HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS

Thesis title: Improving Vietnamese Question Generation Using Reinforcement Learning

Nguyen Hoang Dang
dang.nh204873@sis.hust.edu.vn

Major: Computer Science
Specialization: Data Science

Supervisor:	PhD. Tran The Hung	______________ Signature
Department:	Computer Science	
School:	School of Information and Communications Technology	

HANOI, 06/2024

ACKNOWLEDGMENT

This is my graduation thesis, marking the end of 4 years of study at Hanoi University of Science and Technology. During the past 4 years, I feel that I have had the opportunity to learn and grow through the chance to be exposed to problems that are not only new and applicable but also difficult. With the enthusiasm of the teachers and their dedicated guidance, I am even more passionate about the field of study I have chosen. These are not only useful knowledge but also a solid foundation for me to continue building and pursuing my educational path. I express my thanks to all the teachers who instructed me to complete my subjects here. These lessons always help me in the process of building my future career. Last but not least, I want to express my gratitude to PhD. Tran The Hung, who inspired me to select, complete the thesis, and achieve the expected results.

ABSTRACT

In the field of natural language processing, automatic question generation is reckoned to be a challenge. This is due to the ambiguity where many questions can lead to the same answer. We want our model, after being fine-tuned by on reinforcement learning task, to be able to generate questions that not only resemble the style in the data set but also appear natural and contextual. We introduce a selection method for the preference dataset which is then used for a preference-based reinforcement learning method to align the policy language model. Additionally, we enlarge the training dataset using synthetic data generated from a large language model (LLM) fine-tuned with instruction. Our training method enhances the model stability on different sampling methods, generating better-rated questions in terms of AI evaluation. Quantitatively, our method improves the model performance, showing an increase in Rouge-L (+3.44), BLEU-4 (+3.18), and Win rate (+12.73). The advancement made by our method on the model is also explained in detail.

Student
(Signature and full name)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Definition**
API	Application Programming Interface
AQG	Automatic Question Generation
DPO	Direct Preference Optimization
LLM	Large Language Model
PPO	Proximal Policy Optimization
RL	Reinforcement Learning
RLAIF	Reinforcement Learning from Artificial Intelligence Feedback
RLHF	Reinforcement Learning from Human Feedback
SDPO	Self-play Fine-tuning with Direct Preference Optimization
SFT	Supervised Fine-tuning