



INFORMATICS  
INSTITUTE OF  
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**Auto Commenter: A hybrid modelled transformer-based  
approach to generate source code comments**

A dissertation by

Mr. Ahamed Muaaz

Supervised by

Ms. Minoli De Silva

Submitted in partial fulfilment of the requirements for the  
BEng (Hons) Software Engineering Degree Department of Computing

May 2021

© The copyright for this project and all its associated products resides with the  
Informatics Institute of Technology



## **Abstract**

Source code summarization or comment generation is the task of producing readable natural language annotations for the given source code which can be used to comprehend source code with less effort. Software engineers use comments while programming to understand source code but unfortunately writing comments remains a time-consuming task therefore there has been substantial interest in automating the process by the software engineering and machine learning community over the period. Even though there were several approaches they failed to catch the source code's long-term dependencies and the pairwise relationship between code tokens therefore in this research a novel transformer-based hybrid modelling approach was used to solve the issue of generating comments.

Auto Commenter system was built with the intention of generating accurate comments for multiple programming languages. This dissertation act as proof for the success of the Auto commentator in achieving its aim. Auto commenter uses a hybrid modelling mechanism to learn different information of source code such as the syntactical and structural information. The combined source code modelling along with the state of the art transformer architecture which uses multiple attention mechanism and position-wise encoding gave a promising result and made it a success above other available solutions.

**Keywords:** Automatic comment generation, Code comprehension, Natural language processing, Neural machine translation, Sequence to sequence modelling.