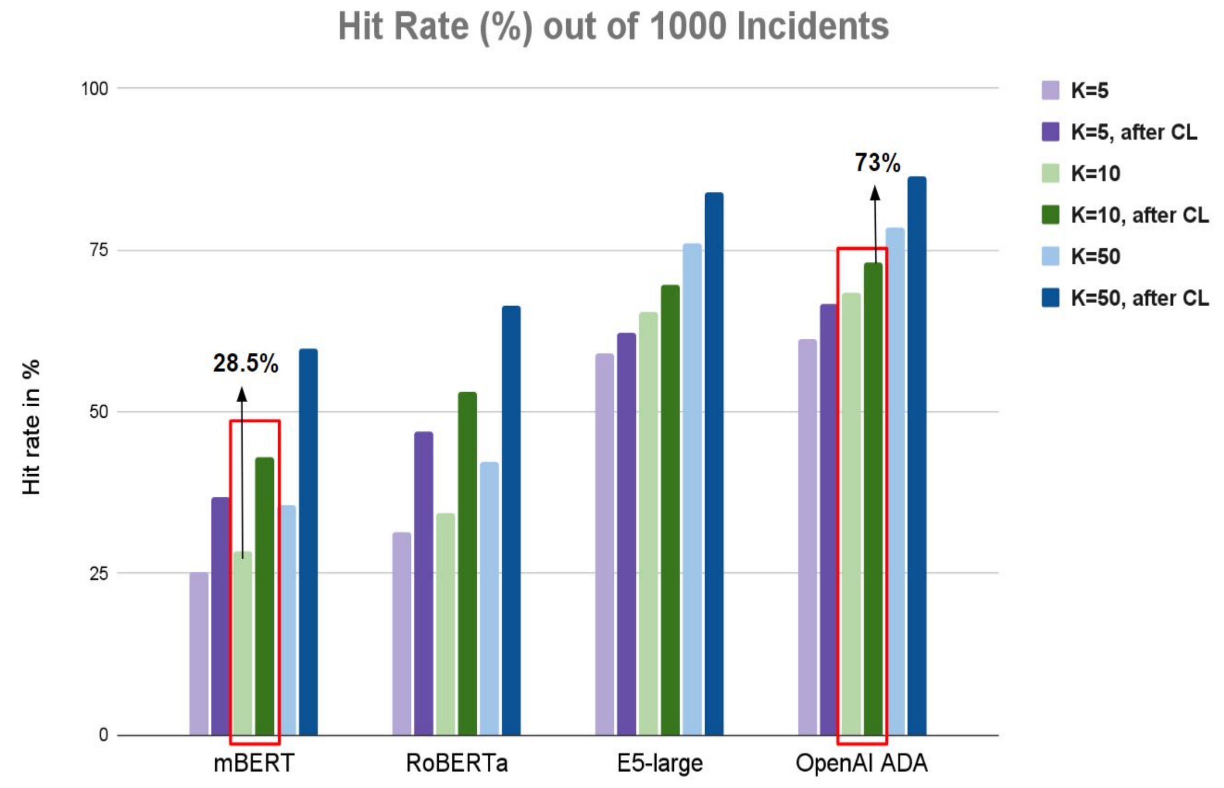

### KL Divergence Improvement:

| | mBERT | RoBERTa | E5-large | OpenAI ada |
|---|---|---|---|---|
| Before | 2.10 | 1.42 | 6.84 | 11.95 |
| After | **9.70** | **8.62** | **16.15** | **20.39** |

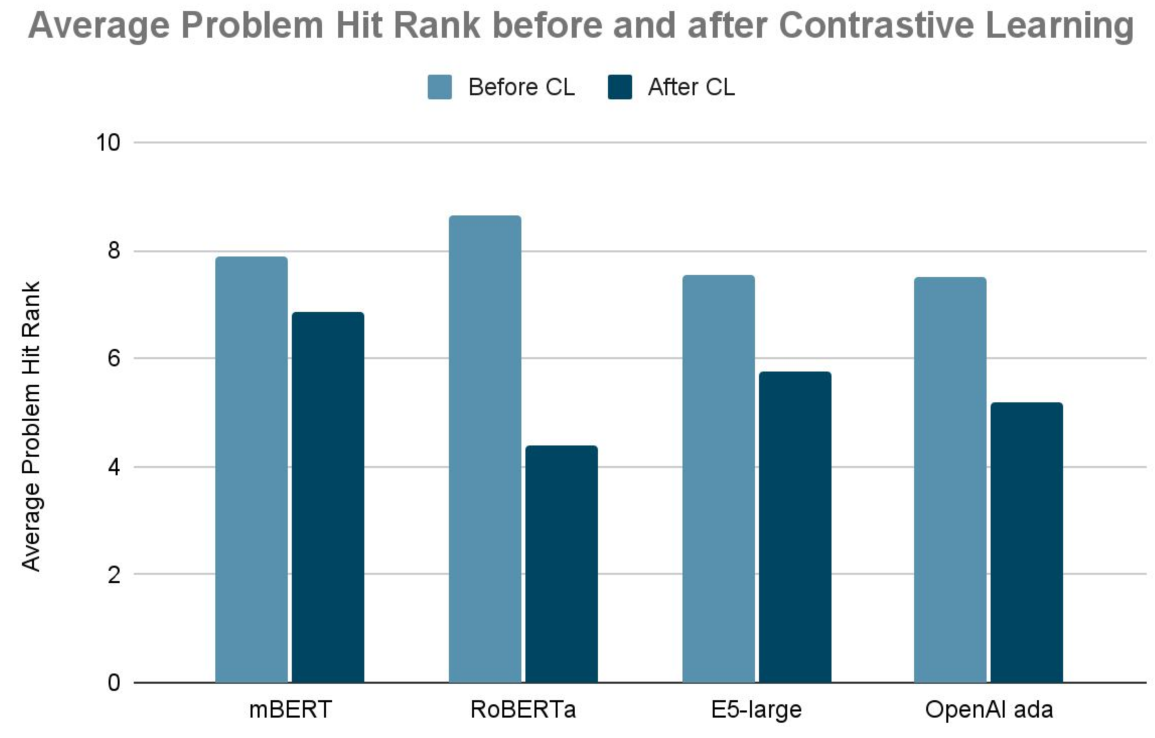
Contrastive learning **pushes negative cosine distance distribution away**, thereby generating better hit rates!

## Results

### Model Hit Rate Comparison




### Model Hit Rank Comparison




At K=10, our best model improved the hit rate by **44.5 percentage points** compared to baseline, ensuring the model **accuracy**.

After contrastive learning, users will read **3** documents fewer on average, effectively improving the **efficiency**.

## Business Impact

Technicians read less documents to find the correct matching, thereby **improving the working efficiency**

**Avoid problem duplication** and **improve database quality**

Boost the RLA within PMP process to ensure **product quality** and improve **customer satisfaction**

## Future Work

- **Re-ranking** when K is large to get more accurate ranking
- Develop **Q&A chatbot** features using Retrieval and Generation pipeline
- **Incorporate root causes and measures information** to further improve the matchings