MUHAMMAD AZRIL SHUKOR

2024736165_MUHAMMAD AZRIL_20252_5A_CHECK.pdf

FYP 2026

FYP 2026

Universiti Teknologi MARA

Document Details

Submission ID

trn:oid:::13381:126845655

Submission Date

Jan 25, 2026, 3:21 PM GMT+8

Download Date

Jan 25, 2026, 3:26 PM GMT+8

File Name

2024736165_MUHAMMAD AZRIL_20252_5A_CHECK.pdf

File Size

2.9 MB

101 Pages

26,595 Words

147,532 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

CHAPTER 1

INTRODUCTION

This chapter provides the background and rationale for the study. It also gives details of the significance of privacy over the Internet, the issues and problems that led to this research.

1.1 Background of Study

Financial Literacy is an ability to effectively understand the use of different financial skills, including personal financial management, budgeting and investing. It was an important key to have a good relationship with money when people are having knowledge. According to Mihalčová B, Csikósová A, Antošová M (2014) financial literacy is an ability to use knowledge, skills and experience of an individual to make effective decisions. Financial literacy is a key in decision making in our lives. Financial Literacy important because it is related to our day-to-day life. According to (Mihalčová et al., 2014), it is essential to avoid the excessive risk when in need of finances. By having basic financial literacy can make someone to handle daily management tasks. Financial literacy also can make strategic financial decision with confidence.

Financial Literacy does cover understanding basic financial concepts. This includes knowing how to manage how money works. According to Financial Industry Collective Outreach (Finco) ensuring young people in Malaysia understand how money works so they empowered to make sound financial decision. Budgeting includes tracking income and expenses to avoid overspending. Another area financial literacy covering is saving. According to Mihalčová et al. (2014), saving is to incorporate future needs, which includes emergency funds and long-term goals. Investing is a core component, because it is associated with risk and potential returns. According to Klapper & Lusardi (2019) insufficient knowledge about risk and interest can lead to market instability. Debt management also covered in financial literacy. The understanding to interest rates, repayment terms and how to manage debt responsibly.

In the digital age, digital financial literacy is a key important that must be taken seriously. Digital financial literacy does cover knowledge about safety. Based on Adnan Z. M. Zahid (2024) the practice of sharing their banking passwords exposes them to the increasing rate of online scams. Overall, with the capabilities to manage finances well by making a good choice can reach stability and independence in people.

Malaysian youth are the most affected by the low of financial literacy. Financial Industry Collective Outreach (Finco) stated nearly 60% of Malaysians declared bankrupt from 2018 to 2022 being between the ages of 25 to 44. Individuals with limited knowledge regarding financial literacy are greatly exposed to the possibility of becoming a fraud victim, making poor decision regarding finance and accumulate too much debt. Other than that, majority youth have poor financial behaviour and attitudes in managing their responsibility. According to the survey that has been held by Financial Industry Collective Outreach (Finco), most students tend to prioritise short term gratification over long term planning. This also exposed that majority youth do not have healthy financial behaviour. Insufficient of financial education in schools also becoming a concern. According to Financial Industry Collective Outreach (Finco), financial education needs to be timely and relevant to each student community. This shown that age do matters because a good understanding in earlier exposure can make them adapt to the transition into work or further education.

Financial literacy education appears giving a large of impact when it is provided at early stage. The early stage mentioned specifically at young ages for teenagers and children when they face their first financial decisions. By ensuring young people in Malaysia understand how money works so they are empowered to make sound financial decisions for long-term financial (Financial Industry Collective Outreach (Finco), 2023). The findings from the survey also shown that if the student have the early exposure, they will have a good chance to achieve a good understanding. According to Lussardi & Messy (2023), young people who were exposed to high school regarding financial education are less likely to have problems with debt as young adults. Next, during the transition from student to adulthood. Financial education needs to be convenient and appropriate. The appropriate content means what is the education currently focusing on. For example, basics understanding such as personal budgeting, understanding their salary, student loans and how interest works (Financial Industry Collective Outreach (Finco), 2023). So, financial literacy education does important and had the greatest impact to the youth. It also suggests that building a strong foundation can influence financial well-being in the future.

The fundamental of financial literacy has been injected for preschool, primary and secondary schools. This such effort does align with the priority to educate people in early age. For example, under the national strategy for Financial Literacy 2019-2023, an effort to address has been made by the Ministry of Education (Financial Industry Collective Outreach (Finco), 2023). However, it is not good enough. Not enough attention is given to teach young people within this curriculum. Through the findings from Financial Industry Collective Outreach (Finco) the knowledge should be improved so that students are equipped with the fundamentals of financials a soon they transition into adulthood. Next, from various event and workshops that been held. FINCO did run Financial Literacy programmes that support the National Strategy for Financial Literacy 2019-2023. The survey that been conducted include identify key trends and approaches to improve youth financial literacy. For example, students' knowledge, awareness, behaviour and attitude towards money and financial products and services (Financial Industry Collective Outreach (Finco), 2023). Despite these efforts, the gaps between effective financial literacy education still lacking a formal financial literacy program.

Many Malaysian aged younger than 44 had been recorded for majority in bankruptcy. According to Ahmad & Mohamed Zabri (2023), it was discovered that majority bankruptcy cases in Malaysia between 2019-2023 involved individuals under the age of 44. Malaysian continues to face three main challenges. These includes poor financial management, saving habits and lack of digital financial literacy skills. For example, one in four Malaysians feels their debt is burdensome (Ahmad & Mohamed Zabri, 2023). Bad life choices and impulsive buying worsen this financial problem. This will increase them to make poor financial decision. According to Lusardi & Mitchell (2023), the issue is intensified by lifestyle choices. Next, is to create awareness for young people to make good decision making. With a good decision making will empower individuals. According to Ahmad & Mohamed Zabri (2023), it greatly enhances their capacity to handle their finances. Other than that, financial literacy expected guide individuals' behaviour in society. With financial literacy can equip youth with essential financial management. Financial literacy will enable them to make sound financial decisions and protect their stability. Coincide like been recorded by Ahmad & Mohamed Zabri (2023), It is crucial for the younger generation to have a strong understanding of financial literacy to effectively manage their finances and ensure a secure financial future.

1.2 Problem Statement

Recent years have seen increasing attention toward improving financial literacy. However, Malaysian youth still display lacking knowledge regarding the financial literacy. Youth capabilities still lead to poor in decision making. Several reports demonstrate how concerning this issue. Many young Malaysians also lack sufficient money to pay for even a basic emergency expense. According to Adnan Z. M. Zahid (2024), 61% of Malaysians have difficulty to come up with RM1,000 in case of emergency. This shown that a surprising fact to indicates inadequate financial reserves. Financial stability that be faced lead to serious outcome. For example, with nearly 60% of Malaysian between the ages of 25 to 44 were declared bankrupt from 2018 to 2022 (Financial Industry Collective Outreach (Finco), 2023). Young Malaysians do faces serious financial instability. Financial instability was because of their lacking knowledge in financial literacy education and behaviour.

Firstly, the lack of knowledge among youth of basic money concepts and practical skills to manage a good financial management. Many young people find that it is hard to handle the complexities of modern financial products. Financial product can be understood as various services and accounts provided within the financials system. The knowledge gap impacts basic aspects. For example, budgeting, saving, investing and debt management. With the lack of knowledge, skills and behaviour make it as an important to young people. Most youth lack the basic to use financial skills effectively. Which made them think that financial management is a challenging and saving is too hard. The absence of fundamentals will add negative outcomes. For example, high levels in debt and poor in making decision regarding financial. This will be ignoring the youth in multiplying their wealth opportunity. The youth are indeed the most affected by low financial literacy.

Other than that, the low levels of digital financial literacy currently form a serious challenge for youth in current globalisation. With the growth of the concept of financial literacy in digital era they are easily becoming a target. For example, in online fraud and scams. According to Adnan Z. M. Zahid (2024) 15 per cent of the population, mostly teenagers, sharing their banking passwords. Indirectly, makes them vulnerable to the increasing of online scams. The financial environment has been complicated. Today digital era and services are important. However, many young people do not have the exact digital financial literacy knowledge they require. This is a great weakness in their skills preparation. They need to understand how to handle their finances safely and practical.

1.3 Research Questions

Based on the stated problem statement, the project will address following research questions.

i. What is the educational material on personal finance?
ii. How can we fine-tune a LLM to explain financial literacy?
iii. How can a web-based chatbot be developed to teach financial literacy to Malaysian youth?

1.4 Research Objectives

The research objectives of this project are as formulated below:

i. To identify educational material on personal finance.
ii. To design a fine tune approach for LLM model on financial literacy.
iii. To develop a web-based application for educational chatbot of personal finance using LLM model.

1.5 Scope and Constraint

This initiative is focusing around creating AI-generated financial literacy chatbot that makes use of the capabilities of a pre-trained Large Language Model (LLM). This system is to help overcome the growing challenges of managing finances. The system solves issues of the various aspects of finance, lack of financial knowledge, and inability to follow expenses and budgets. Our goals informed by the sources and the dialogue center around building an AI-based financial literacy. The basic objective is to assist individuals in dealing with their finances, complexities of modern finance, what to learn about finance, and how to monitor spending and budget. A pre-trained Large Language Model (LLM) is the underlying technology of this project. At the heart of the objectives of this project are the offering of budgeting tips as well as financial planning practice all aimed at the guidance of the users towards a better management of their finances. Automating expense tracking within the scope of the project, personalized budget recommendations through machine learning, customized investment guidance using predictive analytics, and the provision of educational resources, which will increase the users' understanding of finances, are all included in the project's scope. The end goal of the project is that users will be empowered, through their financial education and their improved decision making in finances, their pursuit of their financial dreams, and cultivating a basis for sound financial planning.

The goal of this initiative is to create AI-based personal finance assistant or chatbot with pre-trained large language model (LLM) as the primary technology. The biggest problem is that, despite the ability to generate natural language, LLMs can't deliver reliable accuracy in financial facts, and it does so consistently. They are often not able to perform logical and symbolic tasks well and can result in erroneous or non-existent outputs thus, making them not very suitable for precise financial situations. One of the major problems with LLMs in the finance domain is that they are only marginally capable of doing careful calculations and logical numeric reasoning. Hiring external numeric solvers may mitigate this problem. Besides, the restrictiveness of training data cutoff date of LLMs makes it difficult to leverage them on applications that require up-to-date financial data. It is claimed that it is a costly and time-consuming affair to keep models up to date with developments in finance. Building a successful domain adaptation for LLMs in context-specific-finance is a big issue itself. Fine-tuning, although an improvement in performance, frequently includes substantial cost, and can lead to compromising the integrity of the base model. The differences between LLM responses while updating pose considerable hurdles in developing reliable static test cases for financial system evaluation. To systematically examine the response quality of the user, it may be especially difficult. It is being feared that some benchmarks do not cover adequately domain specific needs or approaches to evaluate consistently. Because of the sensitive nature of finance, there is a need to ensure that whatever is created and released is free from toxicity or bias. Ensuring ethical use requires that deployment can only occur after thorough risk and test assessment have been made.

Design considerations for this initiative revolve around building an AI-based personal finance assistant or chatbot on the skills of a pre-trained Large Language Model (LLM). The Platform is supposed to help people to overcome the challenges of personal finance management, such as the complexity of financial systems, low financial literacy, and difficulty managing their saving and expenses. Evaluation of the system, according to the OECD-INEF Toolkit for Measuring Financial Literacy and Inclusion of Others 2022.

1.6 Project Significance

The significance of this project to come out the address about the low financial literacy levels among youth. The financial ignorance of young Malaysians that may have major negative impacts not only on people but on society with the passage of time. Malaysian youth are prone to face the same issue recurring to financial management. For example, acquiring debt, deficient saving and having no savings to cope when there are unexpected expenses. According to RinggitPlus Malaysian Financial Literacy Survey 2023,50% of Gen Z respondents are month-to-month, 26% of Malaysians under 35 have limited savings to help emergency situations. With these figures, it is quite easy to see that youth are still not prepared. There is still a long way for them to manage themselves and set them up to suffer from possible economy instability. Notably, if this financial education deficit continues, it will only make it harder for them to manage their finances, thus complicating achievement of financial freedom and long-term prosperity.

In addition to that, this vulnerability goes beyond individuals and contributes to the worsening of bigger economic issues too. Where so many citizens are failing in financial literacy, it adds pressure to national safety nets and put at risk to the stability of the economy. When youth are not able to save or manage their debt well, they tend to have other financial issues that then affect them in terms of their health, their personal lives, and opportunities at their job. Such issues can increase demands on public finances and thereby cause governments to turn to subsidies and vital welfare schemes to cushion people against financial distress. The consequences of ignoring this situation can be far reaching, with impacts well beyond personal finances and an ability to undermine the long-term economic well-being of the nation.

By adjusting its content to Malaysia, the chatbot ensures that recommendations are relevant and easy to understand. The chatbot is concerned with solving crucial Malaysian issues, like EPF, PTPTN student loans, and Islamic finance products that are central to our financial sector. This service ensures that a given user is presented with advice that would be in accordance with their cultural setting. This will be resulting in informed decisions based on reliable information that is relevant to Malaysia's financial scene. Moreover, it will use an evaluation such as the OECD/INFE Assessment of Financial Literacy. By using this evaluation, can measure the effectiveness of the chatbot in raising financial literacy among Malaysian youth. One can gain valuable insights on how wealthy technology could support money education. The main function of the chatbot is to instruct on financial ideas, but it will also encourage more responsible financial decisions, such as regular savings, expenditure control, and wise use of credit.

The project is intended to achieve SDG 4 (Quality Education) and SDG 10 (Reduced Inequalities). As well as advancing inclusive and sustainable economic development by making financial education available to marginalized youth. With a digital platform, the chatbot empowers the Malaysian youth with just-in-time solutions of financial educations. The dialogical and approachable nature of the chatbot serves as a tool to allow people to take responsibility over their money management and make sound financial decisions. Apart from enhancing financial literacy, it will develop a money-conscious, empowered generation. Who is crucial in maintaining and progressing Malaysia's economic future. In the light of the changing nature of financial markets, this chatbot becomes an essential tool for setting young people on the path to achieving financial independence and success.

CHAPTER 2

LITERATURE REVIEW

This section reviews on the previous studies and articles that have been done by other researchers that is related to this project. This section will also discuss on the techniques to be used in the research project.

2.1 Introduction to Financial Literacy

The term financial literacy describes the skills and knowledge needed to manage your money well. According to Aleksandrova et al. (2024), It is a much broader concept that includes both the knowledge and skills needed to understand financial matters and make the right financial choices In most cases, it refers to using knowledge and skills in managing money, such as saving, budgeting, investing, and borrowing, to help you handle your finances more wisely. Individuals must navigate these options wisely to ensure financial stability and growth (Lakshmi, 2024). With a proper education regarding financial, people can prepare for what comes next and deal properly with problems related to money. Financial literacy is very important. As example, it teaches people to use money like loans, credit cards, or insurance safely. According to Lussardi & Messy (2023), Financial literacy is an essential skill for making savvy financial decisions, understanding the world around us, and being a good citizen. This can prevent them from moments of financial stress. Basically, understanding finances well gives individuals the strength to maintain their financial security and that of their families going forward.

2.1.1 What Is Financial Literacy

A well-informed person with financial literacy is equipped to handle their money properly. Financial literacy matters on many levels (Klapper & Lusardi, 2019). This also means you should know about simple financial terms. For example, interest rates, inflation, and the need to spread your risks in investments. This also include as well as how to handle your personal money by planning a budget, saving for the future, and choosing where to invest. An example of financial knowledge is realizing that the interest and fees are what adds up to the actual cost of the loan, using the Annual Percentage Rate (APR) to measure this. According to Klapper & Lusardi (2019), consumers who are more financially literate may be better equipped to make investment and product decisions Basically, it involves applying your understanding to practical financial tasks. Understanding finances properly helps people avoid getting too deeply into debt and having too little saved. Financial literacy becomes a must for informed consumer use along with adequate financial protection (Lussardi & Messy, 2023). On the other hand, those who are financially literate often budget more wisely and financially prepared, enhancing their own well-being and security.

2.1.2 Traditional Vs. Digital Financial Literacy

With the pressure of digital technology, the definition of financial literacy is changing. In its place is “digital financial literacy”, encompassing new areas (RinggitPlus Malaysian Financial Literacy Survey 2023, 2023). Differences between traditional and digital financial literacy also can be seen. Financial literacy, at its core, encompasses the knowledge and skills necessary to make informed financial decisions (Lakshmi, 2024), such as how interest and inflation work, and how to make a budget and save. These are classic principles, such as being able to estimate how much interest you will pay on a loan or understanding why diversifying can decrease investment risk are. However, now on the agenda and is becoming an increasingly important aspect of education in the FinTech era (Aleksandrova et al., 2024). Digital financial literacy does apply those principles to the contemporary technology-oriented financial environment. According to Norzitah Abdul Karim, Zainora Ab Wahid, Sharifah Heryati Syed and Siti Aminah Mainal (2025) ,digital tools can make financial education more interactive and engaging, they also expose teenagers to online scams and fraudulent schemes. Not only does it demand an acquaintance with the rudiments, but also the digital competence to operate electronic devices.

As digital tools become an integral part of personal finance management, Malaysia has shifted its focus towards enhancing digital financial literacy (Sharifah, 2024). There are a few differences that explain why digital financial literacy is viewed as a necessary skill in the modern world. The pandemic has sparked increased use of digital technologies, including apps and platforms integrating AI to assist with transactions and financial planning (Aleksandrova et al., 2024). On the one hand, digital finance provides brand new financial products and services. For example, mobile payment applications, and online banks. Individual that digitally literate top up an e-wallet, compare online investment platforms, or an individual's credit score and offer tips on how to improve it (Lakshmi, 2024) on a fintech app. Second, digital financial literacy places a large focus on cybersecurity and fraud prevention. The focus will reflect the growing need for awareness in the face of increasing cyber threats (Sharifah, 2024). In a cash-based society, there are pickpockets to fear. However, in the digital society, there are phishing emails, hacking and scams to watch out for. According to Financial Education Network (2025), in this digital era, consumers face new challenges in managing finances, such as online fraud, artificial intelligence (AI) complexities, and deepfake technology. For example, online scams, internet security risks, identity protection and even the difficulty of data privacy. It is also about being aware of the digital consumer rights. The importance being aware of who to approach in the event of unauthorized transactions being made, or where to turn to in case of digital payment errors.

2.2 Financial Literacy Among Youth

2.2.1 Challenges Faced by Youth

Reaching financial literacy poses a complex challenge for many young people around the world. According to Murugiah et al. (2023), financial literacy is an essential lifelong skill that should be taught to children at any age. A key problem is not giving children enough guidance early on with managing finances. Most youths are not properly taught about budgeting, saving, or investing at school or home. As example, not enough attention is given to teaching young people about how to manage their finances and nurture the right values from young (Financial Industry Collective Outreach (Finco), 2023). This problem causing them to be unprepared when they earn their first money. According to experts, many young adults do not know about important principles like inflation and interest rates. The number of young individuals who are experiencing financial challenges has been rising, which puts them at risk of bankruptcy (Ahmad & Mohamed Zabri, 2023). According to Klapper & Lusardi (2019), just one in three adults are financially literate. The lack of understanding often leads to poor choices with money. Money guidance for youth ensures that they do not develop any bad habits. For example, follow the most recent digital lifestyle trends, which leads to increased borrowing, personal loans, and credit card indebtedness (Murugiah et al., 2023). In fact, owing more money is a common issue for many young people today. Those with minimal knowledge of finances often end up with debt issues and costly balances on their credit cards. Traditional financial planning tends to be less interesting for young people, as saving for retirement can feel like something far into the future. The development of financial literacy in young people is determined by several behavioural characteristics and agents of socialization. These define the developmental expectations in the learning of the youngsters on saving, spheres, and consumer ample money.

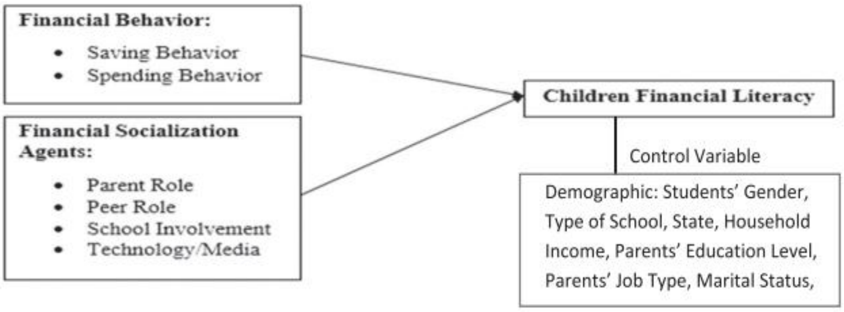


Figure 1: Theoretical Framework (Source: Murugiah et al., 2023)

Figure 1, financial socialization agents like parents, peers, schools and media influence development of financial literacy in children. It was their roles in shaping their financial behaviours, such as expenditure and savings. These influences are also muted by the control variables like demographics and socioeconomic status.

These matters are further complicated by the challenges of managing finances in our modern digital world. Young people especially those from Gen Z and digital natives, experience visual norms and digital technologies as integral to their upbringing (Yaacob M, Farhana A and Rohani H, 2025). Despite their skills with technology, they need to learn about these tools and how to avoid dangers. For example, with digital finance, it's easy for people to buy anything online with one click. Which makes it easier to spend impulsively without immediate financial consequences (Financial Education Network, 2025). Online loans and other credit options can lead young people into debt without them fully knowing the risks. Moreover, people who use online banking and payment services should protect themselves from internet threats like phishing scams and data privacy. According to (Sharifah, 2024), One of the key findings from the 2023 survey was that 94 percent of respondents reported encountering potential fraud or scams. Extra training and understanding are necessary for young people to stay safe on these websites. Many youths in Malaysia are not sure how to protect themselves. Being exposed to fraud is rather common among them. Simply put, youths tend to be uncertain about managing their money, have not built strong financial habits, and feel the pressure of using finances online. Because of these reasons, creative and personalized methods are needed to help the young generation improve their financial literacy.

2.2.2 Financial Education in Schools and Digital Platforms

Because of these challenges, experts in education and policy have aimed to raise financial literacy with both traditional programs and digital technologies. According to Klapper & Lusardi (2019), Promoting financial education in school seems also important, to promote financial literacy among the young. Now, financial education is starting to play a role in the education system in Malaysia. According to Norzitah Abdul Karim et al. (2025), important for teenagers and helps them understand and apply financial knowledge as they transition into adulthood. For example, under the National Strategy for Financial Literacy 2019-2023, a concerted effort to address persistent low levels of financial literacy has been made (Financial Industry Collective Outreach (Finco), 2023). By including discussions on budgeting, savings, and the significance of money into our teaching, their purpose is to guide our students toward financial awareness at an early stage. While still in secondary school, some students get the chance to learn practical skills in financial planning and investing through modules or special programs. These actions line up with national plans. For example, Financial Industry Collective Outreach (Finco) that highlight “inspiring values in youth” as a main point.

Besides classroom lessons, digital tools are now widely used to teach youth about finances. The availability of financial literacy resources through apps and on the internet, these platforms make financial education more accessible and effective (Lakshmi, 2024). For instance, in Malaysia, the Credit Counselling and Debt Management Agency (AKPK) created an online learning platform (learn.akpk.org.my) to educate people for free. Gamified learning is also gaining popularity, as Visa has introduced the “Mind Your Ringgit” web game (Financial Education Network, 2020), where young people can get experience making financial decisions without risking any real funds. Financial teaching tools convert learning moments into games or quizzes that show different scenarios, attracting the engagement of tech-savvy youth. Moreover, organizations including NGOs, businesses, and public officials are collaborating to spread financial literacy further. According to (Norzitah Abdul Karim et al., 2025) , this initiative works to add financial education to school curricula and to improve financial literacy for all age groups including teenagers. For example, Bank Negara Malaysia has partnered with others financial industry on programs like "Kempen DuitSmart" and initiatives in schools. Most recently, an initiative brought together a social enterprise called Little Tauke, an insurer known as FWD Insurance, and a bank, Bank Simpanan Nasional (BSN), to roll out the Camp Millionaire project (Empowering Malaysia's Youth with Financial Literacy, 2025). These programs do teach young students about finances alongside games and other engaging activities. Reaching out to students across the country, the program made use of workshops and educational content online. This initiative shown suggests that using digital tools could significantly enhance financial literacy and well-being (Murugiah et al., 2023).

These new tools and programs are valuable because they offer financial education to youths where they spend most of their time. For example, small videos, apps, and even chatbots are being used to give quick financial advice on Instagram or TikTok. Although Malaysia is still working on these methods, it's clear that mixing learning in schools with online tools can help improve the financial literacy of young people. It is crucial to be sure that the details provided on these websites are true and appropriate for youth.

2.3 Digital Tools and Large Language Models in Financial Literacy

2.3.1 Rise of AI in Financial Literacy

Large Language Models (LLMs) can be defined as a group of artificial intelligence models that process and generate natural text. According to Hean et al. (2025), a significant development in artificial intelligence (AI), have emerged as a transformative technology. LLMs, trained on vast datasets comprising both text and code, have shown exceptional proficiency across a wide array of language processing tasks(Razafinirina et al., 2024). They are characterized by their use of massive scale typically comprising billions of parameters trained on large compilation of written language. As an example, the GPT-3 model developed by OpenAI has 175 billion parameters. According to Yang et al. (2024), With the advent of GPT-3, the potential of causal language models for in-context learning was fully realized. This showed that model size can be increased to achieve good performance in a wide range of language-related tasks without task-specific training. With the rise of deep learning the latest advances allow modern LLMs to capture linguistic patterns and context in a more complex way. Enabled systems to improve their performance over time without being explicitly programmed (Lakshmi, 2024). Later advances in natural language processing (NLP) thanks to the Transformer architecture and increased computing power. For example, short-term context from real-time interactions, enhancing contextual understanding and ensuring personalized learning (Chu et al., 2025). Combined with much larger training data, have allowed LLMs to reach human-level proficiency on a wide variety of language tasks. Figure 2.2 points out the detail of how LLM agents can be categorized into pedagogical and domain-specific agents who are contributing into personalized and adaptive learning. They can plan, remember and recommend knowledge, which perfectly correlates with the intention to provide an individual approach to financial education by means of smart chatbots.

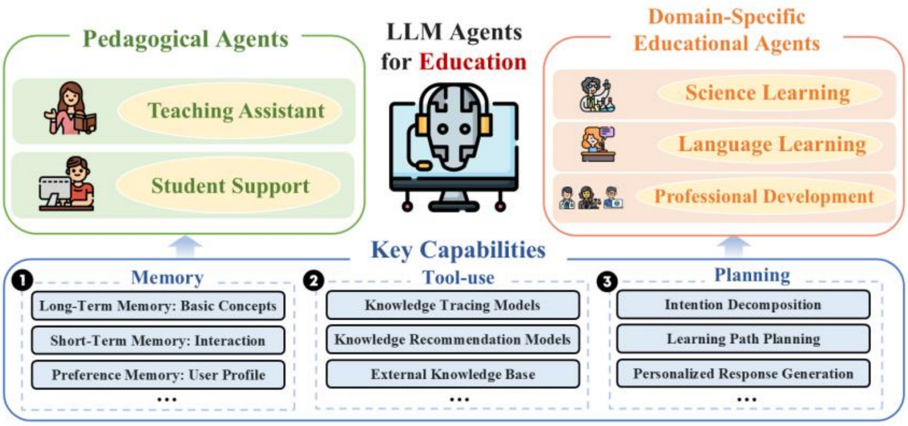


Figure 2: The Overview of LLM agents for Education (Source: Chu et al., 2025)

In their essence, LLMs train to predict text. An LLM is trained in a language modelling task, in which words are fed to the model, and it learns to predict the next word in the sequence. For example, these models leverage self-attention mechanisms, which allow them to capture long-range dependencies within text sequences efficiently (Razafinirina et al., 2024). As the model is trained on a massive amount of text, it continuously changes its parameters to reduce the errors. This self-supervised learning helps the LLM to have an implicit grasp of grammar, facts, and even finer aspects of a language. Because one of the main advantages of AI is its ability to efficiently manage vast amounts of data and offer personalized guidance (Aleksandrova et al., 2024). Most LLMs are based on a structure called the Transformer architecture, its unique self-attention mechanism not only introduces an exceptionally performant parallelism to the training and inference processes but also makes capturing long-range dependencies (Yang et al., 2024). A significant fine-tuning method of chatbot applications is alignment training which refers to training an LLM to execute instructions and generate useful, safe answers by training on human feedback. Development in plans to build large language models (LLMs) has been gaining momentum over recent years, with the largest technology providers providing more powerful models. A review with a comparative overview of the top LLMs applied in conversational AI is given in Table 2.1, including the version and license type.

Table 1: Overall Performance of AI Models (Percentage of Correct Answers), (Source: Hean et al., 2025)

Provider	Model	Version	License Type
OpenAI	ChatGPT 3.5	Gpt-3.5-turbo-1106	Proprietary
OpenAI	ChatGPT 4	Gpt-4-turbo	Proprietary
OpenAI	ChatGPT 4o	Gpt-4o	Proprietary
Google	Gemini	Gemini-1.5-flash	Proprietary
Google	Gemini Advanced	Gemini-1.5-pro	Proprietary
Anthropic	Claude 3 Haiku	Claude-3-haiku-20240307	Proprietary
Anthropic	Claude 3.5 Sonnet	Claude-3.5-sonnet-20240620	Proprietary
Anthropic	Claude 3 Opus	Claude-3-opus-20240229	Proprietary
Meta/Facebook	Llama 3 8B	Llama3-8b-8192	Open Source
Meta/Facebook	Llama 3 70B	Llama3-70b-8192	Open Source

As illustrated above, the proprietary license of models like ChatGPT and Gemini can be replaced by the open-source versions of LLaMA 3 series published by Meta, which makes them appropriate in terms of fine-tuning in academic or experimental settings, e.g., the chatbot project of this financial literacy round. According to Yang et al. (2024), This involves training models to avoid generating harmful, biased, or misleading content while enhancing their ability to produce responses that are truthful, appropriate. All this culminates in an AI model that is not only statistically good at predicting text, also in a manner that frequently seems to know something.

2.3.2 Related Work on Financial Literacy

Some of these studies specifically focus on chatbots in finance applications. Agarwal et al. (2024) suggest a prototype of an AI personal finance assistant combining machine learning, NLP, and visualization. To enable the user to keep track of the expenses, create a budget and obtain custom financial advice. They suggest such an assistant could improve users' financial awareness and long-term planning. Akintunde et al. (2023) developed a customer-support chatbot in a web-based payment service. Their system, written in AIML, could answer common questions (e.g. "How to reset my PIN?") and demonstrated high reliability and user satisfaction. Ramjattan et al. (2021) overview financial chatbot technology and observe that contemporary systems can perform services. For example, such as budgeting and bank inquiries, which may offer users the advantage to make personal financial choices.

These, in addition to others, have dug into areas of finance. As an example, a bot to manage personal investments such as ChatFinance: Tracking Wealth with a Conversational Bot (Johnson et al., 2024), this study proposed a new design and implementation of an intelligent chatbot to address the complex requirements of financial management systems. Although there are numerous domain-specific chatbots, earlier systems have been constrained by either fixed Q&A or limited domain. Dynamically in Malaysia alone even among the published projects there are limited talks of applying youth financial education through chatbots, there is a localization issue of projects.

2.3.3 Cross-Domain Studies Using LLM

The analysis of chatbot projects in other sectors, as well, helps to find promising strategies. LLM has been heavily used in education and mental issues that could be transferred to finance. Reviewing AI tutors in general education, Labadze et al. (2023) retrieve the fact that chatbots can personalize the learning experience and relieving teachers of their burdens considerably. These results indicate that financial chatbots may as well be adjusted to the level of knowledge of each learner. Chen et al. (2024) discovered in the healthcare system that a conversational agent was efficient in enhancing the knowledge of adolescent patients about health insurance concepts, which implied a design, a template of health literacy conversational setting.

The SuDoSys chatbot demonstrates that LLM interactions with a high level of structure can be empathetic and coherent across turns. Tourism planning (Yoshikawa et al., 2023) employed GPT-4 to respond to multi-step queries. For example, booking flights and hotels and booking attractions. It may provide financial advice chatbots with information on structuring multi-concept dialogues. The example of education platforms such Jaybot, (Odede & Frommholz, 2024) designed to aid university students and admissions by providing information and enhancing user experience through conversational interaction.

In sum, cross-domain analyses have shown that LLM chatbots have the potential to reproduce tutoring roles effectively when well-designed. According to Razafinirina et al. (2024) , LLMs can revolutionize online education by understanding a wide range of student questions. As an illustration, stage-based prompts have been used with knowledge bases in systems to give context, such as SuDoSys or travel agents. The suggestions are that methods such as retrieval-augmented prompts, memory storage, and hybrid systems. It can increase the reliability and usability of chatbots. These will drive the creation of a versatile financial chatbot able to provide answers to various questions and explaining using simple terms and gaining user confidence.

Table 2: Table of Related Studies

No	Author(s) and Year	Title of Paper	Objective/Purpose	Techniques	Findings/Results
1	Visesh Agarwal, Ravi Ray, Nisha Varghese (2024)	An AI-Powered Personal Finance Assistant: Enhancing Financial Literacy and Management	To develop an AI-powered finance assistant to improve financial literacy and management	Machine Learning, NLP, React.js, Flask, MongoDB, Firebase; evaluation through focus groups and usability studies.	Potential improvement in financial decision-making, financial literacy, and user satisfaction with secure and personalized insights.
2	Roqib Akintunde Akinyemi, Wumi Ajayi, Ayuba Atuman (2023)	Automation of Customer Support System (Chatbot) to Solve Web Based Financial and Payment Application Service	To automate customer support for financial/payment applications using a chatbot based on AIML and web technologies.	AIML with Program AB, Vaadin-based web UI, Google Script for translation; tested for load, stability, reliability, usability.	70-90% scores for performance metrics; chatbot compatible across systems, which noted dependency on network quality.
3	Sandra Johnson, S Srijayanthi, T V N L HARIKA JHANSI, Pavithra K, Sowmya S, Murarisetty Mahathi Devi (2024)	ChatFinance: Tracking Wealth with a Conversational Bot	To design a chatbot for wealth management leveraging AI and NLP for personalized and efficient financial services.	NLP, ML algorithms for intent recognition; user behavior adaptation; implemented via real-time processing technologies.	Enhanced personalization, improved customer experience, efficient handling of simple to complex financial queries.
4	Katsumasa Yoshikawa, Takato Yamazaki, Masaya Ohagi, Tomoya Mizumoto, Keiya Sato (2023)	Developing Interactive Tourism Planning: A Dialogue Robot System Powered by a Large Language Model	To develop a dialogue robot system using LLM for efficient tourism planning by breaking tasks into stages.	Scenario-based dialogue phase control using OpenAI GPT-4 API; Streaming API, image	Effective phase management for dialogue flow; system achieved 4th place in DRC2023 preliminaries.

				and map viewers for user guidance.	
5	Julius Odede, Ingo Frommholz (2024)	JayBot – Aiding University Students and Admission with an LLM-based Chatbot	To enhance university student support using JayBot, an LLM-based chatbot addressing admission and academic queries.	Combination of OpenAI GPT-3.5 Turbo, embedding models, vector database, and prompt engineering for domain-specific Q&A.	JayBot reduced hallucinations using vector databases and provided focused, efficient information retrieval for university inquiries.
6	Yixiang Chen, Xinyu Zhang, Jinran Wang, Xurong Xie, Nan Yan, Hui Chen, Lan Wang (2024)	Structured Dialogue System for Mental Health: An LLM Chatbot Leveraging the PM+ Guidelines	To develop SuDoSys, a structured dialogue system for psychological counseling based on WHO's PM+ guidelines.	Stage-aware multi-turn dialogue with LLMs using structured prompts based on PM+; includes automatic evaluation mechanism.	SuDoSys improved coherence and direction in multi-stage counseling sessions compared to stage-unaware methods.

2.4 The Role of Digital Financial Tools and Chatbots

2.4.1 AI And Chatbots in Financial Education

With the rise of AI and improved language models, it is now easier to provide financial education on a personal level and to a bigger audience. For example, unlike traditional rule-based learning systems, they deliver interactive, tailored feed-back, allowing students to learn at their own pace (Chu Z, Wang S, Xie J, Zhu T, Yan Y, Ye J, Zhong A, Hu X, Liang J, Yu P and Wen Q, 2025). AI-based chatbots are gaining importance in helping people access financial advice and guidance. With a chatbot, users can interact and discuss their money needs in an individual, engaging way. For example, offer tailored financial advice to a wider audience, including individuals who may not have had access to traditional wealth management services (Mathew, 2025). For this reason, a user could press their chatbot and request help by asking, "What is the best way to budget with RM2,000 per month?" to receive personal guidance. AI powered by Large Language Models (LLMs) such as GPT can answer these natural language questions by providing the necessary steps or examples. The ability to personalize questions and answers in finance education makes it like having personalized financial advisor. This shown that, LLMs empower educators and learners alike, fostering adaptive learning paths and addressing individual needs at unprecedented scales (Sharma S, Mittal P, Kumar M and Bhardwaj V, 2025). To comprehend in more detail how the architecture of a chatbot system would work in the environment of financial education, we should discuss its main parts and the way these parts interact with the people who are its users.

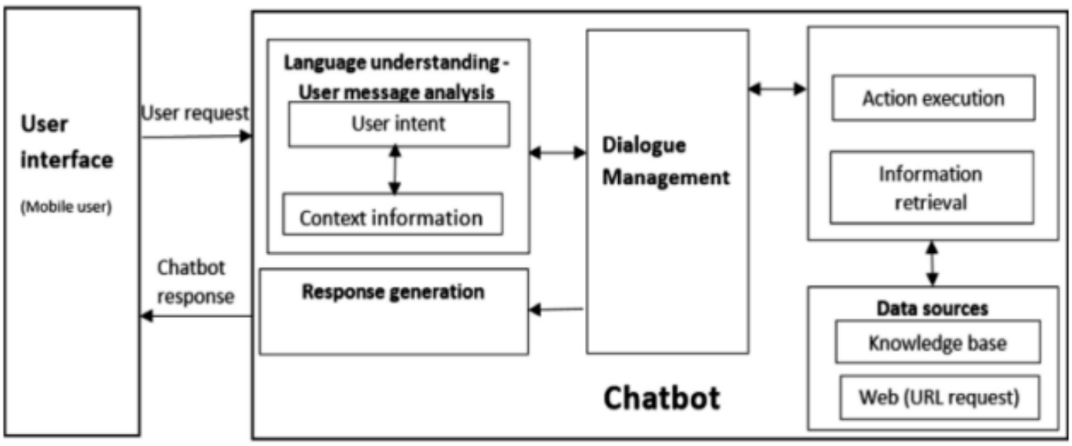


Figure 3: General Chatbot Architecture (Source: Akintunde et al., 2023)

As Figure 3 shows, the chatbot runs on several core modules, comprising language understanding, dialogue management, and response generation, each of which establishes communication with the user inputs and external databases. Dynamic nature and context-specific conversations are supported with this modular structure, which makes it appropriate to convey the financial literacy material to any user leveraging any kind of personal interactions.

One important advantage of AI chatbots is that they can offer advice and education to many users. According to Hean et al. (2025),evaluation from their investigation shown that LLM do assess their comparative efficacy in delivering accurate and actionable financial advice. Their ability to handle many requests at the same time ensures more young people can receive financial support, as not all of them can see a human advisor. For example, not only make financial literacy more accessible but also integrate financial education into the user's daily life, reinforcing learning through practical application (Lakshmi, 2024). Chatbots can address many issues, including talking about simple compound interest, teaching savings tips, budget creation, and providing a loan overview. Simplifying terms and using examples, AI tutors make learning finance less complicated for students who struggle with the standard methods. Advice and solutions not only enhancing service offerings but also reducing human error and providing real-time support (Mathew, 2025), always accessible through 24-hour availability, letting a young user address questions late at night or get instant help with financial decisions whenever needed.

2.4.2 Current Use of Chatbots in Financial Literacy

The development of AI-driven chatbots and financial services has greatly altered the state of personal finances. According to Agarwal et al. (2024), this assistant can empower users to manage all aspects of their finance effectively. Making knowledge and financial management more accessible and improving the user experience. Examples of successful applications are robo-advisors, democratized access to investment management by offering automated, algorithm-driven financial planning services with low fees and minimum investment requirements (Lakshmi, 2024). Other example, the chatbot "Cleo" offers 24/7 financial assistance, helping users manage their budgets and savings (Mathew, 2025) and Erica by Bank of America the chatbot provides personalized financial advice, tracks spending habits, helps users manage bills, and automates numerous financial tasks (Mathew, 2025). Besides providing 24/7 access and affordable services, such tools assist users in developing financial literacy skills. As they are combined with instructions being fused into basic daily financial routines. According to Aleksandrova et al. (2024), AI tools in banking applications reduce barriers to financial knowledge, benefiting various audiences, including younger generations. They are especially popular with younger users who rather have digital, convenient options than a traditional financial advisor.

The potential of Large Language Models (LLMs) such as GPT-4 to deliver financial advice is increasingly promising. As example Hean et al. (2025) said, their potential applications, particularly in personal finance, become increasingly relevant and practical, extending well beyond academic research. With LLMs achieving 99 percent accuracy on standardized financial literacy tests and the ability to offer personalized advice on a wide range of financial topics. AI and ML technologies have been integrated into the model to improve the advisory process to the effect of speed that helps advisors to process vast quantities of data in record time (Challagundla &. B, 2024). They provide data analysis and insights to human financial advisors. There are some exceptions, such as those focused in a specific area like student loans or financial advice focused on women. Nonetheless, these models have limitations, which include difficulties with complex queries, arithmetic operations, and the possibility of the production of nonsense. For example, one of the primary concerns is their tendency to generate hallucinations or spurious outputs, which can be particularly problematic (Yang et al., 2024). According to Yang et al. (2024), updating these models to keep pace with the rapidly changing financial landscape poses another significant hurdle.

Despite their drawbacks, LLMs and AI-based platforms have enormous potential. When used responsibly, it can improve financial literacy and encourage more people to make better financial decisions. For example, with the integration of AI will continues to shape the future of how individuals engage with financial information, manage their finances, and make decisions (Lakshmi, 2024). According to Lakshmi (2024), it is also important to recognize the potential risks and challenges associated with this technology. The critical issues arise due to data privacy concerns, bias control, enhancing model clarity, and preventing the excessive reliance of users on AI-provided recommendations. To achieve real efficacy, these tools should be developed fairly, understandably, and accessibly, particularly in schools and underrepresented groups. According to Lakkaraju et al. (2023) , The application of language models in the finance industry has witnessed a surge in recent times due to their ability. The future work should be aimed at incorporating real-time data, inclusion design practices, and frameworks that promote ethical and sustainable use of AI in financial education.

2.5 Discussion and Solutions

2.5.1 Identified Gaps in Current Financial Literacy Programs

Looking at what is already being done in financial literacy, some of the gaps that still exist has been identified. Firstly, engagement is a significant problem there are just not being engaged enough with the conventional methods of financial education. Such low levels of engagement are supported by the low knowledge retention rates and little behaviour change upon existing programs. As an example, thousands of students may get some financial education at school. However, surveys continue to indicate a significant proportion of students who cannot do simple financial decisions. The traditional method like seminar based is not engaging the audience to be memorable. Second, there exists a personalization gap. The existing programs mostly provide the same information in the same manner to all people. Although, as we all talked about, youths need different things. There is no personalized attention and so individual doubts or weaknesses are not dealt with. A learner who has failed to understand the concept of compound interest in the classroom might transfer that weakness into adulthood. Thus, youth are making unfortunate choices such as undervaluing the cost of loans or the power of saving early in life.

The other gap is on the coverage on digital financial skills. Although, Malaysia has begun to recognise digital financial literacy as a priority area. A lot of the education on the ground is yet to change with traditional matters being taught. It is indicated that there is widespread evidence that youth lack the necessary knowledge. They always thought that they have no time or are unable to properly utilize the existing information in matters relating to financial services. It shows that the information may be available regarding the online scams but is not being put into practice. The old programs have not quite figured out how to achieve comprehension and proper use in the cyber world. What we have instead is which young people may be aware of the concept of budgeting but will still be sucked into an Internet money-making opportunity simply because they were not taught to be critical of online financial propositions.

There's also the issue of scalability and reach. We are aware of some fantastic programs being out there, such as special camps and counselling sessions. The problem is that many of those are rather narrow in their effectiveness. As they are restrained by the number of available resources. They cannot possibly offer to every single youth a human financial advisor or place them all in an intensive workshop. In this way a substantial proportion of the youths particularly in the bound of major cities might not be reached by the programs at all. The difference between the beneficiaries and non-beneficiaries of the complete financial education might increase inequality in the financial performance.

Last but not the least, trust and relatability also have a gap. The problem with many youths is that they do not trust financial information when it is presented as corporate sales or adult lecturing. When the content on financial literacy is provided by an individual who could also be trying to sell a product, the youths might think that there is a conflict of interest. When it is presented in an incongruous manner, they can reject it. The existing programs occasionally lack the ability to engage with the youths at a cultural. They do not talk about the role of social media in spending, and they do not sound authentic. This can leave a gap where the youths are not taking the messages inwards since they do not consider the messenger as credible or relatable.

2.5.2 How AI Can Address These Gaps

AI-based applications, especially intelligent chatbots and adaptive learning platforms are in a good position to mitigate most of the above gaps in financial literacy programs. Firstly, on the aspect of engagement and relatability. A well-developed AI chatbot can coordinate in a conversational tone that youths can find interesting. In contrast to a dry pamphlet, a chatbot may make jokes, may ask questions in turn, and may generate a conversation. Such an interactive process naturally attracts the users towards it. It is more like chatting with an informed friend than reading a textbook. AI tools can be accessed via channels that youths frequently used. Thus, bringing financial guidance to the learners instead of the other way round.

AI is especially brilliant in terms of personalization. In real time, machine learning algorithms could analyse the inputs of a user and personalize the contents. In case the user poses simple questions, the bot will adapt to simple explanations. In case the user demonstrates his advanced level of knowledge, the bot can discuss more complex issues. The AI assistants can have a dynamic conversation with follow-up questions to narrow down on what the user must know. What this translates into is that each learner received a personalized curriculum on-demand. In the long run, the AI will be able to create a profile of the user financial habits and learning progress and become increasingly precise in its advice giving. This closes the personalization gap because it provides individually tailored guidance at scale.

With digital financial literacy, the AI tools are naturally positioned to keep the content constantly updated. An internet-linked chatbot, with access to the right databases, can draw down the very latest news on scam patterns, new financial products, or changes to regulations. It can then inform users of these emerging issues, and the knowledge is up to date. As an example, when a particular new form of online fraud is hitting the news, an AI might be instructed to notify users, to avoid falling victim to it. This agility assists in bridging that gap that in which conventional programs tend to fall behind the rapid digital changes. Then, since AI is also a digital tool, it can also teach by example. By guiding the user through the usage of an app, or showing them digital habits, such as enabling two-factor authentication in the conversation interface, authentication.

On the issue of the sustainability of learning, AI chatbots provide an outside-the-classroom support system. They do not provide a single lesson and leave, but they are always there when a question or choice needs to be made. So, they will serve as a sort of lifelong mentor. This can strengthen and uphold positive practices. An example of this would be a young adult who was taught about budgeting at school but then months later they do not know how to go about setting up a budget to their new job. This is where an AI assistant can intervene and assist the individual through the process step-by-step therefore closing the gap of when the information was learnt and when it can be applied in real-life. As the financial circumstances of the person changes through the years. The AI can adjust advice on the same, offering a continuity of learning that is not available with conventional programs.

AI solutions have scalability and reach as their inherent strengths. A chatbot system, once developed, can engage an essentially unlimited number of users at the same time without any deterioration in the quality of the interaction. The implication is that rural or disadvantaged adolescents who have access to a smartphone and the internet can receive the same level of guidance as a person in an urban area who participates in a workshop. AI has the potential to democratize financial guidance and leave no one behind just because they are not in the right place or human coaches are not available. It can do so at relatively low marginal cost per extra user, which is appealing in policy terms.

Lastly, on the topic of trust, some may distrust a machine, but a significant number of young people may trust an AI that they believe is non-judgmental and neutral. They are aware that the bot is not spreading any gossips about their stupid questions. What is more, the transparently designed AI, aimed at educating might establish credibility. Challenges exist when the AI should always provide accurate information to be trusted. That should be acknowledged through thoughtful design and maybe human monitoring. This can be combatted, though, through training the AI with confirmed information and teaching it to reference sources or give disclaimers when unsure about something. The AI reinforces trust by using valid sources, but it additionally trains the user to locate valid sources.

2.5.2 Hybrid RAG and Fine-tuning Approach Justification

2.6 Summary

Financial literacy has been widely acknowledged as a very vital life skill that helps people to make responsible and responsible decisions in their financial matters. The literature brings out the fact that traditional financial education which includes basic topics like budgeting, saving, investing and dealing with debts is in most cases not interactive and is not in touch with the digital realities of the modern life. The development of online transactions, e-wallets, and fintech apps has placed a new level of digital financial literacy on young people who not only need to know the financial principles but also how to overcome cybersecurity risks, online scams, and online financial practices without harm.

Nevertheless, regardless of the numerous programs, including the National Strategy of Financial Literacy (20192023), or programs of AKPK, FinCo, and other educational facilities, most Malaysian youths still show a low level of financial literacy. Literature has shown that this gap is caused by little exposure to apply financial education, no motivation, and the underutilization of technology-based learning tools. The current programs usually depend on fixed content that cannot offer personalized learning opportunities or real time feedback which is essential in inspiring meaningful comprehension and lifelong behavioral transformation.

To address these issues, recent studies indicate the possibilities of artificial intelligence, specifically Large Language Models (LLMs) to transform the financial education process. Chatbots that are powered by AI could provide a contextual, interacting and adaptive learning experience to each individual user. The systems can increase the engagement and trust by incorporating validated local financial information of reliable sources like Bank Negara Malaysia and OECD. All in all, the literature summarizes that an AI-powered, two-language and gamified financial literacy chatbot has the potential to reduce existing educational gaps, enhance financial literacy throughout people's lives, and help the nation achieve its objective of empowering financially responsible Malaysian youth.

CHAPTER 3

METHODOLOGY

This chapter discusses the research methodology adopted used in the completing of this project. It begins with an overview of the methodology, theoretical study, empirical study, architecture design, development, finally evaluation and documentation.

3.1 Introduction

The given chapter describes the process of creation of a web-based chatbot that is supposed to help improve financial literacy by utilizing Large Language Models (LLMs). Research methodology is designed to addresses the research aims. Which are to identify the relevant content when it comes to financial literacy, come up with an efficient approach of LLMs, and design, implement, and evaluate the chatbot prototype. All the steps start with the Preliminary Study, during which available financial literacy frameworks and resources are analysed to determine major topics and gaps. During the Knowledge Acquisition phase, the relevant financial information is collected upon reliable sources. It serves as the initial knowledge of the chatbot. In the Data Collection stage, the datasets, consisting of financial scenarios and user queries, are gathered to train the LLM to respond correctly and contextually. System Design phase aims at the development of the architecture of the chatbot, the smoothness of the user interface and the backend. System Development comes next, in which the chatbot is constructed, including the LLM and essential functions. Once it has been developed, Evaluation & Refinement is carried out by usability testing and successive versions are improved according to the response given by the users. Lastly, during the Documentation phase, the research process, and system design are documented, as well as the evaluation results, to give an all-around view of the conducted research.

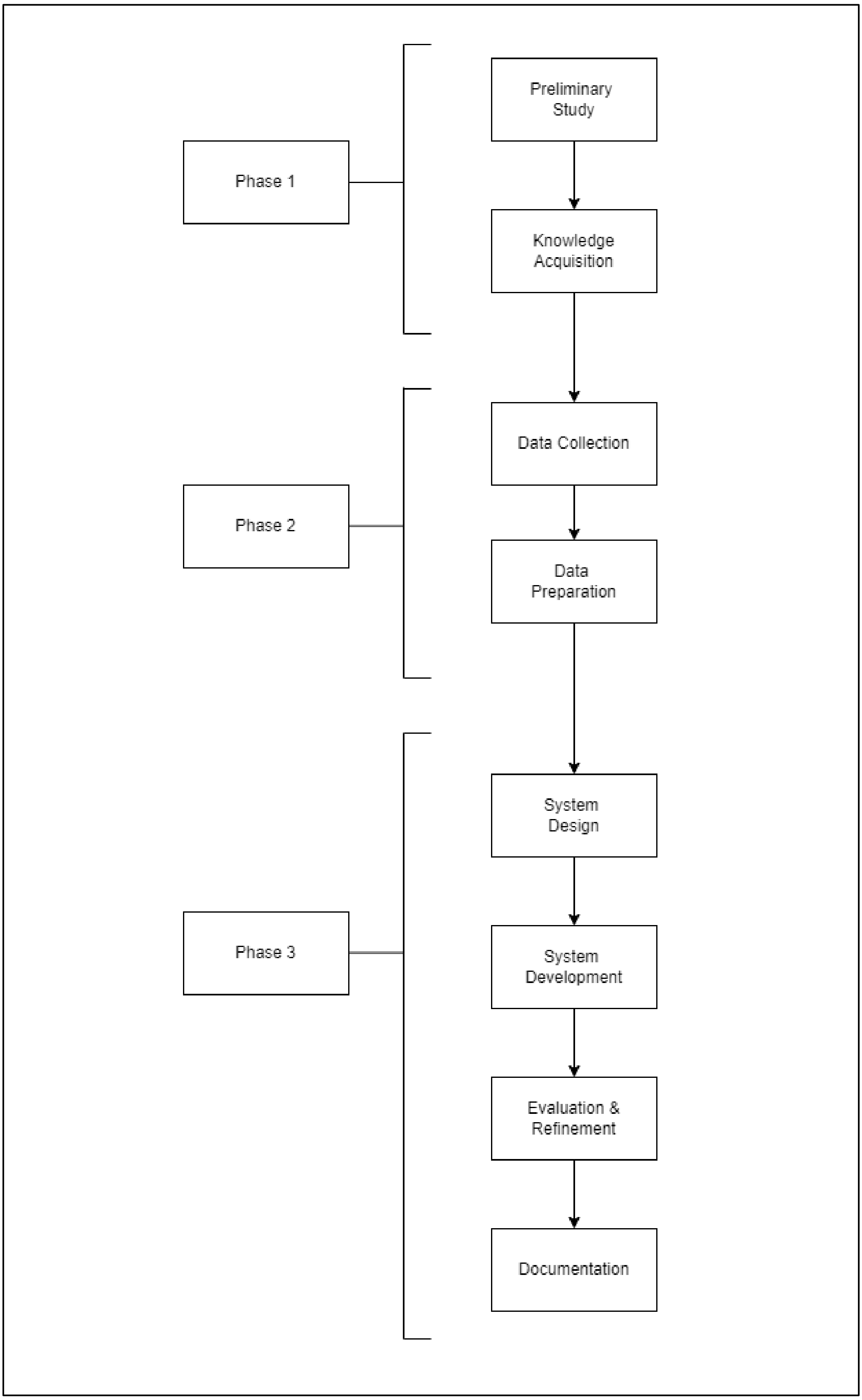


Figure 4, Research Framework

3.2 Research Objectives

Table 3: Research Questions and Research Objectives

Research Questions	Research Objectives
What is the educational material on finance?	To identify educational material on finance.
How can a fine-tune LLM explain financial literacy?	To design a fine tune approach for LLM model on financial literacy.
How can a web-based chatbot be developed to teach financial literacy to Malaysian youth?	To develop a web-based application for educational chatbot of personal finance using LLM model.

Table 3 shows, the methodology will cover each goal of the research by matching certain steps with the expected outcomes and make the development of the web-based chatbot on financial literacy organized.

3.3 Research Methodology Table

Table 4: Research Methodology Table

<table>
<tr><th>Objectives</th><th>Approaches (Phases)</th><th>Activities</th><th>Deliverables</th></tr>
<tr><td rowspan="2">To identify educational material on finance.</td><td>Preliminary Study</td><td>➢ Problem identification of issue in financial literacy among youth in Malaysia.
➢ Background study of financial literacy in Malaysia.</td><td rowspan="2">Project Formulation
• Project Title
• Problem Statement
• Research Question
• Research Objectives
• Scope and Constraints
• Project Significance
• Expected outcome

Literature review analysis report</td></tr>
<tr><td>Knowledge Acquisition</td><td>➢ Do findings based on research papers using tools like Sciscpace, NotebookLM, Google Scholar, Scopus, IEEEExplore, Science Direct, Web of Science and UiTM Online Database.
➢ Reviewing the literature of the domain and technique been used.
➢ Extract key financial concepts for chatbot development.</td></tr>
</table>

Objectives	Approaches (Phases)	Activities	Deliverables
To design a fine tune approach for LLM model on financial literacy	Data Collection	➢ Collect datasets including financial scenarios, FAQs, and real user queries. ➢ Organize and structure data for LLM training.	Cleaned dataset of financial queries, scenarios, and FAQs for LLM training.
	Data Preparation	➢ Preprocess and clean data for LLM use by normalizing and structuring the dataset. ➢ Ensure that data aligns with the chatbot's educational goals.	
	System Design	➢ Design the system architecture, including frontend and backend systems. ➢ Define integration between the LLM and chatbot functionalities.	• System design architecture • User interface of chatbot • System Design Flowchart
To develop a web-based application for educational chatbot of personal finance using LLM model	System Development	➢ Develop the chatbot's frontend (UI) and backend (LLM integration). ➢ Implement features for user interaction and personalized financial responses.	A web-based Financial Literacy Chatbot
	Evaluation & Refinement	➢ Conduct usability testing with real users. ➢ Collect feedback on performance, accuracy, and user experience. ➢ Refine the chatbot based on testing results.	
	Documentation	Document the entire research process, model development, system design, and evaluation results. Provide recommendations for future improvements.	Final Year Report

3.4 Preliminary Study

Table 5: Preliminary Study Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To identify educational material on finance.	Preliminary Study	➢ Problem identification of issue in financial literacy among youth in Malaysia. ➢ Background study of financial literacy in Malaysia.	Project Formulation • Project Title • Problem Statement • Research Question • Research Objectives • Scope and Constraints • Project Significance • Expected outcome Literature review analysis report

Preliminary study phase acts as preparations to determine the scope and direction of the research. At this point, the key objective is to discover current problems with financial literacy in Malaysia. The focus is especially among the youth in terms of both exposures, as well as obscurities and habits, to personal financial issues. To obtain trustworthy facts about financial literacy, online learning resources, and chatbots in education, a thorough review of professional literature and satisfactory academic databases (Scopus, ScienceDirect, Web of Science, and IEEE Xplore) was performed.

From the research showed that there are significant insufficiencies in existing financial literacy programs. Among insufficiencies was a lack of interactivity, contextualization, and learner support. The results were used to develop a relevant problem statement and were used to guide the logical establishment of main project elements. Consequently, this stage resulted in developing project title, a well-defined problem statement, research questions, research objectives, and scope and constraints. The project significance was further determined by the need to provide more interesting and dynamic financial education solution, which was not of a surprise, as the need was observed.

3.5 Knowledge Acquisition

Table 6: Knowledge Acquisition Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To identify educational material on finance.	Knowledge Acquisition	➢ Do findings based on research papers using tools like Sciscpace, NotebookLM, Google Scholar, Scopus, IEEEExplore, Science Direct, Web of Science and UiTM Online Database. ➢ Reviewing the literature of the domain and technique been used. ➢ Extract key financial concepts for chatbot development.	Project Formulation • Project Title • Problem Statement • Research Question • Research Objectives • Scope and Constraints • Project Significance • Expected outcome Literature review analysis report

This step is aimed at acquiring research knowledge relating to the desired topic in the development of a financial literacy chatbot among the Malaysian youth. Two knowledge areas were given priority. Firstly, familiarisation with the financial literacy issues youth experience. For example, many secondary students demonstrate low or medium financial knowledge (Financial Industry Collective Outreach (Finco), 2023) and many young individuals nowadays have a stronger desire to follow the most recent digital lifestyle trends, which leads to increased borrowing, personal loans, and credit card indebtedness (Murugiah et al., 2023). Secondly researching the possible use of Large Language Models (LLMs) in chatbots within the context of education. The literature review was conducted through the methodological process of finding and evaluating peer-reviewed literature research articles via the database and tools. For example, Scispace, NotebookLM, Google Scholar, Scopus, IEEE Xplore, and ScienceDirect, Web of Science, and the UiTM Online Database.

The study utilizes two major and influencing areas of knowledge collection. Studying the material on financial literacy and understanding the application of Large Language Model (LLM) in chatbots as an educational tool. First, it is critical to become acquainted with the problems that bother the young generation when it comes to financial literacy before coming up with an AI-driven personal finance assistant that would meet the demands of the youthful population in Malaysia.

Lastly, this phase yielded a literature review analysis report having consolidated critical academic insights, defined research gaps, and substantiated the argument. This phase purpose is to create a chatbot-based product to support the financial literacy of young people in Malaysia.

3.6 Data Collection

Table 7: Data Collection Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To design a fine tune approach for LLM model on financial literacy	Data Collection	➢ Collect datasets including financial scenarios, FAQs, and real user queries. ➢ Organize and structure data for LLM training. ➢ Source data from Bank Negara Malaysia Open API and Hugging Face datasets	Cleaned dataset of financial queries, scenarios, and FAQs for LLM training.

The data collection plan of the project has major changes as compared to the plan obtained in the original proposal. The previous methodology had proposed to obtain data on the general Bank Negara Malaysia (BNM) Open API and publicly available Hugging Face datasets. This method was revised on a critical analysis to come up with a more specialised, high-quality, and localised knowledge base, which was deemed imperative in the specific focus of the chatbot on Malaysian financial literacy, and therefore guaranteeing users confidence in the responses provided by the system.

The reformulated plan involved the creation of a bespoke corpus, based on one, authoritative, and expertly written source, the official site of the Kumpulan Wang Simpanan Pekerja (KWSP), or Employees Provident Fund (EPF). This source has been selected because it has a large, free library of financial-literacy articles, guides, and infographics. These are also specifically designed to appeal to a Malaysian audience and go in detail to discuss the issues necessary to this initiative, including retirement planning, budgeting, investment, and fraud prevention.

To access this information, a shell python script called scrapeToPdfs.py was written based on the Selenium web-automation framework. The script was coded to automatically loop through a set of 42 article URLs which were located within the kwsp.gov.my domain. On each URL, the script used a browser instance to load the article and used the browser feature to print to PDF to save a clean and well-formatted PDF of the entire article content. The results of this automated process were a list of 42 quality PDF-documents with the topics of the First Salary Tips, the How to Use Akaun 3, and the How to Avoid Online Scam. The resultant filtered collection of PDFs was next placed in the source_pdfs directory to be used as the input basis of the next Data Preparation step.

The data gathered contains believable financial situations and commonly searched questions as well as user queries concerning personal finance. For example, loan, budgeting and saving. Credible sources like official publications, academic reports, and platforms of financial institutions were also used to compile educational content.

3.7 Data Preparation

Table 8: Data Preparation Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To design a fine tune approach for LLM model on financial literacy	Data Preparation	➢ Preprocess and clean data for LLM use by normalizing and structuring the dataset. ➢ Ensure that data aligns with the chatbot's educational goals.	Cleaned dataset of financial queries, scenarios, and FAQs for LLM training.

After obtaining the raw PDF files, a Data Preparation Pipeline was designed in several stages that would convert the raw information into a clean, searchable and structured storage. The following automated pipeline, which is described in the README.md of the project and which is captured in executable scripts, is a vital part of the building of the chatbot knowledge base. The process was segmented into three major stages: Text Extraction and Chunking, Text Cleaning and Refinement and Vector Embedding and Storage.

Text Extraction and Chunking was the first stage that was organized by the processpdfs.py script. All the documents in the sourcepdfs directory were ingested with this script which used the pdfplumber library to find digital text in each page. Another major aspect of this script was that it had an OCR (Optical Character Recognition) fallback ability. With the use of pytesseract and PDF2image, it was programmed to automatically run OCR when the length of text that had been extracted on a page was below 60 characters, hence making sure that the text of image-based pages or scanned pages was also being retrieved. After extraction, the text of all pages of a document was assembled and divided into smaller and overlapping fragments. According to the parameters of the script, the chunk size of 1500 characters and an overlap of 200 characters were used, a design solution that helps to maintain the semantic context at the edges of a chunk. The resultant raw text chunks were then saved as JSONL files in the data/chunks directory.

The second step was Text Cleaning and Refinement, which utilized cleanjsonl.py script to work with raw chunks created in the first step. This script utilized a sequence of normalization rules to improve the quality of data, such as Unicode Normalization (NFKC), the removal of common document header and footers (such as page numbers or a note that this document is confidential) using regular expression patterns, the standardization of different forms of bullet point to a standard dash and the removal of noise (lines with a small ratio of alphabetic characters to total characters or repeated sequence of characters). Any fragment that in its cleaned form had less than 50 characters was thrown out on the grounds that it had too little information content. The cleaned-up data were later stored in another data directory, data/cleanchunks, thus concluding the cleaning process.

The last preparation step, Vector Embedding and Storage was taken care of by the databasebuild.py script. This script loaded the clean chunks of text in the data/cleanchunks directory. The text of every chunk was then embedded in a high-dimensional vector with the HuggingFaceEmbeddings library, which was instantiated with the intfloat/multilingual-e5-small model. This embedding model was chosen because it has a better performance and most importantly ability to understand both English and Malay. These vectors along with the metadata in terms of the original file (source file) were added to a ChromaDB vector database. The database was turned on to be persistent, and the contents of the database was stored in a local directory which was given financedb falling under the collection finance knowledge, so the knowledge base was now ready to be incorporated into the chatbot application.

3.8 System Design

Table 9: System design Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To design a fine tune approach for LLM model on financial literacy	System Design	➢ Design the system architecture, including frontend and backend systems. ➢ Define integration between the LLM and chatbot functionalities.	• System design architecture • User interface of chatbot • System Design Flowchart • Tools that will be used are Python, Streamlit, Hugging Face, LLAMA 3.2 1B, ChromaDB, GitHub

The engineered system is built based on a Retrieval-Augmented Generation (RAG) architecture, which represents a significant development of the original fine-tuning architecture defined by this thesis. The design shown in Figure 5 was taken to ensure that any response that is generated is anchored to the provided documents. The decision is of special importance to a financial literacy chatbot since it mostly prevents the Large Language Model (LLM) to hallucinate, or produce artificial material, which helps to increase the accuracy of facts and build confidence in users. The whole framework is organized using the LangChain framework that provides the essential elements of the integration of the language model, knowledge base and user interface.

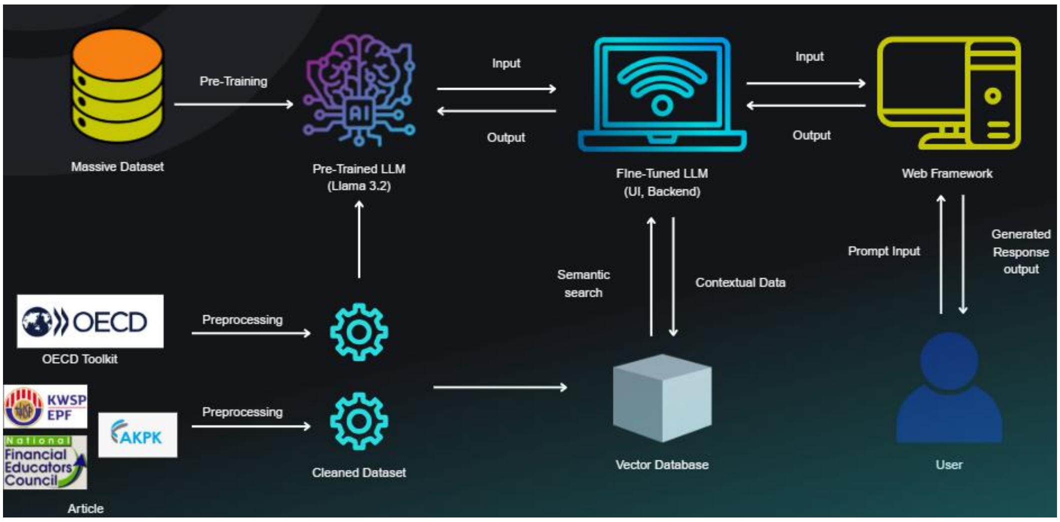


Figure 5: System Architecture of Financial Literacy Chatbot

The architecture, Figure 5, consists of a few important parts that are consecutive in their work:

Data Ingestion Pipeline: This is the offline, foundational, procedural (explained in Section 3.7) which feeds the live system. It reads the source KWSP (EPF) Articles and both parses and cleans and chunks the text with a series of scripts, which finally populates the vector store.

ChromaDB Vector Store: It is the long-term memory of the chatbot. It stores the chunks of text, which are ingested by the ingestion pipeline, but which are in the form of vectors. It is shown that it trains the intfloat/multilingual-e5-small model to generate English- as well as Bahasa Melayu-first language compatible embeddings because of the databasebuild.py script. This database is consulted when a user poses a query to it to identify the most semantically interesting passages of text (the Context). The database is also made persistent with its information stored to the directory, ./financedb of the collection finance knowledge.

Llama 3.2 1B (Ollama Local LLM): It is the chatbot generative brain. It does not store knowledge, as the vector store does, but only reasons and produces language. It is fed the Relevant Chunks (Context) of the vector store and the original User Query, and it is only left to generate a useful answer to it, using that context alone. The model employed in the system is llama3.2:1b, which is locally executed with Ollama and made sure that the model is not expensive but ensures data privacy.

Streamlit Chatbot UI (LangChain Orchestration): The user interacts with this application. Developed using Streamlit, it has the chat interface, text input box, and "Sources Display". Under the hood, it runs the LangChain library to coordinate the whole RAG flow: It feeds user Prompt / Query Input to the ChromaDB, packages the results to the Llama 3 LLM, and lastly, returns the results Generated Response Output to the users or their screen.

3.9 System Development

Table 10: System Development Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To develop a web-based application for educational chatbot of personal finance using LLM model	System Development	➢ Develop the chatbot's frontend (UI) and backend (LLM integration). ➢ Implement features for user interaction and personalized financial responses.	A web-based Financial Literacy Chatbot

The System Development phase was aimed at the implementation of RAG architecture designed in the previous section. This entailed the coding of the application logic that links all the parts together to create a working interactive chatbot. The major work was implemented in Python. The frontend was built with the help of the Streamlit library, and the backend RAG pipeline was organized with the help of the LangChain library. The entire process, user query through to the response created, is packaged at the appfinancestreamlit.py script.

Workflow of the application starts with a startup. Upon starting the Streamlit application, the persistent ChromaDB vector store present in financedb directory is loaded and made an initial object of LangChain retriever. This retriever is set to do a similarity search and retrieve the top 5 best document chunks to any query. This whole retriever object is stored as a streamlit caching decorator, st.cacheresource, which is a highly important optimization measure ensuring that the database is not loaded each time the user taps the UI.

The interactive RAG process is ensued. First, the Retrieval step is done. The query of the user is sent to the retriever in the cache, where the query is embedded and the vector store is searched with it, giving a list of the top 5 relevant document chunks to the KWSP articles. Then, during the Augmentation step, these retrieved chunks are structured.

Lastly, there is the Generation step. The augmented prompt is inputted into the local lllama3.2 1B model through the ChatOllama integration. The temperature of the model will be set to 0.1 that the model responses are factual, deterministic and not too creative. A response is produced by the model, and the response is relayed back to the user on his screen token-by-token to give the impression of a real-time experience. The application then shows a section to complete the feedback loop and gain the trust of the user to enable the user to see the verbatim pieces of text that were fetched by the knowledge base and utilized to answer.

3.10 Evaluation and Refinement

Table 11: Evaluation and Refinement Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To develop a web-based application for educational chatbot of personal finance using LLM model	Evaluation and Refinement	➢ Conduct usability testing with real users. ➢ Collect feedback on performance, accuracy and user experience ➢ Refine the chatbot based on testing results	A web-based Financial Literacy Chatbot

The last and the most important step of the methodology is the Evaluation and Refinement that is aimed at validating the effectiveness, accuracy, and usability of the chatbot in a systematic way. This is per the initial purpose of carrying out usability testing and obtaining user feedback. A three-pronged evaluation strategy was developed to evaluate the performance of implemented RAG architecture appropriately. This methodology is a technical quantitative test, user study (qualitative) and domain expert review. This is a multi-layered plan that will evaluate the technical accuracy of the system, its practical applicability and professional level of the financial guidance.

This evaluation consists of the first part, which is a Quantitative Test. This is a test that is intended to test the core accuracy of the RAG pipeline. The questions will be developed manually as a set of 20-30 test questions with each question having an answer that is known and verifiable within the KWSP, AKPK and NFEC source articles. There will be two important metrics to be determined regarding each question. First, there is retrieval hit rate that measures whether the system was able to access the correct document fragments in the ChromaDB vector store and factual accuracy that measures whether the answer generated by the LLM is an accurate and faithful summary of the context retrieved. This is a technical test that is needed to confirm the functionality of the data pipeline and the timely engineering.

The second and the third sections of the assessment include Qualitative Reviews. A Qualitative User Study shall be done to respond directly to the usability objectives of the project. The Streamlit application along with a set of realistic tasks will be provided to a test group of 5-10 target users (including fellow university students). These users will also fill out a feedback survey after completing the tasks, to gauge the important metrics on a 1-5 scale, e.g. "Ease of Use," trust in the information, and the helpfulness of the help feature, i.e. the sources feature. In line with this, Qualitative Domain Expert Review will be conducted.

A Subject-Matter Expert (SME) will be offered the same set of questions as the quantitative test and will be asked to respond to these questions (e.g., a financial literacy lecturer). This professional will look at the responses that the chatbot produces since their correctness to understand whether the financial advice accurate or not. Next, completeness to determine whether critical details are missing. Finally, suitability to ensure that they use the right tone to attract Malaysian young people. This professional verification gives the chatbot content an imperative touch of confirmation that the content is not only factual, but also professionally viable.

The final stage will be the Refinement plan, which will be informed by a combination of the results of all three assessments. The Quantitative Test and the Expert Review will inform technical improvement, including changing the chunk size of the data pipeline or making the prompt of the LLM more robust. User Study Feedback will mainly be used to guide the UI/UX change to the Streamlit application.

3.11 Documentation

Table 12: Documentation Phase Table

Objectives	Approaches (Phases)	Activities	Deliverables
To develop a web-based application for educational chatbot of personal finance using LLM model	Documentation	Document the entire research process, model development, system design, and evaluation results. Provide recommendations for future improvements.	A web-based Financial Literacy Chatbot

The Documentation phase addresses the process of the documentation of the whole development process of the web-based financial literacy chatbot. This also including initial research of the problem and problem identification, system design, data preparation, and model integration, evaluation. This covers records of the use of Large Language Models (LLMs), front and backend development activities, chatbot interaction patterns, and user testing results. The idea is to present a transparent, interpretable account of the chatbot construction and iteration towards achieving its educational intentions.

This stage goes beyond the development details by assembling the findings of the various operations of the evaluation of systems. This also including the feedback of users to define limitations and improvement areas. These insights will be used to present practical improvement recommendations in the documentation. For example, refining response accuracy, broaden the scope of the knowledge base by including more financial information, or multilingual capabilities.

The essential outcome of this phase is a final report, which will capture the lifecycle of the project besides delivering a fully documented, web-based Financial Literacy Chatbot and a future course of action of improvement.

3.12 Summary

In this chapter, the systematic methodology of the construction of the financial literacy chatbot is described. The process has developed out of the plan in the first proposal of fine-tuning the proposal to a more realistic and precise Retrieval-Augmented Generation (RAG) architecture. The process started with its updated Data Collection phase that turned off course toward generic APIs. A bespoke Python code was created to harvest a high-quality and regionalised article on financial literacy of many sources including KWSP, AKPK and NFEC publications and article.

After the collection, a Data Preparation process comprising of multi stages was carried out. This automatic process entailed text extraction of the PDFs with an OCR back up to extract text of any images. The text was then extracted and broken down into 1500-character portions and cleaned with a text cleaning program to eliminate noise and normalize text. Lastly, such clean chunks were transformed and put in a persistent ChromaDB vector database.

This RAG architecture was described in the System Design section and is controlled by a central orchestration architecture. The fundamental elements of the architecture. The knowledge base is encoded in a vector database, a large local language model (locally run) will provide free and unlimited text generation, and the user interface will be interactive through a web application. The System Development phase then described the live workflow: a query by a user of the system retrieves the most relevant information in the vector database and injects this information as a strict prompt known as context. This augmented prompt is then fed to the language model to produce a fact-based response, and it is then fed back to the user screen with its sources.

The chapter ends with the description of a strong, 3-pronged Evaluation plan. This plan provides a quantitative test of technical accuracy of a golden set, a qualitative test of usability of a user study, and a review by a domain expert to ensure that the financial advice is of a professional standard. Finally, the Documentation stage makes sure that the project is reproducible, so that it can be maintained not only in the context of this thesis but also with the help of such all-encompassing technical documentation as the README.md file, executable pipeline scripts, and much-commented code.

CHAPTER 4

RESULTS AND FINDING

This chapter outlines the findings and conclusions of the development processes and evaluation procedures involved in the financial literacy chatbot using Large Language Models.

4.1 Data Preprocessing Results

This section outlines the results of the preprocessing stage of data processing that will be done before fine-tuning of the model and deployment of the system. The output of preprocessing is explained as a way of demonstrating how raw financial literacy data were converted into structured and reliable inputs that can be trained on a large language model and those that can be retrieved by operations. The following subsections elaborate on the results of data cleansing and formatting, organization of data set through text chunking, creation of vectors-database and implementation of Retrieval- Augmented Generation (RAG). These results are the basis upon which the fine-tuning experiments and performance tests shall be argued in subsequent sections.

4.1.1 Data Cleaning and Formatting

The process of data cleansing and data formatting produced a highly filtered and standardised body of financial-literacy content that can be further organised and indexed. To begin with, the data consisted of raw unprocessed text snippets found on the web and as PDF files, and it included the noise elements of duplication of material, metadata tags, formatting artefacts, and the irregularities caused by the OCR. These contaminations may have a negative impact on semantic interpretation when they are not treated. As a result, the preprocessing effort focused on converting raw inputs of heterogeneous nature into clean, uniform and human readable textual forms.

At a conceptual level, the cleansing workflow involved document extraction, text normalisation, noise removal and aggregation to textual outputs in form of structures. This did not only make sure that only instructionally relevant instructional content related to financial literacy was retained, but also extraneous, non-informative information was removed. The result of this step is a set of purified pieces of text that are structurally homogenous and ready to be organised into the next data structures and chunked.

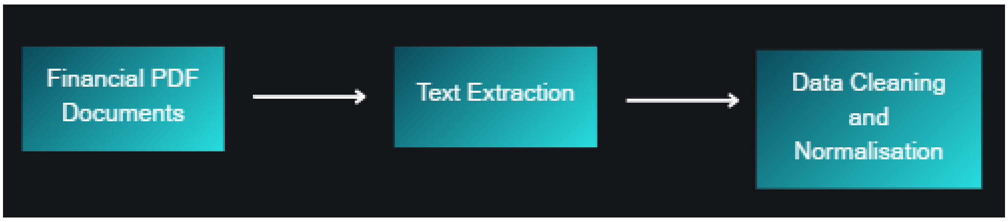


Figure 6: Data cleaning and normalization pipeline

The data cleaning and normalisation pipeline used in this investigation is simplified as shown in Figure 6. The PDF financial documents are first transformed into text-based information through a text extraction process as illustrated. The text is then extracted, cleaned and normalised, to produce standardised text that contains no formatting noises and extraction artefacts. This number reinforces the fact that the data cleaning phase is completed with a normalised textual output, which serves as the foundation of datasets structuring and text chunking in the further parts.

To support the effectiveness of the cleansing and formatting process further, a qualitative comparison of raw and processed texts is provided. The given comparison highlights the fact that irrelevant metadata, discontinuous sentences, and formatting discrepancies of the original outputs have been trimmed, and the main instructions material survived. Substituting it with the resulting cleaned text exhibits improved readability, coherence, semantic clarity, which are central to the language-model fine-tuning as well as retrieval-based response generation.

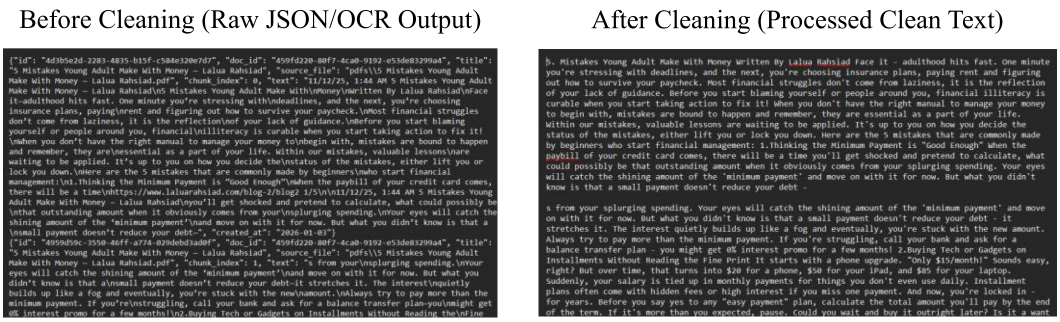
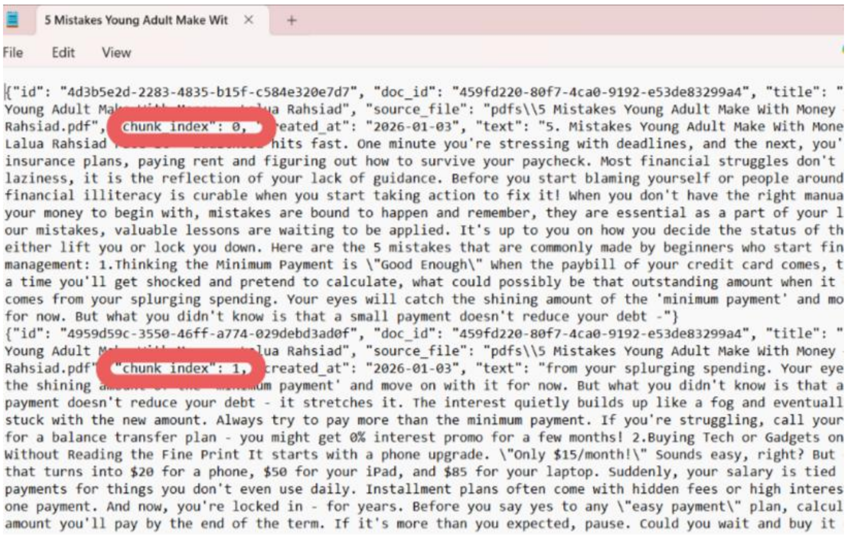


Figure 7: Sample raw JSON/OCR output before cleaning vs. processed clean text after formatting

The sample raw JSON/OCR output before cleansing and the processed clean text after formatting is shown in Figure 7. The juxtaposition presents a significant decrease in noise and structural clutter, with the sanitized one showing one continuous and meaningful financial explanation that can be used in teaching. This result justifies the assertion that the data cleansing and formatting phase has effectively transformed the unstructured raw information into quality textual information, which can be used as a dependable basis of the further structure of data and semantic processing.

4.1.2 Dataset Structure and Text Chunking

After the data cleaning and formatting step, the cleaned textual information was restructured into a structured data through a process of text-chunking. This structuring is aimed at dividing continuous cleaned text into smaller semantically coherent units associated with the single financial concepts or even closely related ones. This approach avoids the way long texts are treated as a block of text that may simplify the precision of context in terms of the subsequent processing. Alternatively, chunking maintains topical attention in every bit and keeps relevant instructional information.




```
{"id": "4d3b5e2d-2283-4835-b15f-c584e320e7d7", "doc_id": "459fd220-80f7-4ca0-9192-e53de83299a4", "title": "
Young Adult Ma [illegible] ua Rahsiad", "source_file": "pdfs\\5 Mistakes Young Adult Make With Money
Rahsiad.pdf", "chunk_index": 0, "created_at": "2026-01-03", "text": "5. Mistakes Young Adult Make With Mone
Lalua Rahsiad [illegible] hits fast. One minute you're stressing with deadlines, and the next, you'
insurance plans, paying rent and figuring out how to survive your paycheck. Most financial struggles don't
laziness, it is the reflection of your lack of guidance. Before you start blaming yourself or people around
financial illiteracy is curable when you start taking action to fix it! When you don't have the right manua
your money to begin with, mistakes are bound to happen and remember, they are essential as a part of your l
our mistakes, valuable lessons are waiting to be applied. It's up to you on how you decide the status of th
either lift you or lock you down. Here are the 5 mistakes that are commonly made by beginners who start fin
management: 1.Thinking the Minimum Payment is \"Good Enough\" When the paybill of your credit card comes, t
a time you'll get shocked and pretend to calculate, what could possibly be that outstanding amount when it
comes from your splurging spending. Your eyes will catch the shining amount of the 'minimum payment' and mo
for now. But what you didn't know is that a small payment doesn't reduce your debt -"}
{"id": "4959d59c-3550-46ff-a774-029debd3ad0f", "doc_id": "459fd220-80f7-4ca0-9192-e53de83299a4", "title": "
Young Adult M [illegible] lua Rahsiad", "source_file": "pdfs\\5 Mistakes Young Adult Make With Money
Rahsiad.pdf" "chunk_index": 1, created_at": "2026-01-03", "text": "from your splurging spending. Your eye
the shining [illegible] um payment' and move on with it for now. But what you didn't know is that a
payment doesn't reduce your debt - it stretches it. The interest quietly builds up like a fog and eventuall
stuck with the new amount. Always try to pay more than the minimum payment. If you're struggling, call your
for a balance transfer plan - you might get 0% interest promo for a few months! 2.Buying Tech or Gadgets on
Without Reading the Fine Print It starts with a phone upgrade. \"Only $15/month!\" Sounds easy, right? But
that turns into $20 for a phone, $50 for your iPad, and $85 for your laptop. Suddenly, your salary is tied
payments for things you don't even use daily. Installment plans often come with hidden fees or high interes
one payment. And now, you're locked in - for years. Before you say yes to any \"easy payment\" plan, calcul
amount you'll pay by the end of the term. If it's more than you expected, pause. Could you wait and buy it
```


Figure 8: Conceptual representation of text-chunking process

Figure 8 shows the conceptual representation of the text- chunking process which illustrates how a continuous block of text containing financial literacy information is broken down into several smaller text chunks. Every chunk itself represents a specific idea or subtopic and allows organising the information into specific and specific units. This theoretical model emphasizes the use of chunking to make the structure clearer without losing the semantics in the entire dataset.

To ensure continuity of meaning within neighbouring chunks, the overlap strategy was used in the chunk segmentation. Intertextuality between consecutive units is used to maintain the contextual flow especially when the financial statements are broadly defined in several segments. This design will reduce case loss of context and will make vital transitional information available in the next level of retrieval.

```
Text: "ABCDEFGHIJKLMNOPQRSTUVWXYZ..." (5000 chars total)

Chunk 0: chars 0-1500 (A-1500)
Chunk 1: chars 1300-2800 (1300-2800)  ← 200 char overlap with Chunk 0
Chunk 2: chars 2600-4100 (2600-4100)  ← 200 char overlap with Chunk 1
Chunk 3: chars 3900-5000 (3900-end)   ← 200 char overlap with Chunk 2
```

Figure 9: Structured chunked dataset with indices and overlap

Figure 9 represents an example of the structured chunked dataset, with the indices of chunks, the ranges of characters covered by the chunk, and the overlap of the two consecutive chunks. The figure illustrates how the individual chunks are indexed in a different manner and how overlap is also kept in control to ensure semantic continuity. This chunk-based, structured output of the dataset-structuring phase is the immediate input to the system of creating vectors embedding and database creation detailed in the next subsection.

4.1.3 Vector Database Construction

After the organization of the dataset and the division of text, every indexed element was translated to a semantic representation and then stored in a vector database to allow similarity-based retrieval. A numerical encoding of each chunk was done to encode semantic content instead of depending on word for word matching. Such a transformation enables the system to locate conceptually meaningful financial text when the query made by the user varies by words with the original text. The development of the vector database marks an important shift in the structured textual information to the form of representation that can be searched and retrieved semantically.

The success in consuming the structured text chunks into the vector database is the major result of this step. In the process, all the indexed chunks were stored and encoded, hence establishing that the dataset was completely absorbed into the semantic layer of storage. Such a consummation outcome shows that the preprocessing pipeline was able to fill the financial literacy contents into the vector database to make them accessible to the retrieval operations.

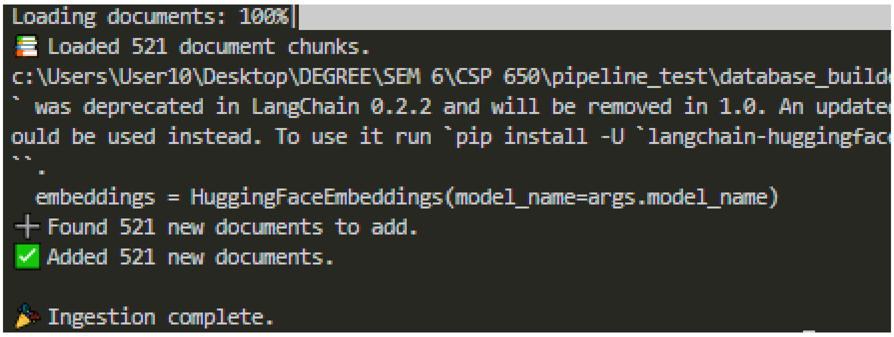


Figure 10: Vector Database ingestion output which demonstrates a successful embedding and document loading.

The output of the ingestion process of the database construction process is shown in figure 10. The result of the output showed that all the structured chunks of texts were loaded, embedded, and inserted into the database, and the ingestion process finished without interruption. This finding corroborates the assertion that the financial literacy data was completely interpreted into the vector database, and this forms the basis upon which semantic retrieval is carried out.

After a successful ingestion, a system-level test was performed to verify that the created vector database was there and intact. This check is done to confirm the data embedded is stored in a persistent mode and can be accessed by downstream operations of retrieving. The outcome of the check-up indicates that the database directory has been created, and the related embedding files have been created and stored in the right way.

```
======================================================================
VECTOR DATABASE VERIFICATION
======================================================================
✅ Database directory exists: finance_db(1)

📁 Database contents:
  - chroma.sqlite3 (18,714,624 bytes)
  - f6cef115-5613-4999-8aca-9c389ad4fab1\data_level0.bin (167,600 bytes)
  - f6cef115-5613-4999-8aca-9c389ad4fab1\header.bin (100 bytes)
  - f6cef115-5613-4999-8aca-9c389ad4fab1\length.bin (400 bytes)
  - f6cef115-5613-4999-8aca-9c389ad4fab1\link_lists.bin (0 bytes)
```

Figure 11: Verification Result of Vector Database Creation and Stored Embedding Files

The verification output that justifies the successful creation of the vector database is depicted in figure 11. The fact that the database directory and embedding files stored proves that the semantic representations are continuously stored and can be accessed. In turn, this check ensures the successful implementation of the stage of the construction of the vector-database and that the system is ready to accommodate the Retrieval-Augmented Generation mechanism outlined in the next subsection. The vector database is a knowledge layer of the chatbot. In the process of inference,

4.1.4 RAG System Implementation

Once the vector database is constructed and verified, a Retrieval- Augmented Generation (RAG) was rolled out to enable synthesis of context-sensitive responses in the financial literacy chatbot. The RAG paradigm combines semantic retrieval and large-language-model generation and thus ensures outputs get pegged on empirically verified material on financial-literacy rather than on pretrained knowledge.

The inquiries of the user are converted into a semantic embedding in the context of the operational workflow and compared against the vector database through similarity-based retrieval. The most relevant fragments of context extracted out of the database are then processed into the language model to influence the construction of responses thus producing outputs that are contextually relevant and factually consistent.

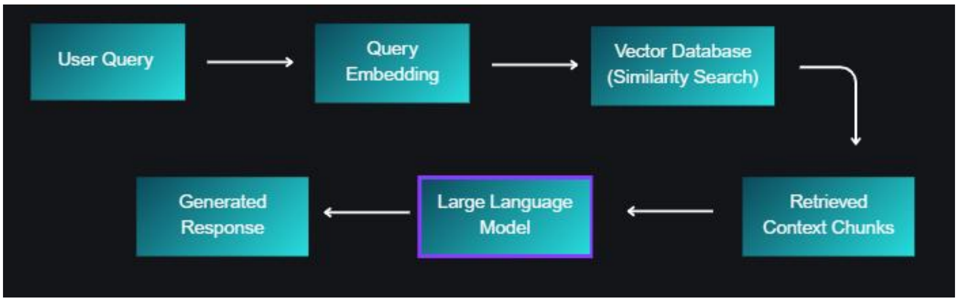


Figure 12: Retrieval-Augmented Generation (RAG) Workflow in Financial Literacy Chatbot

The figure 12 is the RAG workflow that will be utilized during the current research project, which demonstrates the series of steps starting with the user query input, query embedding and similarity-based search within the vector database, and the final stage of response synthesis using the language model with the retrieved context. To ensure the successful implementation of the workflow depicted by Figure 12, a demonstration was performed on the terminal to ensure that the database was ready and that semantic retrieval was effective.

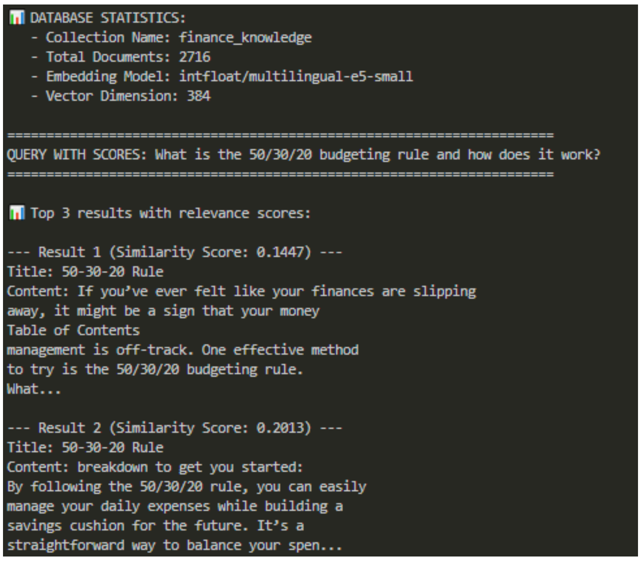


Figure 13: Terminal Output of RAG Workflow Engine showing Semantic Retrieval and Relevance Scoring

Figure 13 shows the terminal output when a typical query is made relating to financial literacy. The results verify the fact that the 2,716 textual blocks embedded in the vectors were initially loaded in the database and that the process of similarity-based retrieval was properly implemented. The system makes use of the L2 distance measure, with the limits being the lower the value, the higher the semantic similarity. The minimum distance score of 0.1447 indicates a large level of relevance between the query by the user and the financial content that has been retrieved and hence support effective semantic retrieval and contextual grounding.

4.2 Fine-Tuning Results and Model Comparison

The section outlines the experimental setting used to fine-tune the language models used in the current research. The fine-tuning regimen aims to adjust a general-purpose language model towards the financial literacy domain to make the model responses more relevant, understandable, and consistent to chatbot-based interactions. The set-up includes the chosen models, fine-tuning approach, dataset characteristics, computational environment and design.

4.2.1 Model and Fine-Tuning Approach

Several smaller pre-trained ones were tested during the experimentation (Qwen2.5-1B, Phi-4, and LLaMA 3.2-1B), and the final system was using LLaMA 3.2-1B. Such decision is consistent with the task of building a domain-adapted chatbot with practical computational requirements. The larger models might have better general capabilities, but the smaller ones will have better possibilities of trial and error, fine-tuning of the models and deployment. In this regard, the main thing is not an open-domain creativity but rather providing the correct and consistent answers to the questions in accordance with financial literacy lessons.

Fine-tuning was done with Low-rank Adaptation (LoRA), which is a parameter-efficient approach that adapts a small number of trainable parameters injected into the model rather than modifying all the model weights. This method is suitable to domain-specific adaptation since it lowers training costs and memory requirements but still allows training of domain-relevant patterns of responses and vocabulary to the model. In addition, LoRA reduces the risk of overfitting the training data in the case of limited training data, which is also a factor in educational corpora based on lesson material collections.

4.2.2 Fine-Tuning Dataset Description

Fine-tuning was done using a set of instructions consisting of about 1,000 chat-formatted examples based on the processed financial literacy corpus. Chatbot objectives should be formatted using instruction-style formats because they teach the model to reply to user prompts in a conversational way, which will facilitate the consistent structure of responses and improve the model-question interaction. The content of the dataset represents the essential themes of personal finance budget, saving, how to spend money, and avoid the most common financial pitfalls. This topical focus is related to the objective of the chatbot: to give learners easily available grounded guidance, which is in line with the lesson contents.

A train-validation split was used with a 90:10 split so that it would keep the training signal intact and have a significant validation split to monitor the generalisation. The validation set delivers specific significance especially in identifying when the model is not training on memorising training prompts but rather on actual domain alignment. Generalisation is crucial when using domain-education tasks since actual user queries will not be identical to training components; rather they will be presented in different wording, incomplete questions and at different degrees of financial literacy.

4.2.3 Environment for Training and Computational Resources

Cloud-based notebook environments were used to do experiments. Initial experimentation and rapid iteration were performed on Google Colab, which has an NVIDIA T4 GPU, and initial runs were made. Kaggle Notebook with dual T4 GPUs (T4x2) was used in longer runs when it needs more stability and less interruption in the session. This two-platform strategy enables testing efficiency. Experiments can be performed in Colab where faster ones are taken, and longer training opportunities and comparisons can be made in Kaggle.

The training pipeline used was the LoRA training pipeline that used Unsloth to offer optimised implementations in saving memory and training faster. This is especially true when small modifications are applied to large language models to fit into the restrictions of the GPU, where the memory pressure of model weights, activations and states of optimiser can cause a constraint on the size of a batch and stability of training. Using an optimised fine-tuning framework enhances pragmatism without changing the conceptual approach (LoRA), thus guaranteeing that methodological stability is maintained, yet giving potential to perform practical experimentation.

4.2.4 Design of Experiments and Hyperparameters Preferences

The experiments with hyperparameters consisted of several experiments aimed at identifying the settings which result in stable convergence and strong validation performance. The hyperparameters under study included batch size, number of epochs, learning rate, optimizer, as well as the ratio between training and validation split. This can be done by systematic monitoring of training and validation loss in these experiments to make an evidence-based conclusion about the final configuration, rather than arbitrary choice of parameters. The final one is especially relevant to domain-adaptation settings, where poorly optimised hyperparameters can cause either underfitting, which is defined as the poor learning of the domain, or overfitting, the consequences of which are reduced generalisation to user prompts.

4.3 Fine-Tuning Results and Analysis

4.3.1 Overview of Training Experiment

In this subsection, a brief overview of the experiments performed on fine-tuning in the context of this investigation is provided followed by an in-depth account of parameter-specific analyses. It seeks to specify the extent, structure and systematic nature of the experimental design, demonstrating how the different models and hyper-parameter settings were tested to identify an ideal setting of financial literacy chatbot.

A comprehensive set of training experiments have been conducted at various stages including baseline model comparison, epoch number sensitivity testing, batch size sensitivity testing, learning-rate sensitivity analysis and optimizer sensitivity analysis and train-validation split testing. Three candidate models, namely, Phi-4, Qwen 2.5-1B, and LLaMA 3.2-1B, were experimented upon systematically sweeping principal training hyper-parameters, monitoring training loss, validation loss, and training duration. This systematic design approach made sure that model-selection and configuration decisions were based on empirical findings and not implementation.

The summary provides the background of the experiment in the subsequent subsections. Instead of just stating parameters in isolation, the training-experiment summary brings out the fact that the performance of models is due to combination of various design choices. In this connection, later sections examine each of the key parameters separately with the support of numbers and dedicated analyses that would help justify the choice of the final model and fine-tuning configuration used within the chatbot system.

4.3.2 Baseline Model Comparison

The subsection will compare the performance of three large size language models, Phi 4, LLaMA 3.21B, and Qwen 2.51.5B, to provide a baseline performance and proceed with a more detailed hyperparameter search. This experiment aims at determining the most appropriate model architecture that can be adopted to be used in the financial literacy chatbot according to predictive performance, computational efficiency, and the ability to generalize, under controlled training conditions.

To make a fair and unbiased comparison, the same experimental setup was used to fine-tune all the models. Every model was trained to epochs of 3 with a 90:10 training-validation split, learning rate of 0.0002, a batch size of 2, and the paged AdamW 8-bit optimizer. Keeping these parameters fixed, any observed performance differences can be blamed mostly on intrinsic model characteristics and not differences in training set up.

Table 13: Overall Performance of AI Models (Baseline Comparison)

Model	Train Loss	Eval Loss	Time (minutes)
Phi-4	0.0609	0.0664	58.68
Llama 3.2 1B	0.2469	0.0647	5.77
Qwen 2.5 1.5B	0.3087	0.32022	15.48

Its findings show that LLaMA 3.2-1B was the lowest in evaluation loss (0.0647) out of the three models, which is better at generalizing on unseen financial text. Although the evaluation loss of Phi-4 was also very similar (0.0664), it took much longer to compute, and it took it over 58 minutes to train, which was a lot longer than LLaMA 3.2-1B (5.77 minutes). Such a big difference emphasizes that the cost of resources to run Phi-4 is considerably higher than to run Phi-4, which makes it less feasible to experiment with it and deploy it in resource-limited settings.

By comparison, Qwen 2.5-1.5B performed significantly worse, with an evaluation loss of over 0.32, indicating that it is not very effective in terms of adapting to the financial literacy domain, given the same conditions of fine-tuning. Although moderate training time, the high validation loss means poorer domain compatibility as well as poorer generalization ability than the other models.

Altogether, the base comparison shows that the LLaMA 3.2-1B offers the most reasonable trade-offs between predictive performance and computational efficiency. Its lower evaluation loss, coupled with a significantly lower training time, implies that it will be the best candidate in the further experiments of hyperparameters optimization. As a result, the LLaMA 3.2 -1B became the first model to be targeted in further fine-tuning tests, i.e., the batch size, the number of epochs, the sensitivity of the learning rate, the choice of an optimizer, and the evaluation of train-validation split that are depicted in the subsequent subsections.

4.3.3 Epoch Size Comparison

The subsection looks at how the number of training epochs affects both the model performance and training performance. The evaluation is done on the relationship between the number of training iterations and convergence behavior, generalization ability and computational cost among the investigated models. The comparison is carried out on experiments of fine-tuning with the same configurations with the only difference in the number of epochs.

Table 14 shows the training and evaluation outputs of Phi -4, LLaMA 3.2 -1B and Qwen 2.5 -1.5B trained with 1, 3 and 5 epochs, a fixed 90:10 train validation split, learning rate of 0.0002, batch size of 2, and paged AdamW 8-bit optimizer.

Table 14: Epoch Size Comparison Results

Model	Epoch	Train Loss	Eval Loss	Time (minutes)
Phi-4	1	0.0822	0.09188	58.68
Phi-4	5	0.0582	0.0642	84.94
Llama 3.2 1B	1	0.6404	0.1522	2.53
Llama 3.2 1B	3	0.2469	0.1522	5.77
Llama 3.2 1B	5	0.1709	0.0647	9.66
Qwen 2.5 1.5B	1	0.3795	0.0604	5.06
Qwen 2.5 1.5B	5	0.2974	0.3959	26.07

In the case of LLaMA 3.2 -1B, the model performance increases steadily with the number of training epochs with the training loss of 0.6404, 0.1709, 0.0409 at the end of 1, 5, and 10 epochs with a corresponding evaluation loss of 0.1522, 0.0604 respectively, which implies that the model has been successfully trained and the generalization is enhanced. Nevertheless, the rather insignificant difference between 3 (0.0647) and 5 (0.0604) epochs, and the apparent negative impacts of the increasing training time, indicates a decreasing marginal trend after the third epoch. Phi 4 has a comparable downward movement in evaluation loss (0.09188 at 1 epoch to 0.06421 at 5 epochs) but has a significantly higher computation cost, reaching training time of 84.94 minutes, making it not so practical. On the contrary, the performance gains of Qwen 2.5-1.5B are relatively lower because evaluation loss does not decrease to less than 0.31 with increased epochs, which suggests that it is not well-adapted to financial literacy tasks.

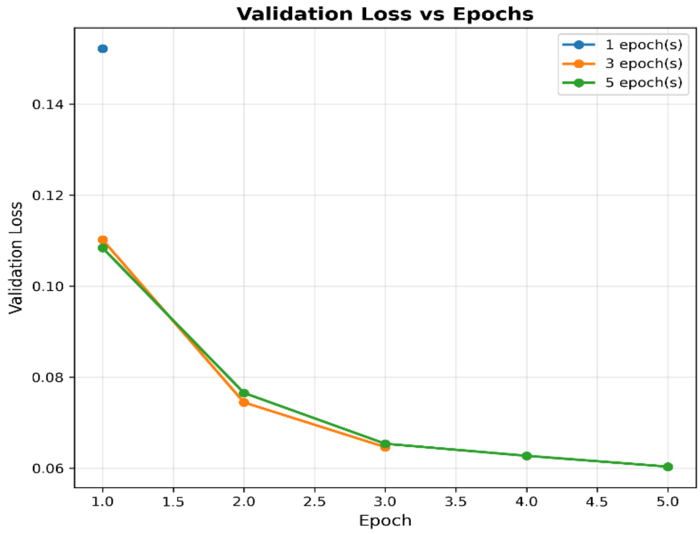


Figure 14: Validation Loss Across Training Epochs

Figure x

Figure 14 shows that there is a significant reduction in the validation loss after the early training phase and the sharpest reduction is at the first few epochs. Validation loss in the 3 epoch and 5 epoch designs decreases significantly after the second epoch and then levels off, hence suggesting that learning is best concentrated in the initial stages and further epochs are not offering much or no benefit. This plateauing trend supports the idea that there is a decreasing return after the third epoch.

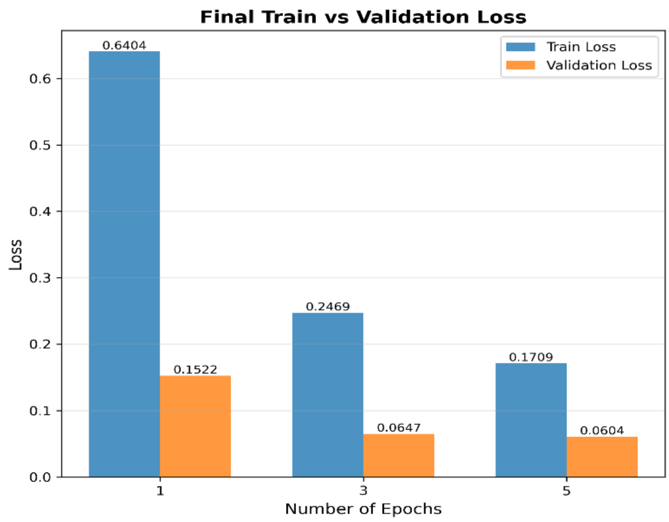


Figure 15: Final Training and Validation Loss by Epoch Configuration

This trend is also supported by the summary of final loss values in Figure 15. The loss to validation of 1-epoch configuration is the highest (0.1522), which is a sign of the lack of training. Further increase of the training epochs to 3 and 5 decreases the performance of the validation loss to 0.0647 and 0.0604 respectively, showing that the difference in losses between the 3 and the 5 epochs is rather insignificant.

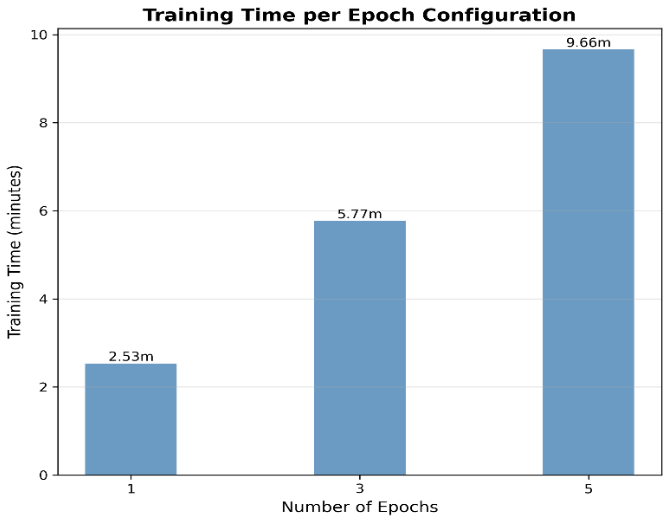


Figure 16: Training Time by Number of Epochs

On efficiency, Figure 16 indicates that the training time increases linearly with the number of extra epochs between 2.53 minutes (1 epoch) and 5.77 minutes (3 epochs) and 9.66 minutes (5 epochs). Within the same analysis as the marginal improvement in the loss of validation across epochs 3 and 5, these results indicate that the 3-epoch setup is associated with more balanced performance and computational cost trade-offs.

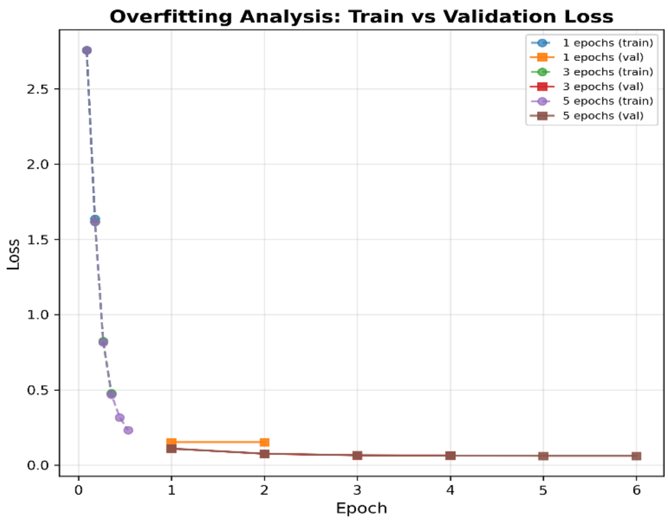


Figure 17: Overfitting Evaluation Across Epochs

The overfitting evaluation in Figure 17 shows that the validation loss does not increase as the training progresses, and not the divergent trend, where the training loss keeps decreasing at the expense of the validation loss, is observed. This implies that going up to 5 epochs of training does not result in a particularly strong form of overfitting in this scenario. However, the fact that the validation loss plateaus after the third epoch indicates that there is little more generalisation leverage.

Figure 18: Convergence Rate by Epoch Configuration

Lastly, Figure 18 quantifies the fact that most of the improvement comes at increasing epochs between the first and third epoch, where the error decreases by an approximate of 41.3%. Adding an additional epoch only results in a small additional error, so on average a 44.3% decrease in the total error, a finding that supports the idea that performance gain is saturating past the third epoch.

In general, the comparison of the epochs shows that the larger the number of epochs, the better the models perform, especially LLAMA 3.2-1B and Phi-4. Nevertheless, LLaMA 3.2-1B 5-epoch trained offers the best trade-off between performance and computation efficiency with low evaluation loss and significantly reduced training time. This was then chosen to be the epoch configuration of the choice in future fine-tuning experiments.

4.3.4 Train Validation Split Ratio Comparison

This subsection examines how varying train validation split ratios impact on the model performance, generalization ability and training efficiency. The aim of this comparison is to find a data split setup that allows to perform effective learning and use sufficient validation data to perform reliable performance evaluation. The same settings were used in all the experiments, and the only variable was the split ratio.

Table 15 provides a summary of the fine-tuning findings of Phi-4, LLaMA 3.2-1B, and Qwen 3.3B with 70:30, 80:20 and 90:10 train-validation splits. The experiments were conducted using 3 epochs, learning rate of 0.0002, batch size of 2 and paged AdamW 8 -bit optimiser.

Table 15: Train-Validation Split Ratio Comparison Results

Model	Split Ratio	Train Loss	Eval Loss	Time (minutes)
Phi-4	0.2	0.061	0.06749	0.061
Phi-4	0.3	0.0605	0.06877	-
Llama 3.2 1B	0.2	0.2970	0.0644	4.96
Llama 3.2 1B	0.1	0.2703	0.0628	5.9
Llama 3.2 1B	0.3	0.3266	0.0672	4.58
Qwen 2.5 1.5B	0.2	0.3061	0.3256	14.18
Qwen 2.5 1.5B	0.3	0.3056	0.328	13.29

The 90:10 split has lowest evaluation loss (0.0628) followed by 80:20 (0.0644) and 70:30(0.0672) in LLaMA 3.2 -1B. Although the differences are not quite large, the trend is that the higher the share of data set to training, the greater the capacity of the model to acquire domain-specific financial patterns. The same pattern can be found with Phi 4, with evaluation loss changing less significantly with the size of the validation portion. Conversely, Qwen 3.3B has a consistently high evaluation loss with all split ratios, indicating lower generalisation despite data partitioning.

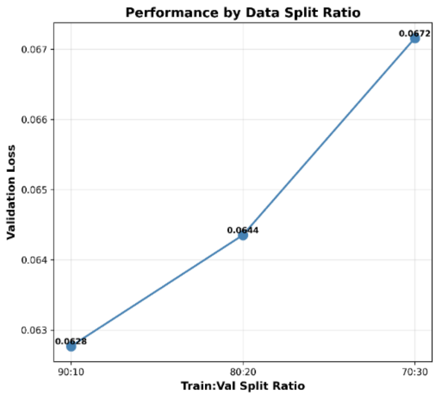


Figure 19: Performance by Data Split Ratio

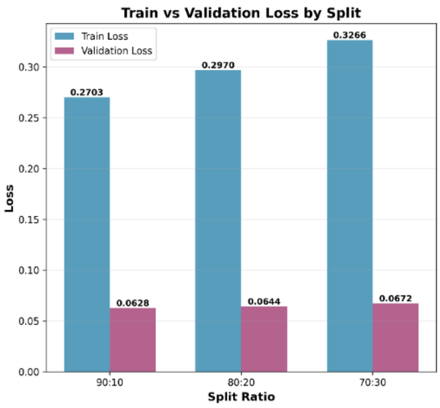


Figure 20: Train vs Validation Loss by Split Ratio

Figure 19 shows a gradual rise in the validation loss with an increase in validation data beginning at 10 per cent to 30 per cent, which supports the fact that decreasing the size of training data constrains the performance of the model. This tendency is also justified by Figure 20, which demonstrates that training loss is growing significantly with increased validation splits, whereas validation loss is changing in a relatively minor way, which also implies that the learning capacity, and not overfitting, is being observed.

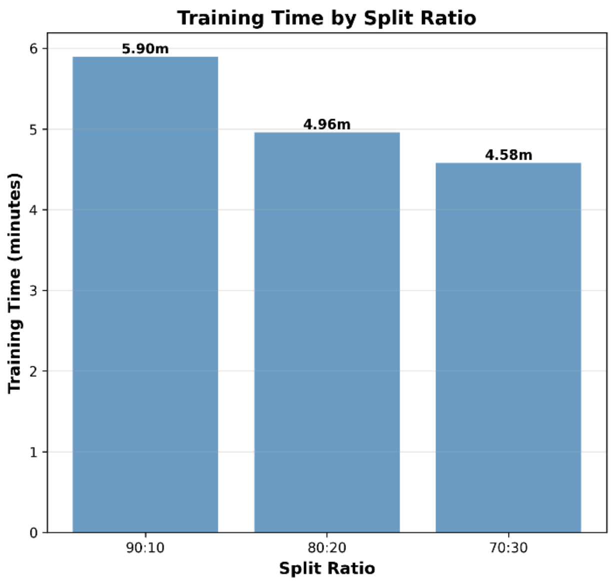


Figure 21: Training Time by Split Ratio

Computationally, in Figure 21, it was observed that the training time reduces a little when the size of the training dataset is reduced. The minimum training time of 4.58 min, 4.96 min, and 5.90 min is obtained with the 70:30 split, 80:20 split, and 90:10 split respectively. Nevertheless, the efficiency in terms of time spent on training is at the expense of the evaluation loss implying a trade of between efficiency and model accuracy.

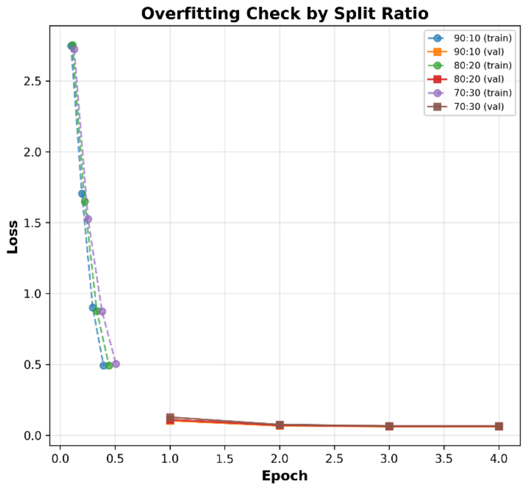


Figure 22: Overfitting Evaluation Across Split Ratio

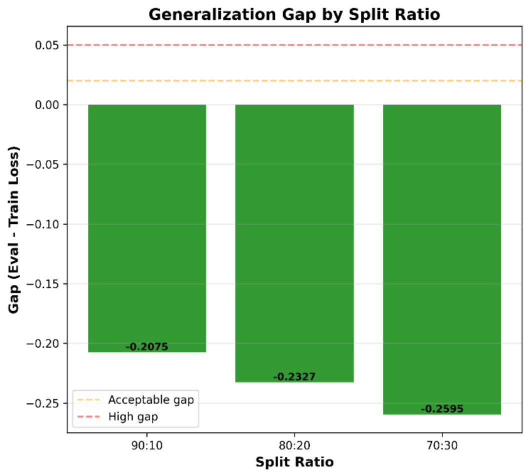


Figure 23: Generalization Gap by Split Ratio

The overfitting curve in Figure 22 indicates that the validation loss does not increase with the epochs in all split ratios, and there is no trend of divergence in which the validation loss rises and the training loss falls. This implies that no split ratios of the tested ones lead to considerable overfitting. Moreover, in Figure 23 the gap in generalisation stays in a reasonable range regardless of the configuration, however, as the proportion of validation increases, the gap also increases.

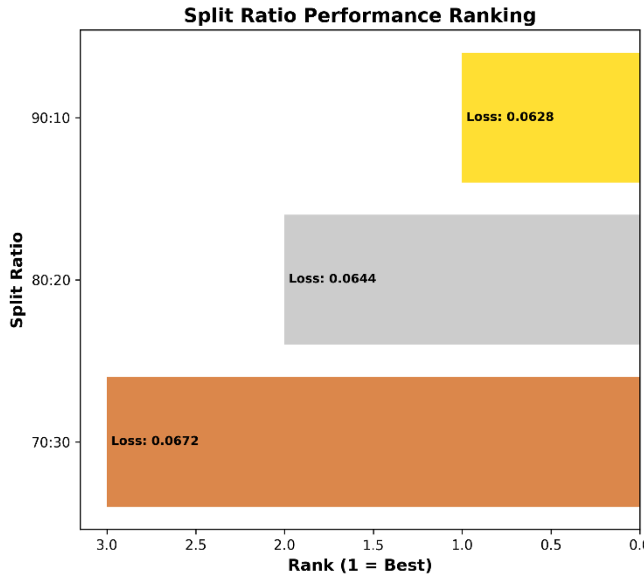


Figure 24: Performance Ranking by Split Ratio

Lastly, Figure 24 performance ranking indicates that the 90:10 split has the best performance of validation, followed by 80:20 and 70:30. The 90:10 train-validation split would offer the best balance in the financial literacy chatbot task when it comes to predictive performance as well as training efficiency.

Following these results, the 90:10 split ratio was chosen as an object of future experiments, the result of which would provide better performance to the model and still have enough validation data to be used as a reliable evaluation instrument.

4.3.5 Learning Rate Comparison

This sub-section carries out a methodical analysis of the tendency of model execution to distinct learning rate configurations throughout the fine-tuning stage. The latter hyperparameter is the learning rate, which has the direct impact on the convergence speed, the stability of training, and the ability to generalize. Fine-tuning experiments were conducted with learning rates 0.0001, 0.0002 and 0.0005 and all other parameters held constant, i.e., three training epochs, 90:10 train-validation split, 2-batch size and paged AdamW 8-bit optimizer.

Table 16: Learning Rate Comparison Results

Model	Learning Rate	Train Loss	Eval Loss	Time (minutes)
Phi-4	1e-4	0.0617	0.06919	52.91
Phi-4	5e-4	0.0608	0.06493	53.32
Llama 3.2 1B	1e-4	0.3869	0.0737	5.54
Llama 3.2 1B	2e-4	0.2703	0.0628	5.39
Llama 3.2 1B	5e-4	0.1957	0.06	5.4
Qwen 2.5 1.5B	1e-4	0.3123	0.33151	15.81
Qwen 2.5 1.5B	5e-4	0.3072	0.3146	16.21

Table 16 lists the training and evaluation results of the training scenarios of the evaluated models regarding the various learning rate settings. In the case of LLaMA 3.2 -1B, it shows an explicit improvement in performance with an increase in the learning rate. At 0.0001, the model obtains a relatively high training loss of 0.3869 and evaluation loss of 0.0737, which is a sign of slower learning and non-convergence. When the learning rate is increased to 0.0002, the training loss (0.2703) is lower and evaluation loss (0.0628) is also lower and indicates that the optimization is more effective. The best performance is 0.0005, in which the training loss further decreases to 0.1957 and the evaluation loss reaches the lowest level of 0.0600, indicating better domain adaptation and generalization.

The same can be found with Phi-4 where an increase in the learning rate will result in a reduction in the evaluation loss of 0.06919 at 0.0001 to 0.06493 at 0.0005. However, even with such performance improvements, Phi-4, on average, has much higher training time, which exceeds 52 minutes, which makes it less applicable to the iterative fine-tuning. In contrast, Qwen 2.5 -1.5B is less sensitive to the learning rate changes, and evaluation loss is more than 0.31 in all the settings, which means that it is less adaptive to the financial literacy domain.

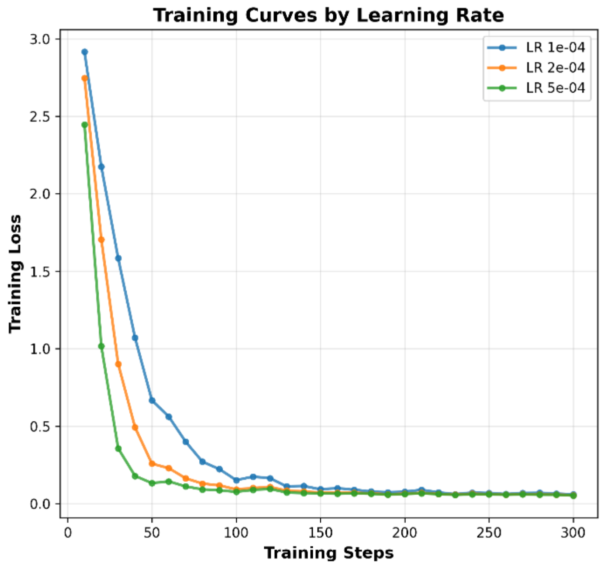


Figure 25: Training Dynamics Across Learning Rates

The training dynamics of Figure 25 reveal that when the learning rates are high, the training loss decreases faster especially at the first training steps. The models trained at a rate of 0.0005 converge faster than those trained at a rate of 0.0001 which have slower loss reduction. The behaviour herein verifies the fact that low learning rates can hinder the process of optimal optimization of this task.

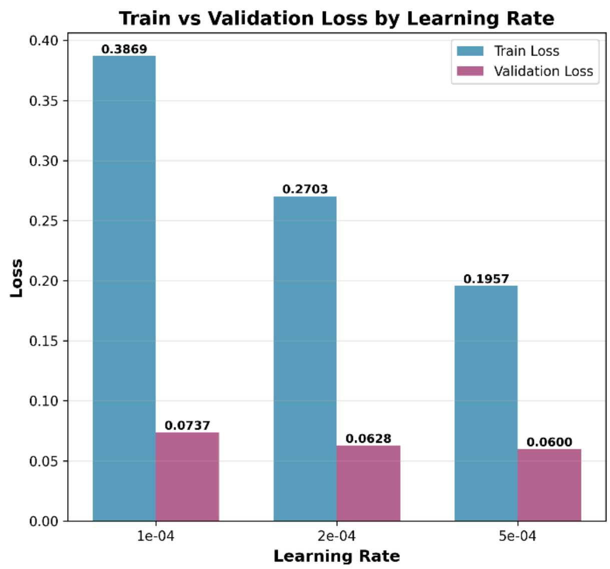


Figure 26: Final Training and Validation Loss by Learning Rate

Figure 27: training Time by Learning Rate

These findings are further supported by the comparison of training and validation loss in Figure 26. Although all the learning-rate settings have a reasonable distance between training and validation loss, the 0.0005 one achieves the lowest validation loss without causing instabilities, so it is not overfitting but instead effective learning. Also, Figure 27 shows that training time does not vary significantly between learning rates of LLaMA 3.2 -1B, indicating that the performance improvements at faster learning rates can be obtained at the same level of cost.

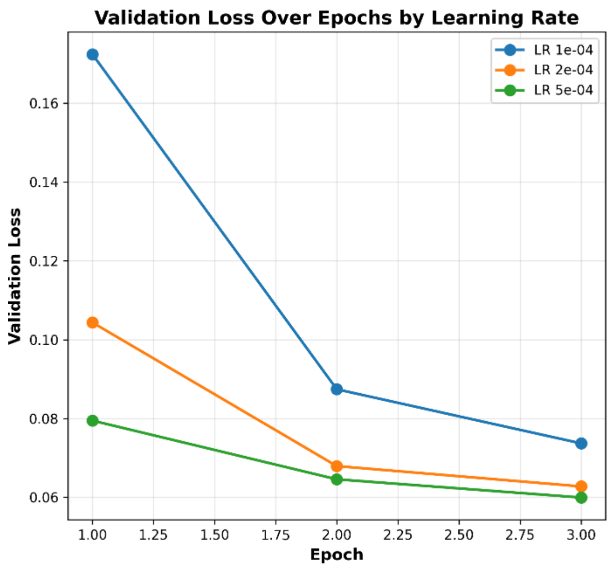


Figure 28: Validation Loss Trends by Learning Rate

Trends in validation-loss in Figure 28 show that higher learning rates in all epochs result in lower validation loss, thus supporting the conclusion that higher learning rates are more efficient in convergence.

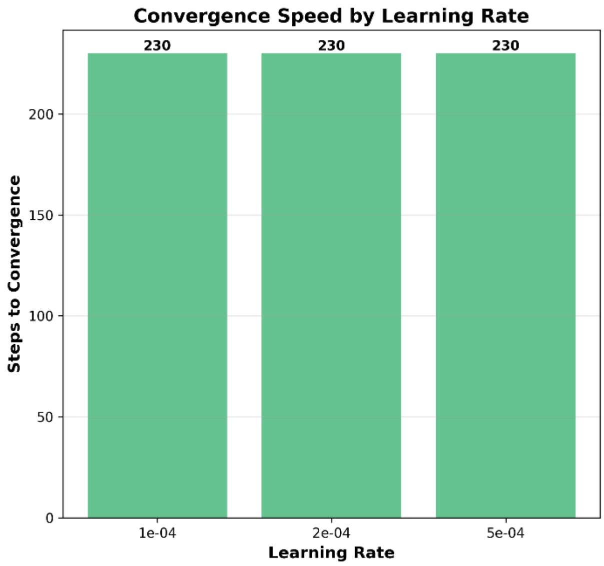


Figure 29: Convergence Rate by Learning Rate

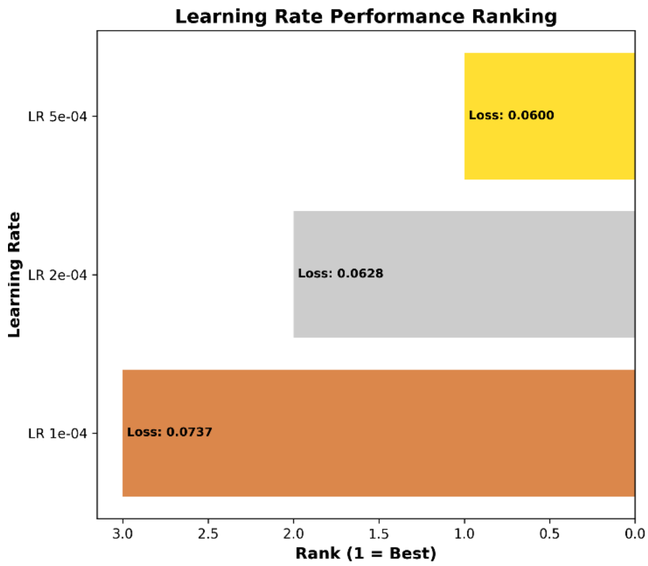


Figure 30: Performance Ranking by Learning Rate

In addition, Figure 29 shows that the convergence rate in all learning rates was similar, which means that the difference in performance was mainly due to the effectiveness of optimization and not the duration of training. Figure 30 clearly shows that the best performance was 0.0005 which was followed by the others 0.0002 and 0.0001.

To conclude, the sensitivity analysis has shown that the convergence and generalization of LLaMA 3.2-1B are significantly enhanced by increasing the learning rate without causing any instability or adding extra computation. The best trade-off between performance and efficiency of all the assessed configurations is a learning rate of 0.0005, and thus it was chosen in this study to be fine-tuned in further experiments.

4.3.6 Batch Size Comparison

In this sub-section, the effect of batch size differences on the model convergence dynamics, generalization, and training efficiency will be explored. The batch size has a drastic impact on gradient estimation stability and computational throughput particularly during fine-tuning of large language models on low resource hardware. To measure these effects, we ran experiments with batch sizes equal to 1, 2 and 4, and fixed a training configuration of 3 training epochs, 90:10 train-validation split, learning rate equal to 0.0002 and paged AdamW optimizer with 8-bit precision.

Table 17: Batch Size Comparison Results

Model	Batch Size	Train Loss	Eval Loss	Time (minutes)
Phi-4	1	0.0614	0.06692	66.77
Phi-4	4	0.0637	0.07579	38.92
Llama 3.2 1B	1	0.1920	0.0619	10.41
Llama 3.2 1B	2	0.2703	0.0628	5.42
Llama 3.2 1B	4	0.4376	0.0738	3.99
Qwen 2.5 1.5B	1	0.3177	0.3335	30.64
Qwen 2.5 1.5B	4	0.3146	0.3452	11.66

Table 17 provides a summary of the training and evaluation results of Phi- 4, LLaMA 3.2-1B, and Qwen 2.5-1.5B under the various settings of the batch-size. In the case of LLaMA 3.2 -1B, a batch size of 1 produced the smallest evaluation loss of

0.0619, with a training loss of 0.1920. A tenfold increase in the batch size to 2 slightly increased the evaluation loss to 0.0628, and a batch size of 4 further increased the loss in the performance, to 0.0738. This pattern implies that smaller batches will be used in making gradient updates more efficient and they will also improve generalization in this task.

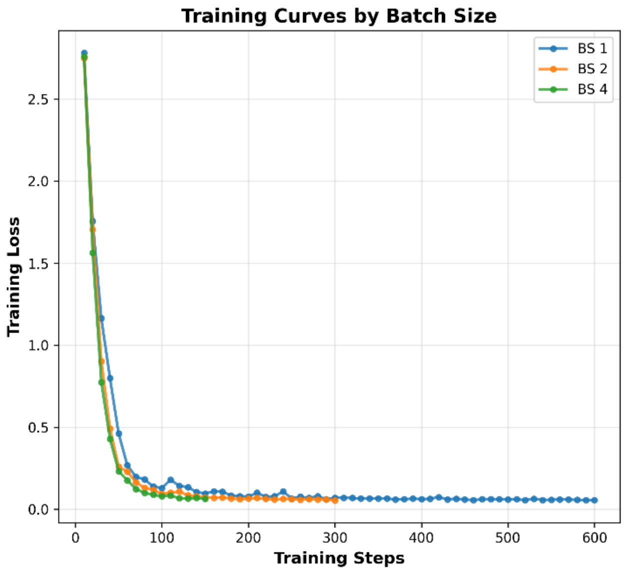


Figure 31: Training Dynamics Across Batch Sizes

Figure 31 shows the dynamics of optimization over train stages, which show that a batch size of 1 has smoother and more constant convergence, and a batch size of 4 leads to more variance at the beginning of training, which is reflective of less reliable gradient updates. With a two-batch size, there is intermediate behaviour, just as a trade-off is achieved between convergence stability and efficiency.

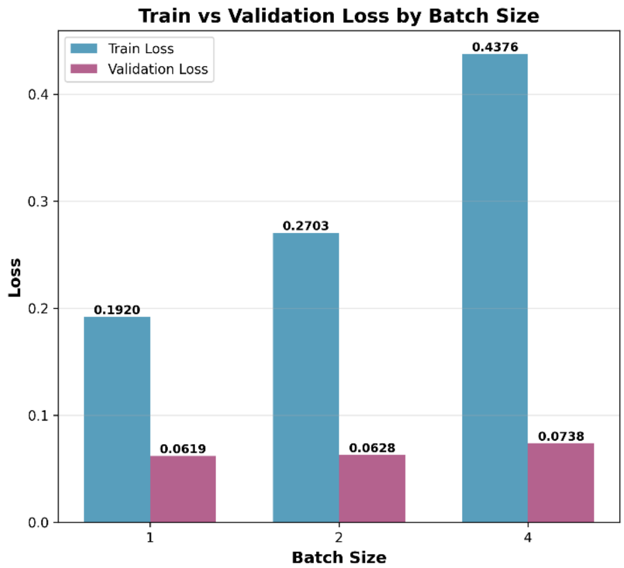


Figure 32: Final Training and Validation Loss by Batch Size

Figure 32 is a comparison of the last training and validation loss of the batch sizes. Although a decrease in training loss is generally expected with a larger batch, more importantly the corresponding increase in validation loss is an indication of poorer generalization, suggesting that less gradient noise during training on a larger batch can result in poor adaptation to their domain-specific financial data.

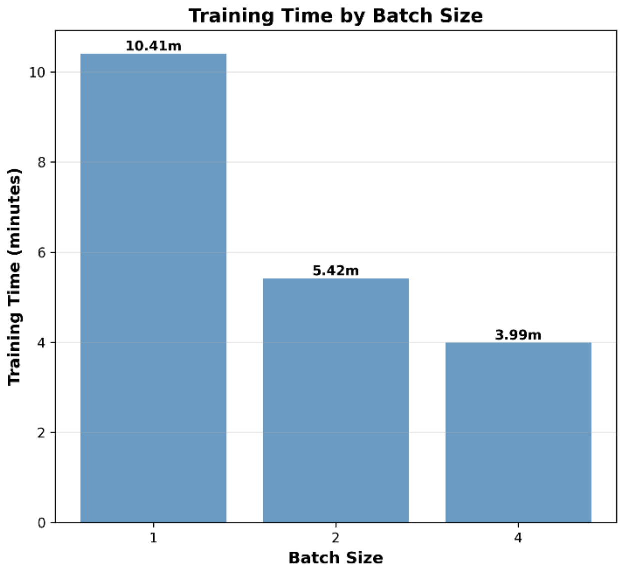


Figure 33: Training Time by Batch Size

Figure 33 illustrates training efficiency, that is, increasing batch size in the training process decreases the training time: a batch size of 1 takes around 10.41 minutes, a batch size of 2 takes around 5.42 minutes, a batch size of 4 takes around 3.99 minutes, which is as expected the computational advantage of processing more samples in a single iteration. This efficiency is however compromised by a loss in evaluation performance.

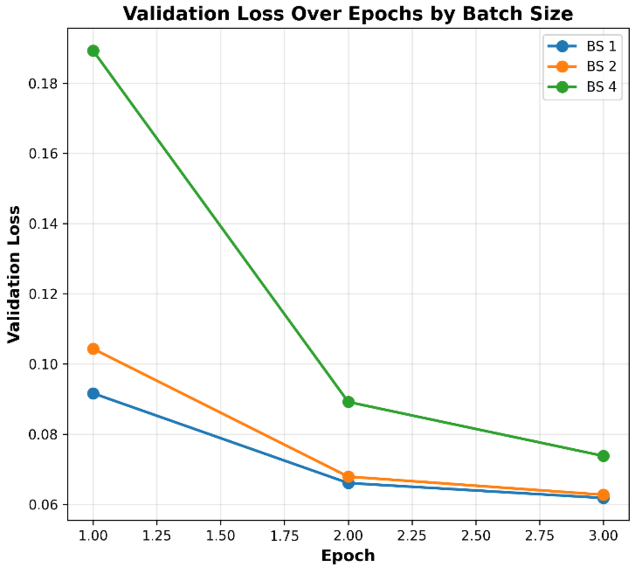


Figure 34: Validation Loss Trends by Batch Size

Figure 34 indicates the trend in validation loss with epochs, and it is observed that a batch size of 1 always has the lowest validation loss in training, and a batch size of 4 always has a higher validation loss, which further supports the evidence of performance degradation.

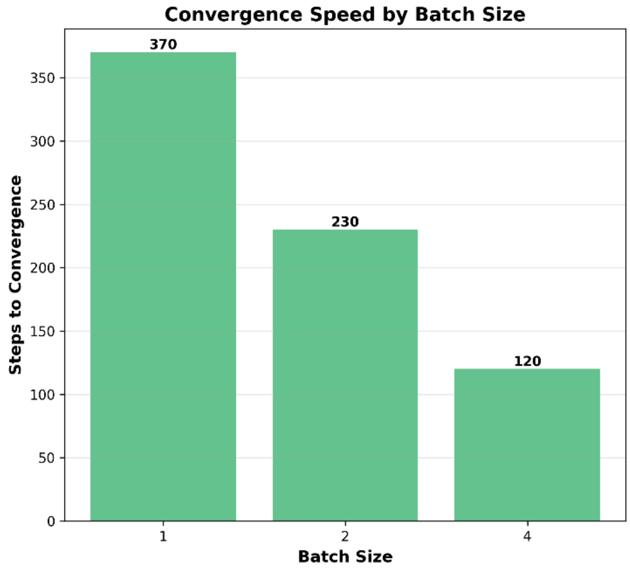


Figure 35: Convergence Rate by Batch Size

Figure 35 compares the rate of convergence, which shows that bigger batches have less steps to converge; however, the rate of convergence does not convert to better generalization.

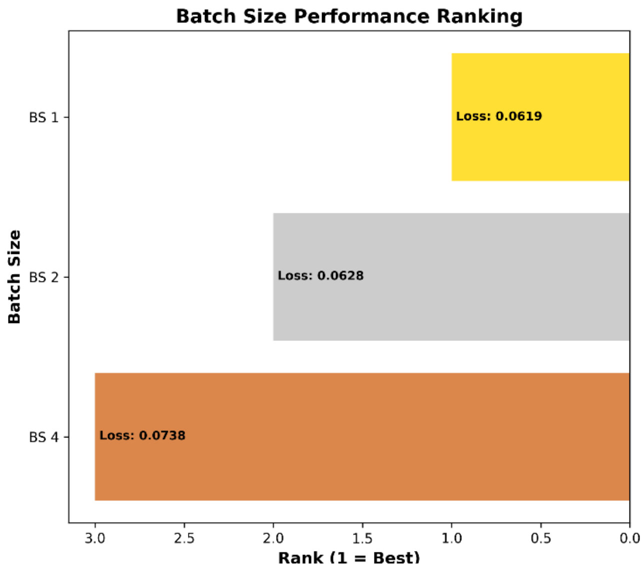


Figure 36: Performance Ranking by Batch Size

Figure 36 shows the performance ranking of the batch sizes, and it can be seen clearly that the best configuration is the batch size 1, then the best size is 2 and the last one performance is the 4.

Overall, the sensitivity analysis of the batch-size indicates that there is an evident trade-off between the efficiency of the training and the performance of the model. Even though a larger batch size would minimise training time, smaller batch sizes would realise better generalisation and minimal evaluation loss. In the case of the financial literacy chatbot, batch size, 1, provides the best trade-off between learning stability and performance quality and was hence the choice of the batch size in further fine-tuning experiments.

4.3.7 Optimizer Comparison

The subsection assesses the effectiveness of the various optimization algorithms on the convergence of the model, generalization and training efficiency. The problem here is to find an optimizer that gives stable learning behaviour with low evaluation loss and at a reasonable cost of computation. Three optimizers, namely, AdamW (Torch), Paged AdamW 8-bit, and Adafactor are compared in the same experimental conditions.

Table 18: Optimizer Comparison Results

Model	Optimizer	Train Loss	Eval Loss	Time (minutes)
Phi-4	Adamw_torch	0.0609	0.06639	50.95
Phi-4	Adafactor	0.0611	0.06548	51.69
Llama 3.2 1B	Adamw_torch	0.2702	0.0628	5.41
Llama 3.2 1B	Paged_adamw_8bit	0.2703	0.0628	5.48
Llama 3.2 1B	Adafactor	0.2440	0.0618	6.46
Qwen 2.5 1.5B	Adamw_torch	0.3089	0.3208	16.84
Qwen 2.5 1.5B	Adafactor	0.3084	0.32141	16.03

Table 18 provides a summary of the training sets and training outcomes of each optimizer. Each of the experiments was trained with a fixed train-validation split of 90:10, a learning rate of 0.0002, a batch size of 2, and three training epochs. This controlled configuration also makes it possible to ascribe any difference in performance which is observed to be due in the first instance to the choice of optimizer.

In the case of LLaMA 3.2 -1B, the lowest evaluation loss was obtained with Adafactor (0.0618), whereas Paged AdamW 8 -bit and AdamW (torch) have the same evaluation loss (0.0628). Although the difference is rather small, the lower validation loss that was achieved with Adafactor shows somewhat better generalisation. Nonetheless, Adafactor will still have a slightly longer training time than the rest of the optimisers.

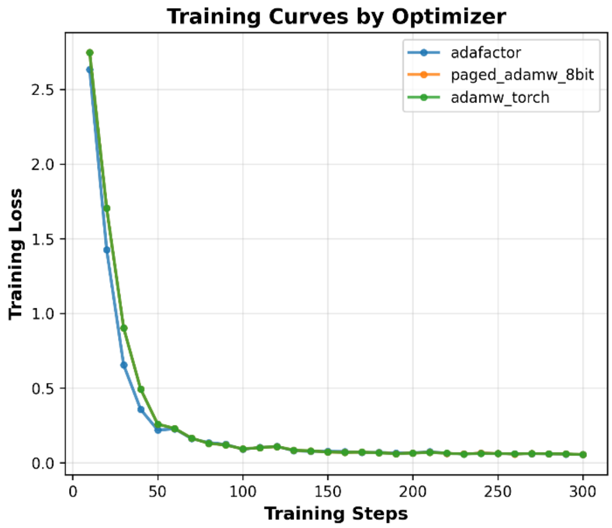


Figure 37: Training Curves by Optimizer

Figure 37 training curves reveal that all the three optimisers display steady convergence, characterized by rapid loss reduction in the initial training steps and subsequent smooth stabilisation. There are no indicators of instability or deviation, which means that all considered optimisers can be used to fine-tune with the specified setup.

Figure 38: Final Training and Validation Loss by Optimizer

The loss in training differs marginally between different optimisers, with Adafactor having a lower final training loss than AdamW variants as shown in Figure 38. The respective validation losses are tightly concentrated, and it means that the generalisation performance is consistent. The distance between training and validation loss is also small and this further implies that none of the optimisers do introduce serious over-fitting.

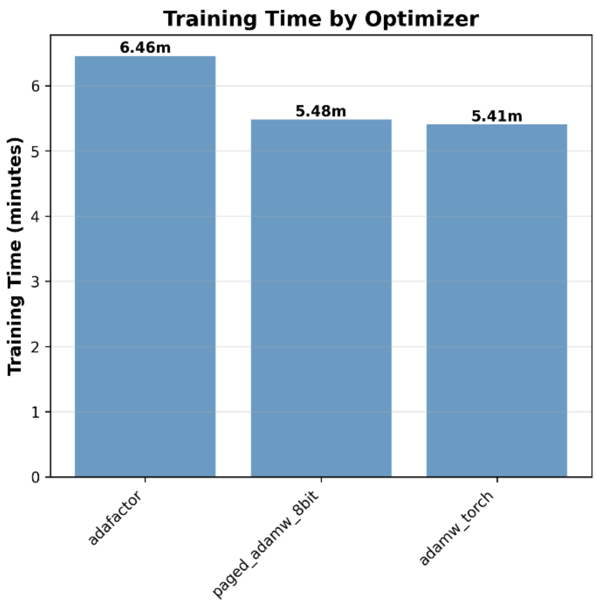


Figure 39: Training Time by Optimizer

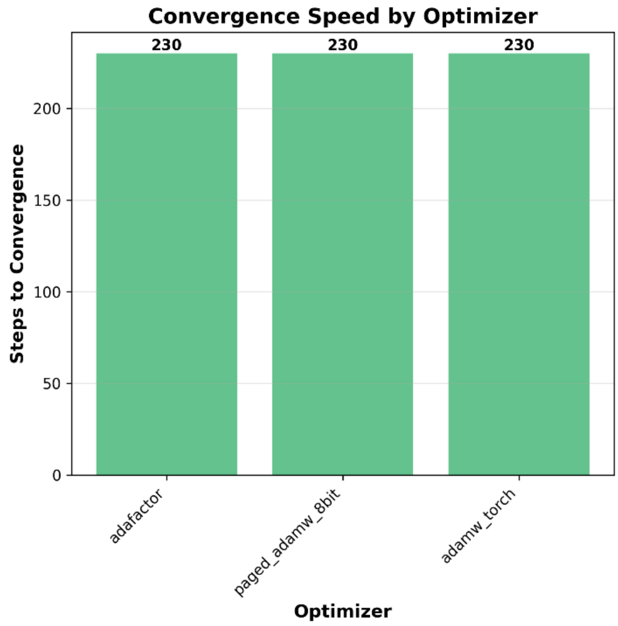


Figure 40: Convergence Rate by Optimizer

Following Figure 39 in terms of computational efficiency, Paged AdamW 8 - bit and AdamW (torch) provide a slightly shorter training time than Adafactor. However, the disparities are not too significant, and all optimisers achieve training in a comparable time interval. The comparison of convergence speed in Figure 40. also indicates that all the optimisers achieve convergence in the number of training steps, which is roughly equal.

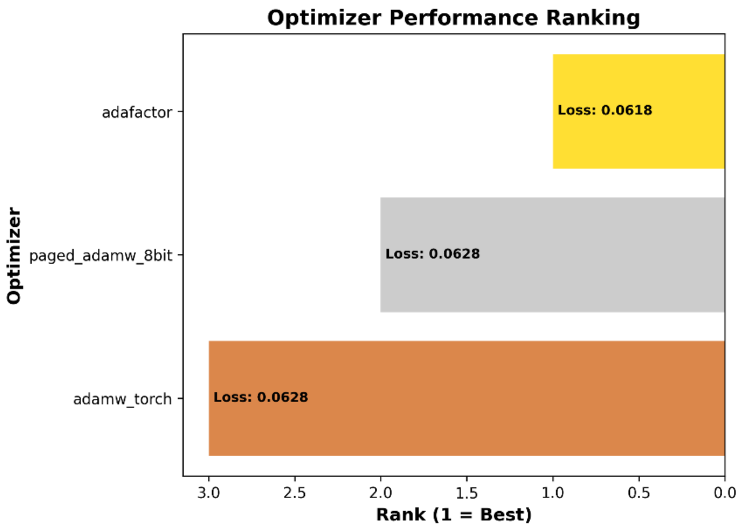


Figure 41: Performance Ranking by Optimizer

The overall optimiser ranking as shown in Figure 41 ranks Adafactor as the most successful optimiser in terms of evaluation loss, trailed by Paged AdamW 8-bit and AdamW (torch). The performance differences are minimal; however, Adafactor has a slight advantage in generalisation, which is why it is an appropriate option when the most attention is paid to the validation performance, as opposed to minimum training time.

According to this analysis, Adafactor was chosen as the optimiser of choice to be used in the following experiments because it offers the best evaluation loss and does not change its convergence characteristics or require a high level of computational cost.

4.4 User Interface and System Implementation Results

This section shows a design and functional structure of the financial literacy chatbot interface. It describes the interaction of users with the system as well as the way information is delivered and the way interface facilitates learning, assessment and evaluation. The interface was to be usable, clear and pedagogically effective and be able to support technical and non-technical users. Screenshots have been used to make illustrations on the implemented components.

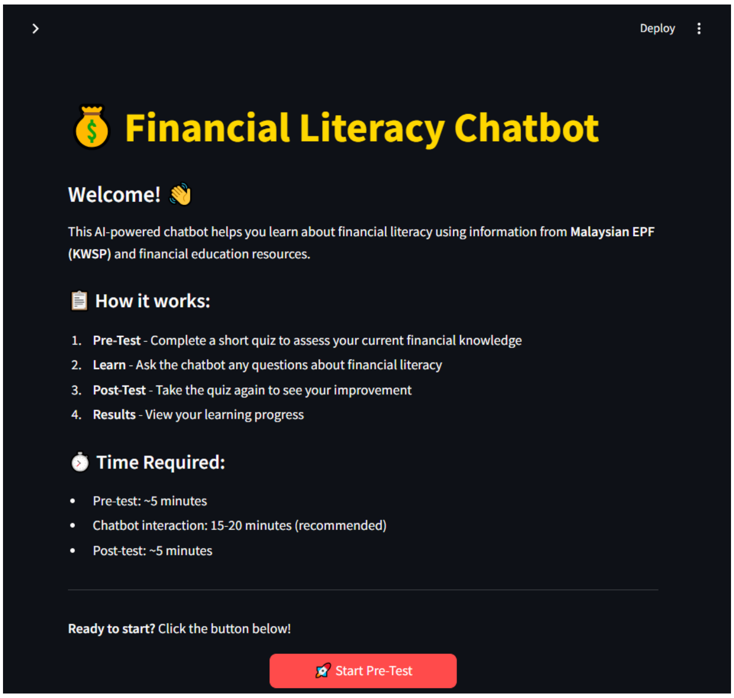


Figure 42: Welcome Interface of Financial Literacy Chatbot

The welcome interface is the first point of the Financial Literacy Chatbot, and it will present the user with the overall learning path of this system. The page explains to the world succinctly as shown in Figure 42 how the chatbot provides financial literacy advice by referring to finances data provided by the Malaysian Employees Provident Fund (KWSP) and other relevant educational materials. It outlines a four-step process that includes a pre-test, chatbot interaction, post-test, results review, and thus, allows users to understand the step-by-step flow in the system. The time estimates are also presented to ensure the user participation can be managed effectively. This clear and simplified layout is meant to make it user-friendly to users who may have different technical backgrounds besides creating a smooth and guided learning process.

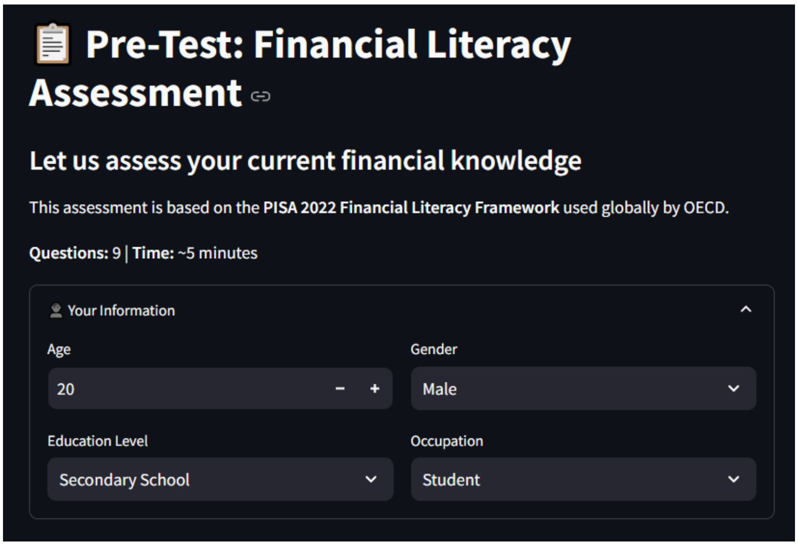


Figure 43: Pre-Test Financial Literacy Assessment Interface

The current page presents the Pre-Test: Financial Literacy Assessment that will become the first diagnostic part of the system that will help to capture the level of financial literacy of the users before the chatbot can interact with the user. The assessment meets the requirements of the PISA 2022 Financial Literacy Framework as shown in Figure 43 and thus demonstrates the conceptual validity and global applicability of the tool. The interface also collects simple demographic information which helps to contextualise analyses of learning outcomes without being cumbersome in terms of a registration process. The inputs allow the system to further decode the results in various learners' profiles.

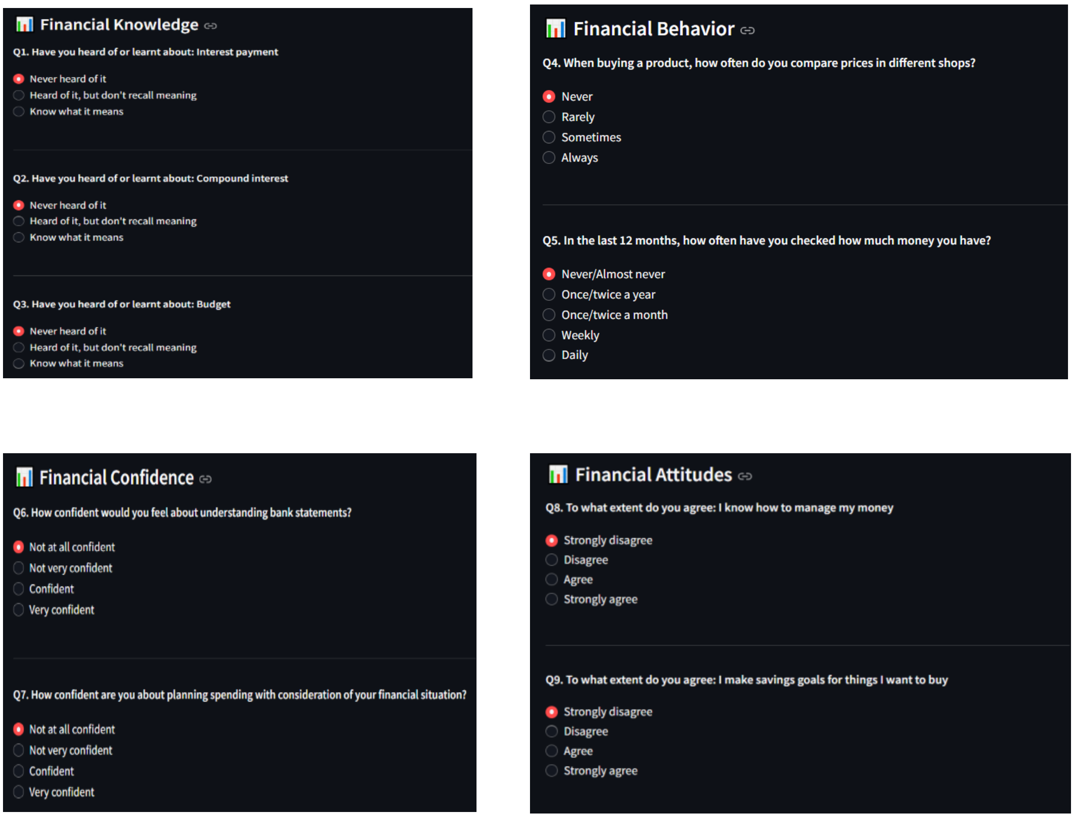


Figure 44: Pre-Test Questionnaire Based on PISA 2022 Framework

Figure 44 in the page displays a structured questionnaire, which derives out of OECD PISA 2022 Financial Literacy Framework. The tool looks at four main dimensions that include financial knowledge, financial behaviour, financial confidence and financial attitudes. The test will include 9 multiple-choice and likert scale questions clearly divided into their corresponding domains to make the test easier to understand and use. This pre-test is structured since it serves the purpose of acting as the baseline of assessing the change in the financial literacy of users after the chatbot interaction and is thus useful in making comparative analyses during the post-test phase.

Figure 45: Chatbot Interaction Interface

The chatbot interaction interface is the main learning environment where the users can make interaction with the financial literacy assistant after the pre-test. The interface design provides that as a user one can query the application in natural language in relation to financial topics like budgeting, saving, or management of spending habits. Every response is presented as a conversation, hence making it easy to understand financial concepts by the users using simple and understandable language.

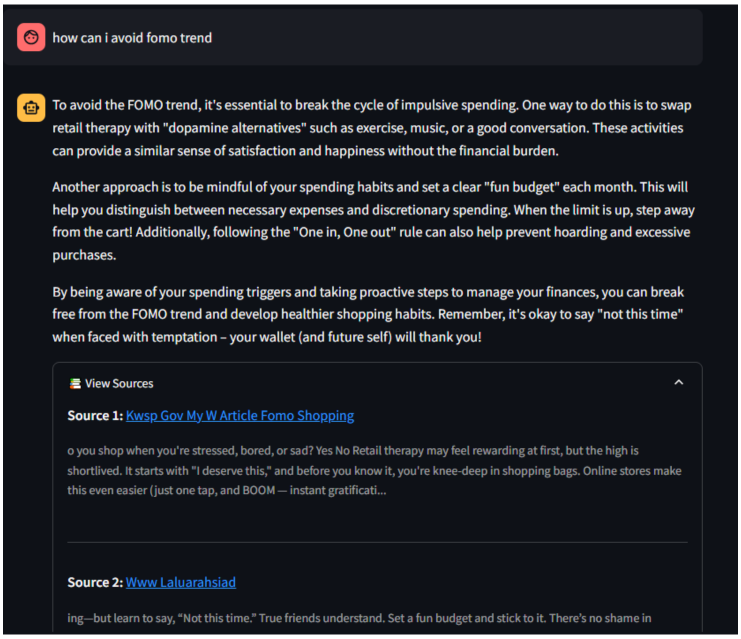


Figure 46: Source References Interface

To enhance transparency and credibility, reference to the source will be shown at the bottom of all answers so that a user may also confirm the information provided and read more. In general, the chatbot interaction page serves the operationalization of the learning-by-inquiry paradigm where the user actively learns the financial knowledge in the form of a dialogue and, thus, reinforces the received understanding before a formal test.

Figure 47: Post-Test Assessment Interface

The post-test interface will evaluate the financial literacy of the users after communication with the chatbot. The learning outcomes and knowledge acquisition will be measured. As illustrated, the post-test is in line with the PISA 2022 Financial Literacy Framework, which is in line with internationally recognised assessment standards.

Figure 48: Learning Results Dashboard - Before and After Comparison

After completion of the post-test, the participants will be taken through an interface that integrates their learning results in the central financial literacy aspects, which are financial knowledge, financial behaviour, financial confidence, and financial attitudes. The interface has an evident before-and-after analysis expressed in percentage indicators, which allows users to visually determine their progress after communication with the chatbot. Both dimensions are displayed together with their corresponding pre-test and post-test score and enhanced by a highlighted sign of improvement, thus making the influence of learning instantly understandable even to someone not possessing a technical knowledge.

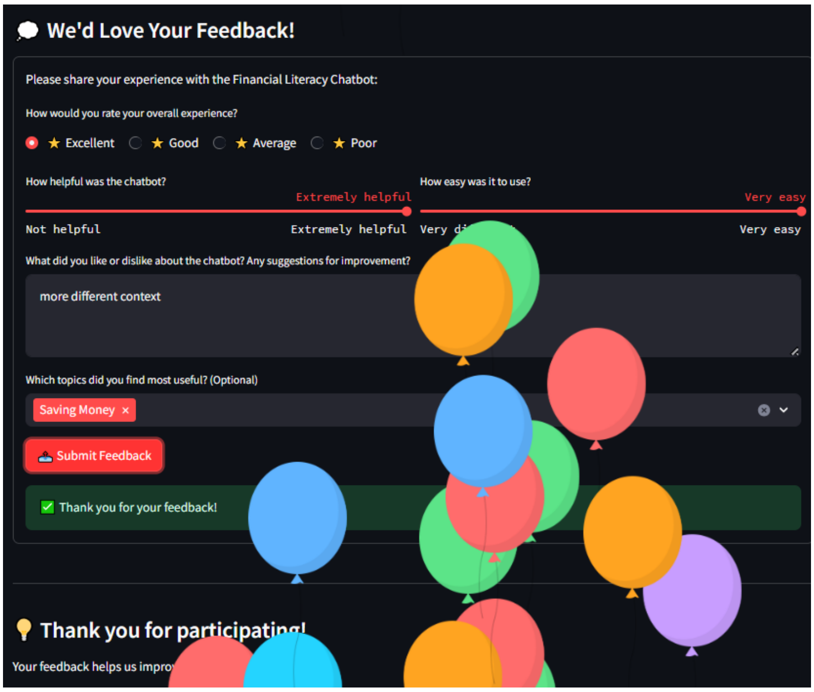


Figure 49: User Feedback Collection Interface

The feedback page will ask the participants to give their overall experience on a Likert-like scale between poor and excellent and then calibrate the sliders that will evaluate perceived usefulness and ease of use. There are open-ended entry fields that solicit qualitative commentaries and suggestions thus, giving deeper information on user satisfaction and quality of interaction. Furthermore, the participants can determine the financial issues which were most useful to them, so the system can focus on the areas of content that were perceived as more relevant. When submitting, a confirmation message is received which indicates that feedback is received successfully. It is this feedback mechanism that hence forms the basis of constant system refinements through the capture of both quantitative and qualitative user evaluations.

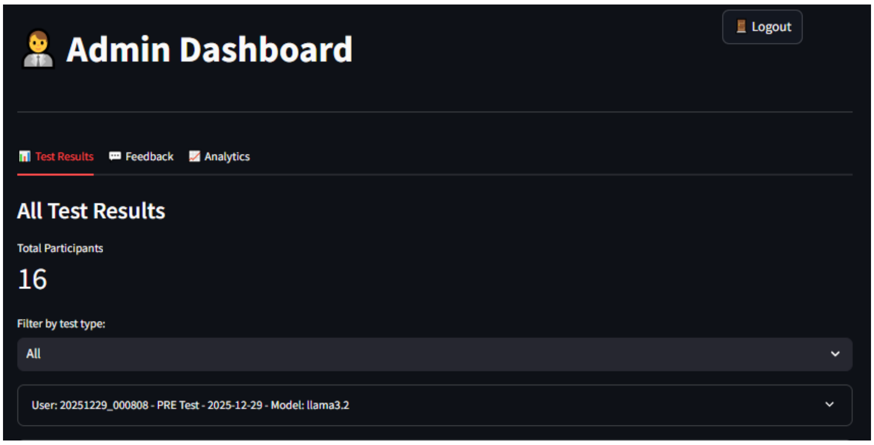


Figure 50: Administrative Dashboard for Monitoring and Evaluation

There is also an administrative dashboard which helps in monitoring and evaluation. This management dashboard is used to present aggregate test results, feedback history and user interaction data in an organized manner. Administrators have the possibility to watch the separate test sessions and feedback information which enables the assessment of the system and the further improvement.

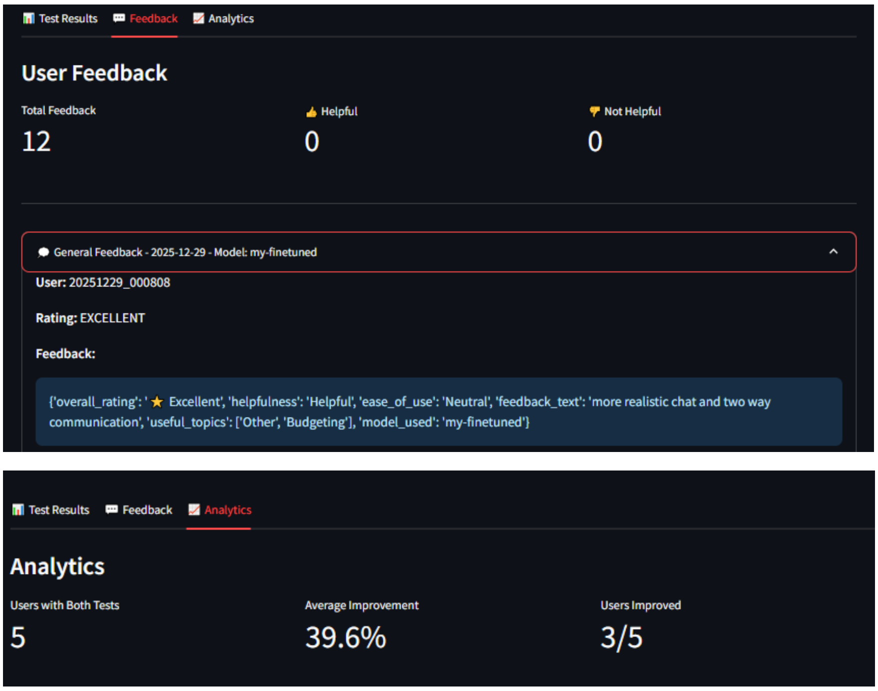


Figure 51: Feedback and Analytics Summary Interface

The system can also offer a feedback and analytics interface where the responses of the users can be collected after the financial literacy activities. Using this interface, users can assess their experience with the chatbot based on their overall satisfaction, perceived usefulness, and ease of use by indicating them on a structured feedback form. Users can also leave open-ended comments and choose the financial subjects that they found the most useful. The feedback system facilitates the process of reflection and the two-way communication between users and the system.

Along with qualitative feedback, the interface displays summarized analytical indicators, including the total number of participants, the number of those users who passed both the pre- and the post-tests, and the average score of improvements. These pointers can be used to determine how effective the chatbot is in enhancing the financial literacy levels of users.

4.5 Evaluation Metrics and Model Performance

The following subsection outlines the evaluation results of the proposed chatbot on financial literacy. The evaluation aims to test the quality, consistency, and reliability of the responses provided by the chatbot using quantitative metrics derived out of model-generated responses. The focus is put on measuring the similarity of semantics between the answers provided by a chatbot and those offered by a reference system, which provides the empirical data on the effectiveness of the system in presenting valid and useful financial data.

To ensure a full assessment, the set of similarity-based measures is employed, which includes cosine similarity, BERTScore, semantic similarity, and Jaccard similarity. These measures summarize the various aspects of text congruence, which go beyond lexical coincidence up to deep context congruence. The visual analytic methods are used to support the interpretation of the results such as distribution plots, boxplots, correlation heatmaps, and mean-score comparisons. A combination of these methodological strategies will make this a systematic and objective evaluation of the chatbot performance in terms of providing financial-literacy material.

4.5.1 Evaluation Framework and Metrics

The assessment paradigm that has been followed in this research will be used to determine the quality of the responses created by the financial literacy chatbot quantitatively. Chatbot outputs are textual and might differ in terms of phrasing without loss of meaning, which does not make them appropriate to be evaluated using traditional approaches to accuracy. Rather, the natural language evaluation measures are based on similarity to assess semantic congruency between chatbot generated responses and pre-established reference answers.

In the given study, four measures are used in evaluation, which are cosine similarity, BERTScore, semantic similarity and Jaccard similarity. Every measure represents a different aspect of quality of responses. Cosine similarity compares the angular similarity of text representations in the form of vectors of generated and reference texts, thus giving a clue about the general lexical and contextual distance. BERTScore is a token-level semantic similarity model based on contextual embeddings, so it is especially appropriate in natural language generation tasks where different, valid expressions can be possible.

This multi-metric analysis model provides a well-rounded and sound measurement of the performance of chatbots. The analysis will provide an overall picture of the effectiveness of the chatbot in providing correct and meaningful financial explanations because the evaluation integrates semantic, lexical, and distribution-based analyses and will, thus, justify the purpose of the chatbot as a teaching tool.

4.5.2 Distribution of Evaluation Metrics

In this subsection, the general distribution of the evaluation scores is explored to determine the consistency at which the chatbot can produce good responses.

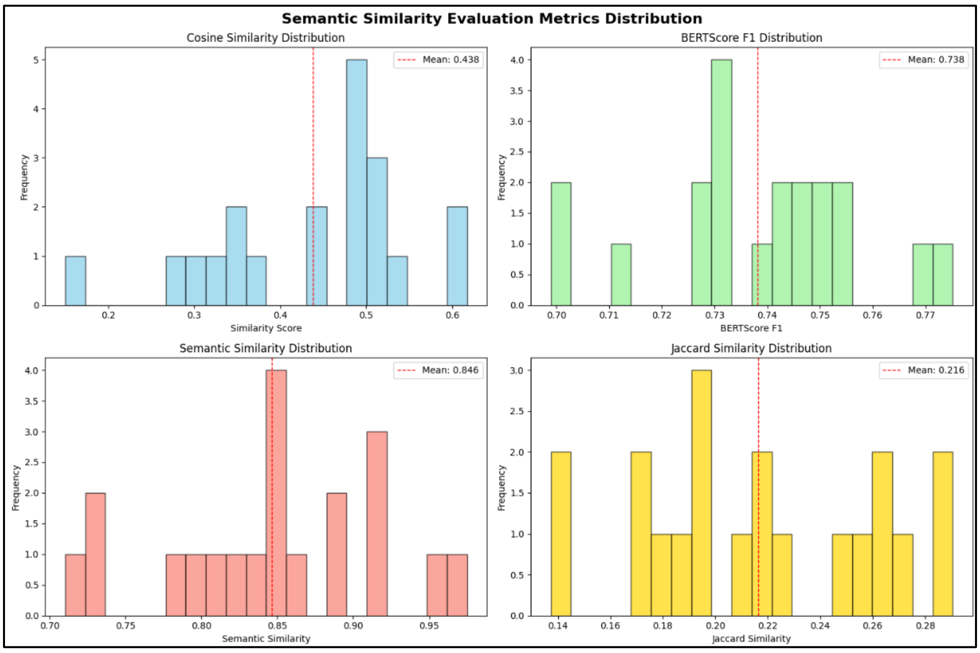


Figure 52: Distribution of Evaluation Metrics

Figure 52 shows the distribution of cosine similarity, BERTScore, semantic similarity, and Jaccard similarity of all responses that were evaluated. The distributions show that the semantic similarity scores are largely clustered at a greater value, which implies that there is a strong conceptual correspondence between the output of the chatbot and the reference responses. BERTScore also has quite a high level of value concentration, which demonstrates good token level contextual matching.

Conversely, the similarity using cosine and Jaccard are more widely distributed indicating more variation in the overlap of lexical relationship. This variability is expected in the case of generative systems where the accurate responses can be phrased in varying words. As the total distributions reveal, the superficial similarity of the responses might vary but the chatbot maintains the same level of semantic quality in response to different questions. These observations are a preliminaries sign of consistency of responses before comparative analysis at a more detailed level.

4.5.3 Comparative Analysis of Evaluation Metrics

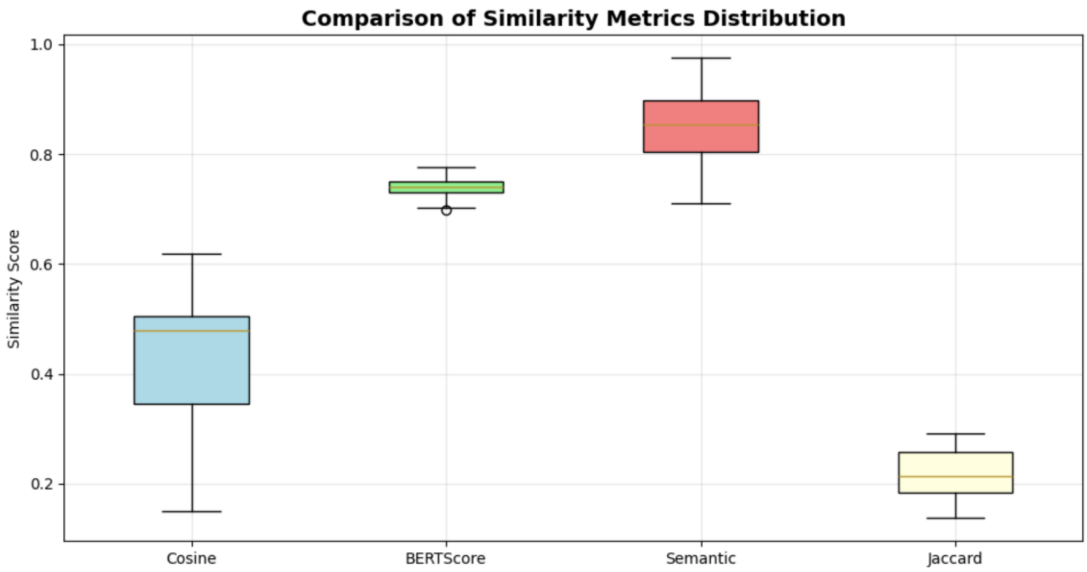


Figure 53: Boxplot Comparison of Evaluation Metrics

A boxplot comparison of the evaluation metrics is presented in figure 53 and it is possible to directly compare the central tendencies and variability of the metrics. The median of semantic similarity is the highest and the interquartile range the smallest, which highlights the lack of variation in conceptual correspondence between the answers. This indicates that the chatbot consistently maintains meaning in the generation of explanations that are connected to matters of financial literacy.

The median of BERTScore is also relatively high and of medium variability, which indicates a high level of similarity between contexts, but linguistic freedom. On the other hand, the median values and distributions of cosine similarity and Jaccard similarity are lower and have broader distributions, indicating that they are susceptible to the lexical surface differences. These metrics show several outliers as the case is expected when paraphrasing is taking place.

In general, the boxplot analysis highlights how semantic-based metrics can give a more informative and stable assessment of chatbot performance compared to the lexical ones. This justifies their appropriateness to evaluate generative educational systems in which preservation of meaning is most important as compared to wording.

4.5.4 Evaluation Metrics Correlation

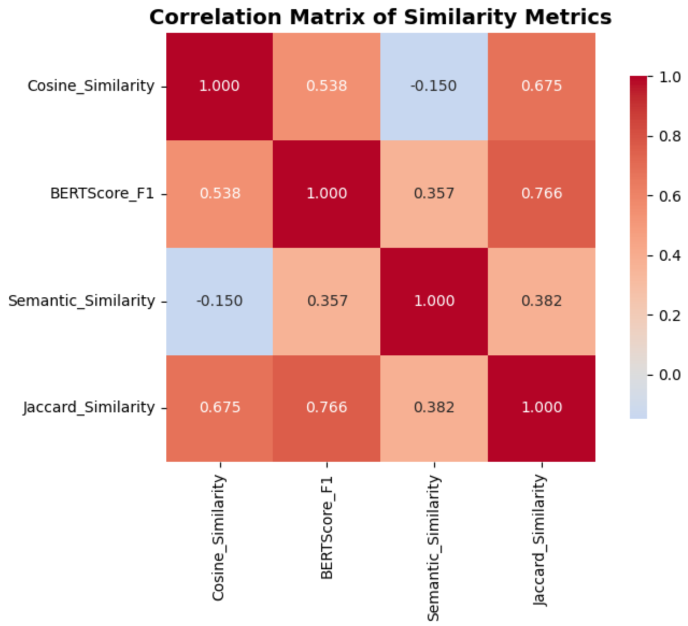


Figure 54: Correlation Heatmap of Evaluation Metrics

Figure 54 represents the correlation between evaluation metrics using a heatmap. As we can see by the analysis, there are very high positive correlations between the semantic similarity and BERTScore, which suggests that the two measures represent overlapping elements of semantic alignment. This is consistent and this makes the framework of evaluation used in this study more reliable.

Intermediate correlations are found between the cosine similarity and the semantic based metrics, which are indications of partial correspondence between the semantic meaning and lexical similarity. Conversely, Jaccard similarity has lower correlations with the rest of the measures, indicating that this metric is dependent on word-for-word overlap, as opposed to semantic interpretation.

These patterns of correlation show that all metrics are valuable perspectives to give, although semantic similarity and BERTScore are the most informative measures of quality of responses. The complementary character of the metrics reinforces the general analysis since it considers various aspects of textual similarity but not one measure.

4.5.5 Mean Performance of the Evaluation Metrics

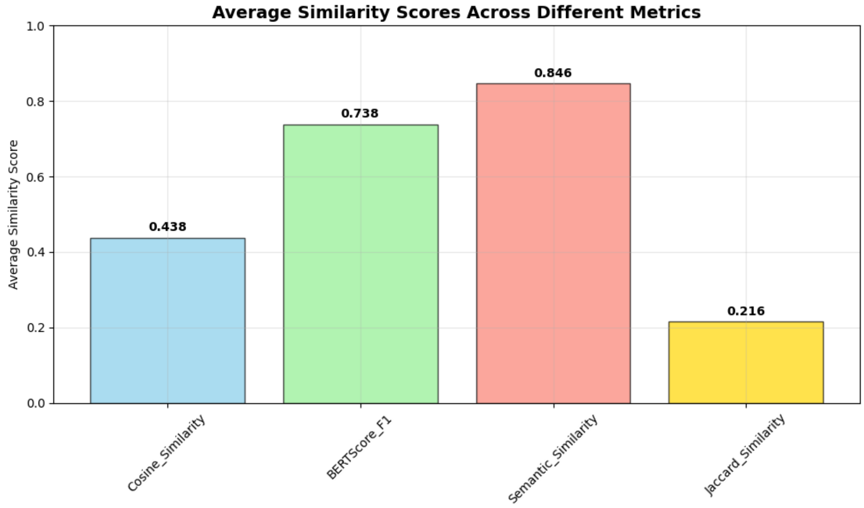


Figure 55: Mean Performance Scores Across All Evaluation Metrics

The mean scores of all the evaluation metrics are provided in Figure 55, which provides a brief overview of the chatbot performance. The highest average value is the semantic similarity then BERTScore which means that generated responses and reference answers have a strong semantic and contextual match. These findings indicate that the Chatbot can provide a meaningful presentation of the core financial concepts.

The average values of cosine similarity and Jaccard similarity are relatively lower and indicate more difference in surface-level phrasing. This is what is required in natural language generation tasks and does not always indicate reduced quality of answers. In its place, it puts emphasis on the ability of the chatbot to paraphrase conceptually accurate.

Combined with the previously mentioned results, it becomes apparent that the chatbot has stable and predictable behaviour along the lines of evaluation. The prevalence of semantic based measures justifies the appropriateness of the system to educational applications, where the emphasis is made on conceptual knowledge as compared to literal matching. The findings are a quantitative piece of evidence that the chatbot serves its purpose as a financial literacy support tool.

4.6 Prototype Expert Validation Result

To determine the usefulness and trustworthiness of the financial literacy chatbot, professional validation was performed with the help of a structured evaluation questionnaire. This validation was conducted by a domain expert who had a long experience in financial literacy Dr. Lalua Rahsiad. The tool was a questionnaire that consisted of questions to assess system functionality, content validity, usability, and educational quality. The assessment was based on the 5-point Likert scale with 1 being Strongly Disagree and 5 being Strongly Agree and categorical response which are Yes, No, and Partially. The obtained data give a reflection of the strengths and weaknesses of the chatbot and justifies its use as a learning support system.

Domain Accuracy and Reliability

1. **System Factual Accuracy**

   Score: 3

   Although this system is mostly effective in delivering correct financial response, sometimes its carry a slight error or incomplete description. Even though the chatbot shows the basic understanding of financial concepts,

2. **Correctness of Domain Logic**

   Score: 4

   Implying that the reasoning process and explanatory sequences of the system are mostly in concurrence with the known financial principles. Therefore, even though some of these factual details may require refining.

3. **Appropriateness of Terminology**

   Score: 3

   Although the words are mostly understandable, there might be certain domain-specific terms that might be considered imprecise or inconsistent. The outcome leads to a possibility of a closer correspondence of language use to traditional financial terms.

4. **Professional Reliability of Outputs**

   Score: 2

   Although the chatbot can be a useful tool of education and information, the results of the tool should not be trusted without the human supervision in the situations where it comes to professional decision-making.

Content Coverage and Completeness

1. **Content Coverage**

   i. **Covers Essential Domain Concepts**

   Response: Yes

   The expert affirmed that it covers important aspects of financial concepts, and this is because the chatbot covers the fundamental areas that should lie in a financial literacy.

   i. **Information Depth is Appropriate**

   Response: No

   On the other hand, the expert indicates that on full content correctness in certain cases implies that some of the answers can be incomplete or untrue.

   ii. **No Critical Omissions**

   Response: Partially

   Despite grasping the key ideas of financial theory, some additional information or background information can be inadequate. This fact suggests that the system provides a general picture of issues and never really gets to the level of rigor generally required by more intensive academic or teaching environments.

   iii. **Avoids Misleading Information**

   Response: Partially

   The expert states that as much as replies are generally applicable, there can be omissions of certain subtleties, illustrative examples, or even edge cases. This can in turn force users to request additional explanation or other materials to gain an all-inclusive knowledge on specific financial topics.

2. Usability Evaluation

i. Easy to Understand

Score: 4

The system is usually understandable, which means that the answers provided by the chatbot can be read and that the user will be able to perceive the information that was delivered to him without any unnecessary efforts. This analysis highlights how the system has succeeded in the communication of financial concepts

ii. Language Used

Score: 4

The tone and language are mostly appropriate to a financial literacy environment. This judgment indicates that the system avoids rather technical language in most cases and maintains its accessibility.

iii. Logical Interaction Flow

Score: 3

Even though the chatbot can be used, it is sometimes that the flow of response or conversational format is not as completely coherent or smooth. The point made by this observation is that users will be able to accomplish their tasks, but overall interaction experience will be polished to look more natural.

iv. Suitability for Real-world Use

Score: 3

It states that even though the system is functioning well, there could be specific areas of improvements to increase the confidence and comfort of the users, including consistency of responses or clarity of the follow-up explanations.

In sum, the professional was quite pleased on a unsatisfactory scale, which means that the chatbot does not go beyond the minimum expectations as it relates to functionality and performance during the prototype phase. The professional also replied with yes, with few adjustments about suitability to be used in real world. The system will work in the educational or support capacity but requires certain adjustments before it could be rolled out to the rest of the world. On balance, the results of the validation show that the financial literacy chatbot has the first competency in understanding financial terms, structuring explanations, and interacting with a person. However, the mediocre and low scores on various domains show conspicuous failures on professional reliability and content accuracy, which is fit to a prototype-type system. These findings provide evidence-based and balanced information on the strengths and weaknesses as they exist today.

CHAPTER 5

CONCLUSION AND RECCOMENDATION

This chapter summarizes that of the research outcomes. The limitations the research is stated and recommendations for future improvements are made.

5.1 Objectives Achieved

This paper has managed to solve the widespread problem of financial illiteracy among young Malaysians and created a web-based Financial Literacy Chatbot by using Large Language Models (LLMs) and a Retrieval-Augmented Generation (RAG) architecture. The project achieved the three research objectives in a systematic approach to the research work that involved a preliminary study, acquisition of knowledge, data collection, data preparation, system design, system development, and evaluation of the complete system. The resulting prototype shows that the use of sophisticated artificial intelligence methods can be effectively used to provide accessible, personalized, and contextually relevant financial education to Malaysian youth to address the long-term shortcomings of traditional financial literacy interventions, which include lack of engagement, personalization, coverage of digital financial literacy, and scalability. The chatbot also provides 24-hour financial advice based on official Malaysian sources, such as KWSP, AKPK, and NFEC publications, therefore, setting answers in accordance with the local financial environment, regulatory landscape, and the issues affecting young Malaysians, such as EPF management, PTPTN student loans, Islamic financial products, and changing fraud trends.

The initial research question, which demanded the need to discover personal finance educational content, was fully fulfilled with the help of a strategic data gathering strategy, which placed emphasis on mainstream authoritative Malaysian financial literacy sources. Instead of using some generic global content or publicly available dataset on the Bank Negara Malaysia Open API and Hugging Face, like initially intended, the methodology adopted was the creation of a custom corpus based on carefully selected materials on the Employees Provident Fund (KWSP) and Credit Counselling and Debt Management Agency (AKPK) official site. A written Python program based on the Selenium web automation systematized retrieval of 42 high-quality articles, guides, and infographics on financial literacy that discuss key topics: retirement planning, budgeting basics, investment strategies, fraud avoidance, and managing first salary. The localization approach means that the chatbot will respond in a way that responds directly to the issues of finance that are pertinent to the Malaysian youth, hence avoiding the spread of general international suggestions that might not be in line with the regulatory, cultural as well as institutional environment of the country. The managed body of knowledge, therefore, includes the basic areas of financial literacy that young Malaysians need to be empowered to make sound financial choices, handle the personal finances, and guard against the ever-advanced Internet-based financial fraud targeting digitally native populations.

The second research target, which entailed the development of a fine-tuning strategy to an LLM on financial literacy, was achieved by conducting a detailed system architecture design and the experiment on hyperparameter optimization that provided an evidence-based model setting. This development of design, despite the final implementation adopting a Retrieval-Augmented Generation (RAG) architecture, but not a purely fine-tuned strategy, was indicative of a better understanding of what a financial literacy chatbot needs and what it should not need, like factual accuracy, source attribution, and prevention of hallucinations should take precedence over creative language generation. The RAG architecture incorporates ChromaDB, a knowledge storage in the form of a vector database, multilingual-e5-small embedding model to support bilingual functionality in the future of English-Malay expansion, semantic similarity-based retrieval to identify the 5 most relevant document chunks according to user query, and a locally-executed Llama 3.2 1B model to generate language translations in a privacy-preserving and cost-free way. Three candidate models, each, were used to run a systematic fine-tuning experiment on three hyperparameter configurations: (1, 3, 5 epochs), (70:30, 80:20, 90:10) train validation split, (0.0001, 0.0002, 0.0005) learning rates, (1, 2, 4) batch sizes, (AdamW, Paged AdamW 8-bit, Adafactor). LLaMA 3.2-1B and 5 epochs, a 90:10 split, a 0.0005 learning rate, a batch size of one, and the Adafactor optimizer were found to give the best validation loss of 0.0604, and thus, offers empirical support to the domain-specific language model adaptation in use beyond the current financial literacy context.

The third research goal, which required the creation of a web-based application of an educational chatbot of personal finance using an LLM, was achieved through the implementation of a complex system, which included an integrated assessment, instruction, evaluation, and feedback collection functionality in a single platform developed in Streamlit. The created application implements the entire financial literacy learning cycle, which includes starting with a pre-test assessment corresponding to the PISA 2022 Financial Literacy Framework to assess the baseline competency in four dimensions financial knowledge, financial behaviour, financial confidence, and financial attitudes. This is proceeded by conversational chatbot interaction which allows natural language question-answering with attribution of sources to increase transparency and trustfulness followed by post-test evaluation to quantify learning gains by a before-after comparison. The cycle is completed by systematic feedback gathering with the measurement of both the quantitative satisfaction rates and qualitative improvement recommendations. The RAG workflow, which is orchestrated by LangChain, retrieves user queries in a cache of a vector database, builds prompt structures with context-enhanced prompt construction and generates responses token by token at a temperature of 0.1, therefore, producing factual but not creative responses. The administrative dashboard summarizes the test results, feedback and user interaction data, allowing monitoring of the system at a system level, and making decisions based on evidence to refine the system. Dr. Lalua Rahsiad verified the functional competency that was suitable to the prototype development phase, and also noted the areas that would benefit development, such as greater factual accuracy, greater breadth of content, greater flow of conversation, in this way creating a clear framework of steps that may be followed to create an educational tool that is production-ready.

5.2 Limitation

The sole use of the content obtained through the KWSP in the knowledge base construction presents a certain bias toward the issues of retirement planning and EPF itself, which may not be reflective of the other vital areas of financial literacy. Such areas include the basics of insurance, taxation rules, diversification of investment, financial management of entrepreneurship and protection of the rights of consumers. Although KWSP articles are authoritative and can be relevant locally, it is only with such a comprehensive financial literacy curriculum that can cover the entire range of personal finance problems that young Malaysians face at different life levels, such as financing their education to owning their own homes, and family financial planning. The prototype includes 42 articles that were gathered, which is a sizeable sample to illustrate the research purpose, but is a small sample of the existing Malaysian financial literacy resources. It is possible that this restriction will omit major perspectives, explanatory strategies, practice cases, and new finance issues, including cryptocurrency, digital banking innovations, services and the development of new types of fraud, which other credible sources, such as Bank Negara Malaysia, the Securities Commission Malaysia, AKPK and insurance industry associations, can provide. The content limitation, in turn, affects the ability of the chatbot to provide a complete response to the general pool of queries that the user might have, which may lead to the insufficient answer to questions which are not covered in the KWSP corpus. This lack can be irritating to users who want to get advice on the underrepresented topics and undermine the usefulness of the entire platform as a full-fledged financial literacy platform.

The existing monolingual English application is also a significant accessibility obstacle to Malay-speaking young people who might desire or even need financial literacy education in Malay. This drawback is inconsistent with the outreach aims of inclusion that is consistent with the multilingual nature of Malaysia, and the equity aims of the national financial literacy strategy. Whereas the multilingual-e5-small embedding model provides a technical platform to support bilingualism, its practical implementation requires professionally translated parallel content corpora, which would make sure that the transparent financial concepts, technical terminologies and idiomatic expressions can be conveyed across the language without distortion, ambiguity and unintentional distortion of financial advice. The language barrier also adds another layer of exclusion to a large portion of the youth demographic, especially in the rural regions, among the less privileged as expressed through the lens of socioeconomic communities, or those with less proficient standings of the English language, and thus, the Malay language is the main medium of communication, and thus, jeopardizes the equity goals that the democratization of financial literacy access primarily aims to promote. Furthermore, the system lacks advanced personalization features higher than self-based exploration of the conversation, and thus, does not actively determine individual gaps in knowledge, the learning process between beginner and advanced concepts, and persistently models users over the course of multiple sessions. This weakness limits the capacity of the system to serve the user who needs to be guided in a structured manner as opposed to reactive query response and may end up acquiring disjointed knowledge by studying material in inefficient order without the necessary preparatory knowledge requirement.

5.3 Future Work and Recommendation

5.3.1 Future Work

The further development of the knowledge base into a comprehensive financial literacy content covering more than the knowledge that is being currently covered by Focus of KWSP with the incorporation of the materials of a wide range of reputable Malaysian sources should take a priority in future research to provide the balance between all the important areas of personal finance. This enlargement should incorporate information provided by the Bank Negara Malaysia Financial Education Network, including the monetary policy, banking rules, payment system, and consumer protection. Next, by the Securities Commission Malaysia investor education resources, including information on the fundamentals of investments, securities market, investment unit selection, and investment fraud prevention. Also, by the Life Insurance Association of Malaysia and General Insurance Association of Malaysia, including the information on risk management, the choice of a particular investment unit, and the procedure to file a claim. Lastly, the Inland Revenue Board Malaysia, where prospective investors can receive the information about taxation and the possibility In addition, the emerging financial technology issues, such as digital banking security, the most effective practices in using e-wallets, risk and opportunity of the cryptocurrencies, and the issues regarding peer-to-peer lending, should be included into the knowledge base. This diversification of content will have to satisfy users with different levels of proficiency, between simple concepts that can be understood by a total beginner and complex ones that can be accessed by users with high financial sophistication and specialization needs, and clear difficulty tagging must be used to support later personalization systems which can adjust the complexity of the content to the level of understanding already demonstrated by the user.

To overcome the critical accessibility drawback, it is necessary to build bilingual capabilities to support both English and Malay using a professionally translated parallel content corpus, intelligent language detection, and culturally competent adaptation, which will guarantee financial literacy education reaches the entire youth population of Malaysia regardless of language preference or language proficiency. Bilingual implementation is to be done with domain experts who are not only proficient in both languages but also understand the terminology of finance and such that subtle financial concepts as compound interest, diversification, liquidity, and risk-adjusted returns can be translated to Malay without distortion and oversimplification. Financial terms, which do not have direct translation into Malay equivalents, should be used with caution, using relevant loanwords, descriptive translations or inclusion of English terminologies in parenthesis to avoid mixing up, yet at the same time developing financial expressions in both languages. The language detector algorithms will be able to detect the query language of the user based on input text without having to explicitly specify the language to use in generating the response, thus, providing a smooth bilingual interaction experience that allows Malaysian young people to mix English and Malay during conversations. Increased accessibility features must also ensure that the system is accessible to a wide range of user populations, such as those with visual, auditory, motor, or cognitive impairments, by being compatible with screen-readers, using high-contrast visual themes, allowing text to be adjusted in size, providing voice input and output options, and having a structure that can be accessed by assistive technologies to support more inclusive financial literacy education to the communities traditionally underserved.

5.3.2 Recommendation

The modernization of the computing infrastructure and the use of cloud-based API services are the key technical advances to expand the capabilities of the chatbot and its implementation scope. Its current implementation is based on local deployment using limited computational resources, which limits the system to smaller language models (1B parameters) that, though efficient, do not have the advanced reasoning capabilities and worldly understanding of larger models like LLaMA 3.2 -3B, Qwen 2.5 -7B or even frontier models available over APIs like GPT-4 or Claude. By investing in increased RAM capacity and GPU resources, it would be possible to experiment with larger open-source models with proven more effective at solving complex financial reasoning tasks, and it can also be integrated with commercial LLM APIs, but the cost of such a strategy must balance carefully the cost of API use, data privacy concerns, and internet connectivity needs. In addition, API-based architectures have been shown to possess high benefits such as automatic model updates as providers release enhanced versions, no local model maintenance overhead, simple horizontal scaling to serve several users concurrently, and decreased technical complexity to implement in institutions with a low level of IT systems. An alternative solution of a hybrid combination of local deployment to privacy sensitive interactions and selective API invocation to address complex queries requiring more sophisticated reasoning would be a dimension of trade-off between performance, cost, privacy, and accessibility. Cloud infrastructure would also make it easier to integrate data in real-time as it is suggested in earlier sections, connecting to financial databases, news feeds, and government APIs with ease that would be impractical in a purely local deployment of systems, and support mobile app backs, user authentication systems, and analytics dashboards that would otherwise demand server infrastructure becoming constantly available.

A general longitudinal study is a critical area of research that must be used to confirm the usefulness of the chatbot in effecting long-term behavioural change and not just estimating short-term knowledge acquisition. The next round of researches should also adopt 12 to 24 month follow-up studies which would trace real monetary actions of chatbot users versus control group which would gauge tangible results like raising rate of savings, decreasing debts, joining formal investment instruments, creation of emergency funds, and individual levels of financial stress to identify whether chatbot engagements translate into practical gains in monetary wellness. Such studies would have to utilize mixed-methods methodologies that involve both quantitative measures of user financial behaviour with proper consent and privacy measures and qualitative interviews on how the user translates learned concepts to financial decisions, issues they had when implementing them, and how they value the chatbot as a learning tool in the long-term. Also, longitudinal study needs to examine differing effects between demographic groups such as socioeconomic status, educational level, initial financial literacy, and urban and rural environment to determine which groups could benefit the most under AI-based financial education, and which ones may need additional measures. The results of such research would serve as important information to policy makers and educational institutions contemplating implementing vast amounts of AI financial literacy tools, directing resource use models of youth financial education programs, and guiding an iterative cycle of improvements in the pedagogical approach of the chatbot, the content priorities, and the personalization models, based on empirically observed learning and behavioural results as opposed to hypothetical proposals on pedagogical efficacy.