Exploring the Power of Generative AI in Enhancing Data Quality: An Overview of Use Cases and Benefits

In a world heavily influenced by data, where data analytics is becoming extremely prevalent in formulating strategic decisions, the quality of data emerges as a cornerstone for an organization's success.

Neglecting to enhance data quality can result in flawed decisions, inefficient strategies, operational dysfunction, missed business opportunities, and customer abrasion.

As organizations pivot towards transformative methods like Generative AI, numerous use cases can surface that could facilitate the enhancement of data quality.

This article illustrates the application of Generative AI in maintaining data quality through practical use cases.

Data Auditing:

Organizations need to assess their current state of data quality and identify potential issues. This allows them to take preemptive measures in fixing on time without disruptions. Gen AI predictive analysis capability can highlight current or future audit risks and generate recommendations.

Detection of Data Anomalies:

Data Anomalies are described to identify data points that deviate from expected patterns in a dataset. Data Anomalies can occur due to several reasons like data entry errors, Stale applications, malfunctions in the process that captures data, or even fraudulent activities. Training Data models with clean data can highlight data points that stand out from the distribution of normal Data points.
For example, Fraud in Billing. Identify abnormal upsurge in claims billing amount. Gen AI can be trained with patterns of real-time data. If there is a billing amount significantly different from learned patterns of amount, the system can raise a flag and notify the user of a potential detection of fraud.
Identify Patient condition: Anomaly detection can predict the health condition of a patient or the onset of a disease. Patient data can be fed into AI models and predictive algorithms. Health Insurers can improve patient safety and save many lives by implementing Gen AI.

Data Cleansing:

Gen AI use case in Identifying fulfilling missing data elements in a dataset.

Overall Data completeness is essential for efficient business operations, accurate decision-making, enhancing predictive Data Models, and Improving customer satisfaction.

Innovations in the healthcare industry can develop Gen AI's capability to find and correct missing information without having to reach providers. Simple things like the Physician's Race or ethnicity, and languages spoken can be updated from self-reported publicly available web portals.

Identifying Data errors by simplifying data patterns. Physician health plan network errors are a painstaking issue for claims processing which leads to a great amount of errors and manual intervention to correct. Frequent claims errors can potentially lead to provider or member abrasion. Gen AI model algorithms can be built to identify network errors and correct them improving accuracy.

Often, finding real-time data scenarios to feed models becomes increasingly difficult. An example would be identifying patient data to predict underlying cardiovascular disease. Gen AI models can be empowered to fulfill the lack of datasets. Data Imputation techniques can used to replace missing data elements with Synthetic Data.

Master Data Management:

Master Data Management (MDM) is a methodical approach an organization uses to establish a single source of truth for all its critical Data. Here are a few methods in which Data Quality can be enhanced using Gen AI.

1. Data deduplication:

    Gen AI can learn from existing data patterns and identify potential duplicates. For example, Multiple Data Entries could have been done for the same Physician with similar Names with the same DOB and National Provider Identifier. Gen AI Model can identify them as duplicates and merge them into a single golden data record.

2. Customizable Standardization & Correction

    Gen AI can customize standardization rules based on their specific requirements. This flexibility allows businesses to align data standardization processes with their unique data governance policies and industry standards.

    For example, Formalizing Addresses in a standard format can heavily reduce computing costs. Innumerable formats of the same address can emerge through various channels. In a real-world scenario Health plan contract enrollment platform, formats can bring different versions of the same address.

    Gen AI models can identify different versions of the same address and standardize them to a common deliverable address format. Missing Address components like missing city or postal codes can be corrected.

3. Hierarchy Recognition:

    Gen AI is equipped to recognize hierarchical relationships within datasets. This involves identifying parent-child relationships, identifying levels of hierarchy, and understanding relationships that connect various entities.

4. Probabilistic Record Linkage:

Gen AI employs probabilistic record linkage techniques to calculate the likelihood that two records represent the same real-world entity. This involves assessing the probability of match or non-match based on the similarity of features and historical data patterns.

5. Customizable Matching Rules:
Gen AI can define and customize matching rules based on their specific requirements. This flexibility allows businesses to tailor entity recognition criteria to align with their unique data quality and MDM objectives.

Data Consistency:

Data Consistency in healthcare is extremely important for several reasons like patient safety, quality of care, operational efficiency, and the list goes on. Gen AI brings a perfect use case for identifying and correcting data inconsistencies.

For example, Members often deal with difficulty in finding the right care. A cardiologist who practices in multiple hospitals is listed as a Cardiac specialist or cardiovascular disease specialist or Heart specialist or Cardiac care specialist. A provider who is a board-certified radiologist is also listed as a marriage counselor in the directory which can seem very absurd.

Gen AI can detect Data inconsistencies and can self-correct values within a dataset.

Natural Language Processing (NLP):

NLP is a specialized field in artificial intelligence that involves the interpretation of natural language through computers. The main goal of NLP is to read, decipher, and understand human language in the form of speech, and text in a valuable a useful way.

There are several subfields and techniques in NLP that can be utilized to improve data quality.

a. OCR (Optical Character Recognition) is a method by which handwritten text can be converted into a digitally readable format. In the HealthCare industry, OCR is used to read clinical notes, discharge summaries, and patient enrolment forms.
b. Text Classification is another technique that can be used to analyze data and label them in entity-based datasets.
c. Sentiment Analysis is a very common form of text classification to determine the sentiment or opinion expressed.
d. Named entity recognition: This method involves identifying and classifying names and entities in a text.
e. Speech recognition: Converting spoken language into written text. This is widely used to transcribe speech into written text. The text can be fed into a Text analysis algorithm to analyze and summarize into a meaningful, actionable format.

Overall, the outcomes and benefits achieved by organizations using Gen AI in Data Quality can accelerate growth in leaps and bounds. Here are specific key impacts.

Improved Accuracy: Deploying Gen AI methods can give plenty of opportunities to automate processes reducing time, effort, and human errors.

Enhanced Decision Making: Higher quality of data can facilitate organizations to improve strategies and data-driven decisions.

Cost Savings: Gen AI methodologies can reduce the extensive manual effort of data cleansing, data analysis, and text interpretation.

The administrative burden with maintaining Provider directories is humungous as cited in below article by CAQH.
$2.76 Billion: The cost of directory maintenance to US physician practices
https://www.caqh.org/sites/default/files/explorations/CAQH-hidden-causes-provider-directories-whitepaper.pdf

Automated data quality processes can reduce time and labor costs and the potential cost of decision-making based on poor data quality.

Compliance: AI can assist in adhering to data regulations by maintaining data accuracy and providing records of data processing for audits.

These benefits demonstrate how integrating AI into data quality systems can lead to enhanced efficiency, accuracy, and strategic insights.

Conclusion:

In the current digital age where data analytics is crucial for strategic decision-making, data quality has become fundamental for a company's success. Ignoring it may lead to various consequences such as poor decisions, ineffective strategies, missed opportunities, and customer dissatisfaction. With organizations now utilizing transformative methods like Generative AI, several use cases can enhance data quality.
Harnessing the power of Gen AI can unlock new opportunities for growing businesses fueled by data-driven decisions.