A Personal Virtual Assistant Using a Fine-Tuned Language Model

Trần Xuân Thanh
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student615@ictu.edu.vn

***Abstract*—This undergraduate thesis investigates a personal virtual assistant based on a locally adapted large language model. A conversational assistant needs a model that can answer questions in its intended language and domain. The student studies Transformer and LLaMA-family models and applies four-bit QLoRA adaptation to a roughly seven-billion-parameter model. The manuscript describes training settings and example Vietnamese question-and-answer interactions. The reported training table shows validation accuracy rising to 0.764010 at epoch 7, with validation loss of 0.139995. These are source-reported training statistics, not an independent assessment of factual answer quality. The thesis does not present a blinded benchmark, hallucination rate, privacy assessment, or comparison with other assistants on the same questions. The contribution is a documented student study or prototype whose stated scope can guide further testing.**

***Index Terms*—virtual assistant, large language model, QLoRA, Vietnamese**

I. Introduction

A conversational assistant needs a model that can answer questions in its intended language and domain. This thesis investigates a personal virtual assistant based on a locally adapted large language model. It begins from a practical engineering need and explains the technologies and workflow chosen for the case. The objective is to document a student implementation or comparative investigation, together with the evidence available in the manuscript. Broader operational or economic effects require separate measurements and are not assumed from the existence of a prototype [1]. Related published studies provide context for the subject and methods [2], [3], [4], [5], [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The student studies Transformer and LLaMA-family models and applies four-bit QLoRA adaptation to a roughly seven-billion-parameter model. The manuscript describes training settings and example Vietnamese question-and-answer interactions. The source is summarized through its stated design, implementation material, examples, and conclusion. Features shown in the manuscript are distinguished from proposals for later development. This review does not add a new experiment, fill missing measurements with estimates, or treat a planned capability as an observed result. The source thesis remains the provenance for the reported work [1].

Supervisor: Ngô Hữu Huy, Thai Nguyen University of Information and Communication Technology (nhhuy@ictu.edu.vn).

III. Results

The reported training table shows validation accuracy rising to 0.764010 at epoch 7, with validation loss of 0.139995. These are source-reported training statistics, not an independent assessment of factual answer quality. The report's screens, configurations, examples, or measurements constitute the evidence for this summary. A capability described only as a design intention is not counted as a validated operational outcome. No additional performance, adoption, financial, or safety figure has been introduced here. The result is interpreted within the scope of the student's documented setup [1].

IV. Discussion

The thesis does not present a blinded benchmark, hallucination rate, privacy assessment, or comparison with other assistants on the same questions. Further work could create a held-out domain test set and evaluate factual correctness, safety, and user usefulness separately from training metrics. These limitations affect how the evidence should be read. The current manuscript can support understanding of the design and reported demonstration, while independent tests would be needed for stronger claims. Any later deployment should document its environment and separate observed outcomes from intended benefits [1].

V. Conclusion

The thesis documents a personal virtual assistant based on a locally adapted large language model. The reported training table shows validation accuracy rising to 0.764010 at epoch 7, with validation loss of 0.139995. These are source-reported training statistics, not an independent assessment of factual answer quality. The thesis does not present a blinded benchmark, hallucination rate, privacy assessment, or comparison with other assistants on the same questions. Further work could create a held-out domain test set and evaluate factual correctness, safety, and user usefulness separately from training metrics. Accordingly, the present contribution is a source-grounded account of the student work and its tested or illustrated scope. Claims about broader effectiveness should be assessed in later work using explicit evaluation conditions [1].

References

[1] Trần Xuân Thanh, ”A Personal Virtual Assistant Using a Fine-Tuned Language Model,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] Sudarshan Modhave, ”A Survey on Virtual Personal Assistant,” International Journal for Research in Applied Science and Engineering Technology, 2019. doi:10.22214/ijraset.2019.12050.

[3] Ms. Shruti Sarwate et al., ”Personal Virtual Assistant (“DANI”),” International Journal of Innovations in Engineering and Science, 2022. doi:10.46335/ijies.2022.7.8.14.

[4] MD Mohsin Ali et al., ”Virtual Assistant Using Supervised Learning,” International Journal for Research in Applied Science and Engineering Technology, 2023. doi:10.22214/ijraset.2023.54262.

[5] Sumitkumar Sarda et al., ”VPA: Virtual Personal Assistant,” International Journal of Computer Applications, 2017. doi:10.5120/ijca2017913619.

[6] Impana N.R et al., ”VOICE AND TEXT BASED VIRTUAL PERSONAL ASSISTANT FOR DESKTOP,” International Journal of Engineering Applied Sciences and Technology, 2022. doi:10.33564/ijeast.2022.v07i01.030.

[7] Hai Son Hoang et al., ”Design and implementation of a VoIP PBX integrated Vietnamese virtual assistant: a case study,” Journal of Information and Telecommunication, 2023. doi:10.1080/24751839.2023.2183631.

[8] Andrei Vasilateanu et al., ”Call-Center Virtual Assistant Using Natural Language Processing and Speech Recognition,” Journal of ICT, Design, Engineering and Technological Science, 2018. doi:10.33150/jitdets-2.2.3.

[9] Tummala Sri Ranga Sai Krishna, ”Virtual Personal Assistant for Desktop Automation using Selenium,” International Journal for Research in Applied Science and Engineering Technology, 2021. doi:10.22214/ijraset.2021.35798.

[10] Mayank Raj et al., ”VIRTUAL ASSISTANT USING ARTIFICIAL INTELLIGENCE,” International Journal of Engineering Applied Sciences and Technology, 2023. doi:10.33564/ijeast.2023.v08i02.053.