

Producing Efficient Structured Query Language code using Generative Artificial Intelligence Technologies

George Garman
Metropolitan State University of Denver

Abstract

Organizations have managed their data within relational database systems for many decades. The large enterprise RDBMS have been the norm for creating a repository of operational data and the Structured Query Language (SQL) has been the standard for data retrieval since the 1980's. With the added requirement of storing and processing both structured and unstructured data over the past several decades, data warehouses have become an indispensable tool for data managers. With the past decade, a new and somewhat controversial approach to assist data managers has emerged. This new tool, generative artificial intelligence (AI), is rapidly being integrated into database management system conventions. This paper examines how the data analyst can use generative AI to write the SQL code that will effectively retrieve data from the relational database.

This paper examines some of the commercial AI to SQL code generators that are currently being offered including ai2sql, text2sql, and InsightBase. ChatGpt and Microsoft's Copilot are two of the openAI chatbots that take as input natural language queries and produce human-like responses. ChatCpt is an add on product while Copilot is embedded deep within Microsoft products. This paper also examines the Large Language Models (LLM's) that are trained on large data sets to rapidly produce SQL code.

FVC24002