Optimizing dental practice consolidation with Generative AI-driven data classification

Chain of Healthcare providers specializing in Orthodontics

Developed a Generative AI -based multidimensional data classification solution that efficiently categorizes a transaction based on multiple attributes to reduce the manual processing time. Solution significantly accelerated the M&A processes by reducing the data acquisition and integration turn-around time.

Orthodontics company needs “Gen AI” driven classification

Picture this…

You’re looking for an Open AI based Generative AI solution with functionality to generate mappings for transaction description into various pre-defined categories. You have adopted a profit-sharing model with the acquired Practices, that lacked the capability to automatically harmonize the transaction data of all practices for business reporting. Significant amount of manual mapping efforts was required to align new transaction descriptions with the pre-defined categories for consistent reporting, which led to increase in the onboarding time for new acquisitions

You turn to Accordion.

We partner with your team to develop a Generative AI -based multidimensional data classification solution that efficiently categorizes a transaction based on multiple attributes to reduce the manual processing time. Solution significantly accelerated the M&A processes by reducing the data acquisition and integration turn-around time, including:

1) Developing a Generative AI framework utilizing Azure OpenAI GPT models, achieving data processing and categorization with an accuracy exceeding 95%
2) Curating an optimal dataset for the LLM's contextual analysis through a ML-based clustering algorithm, significantly increasing the accuracy of mappings
3) Integrating specialized orthodontic knowledge into the LLM’s prompts across various categories, enabling the generation of tailor-made mappings
4) Implementing a confidence scoring system, using semantic similarity with related records to ensure a high degree of confidence in automated categorizations
5) Enabling users to review and manually override model responses via Excel reports, providing a human feedback loop to validate data before integration into the reporting suite

Your value is enhanced.

You have integrated practices into a reporting suite ~2 weeks faster, resulting in a notable acceleration (~ 25%) of the acquisition process. You have streamlined new transaction description mappings saving ~60 analyst hours per month. You have also standardized and corrected existing data that led to an accurate categorization of ~ $10 million in the reporting suite.

OPTIMIZING DENTAL PRACTICE CONSOLIDATION WITH GENERATIVE AI-DRIVEN DATA CLASSIFICATION

KEY RESULT

- ~2 weeks faster acceleration of (25%) acquisitions
- ~60 manhours saved per month
- ~$10 million accuracy in the reporting

VALUE LEVERS PULLED

- Generative AI framework utilizing Azure OpenAI GPT models
- ML-based clustering algorithm
- Developed a confidence score metric

Optimizing dental practice consolidation driven by Gen AI

DATA CLASSIFICATION MODEL

Situation

- Client has adopted a profit-sharing model with the acquired Practices, and they lacked capability to automatically harmonize the transaction data of all practices for business reporting. Significant amount of manual mapping efforts was required to align new transaction descriptions with the pre-defined categories for consistent reporting, which led to increase in the onboarding time for new acquisitions
- Partnered with the client to develop an Open AI based Generative AI solution with functionality to generate mappings for transaction description into various pre-defined categories

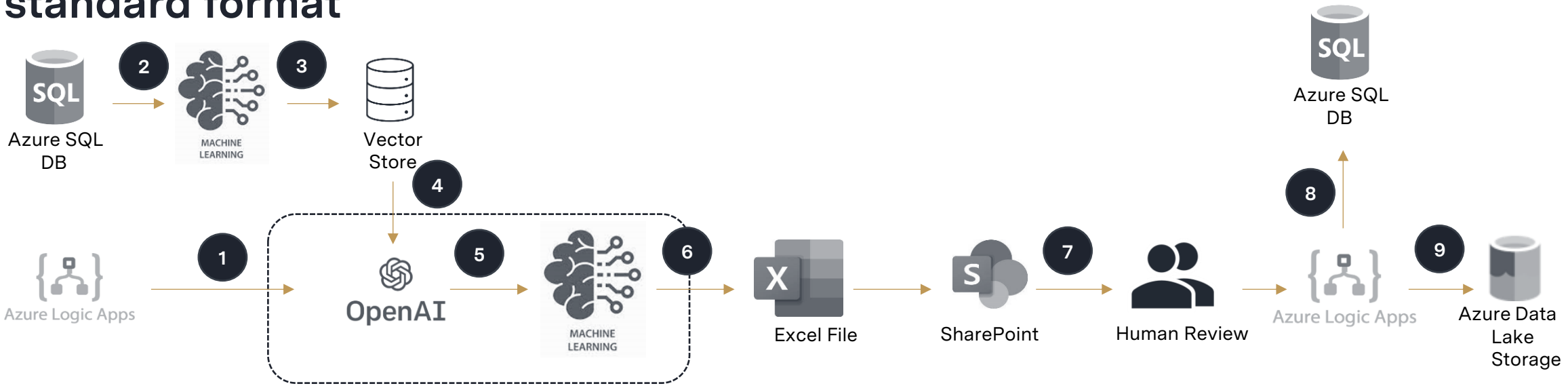
Accordion Value Add

- Developed a Generative AI framework utilizing Azure OpenAI GPT models, achieving data processing and categorization with an accuracy exceeding 95%
- Curated an optimal dataset for the LLM's contextual analysis through a ML-based clustering algorithm, significantly increasing the accuracy of mappings
- Integrated specialized orthodontic knowledge into the LLM’s prompts across various categories, enabling the generation of tailor-made mappings
- Implemented a confidence scoring system, using semantic similarity with related records to ensure a high degree of confidence in automated categorizations
- Enabled users to review and manually override model responses via Excel reports, providing a human feedback loop to validate data before integration into the reporting suite

Impact

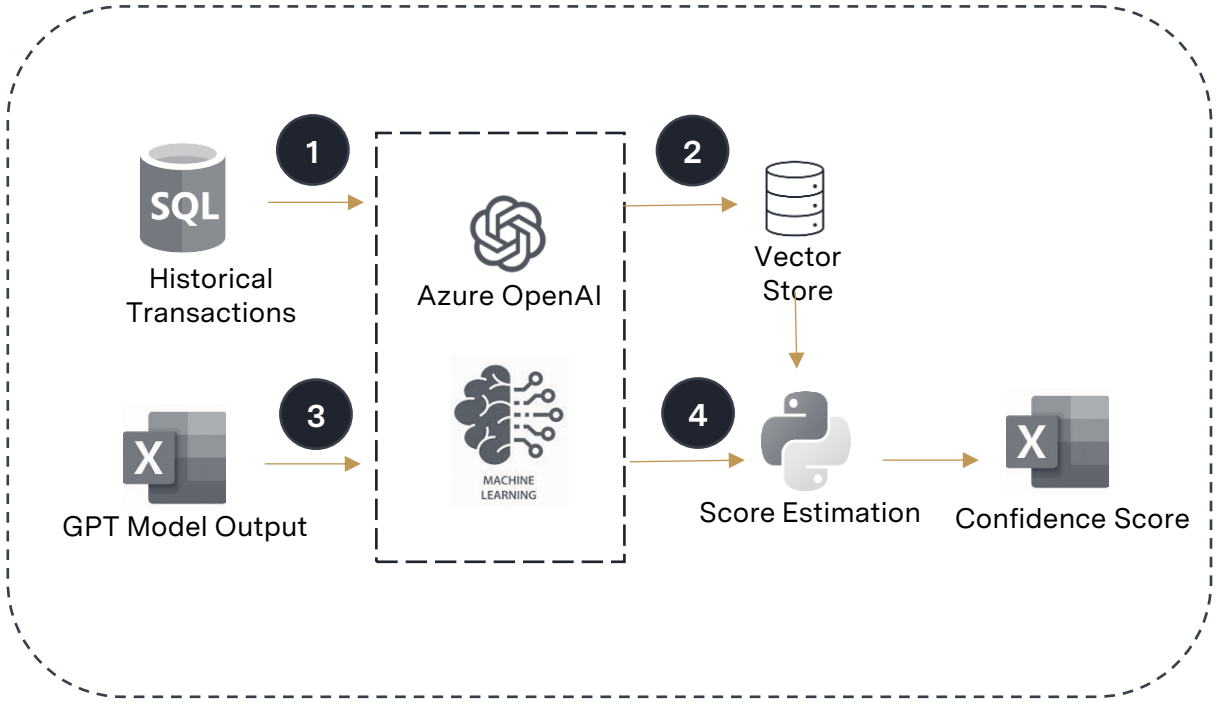
- Practices were integrated into the reporting suite ~2 weeks faster, resulting in a notable acceleration (~ 25%) of the client's acquisition process
- Streamlining new transaction description mappings saved ~60 analyst hours per month
- Standardization and corrections in existing data led to an accurate categorization of ~ $10 million in the reporting suite

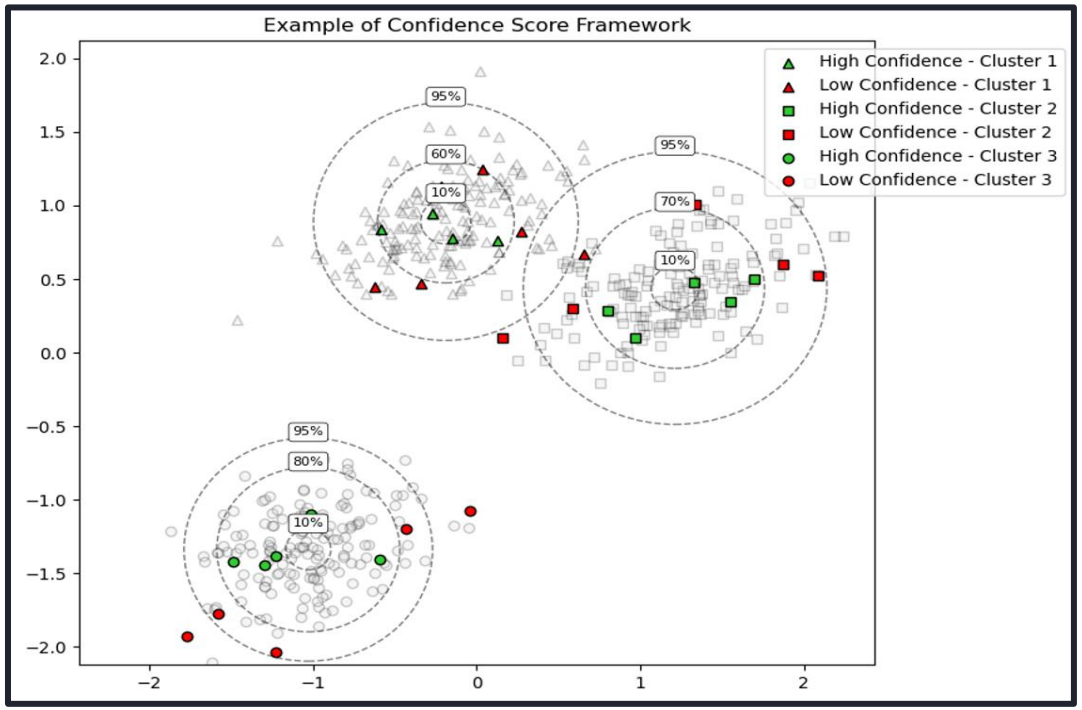
Information flow framework to map new transactions to standard format



1. **Azure Logic Apps** triggers API endpoint of Gen AI based mapping model to identify accurate mappings for **new transactions**
2. **Historical transaction data** hosted in Azure SQL DB would be passed through **ML algorithms** to identify optimal dataset for LLM's prompt
3. **ML based clustering algorithms leveraged** to build data clusters for **each transaction category** and identify relevant records
4. Data points nearest to **centroid of each clusters** would be identified for **LLM training** to identify patterns in data for each category
5. New transactions along with identified optimal dataset would be passed to **Open AI GPT model** to generate mappings for new transactions
6. **Confidence score** will be evaluated for model outputs using clustering algorithm by **estimating proximity** to the tagged category
7. Output excel shared for human review through SharePoint to check output **accuracy for low confidence results**
8. Results after vetting is stored in Azure DB for **reinforcement feedback mechanism** essential for model re-learning
9. Final consolidated and reviewed transaction mapping excel is pushed to **Azure Data Lake** for downstream process and DW ingestion

Methodology: Confidence score





1. Historical transaction data was separated into groups based on **category** and were embedded using **Azure OpenAI** embedding model and stored in a vector database

2. **K-means clustering algorithm** was used to cluster historical transaction data. **Centroid information along with distance** of each data point from center were stored in vector database

3. **Model outputs** containing **new transaction description** along with mapped category was embedded using Azure Open AI embedding model

4. **Distance of new transaction** from centroid of mapped category was leveraged to estimate percentile for each new data point for **confidence score** calculation

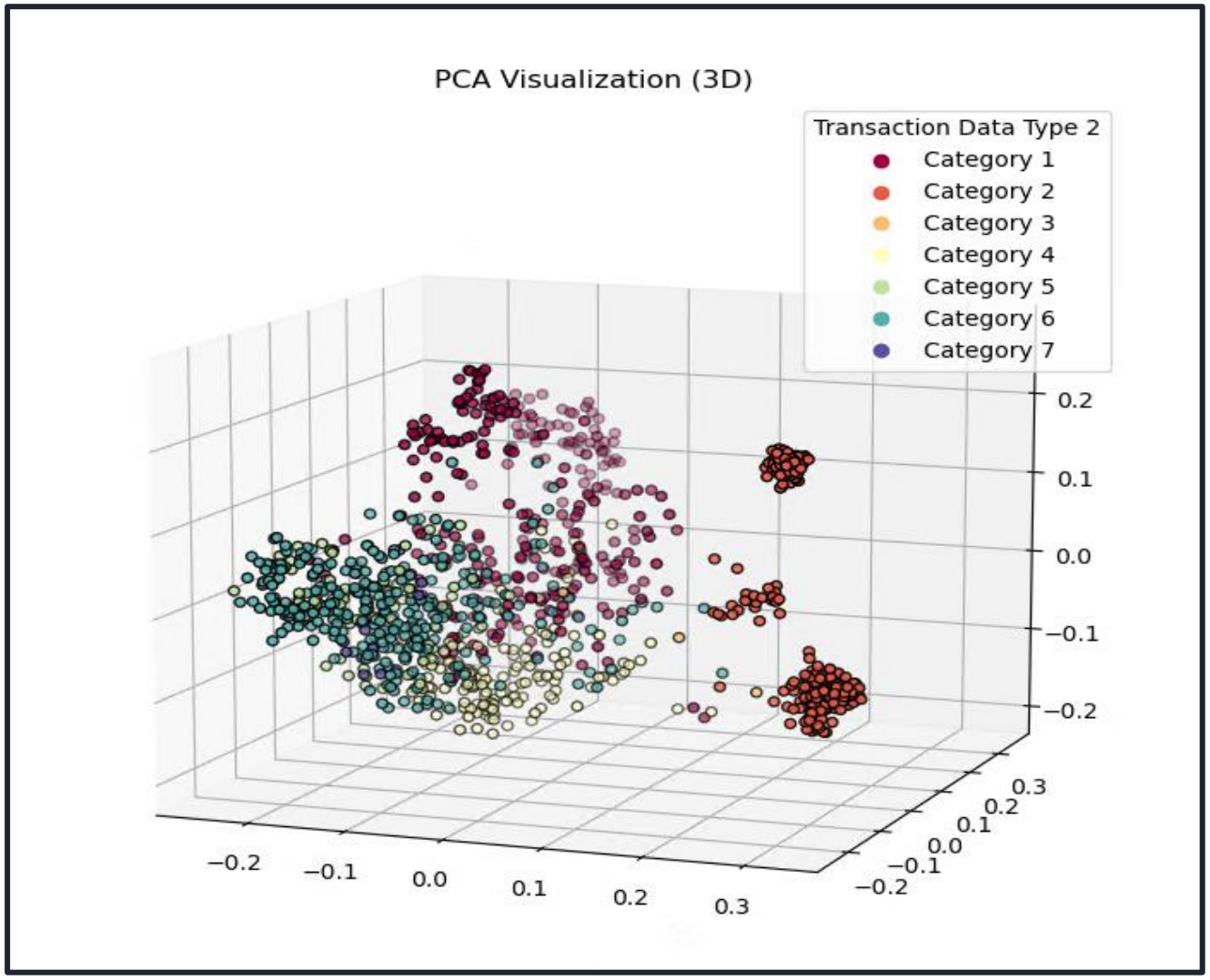
Model output

	Model Input		LLM Input
	Manual Override		Final Output

Practice ID	Practice Name	Description Type	Description	Type 1	Confidence Score (Type 1)	Type 1 Override	Type 2	Confidence Score (Type 2)	Type 2 Override	Transaction Category (Type 1)	Transaction Category (Type 2)
PracticeID 1	Practice 1	Description Type 1	Description 1				Category 1	Low	LLM is correct		Category 1
PracticeID 2	Practice 2	Description Type 1	Description 2				Category 2	High	LLM is correct		Category 2
PracticeID 3	Practice 3	Description Type 1	Description 3				Category 3	Low	Category 7		Category 7
PracticeID 4	Practice 4	Description Type 2	Description 4	Category 1	High	LLM is correct				Category 1	
PracticeID 5	Practice 5	Description Type 2	Description 5	Category 2	High	LLM is correct				Category 2	
PracticeID 6	Practice 6	Description Type 2	Description 6	Category 3	Low	Category 4				Category 4	

User input to enter a transaction category from drop down in case of override

Confidence score clusters



Descriptions for transactions which are similar in semantic meaning, were clustered together in vector space

Dimensions of data represented in vectorized form were significantly reduced to visualize in 3D