



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**PseudoFlow: Generating Algorithm and Flowchart from Natural
Language Input**

A Project Proposal by

Ruvik Perera

W1761096 / 2019378

Supervised by

Mr. Guhanathan Poravi

Submitted in partial fulfillment of the requirements for the BEng (Hons) in Computer
Science degree at the University of Westminster.

February 2022

ABSTRACT

In this thesis, we introduce the design and development of an innovative, intelligent system aimed at enhancing the learning experience for beginners in programming as well as providing an assistive tool for more seasoned developers. The system takes natural language input, representing a problem or task, and outputs an algorithmic pseudocode representation. It further generates flowcharts using Graphviz, providing a visual representation of the algorithm, thereby assisting in understanding complex algorithms. The approach used in developing this system is centered around a Long Short-Term Memory (LSTM) model, a type of recurrent neural network that is proficient in sequence prediction problems. The LSTM model was trained with a dataset of problem-description and corresponding pseudocode pairs, allowing it to learn the mapping from natural language instructions to algorithmic pseudocode.

A comprehensive review of existing literature exposed a significant gap in the availability of tools that perform such tasks, thus underlining the need and significance of this project. The development process was systematic and followed a structured approach, from the elicitation of requirements to the detailed design and implementation of the system's core functionalities. In addition to the system design and development, this thesis also delves into the testing and evaluation of the system. Evaluation metrics such as BLEU, METEOR, ROUGE, and CIDEr were used to assess the LSTM model's performance, while the system as a whole was tested and evaluated against functional and non-functional requirements.

The project faced limitations, including time constraints and lack of access to evaluators, which are candidly discussed. Despite these challenges, the system was successfully developed and has demonstrated potential to be a valuable tool in programming education and practice. This thesis also outlines future improvements and enhancements to the system. These include refining the LSTM model for more accurate pseudocode generation, expanding the scope of problems the system can handle, and developing an interactive user interface for a more engaging user experience. As a unique educational and developmental tool, this system has the potential to revolutionize how programming is learned and practiced.