

ĐẠI HỌC QUỐC GIA
THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC QUỐC TẾ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Phụ lục

(Kèm theo Quyết định số /QĐ-ĐHQT ngày tháng năm 2023
của Hiệu trưởng Trường Đại học Quốc tế)

Phụ lục 1.1 - Mẫu đăng ký đề tài LVTN

Thesis Registration Form

INSTRUCTIONS:

Student: Please fill your name and ID on the form and discuss with your Thesis Advisor to fill the other contents. Afterward, sign the form to verify your agreement and return the form to the Undergraduate Academic Assistant of the Department.

Thesis Supervisor: Please authorize the Thesis Registration Form for this student by signing below.

Student's name: Nguyen Nhat Kiem

ID: ITDSIU21091

Email: ITDSIU21091@student.hcmiu.edu.vn

Phone: 0899325497

Thesis title: AI-Driven System for Analyzing Complex Documents and Automatically Generating Flowcharts

Major: Data Science

Thesis goals and objectives:

(brief description of the project output)

This thesis aims to develop an AI system that simplifies the understanding of complex documents (e.g., legal contracts, financial reports) by converting them into easy-to-read flowcharts. The system will use advanced Natural Language Processing (NLP) techniques to analyze document structure, identify key entities and relationships, and organize the information into a visual flowchart. The system will feature an interactive interface for exploring these flowcharts and will be optimized for scalability through Docker and Kubernetes deployment.

Requirements:

(list all major requirements for the thesis work, refer to the guidance on the next page for details)

- Document Input and Preprocessing: Support for various formats (PDF, DOCX, TXT), with a pipeline for text extraction and normalization.
- NLP for Structure Understanding: Use of NLP models (LLaMA 3, Phi 3) to segment documents, identify entities, and detect relationships.
- Text Embedding and Storage: Generate text embeddings and store them in a vector database for fast retrieval.

- Information Extraction: Identify key clauses, conditions, and decision points, organizing them logically.
- Flowchart Generation: Convert structured data into flowcharts using graph algorithms to represent document logic.
- Interactive User Interface: A web-based interface (Streamlit) allowing users to interact with the flowcharts (zoom, click, etc.).
- AI Model Optimization: Use PyTorch to refine and evaluate AI models for improved accuracy.
- Deployment and Scalability: Containerize with Docker and deploy with Kubernetes for CI/CD and scalability.

Name of Supervisor 1

Name of Student 1

Date of Signed

Date of Signed

Name of Supervisor 2 (*optional*)

Name of Student 2 (*optional*)

Date of Signed

Date of Signed