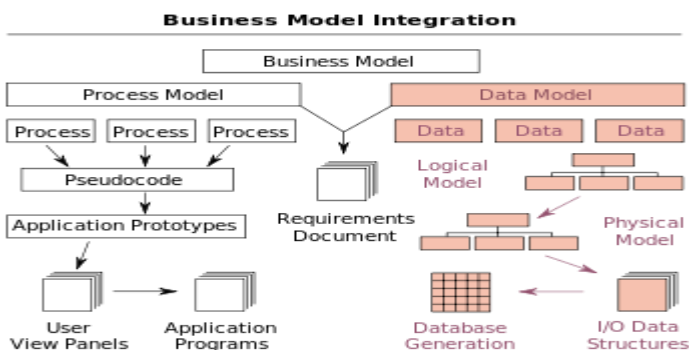


BIG DATA ANALYTIC : Data modeling and source code generation and ai

ARTICLE 1: A Metric Model for Automatic Code Generation using GPT-3

Today, AI technology is showing its strengths in almost every industry and walks of life. From text generation, text summarization, chatbots, NLP is being used widely. One such paradigm is automatic code generation. An AI could be generating anything; hence the output space is unconstrained. A self-driving car is driven for 100 million miles to validate its safety, but tests cannot be written to monitor and cover an unconstrained space. One of the solutions to validate AI-generated content is to constrain the problem and convert it from abstract to realistic, and this can be accomplished by either validating the unconstrained algorithm using theoretical proofs or by using Monte-Carlo simulation methods. In this case, we use the latter approach to test/validate a statistically significant number of samples. This hypothesis of validating the AI-generated code is the main motive of this work and to know if AI-generated code is reliable, a metric model CGEMs is proposed.

Currently, though there are several researches regarding code generation, there aren't any good holistic researches carried out so far to validate the 'AI-generated code', verified for its efficiency, correctness, and usability for the consequences described in Figure 1 . The main utility of using AI is to reduce the burden and ease the completion of tasks. If automation or AI does not aid in this, the whole point of using AI becomes fruitless. These metrics serve to support the hypothesis, as well as a reference to future code generation techniques.



We ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a good metric composed by 4 metrics who are

maintainability, complexity, Halstead, Raw metrics

Raw metrics include: LOC (Lines of code): Total number of lines of code. LLOC (Logical Lines of Code): The number of logical lines of code.

The metric model CGEMs for automatic code generation was studied in this work on a limited number of AI-generated codes. To the best of our knowledge, this is the first work in which ANN has been combined with codemetrics to assess the AI-generated codes.