USABILITY OF NATURAL LANGUAGE PROCESSING TO ASSIST STUDENTS IN LEARNING PROGRAMMING: A CASE STUDY OF GRETSA UNIVERSITY

BRIAN BENEDICT NJAU

A RESEARCH PROJECT SUBMITTED TO THE SCHOOL OF COMPUTING AND INFORMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE OF GRETSA UNIVERSITY

JULY 2023

STUDENT DECLARATION PAGE

STUDENT DECLARATION PAGE

This research project is my original work and has not been presented for the award of a degree or any similar purpose in any other institution.

Signature: _____ Date: 6/12/2023

BRIAN BENEDICT NJAU

ICT-4-4046-22

Supervisor:

This research project has been submitted with my approval as university supervisor

Signature: _____ Date: 6-12-2023

Mr. DENIS WAPUKHA

SCHOOL OF COMPUTING & INFORMATICS

GRETSA UNIVERSITY

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABBREVIATIONS AND ACRONYMS.

AI - Artificial Intelligence

ANOVA - Analysis of Variance

IDE - Integrated Development Environment

NLP – Natural language processing

SPSS - Statistics Package for Social Science

OPERATIONAL DEFINITION OF TERMS

Artificial Intelligence – Computer and machine programs that mimic the problem-solving and decision-making of a human mind.

Debugging – A process of finding and fixing errors in the code of any software.

Learning – This is the acquisition of knowledge and skills through study, experience, or being taught.

Natural language processing – Branch of AI concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Programming – It is commonly referred to as coding, it's a process of analysis, generation, and implementation of algorithms using specific languages of choice.

ABSTRACT

This study concentrates on the use of natural language processing in learning contexts, especially in learning programming which is characterized by a natural linguistic programming branch that implicates the combination of computational linguistics and rule-based modeling of human language. To study programming deals with the aspects of code debugging features, NLP tools, and NLP infrastructure. This study checks on how students use NLP to learn how to do programming and how the process is engaging from the start to the point where they can be up for work, also checks on how university students learn how to program through experience and self-study. This study has the general objective which was to evaluate the use of NLP in learning programming and the specific objectives which were to evaluate the NLP tools in assisting students to learn programming, to evaluate the NLP infrastructure that is needed to assist students in learning programming, to evaluate how NLP code debugging feature assist student in learning programming. The study used a case study research design in the approach of narrowing down from a broader perspective to a single unit and the technique of data collection is simple collection. Yamane's formula for sampling was put into use in this study for the random selection of the selected census. Resultantly, the use of descriptive statistics aimed at analyzing the collected data via questionnaires as the primary collection instrument. The instrument was tested and analyzed for validity and reliability using Cronbach's Coefficient Alpha formula. The use of ANOVA tested the suitability of this study. The SPSS Statistic Subscription version provided a statistical tool for analysis all through. From the research, the data collected showed that most students use natural language processing to learn programming and gain experience throughout the university learning time. The learning time from class was the major concern from the literature and the code debugging features were the main factor for students learning programming using natural language processing.

CHAPTER ONE: INTRODUCTION

This chapter describes the study's background, defines the research's problem, states the study's purpose, highlights the research's objectives and the questions the research sought to answer, and provides the scope of the study and the limitations likely to be faced.

1.1 Background to the Study

Computer programs are vital in the growth and development of technology and the process of it to be made is becoming much more important and very hard to master if one has not enough experience in the field. It is important to see that for one to succeed in programming one should have prior experience, this does not include studying mathematics as the basis of programming concepts (Jenkins et al, 2002). It is noted that one should have an algorithmic thinking capacity and a way to capture programming quests and solve problems easily, this should be done from a young age and the training does not major in mathematical problems but in advancing the logical thinking of the children (Khramova, 2019). From the modern decade and digitalization, there are no ingrained methods and techniques to train students due to the difficulty of understanding different people on the algorithms.

There is an increase in motivation for people to study programming at any level when there is a good visualization and simple challenges that require a person to figure out the solution on their own, this practice shows that the learning methods differ per person in each community and this leads to the new methods of using other software's to teach the concepts (Maloney, 2008). Due to the speedy advancement of technology, it is very viable for the programming community to gather thoughts and their powers paced on the algorithms made on how to increase the growth of the industry in general and give out the best for the future.

Artificial intelligence was once used in the 1950s when a computer generated its code and that was a step in the evolution of artificial intelligence to help in programming. For now, artificial intelligence can be used in multiple sectors such as education which is one of the most important features of development in the relationship between humans and computers.

1.2 Statement of Research Problem

The technology sector is growing at high speed and expectations are set every day, it is intended that NLP be a technology that helps many not only with programming but also with daily life activities. This research aims to foresee the use of NLP in a much broader perspective of programming. Every day we need more people who can invest in programming and its technologies thus, NLP helps to bring in a new and fresh way of learning programming which can increase the number of students and experts in this domain. The production of students in a natural way of teaching might not be acquitted to all of us considering people with disabilities and online classes which do not help much if a person is not used to that, but NLP tools give us the very best and most direct way of learning by the use of hands-on code model (Alam, 2021). The NLP model of study does not limit the student on what to do and ideas, it gives the student time to think and try again attempts which aren't over until the student is very satisfied with the result (Sakar, 2021).

1.3 Purpose of the Study

This study aims to investigate the potential that AI has to assist students in learning to program. The results of this study will provide valuable insights into the use of NLP to help students learn to program. It will contribute to building a society of great programmers by assisting the new programmers to write powerful and clear codes that are time-saving with the use of some NLP tools.

1.4 Conceptual Framework

The conceptual framework for this study explains how artificial intelligence can be used in helping students learn to program. This study can be described as a multi-disciplinary study that involves two different disciplines, education and programming. This study will get theories of computer science and education to help learners know in which way NLP can assist them in learning programming. This conceptual framework consists of two types of variables, the independent and dependent variables respectively. The independent variable is the variable that can be manipulated and influences the dependent variable. AI technology is inspired to be used in learning systems of higher education; it engages teachers and learners in deep learning and teaching processes (Mehrnaz et al., 2018). Programming requires understanding how certain logical flows and algorithms work and NLP can be of use to understand them easily (Mohammed, 2017). The assisted mode of programming with its search and reuse of codes is good according to the use and practical experience but has its complications and properties (Sarkar et al., 2022). The conceptual model can be simplified in the diagrammatic relationship below:

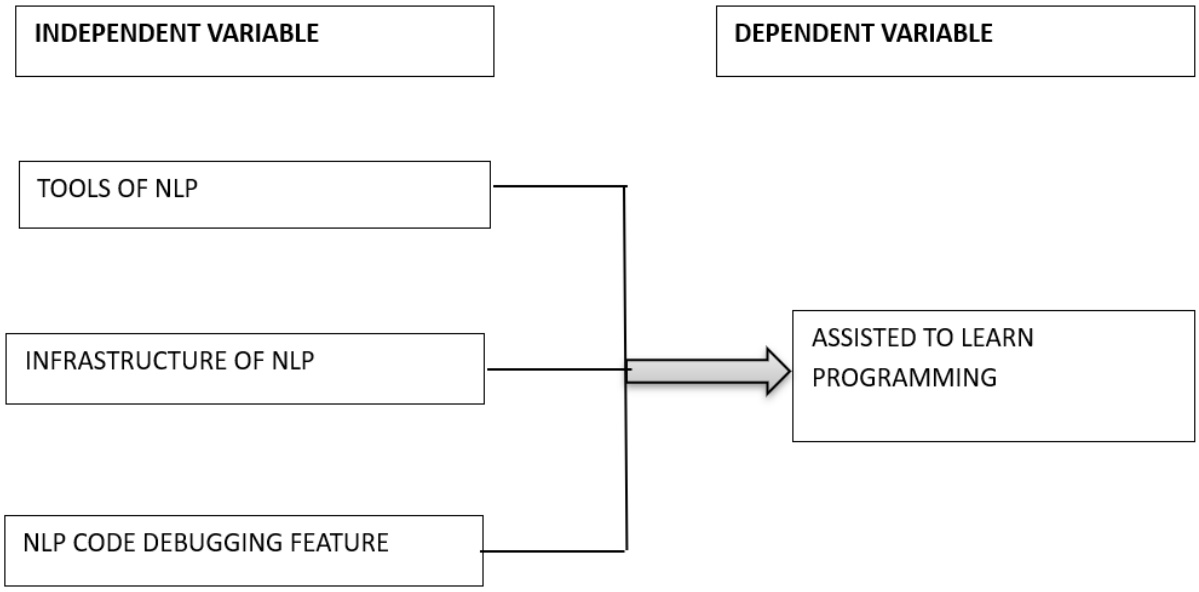


Figure 1: Conceptual framework

The conceptual framework for this study proposes that NLP tools can be used to assist students in tackling some problems in programming and have ease of study. This framework also shows the relationship between the dependent and independent variables. The study will be keen to analyze the effectiveness of these tools in the learning process for a programmer.

1.5 Research Questions

i. How do NLP tools help students to learn to program?

ii. What NLP infrastructures are needed to assist students in learning programming?

iii. How do NLP code debugging features help learn programming?

1.6 Objectives of the Study

1.6.1 General Objective

To evaluate the use of NLP in learning programming.

1.6.2 Specific Objectives

(I) To evaluate the NLP tools in assisting students to learn programming.

(I) To evaluate the NLP infrastructure that is needed to assist students in learning programming.

(II) To evaluate how NLP code debugging features assist students in learning programming.

1.7 Hypotheses of the Study

H01: NLP tools assist students in learning programming.

H02: The NLP infrastructure assists students in learning programming.

H03: The current tools of NLP help students to debug codes in a much more efficient way.

1.8 Significance of the Study

The significance of this study lies in its potential to contribute to developing more effective ways and methods for NLP to assist students in learning to program. The results of this study will provide valuable insights and give realization to students on the use of NLP tools in programming. This study will help new students and people who want to dive into the programming industry to have ease of study and master the languages as fast as possible. This study will also contribute to the wider field of computer science and artificial intelligence by providing a comprehensive evaluation of the performance of natural language processing tools for educational purposes. Overall, this study will have great significance in the programming industry related to the education of programming to university students and newcomers to the industry.

1.9 Scope of the Study

The scope of this study is limited to the use of natural language processing in assisting University students to learn to program. The study will focus on the use of two current IDEs to

write programming codes. The study will not include online or offline platforms that do not have any tools for natural language processing in their systems. The study will also focus on languages that are mostly used by students at Gretsa University which are JAVA, JavaScript, and Dart. The study will be limited to collecting and analyzing collected data only from students of Gretsa University.

1.10 Limitations of the Study

One of the limitations of this study is the potential for biased data sampling. The availability of a collection of data from honest students may also limit the accuracy of this study. The study is based on one University in Kenya that may not give out the general result of the whole programming community.

Additionally, the study will be limited to several editors that are mostly used by university students when learning to program. Note that technologies evolve and this means the programming industry grows and evolves each time and that leads to the irrelevance of this study after some time. These limitations should be considered when interpreting this study's results.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Programming is a complicated domain that requires much concentration and the ability to follow instructions and acquire as much experience as possible. An experienced programmer draws much experience and many skills that have no use in some processes of producing code of a certain application, when learning to program in university a student should be familiar with the use of computers and its basics to get an easy understanding of some of the programming concepts. If a student lacks aptitude it's hard for them to get what is being taught in programming even its basics, that's why mathematics is considered the basis of programming due to its ability to test the logical thinking of a student.

2.2 Learning programming

Students differ in how to do and understand what they are taught in class. It is very hard to make sure that every student understands the things taught in detail and can do them accordingly (Alam, 2022). Using NLP to solve this problem is a good move with great results but not replace teaching instructors because NLP is there to receive instructions and also the instructor helps guide the student on how to go through the materials. There are three stages for students to collaborate with NLP and these are to learn about NLP, learn from NLP, and learn with NLP (Kim, 2022).

Learning about the NLP system involves studying the NLP and training it to its specific specialization, this process gives data to the NLP and not otherwise, also the process of learning with AI involves both training the NLP and gaining the knowledge or using the particular data to solve a problem.

Learning through NLP systems has allowed instructors to increase the quality of what they teach and how they teach it, the curriculum can be modified by the use of simple instructions and materials can be customized to meet certain requirements of students individually (Alam, 2022).

2.3 Tools of learning programming

Several tools can be used to assist in learning programming, especially in the NLP field, this can be done by the use of gaming-based systems and learning platforms which gives the student a better way of understanding how to do programming in a fun way and without straining, there is also the coding direct way of learning to program which involves the use of the integrated development environment. Using the IDE can be fun and also gives the user a chance to figure out the next move, this is among the most efficient ways of learning programming as it involves the very use of the skills from the start and the learner gains much experience in doing so.

2.3 Infrastructure of NLP

The use of NLP in our daily activities grows in a much higher percentage and interests arise from every corner of other fields medicine, education, scientific and mathematics, and so many other fields, but there is low exposure to the functionality of NLP and the infrastructure needed to facilitate the better and best use of it. In developing these NLP systems let's not forget that computer science society also needs the use of this technology to make way for better systems and more students and to produce better experts (Ismail, 2019). There are multiple infrastructures needed in place to facilitate the use of NLP to students who study programming and these are characterized into two which are the personal and the overall, where the personal can be better

machines with good specifications and the overall can be the systems that facilitate the process which mostly are cloud-based (Perchik, 2022).

2.4 NLP code debugging feature

Coding is a process that gives out a lot of errors if not taken seriously and in a concentrated way. Learning how to debug code is very tricky and hard to manage if there is no help from the machine itself, NLP is termed as a processing system that can solve the solution by using the right algorithm and having the right traits and experience to solve the errors.

Solving errors in a large project is a human time-intensive task, that requires the focus of domain experts to develop and execute fact-finding experiments. The use of systems that can handle these tasks is suggested to increase productivity (Nigh, 2021).

2.1 Research gaps

Natural language processing is not a very new field but it has just gained popularity in recent years which makes it seem like a new field of study and research. Much research has been done concerning the NLP topic. Most important is the education of NLP and its significance to the lives of people and their daily activities. NLP has been a game-changing assistant to students especially those in programming, as it has helped them do more work in a very simple way as a click of a button to solve a problem by using NLP. As a computer science society, we have been developing NLP tools and the technology has grown much faster without thinking of the future generation of programmers who are being produced with the help of NLP. This study will look at the help that NLP assists students to learn to program.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Research Design

The proposed study used a case study research design. In this approach, the study was narrowed down to a single unit of an educational institution that has intellectuals and people concerned and affected directly by the study.

3.2 Study Area

The study was focused on the use of NLP to learn to program at Gretsa University. The study was conducted using questionnaires that will be provided per the targeted population and the sample size.

3.3 Target Population

The target population is people, events, and records that encompass the researcher's desired information and can respond to measurement questions (Blumberg, 2014). This study's target population was the 209 registered students of Gretsa University who are in the School of Computing and Informatics.

Table 1: Computing students

Computing Students	Number of Students
Degree government-sponsored	133
Self-sponsored students	11
Diploma students	62

Certificate students	3
Total	209

3.4 Sampling Techniques

The study used a random sampling technique to select students from the School of Computing. The sampling will consider students who are interested in and know to program.

3.5 Sample Size

The sample size was determined by the use of Yamane's formula for calculating the sample size of a research study, which is:

Equation 1 Yamane's formula

$$n = \frac{N}{1+N(e)}2$$

Where by:

n = The sample size

N= The population of study

e = the margin of error in the calculation using the +/-5%

n=209/ (1+(209*0.05*0.05)) =137.27

rounding off to 137

3.6 Measurement of Variables

The study used computing students as the means of acquiring its data. This variable is characterized as a nominal variable as it can be counted and measured.

3.7 Research Instruments

The use of quantitative and subjective data collection methods were used in this study. The semi-structured questionnaires to collect the quantitative and qualitative data have advantages over other types of instruments because of how cheap, do not require much exertion to prepare, and often have standardized answers that ease the process of data compilation. The questionnaires were distributed to the students in the School of Computing and Informatics with a request to complete the questionnaire and that the instrument consider all ethical issues and try to capture the sensitive and appropriate data needed for the research.

3.8 Validity of Measurements

To ensure the validity of the measurement in this study the instruments did not capture personal data and dived deep into the topic without rating anyone involved per their answers or knowledge.

3.9 Reliability of Measurements

The study involved the use of Cronbach's coefficient alpha to measure the reliability and consistency of the study. Cronbach's coefficient alpha is a mathematical formula of

$$C\alpha = \frac{N.\bar{c}}{\bar{v}.(N-1)\cdot\bar{c}}$$

Equation 2 Cronbach's Coefficient Alpha

Where by:

N= Number of the items

$\bar{c}$ = average co-variance between item-pairs

$\bar{v}$ = average variance

$C\alpha$ = Cronbach's coefficient alpha

Table 2 Reliability statistics

Reliability Statistics

Cronbach's Alpha	N of Items
.845	5

3.10 Data Collection Techniques

The study used the primary data collection form which was the questionnaire to make it a simple data collection method.

3.11 Data Analysis

The data collected from this study will be qualitative and there will be a combination of content analysis methods whereby the information can be gathered from text and images and the scrutiny-based technique which uses keywords to analyze the data collected.

3.12 Logistical and Ethical Considerations

The study will adhere to ethical standards for research involving human subjects, including obtaining informed consent from study participants and protecting their privacy. Additionally, the study will comply with all relevant laws and regulations governing research, including data privacy and intellectual property rights. All data collected will be stored securely and destroyed after the study is completed. The study will also adhere to ethical guidelines and rules and gain

authorization from the Gretsa University research committee from Gretsa University as the study is conducted there.

CHAPTER FOUR: FINDINGS AND DISCUSSION

4.1 Introduction

This chapter provides a detailed summary of findings and results therein that were obtained from research done using the survey tool described in the previous chapter. A detailed data analysis obtained in this study is also well documented.

4.2 Preliminary Study

4.2.1 Sample distribution

The study was conducted on a sample of Gretsa University's School of Computing and Informatics students. Online and offline questionnaires were administered to 137 students of all educational levels which are the ones pursuing a bachelor in computer science, diploma in information technology, diploma in computer science, and certificate in information technology which were randomly selected for this study.

4.2.1 Response Rate

The response rate of the research acts as an elementary parameter in evaluating the efforts of the collection of data (Fowler, 2014). This means that the number of people who completed the survey is divided by the number of those eligible who were sampled, including those who did not respond or were unavailable. A census was conducted to the 137 respondents hence the 100% response rate in this study.

4.3 Study on students

4.3.1 Education level

A full study was conducted on the sample of 137 students of computing and informatics which represented different years of study and the courses they pursued with their respective level of education. The students Pursuing a degree (59.9%), the students pursuing a diploma (39.4%), and the students pursuing a certificate (0.7%) according to the educational pyramid which shows

the lower to the highest level of education shows that there are many students in bachelor in degree in Gretsa university which contribute much in this study.

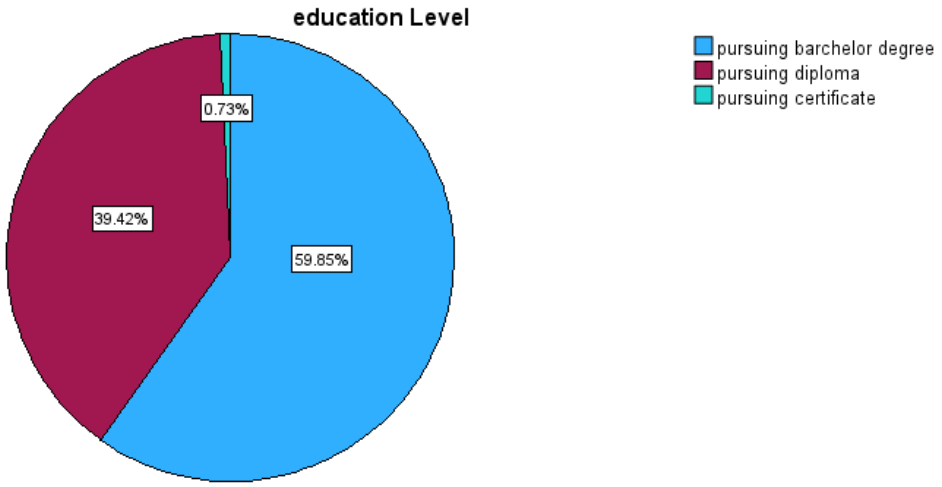


Figure 2 Education level

4.3.2 Gender of Respondents

Census results indicated in Figure 4 show 108 (78.8%) of the respondents were men whereas the remaining 29 (21.2%) were women. These results regarding gender shows the dominance of men in the technology sector especially in the field of IT. The number of women in the ICT field seems to have increased from 20% and below which was said to be the percentage of women in the ICT profession, thus women are still considered as the minority in this sector (Castano, 2011).

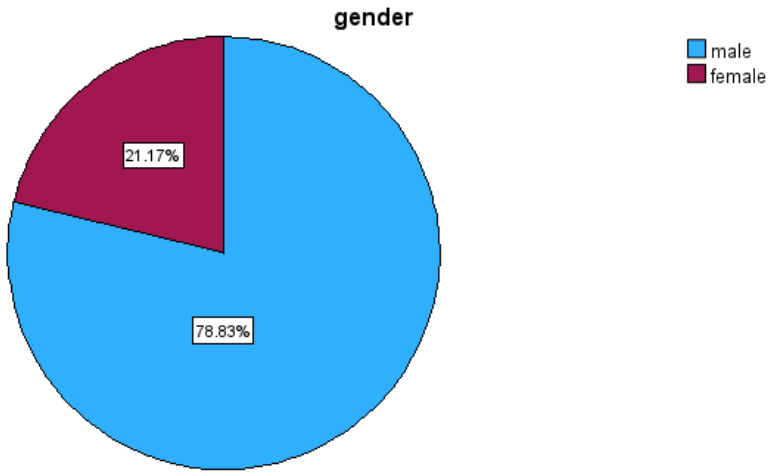


Figure 3 Gender of respondents

4.3.3 Respondents Age Bracket

The findings showed that (62%) which is the largest portion of the Computing and informatics Students respond between ages 17 years and 22 years. 32.8% responded to ages 23 to 25, Findings show that there was only 5.1% which is for the ages of 26 years and above, this implies that there is a younger group of ICT students in universities and leads to the greater, younger, and brighter force in the future of the ICT field. The advancement of technology and the use of simple-handled tools that are fast is one of the factors in this young age percentage (Dúo-Terrón, 2022).

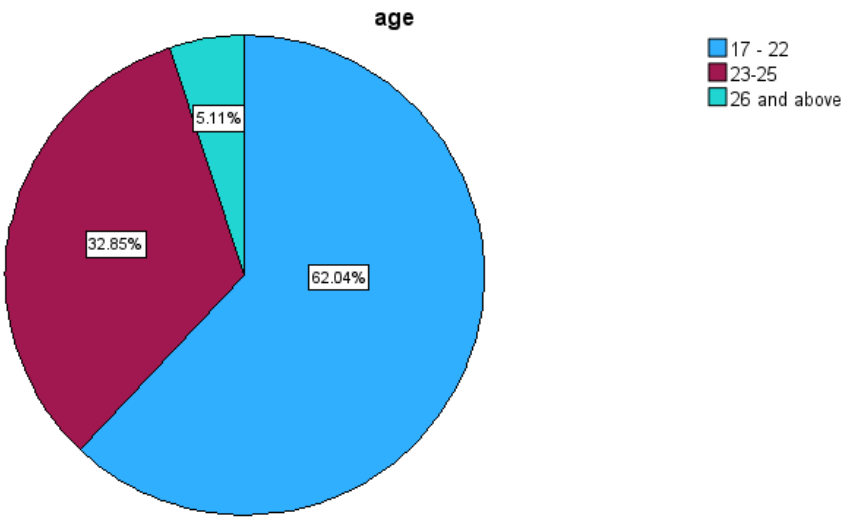


Figure 4 Respondents age bracket

4.3.4 Respondents Year of Study

The data showed that a sizeable number of students in the computing and informatics school are in the second year of their studies at Gretsa University, this means that they already know the basics of computer science and have knowledge of programming. This data is collected from all participants of the survey without considering their level of study.

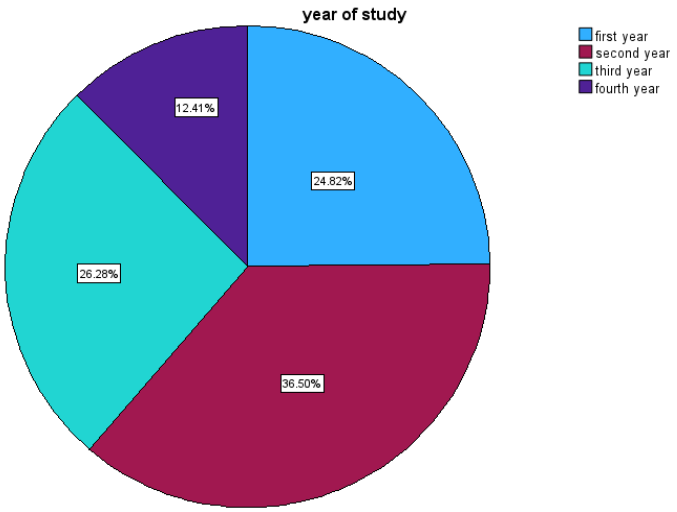


Figure 5 Respondents year of study

4.2.5 Status of students

A large number of students (94.9%) are shown to be full-time students and they have a lot of class time and an instructor to instruct and learn from. The remaining percentage showed the number of distance-learning students who study at they are own pace.

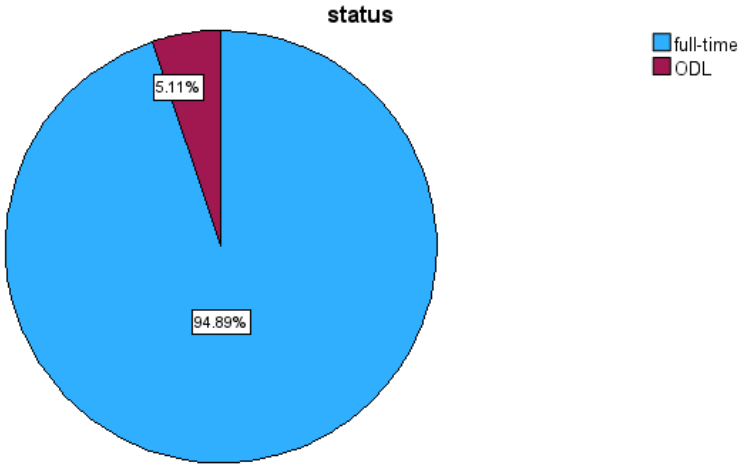


Figure 6 Status of students

4.4 Study of Variables

4.4.1 Findings on Tools of NLP

The average rating of 4.41 for tool availability suggests that respondents generally agreed that NLP tools are accessible, with a majority (54.0%) strongly agreeing. Similarly, the mean score of 4.55 for tool operation indicates that respondents found NLP tools easy to use, with a significant majority (65.7%) strongly agreeing. The mean score of 4.39 for tool integration suggests that respondents agreed that NLP tools can be effectively combined, with a notable proportion (59.1%) strongly agreeing. With a mean score of 4.66 for tool satisfaction, respondents expressed high levels of contentment with NLP tools, with the majority (65.7%) strongly agreeing. Furthermore, the mean score of 4.55 for tool download and installation ease indicates that respondents found it convenient to acquire and set of NLP tools, with a significant majority (58.4%) strongly agreeing, while 38.0% remained neutral.

The table below provides the mean scores for different aspects related to the tools of Natural Language Processing (NLP) and the percentage distribution of responses across different levels of agreement.

Table 3 Tools of NLP results

	MEAN	Tools of NLP in Percentage (%)				
		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Availability of the tools	4.41	0.0	0.0	13.1	32.8	54.0
Operation of tools	4.55	0.0	0.0	10.2	24.1	65.7
Integration of tools	4.39	0.0	0.0	20.4	20.4	59.1
Satisfaction with the tools	4.66	0.0	0.0	0.0	34.3	65.7

Easy to download and install	4.55	0.0	38.0	3.6	0.0	58.4
Overall Mean	4.51	Variance	0.4328	Standard Deviation		0.6476

Overall, the data suggests that respondents perceive NLP tools to be readily available, easy to operate, and integrate, and they express high satisfaction with the tools. The aspect of downloading and installing the tools is also positively perceived.

Table 4 Tools of NLP Descriptive Statistics

Descriptive Statistics

	N	Range	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Availability of the tools	137	2	4.41	.061	.713	.508
Operation of tools	137	2	4.55	.058	.674	.455
Integration of tools	137	2	4.39	.069	.807	.651
Satisfaction with the tools	137	1	4.66	.041	.476	.227
Easy to download and install	137	2	4.55	.049	.568	.323
Valid N (listwise)	137					

4.4.2 Findings on Infrastructure of NLP

The average rating of 4.59 for well-established infrastructure suggests that respondents generally strongly agreed that there is a well-established infrastructure in NLP, with the majority (59.1%) strongly agreeing. Similarly, the mean score of 4.62 for affordability of infrastructure indicates that respondents agreed that the infrastructure in NLP is affordable, with the majority (62.0%) strongly agreeing. The mean score of 4.50 for the availability of infrastructure suggests that respondents have a neutral stance regarding the availability of infrastructure in NLP, but a significant proportion (53.3%) still agreed with this statement. With a mean score of 4.77 for infrastructure requirements, respondents strongly agreed that the infrastructure requirements in

NLP are being met, with the vast majority (83.9%) strongly agreeing. Furthermore, the mean score of 4.50 for students' consideration indicates that respondents held a neutral stance regarding students' considerations in NLP infrastructure, but a significant proportion (53.3%) still agreed with this statement.

Table 5 Infrastructure of NLP results

	MEAN	Infrastructure of NLP				
		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Well-established infrastructure	4.59	0.0	0.0	0.0	40.9	59.1
Affordability of infrastructure	4.62	0.0	0.0	0.0	38.0	62.0
Availability of infrastructure	4.50	0.0	0.0	3.6	43.1	53.3
Infrastructure requirements	4.77	0.7	0.7	3.6	10.9	83.9
Students' consideration	4.50	0.0	0.0	3.6	43.1	53.3
Overall Mean	4.59	Variance	0.3032	Standard Deviation		0.5482

The overall mean score for infrastructure in NLP is 4.59, reflecting a generally positive perception of the infrastructure in this field. The variance of 0.3032 and the standard deviation of 0.5482 indicate moderate variability in responses for the infrastructure aspects.

Table 6 Descriptive statistics table

Descriptive Statistics

	N	Range	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
well-established infrastructure	137	1	4.59	.042	.493	.243
Affordability of infrastructure	137	1	4.62	.042	.487	.237
Availability of infrastructure	137	2	4.50	.049	.570	.325
Availability of infrastructure	137	4	4.77	.053	.621	.386
Students' consideration	137	2	4.50	.049	.570	.325
Valid N (listwise)	137					

4.4.3 Findings on NLP Code Debugging Features

The average rating of 4.69 for "eases the debugging process" suggests that respondents strongly agreed that NLP code debugging features make the process of debugging easier. The majority (72.3%) strongly agreed with this statement. Similarly, the mean score of 4.50 for "codes understanding" indicates that respondents agreed that NLP code debugging features help in understanding codes. A significant proportion (53.3%) strongly agreed with this statement. The mean score of 4.50 for "gaining experience" indicates that respondents agreed that NLP code debugging features assist in gaining experience. A notable proportion (53.3%) strongly agreed with this statement. Lastly, the mean score of 4.50 for "fast and fun" suggests that respondents agreed that NLP code debugging features provide a fast and enjoyable experience. A significant proportion (53.3%) strongly agreed with this statement. The table below presents the mean scores for various aspects of NLP code debugging features, along with the percentage distribution of responses across different levels of agreement.

Table 7 NLP code debugging features results

	MEAN	NLP Code Debugging Features				
		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Eases the debugging process	4.69	0.0	0.0	3.6	24.1	72.3
Codes understandings	4.50	0.0	0.0	3.6	43.1	53.3
Gaining Experience	4.50	0.0	0.0	3.6	43.1	53.3
Fast and fun	4.50	0.0	0.0	3.6	43.1	53.3
Overall Mean	4.54	Variance	0.31625	Standard Deviation		0.56225

The overall mean score for NLP code debugging features is 4.54, reflecting a generally positive perception of these features. The variance of 0.31625 and the standard deviation of 0.56225 indicate moderate response variability for the different aspects of NLP code debugging features.

Table 8 Descriptive statistics table

Descriptive Statistics

	N	Range	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Eases the debugging process	137	2	4.69	.046	.539	.290
Codes understandings	137	2	4.50	.049	.570	.325
Gaining Experience	137	2	4.50	.049	.570	.325
Fast and fun	137	2	4.50	.049	.570	.325
Valid N (listwise)	137					

4.5 Statistical Modelling

4.5.1 Statistical Modelling

A linear regression analysis was undertaken to determine the use of NLP to assist university students in learning programming. The linear regression method models the relationship between the scalar variable denoted as y and another or extra variables denoted x. IBM SPSS Statistics Subscription program was used as a tool for this analysis. A scatter plot was generated for each variable to highlight the kind of relationship that exists between the dependent and every independent variable.

4.5.1.1 Scatter Plot of NLP tools/ Learning programming

A visual examination of the scatter plot for NLP tools suggests that there is a positive relationship with assisting in learning programming, however not that strong.

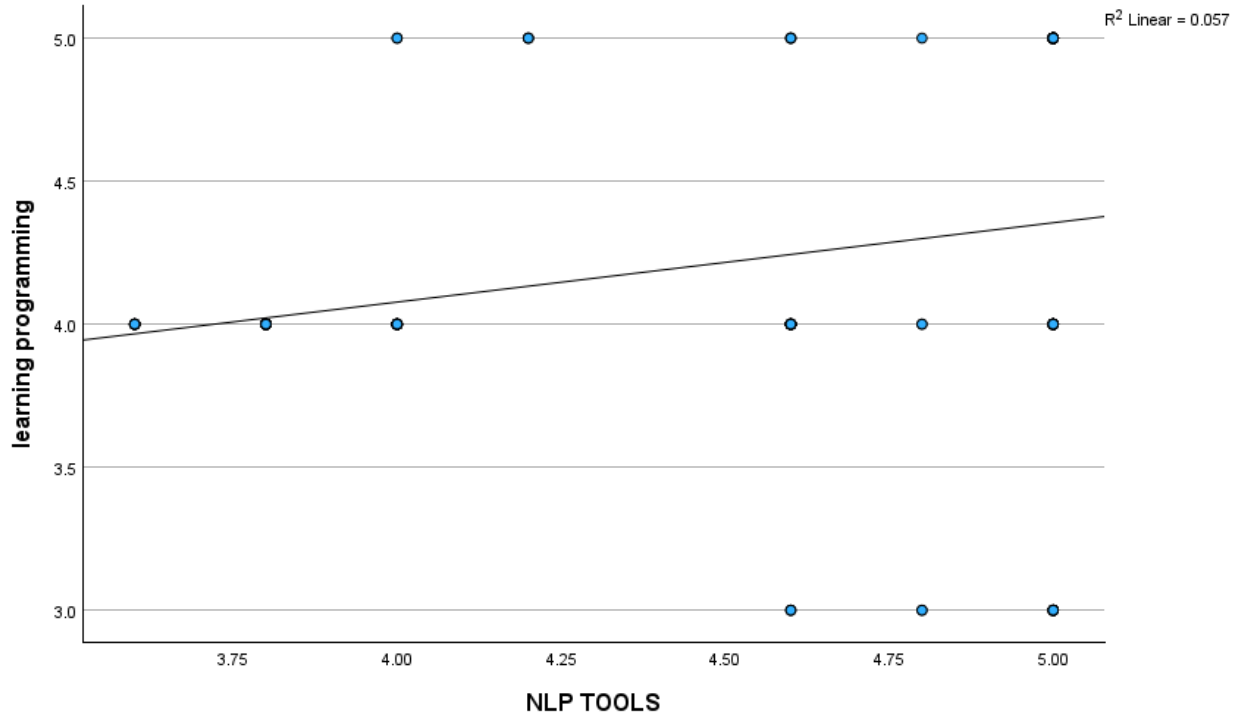


Figure 7 Scatter plot of tools/ learning programming

4.5.1.2 Scatter Plot of NLP infrastructure/ Learning programming

A visual examination of the scatter plot for NLP infrastructure suggests that there is a positive relationship with assisting in learning programming, however not that strong.

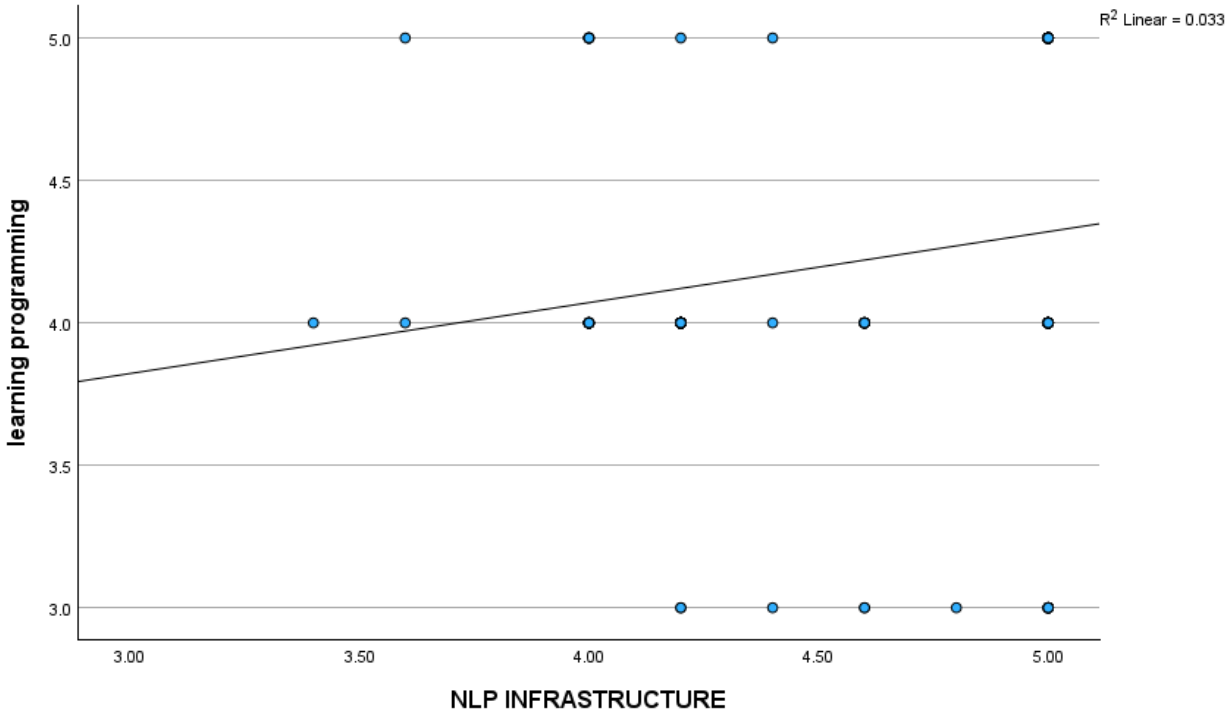


Figure 8 Scatter plot of NLP infrastructure/ Learning programming

4.5.1.3 Scatter Plot of NLP code debugging feature/ Learning programming

A visual examination of the scatter plot for NLP infrastructure suggests that there is a positive relationship with assisting in learning programming, however not that strong.

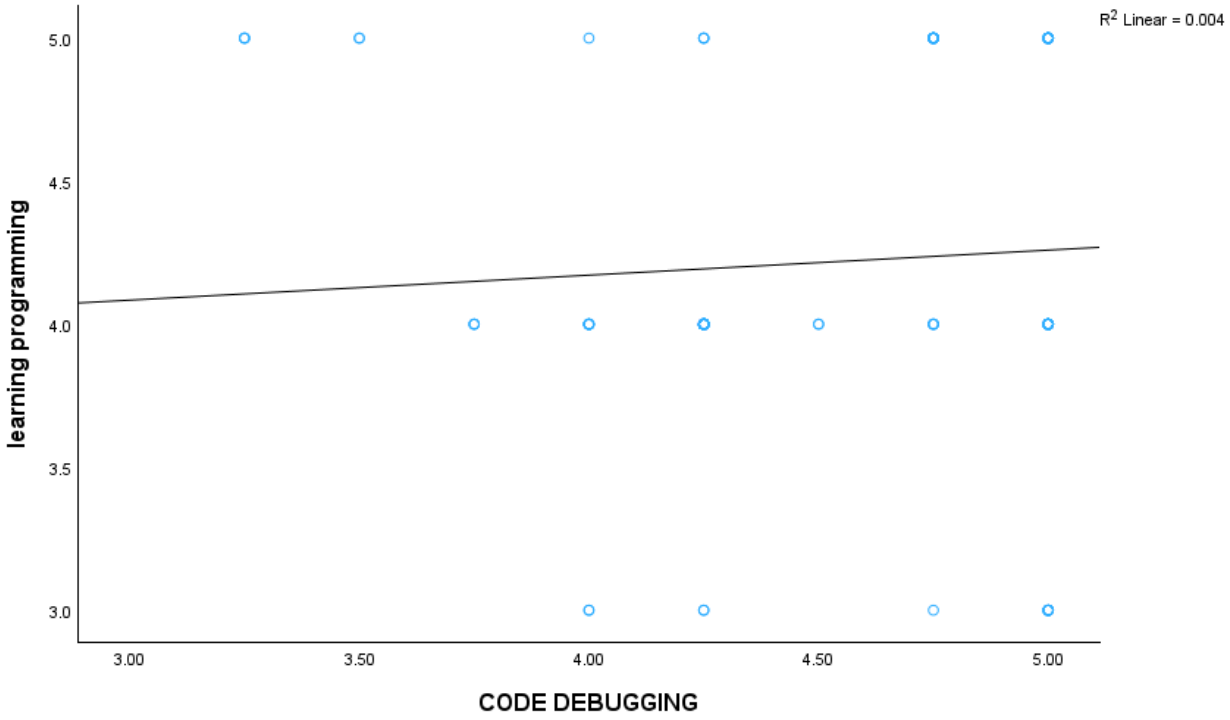


Figure 9 Scatter plot of code debugging/ Learning programming

4.6 Model Summary

Below are summary results from the model

Table 9 Model summary table

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.289[a]	.084	.063	.584

a. Predictors: (Constant), CODE DEBUGGING FEATURES, NLP TOOLS, NLP INFRASTRUCTURE

The correlation coefficient (R) of 0.289 suggests a weak positive relationship between the predictors and the NLP used to assist university students in learning programming. The

coefficient of determination (R Square) value of 0.084 indicates that approximately 8.4% of the variance in NLP to assist university students in learning programming can be explained by the predictors. However, the adjusted R Square value of 0.063, which considers the number of predictors, indicates that the model's explanatory power is limited.

An analysis of Variance (ANOVA) in this study was used to examine the suitability of the developed model for this study. The significance value of 0.09, which is greater than 0.05 hence the NLP learning of programming assessment model shows there is not enough statistical significance in predicting the use of NLP in assisting university students in learning programming.

Table 10 ANOVA

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.129	3	1.376	4.040	.009[b]
	Residual	45.302	133	.341		
	Total	49.431	136			

a. Dependent Variable: NLP assists in learning programming

b. Predictors: (Constant), CODE DEBUGGING FEATURES, NLP TOOLS, NLP INFRA-STRUCTURE

The regression model demonstrates a notable impact on the variability in the dependent variable, as evident from the regression sum of squares. The F-statistic of 4.040 and the p-value (Sig.) of .009 indicate a significant relationship between the predictors and NLP assist in learning programming. The residual sum of squares reflects the unexplained variability in the dependent variable, while the total sum of squares encompasses the overall variability.

Table 11 Coefficients table

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.227	.535		6.030	<.001
	NLP TOOLS	.235	.167	.201	1.408	.161
	NLP INFRASTRUCTURE	.437	.307	.320	1.423	.157
	CODE DEBUGGING FEATURES	-.457	.232	-.343	-1.971	.051

a. Dependent Variable: learning programming

Findings of the analyzed data infer; that keeping the rest of the independent variables at less than zero mark, an upsurge to NLP infrastructure with one unit will result in a .437 increase in learning of

programming using NLP.

The findings show the most significant factor in learning programming using natural language processing is the code debugging feature with 0.051, whereas NLP infrastructure was at 0.157 and the NLP tools were at 0.16.

4.7 Discussion of findings

The majority of the respondents, as revealed by the study are confidently in agreement that code debugging features have the most impact in the learning of programming using NLP, with 72.3% indicating that Learning programming using NLP eases the debugging process, and also the study finds out the majority of students 82% agreed on the availability of NLP tools and also the affordability of infrastructure.

4.7.1 Hypothesis Testing

The testing of the hypothesis was done from the data that was collected from the students and the analysis was done using the SPSS Subscription Program.

The test of the NLP tools assist students in learning programming was done through the use of the coefficient table it shows that the hypothesis is not acceptable since the t value is greater than 1. In this NLP infrastructure to assist students in learning programming hypothesis the statistical data showed that it is not acceptable since the t value from the coefficients table is 1.423 which means it is greater than 1. The current tools of NLP help students to debug codes in a much more efficient this hypothesis was also tested with the use of the coefficients table and showed that the value is -1.971 Which is out of the range of -1 and +1 means the hypothesis is not acceptable.

CHAPTER FIVE: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1 Introduction

The researcher in this study pursued to assess the use of natural language processing to assist university students in learning programming in Kenya through a survey of Gretsa University. The study looked at the tools of natural language programming, natural language programming infrastructure, and natural language programming code debugging features. This chapter gives a summary of the data collected, each statistical handling of analysis through discussing the specific study objectives, interpreting and evaluating results. The conclusions observed here are point-to-point related to the specific study objectives. The chapter's recommendations discuss further studies to be considered or a proposal to change the conclusion observed.

5.2 Summary

Empirical literature revealed the ability of students to learn programming using natural language processing. The literature further showed the ability of students to capture programming knowledge easily using natural language processing, The learning process that is currently used which is the school-based curriculum to educate students does not cover all students and their way of learning. The literature also discussed the infrastructure used by natural language processing together with the tools

In this study there were three main objectives; to evaluate the natural language processing tools in assisting students to learn programming, to evaluate the natural language processing infrastructure that is needed to assist students in learning programming, and finally to evaluate the use of natural language processing code debugging feature assist student in a university in learning programming.

Yamane's formula for sampling was put into use in this study for the random selection of the selected census. Resultantly, the use of descriptive statistics aimed at analyzing the collected data via questionnaires as the primary collection instrument. The instrument was tested and analyzed for validity and reliability using Cronbach's Coefficient Alpha formula. The use of ANOVA tested the suitability of this study. The SPSS Statistic Subscription version provided a statistical tool for analysis all through.

From the research, the data collected showed that most students use natural language processing to learn programming and gain experience throughout the university learning time. The learning time from class was the major concern from the literature and the code debugging features were the main factor for students learning programming using natural language processing

5.3 Conclusions

Guided by the statistical data provided in the previous chapter 72% of the students agreed on the easiness of code debugging features in the use of natural language processing. This percentage shows how code debugging is important in the journey of learning to program, the tools of programming were rated to be good by all of the students with just 38 % of students disagreeing on the easiness of downloading and installing the tools. From the data it showed that the infrastructure is well established and does not require a lot of requirements from users this is also shown from the previous studies.

This study does not cover the entire universe of natural language processing to assist university students in learning programming, since the study sample was taken for consideration, was only that for computing students at Gretsa University. The study can nonetheless be used to provide

insight into the use of natural language programming for students in university learning programming.

5.4 Recommendations for Policy or Practice

The study justifies that the large number of students who are into programming use natural language processing to gain coding knowledge and learn fast compared to the school-taught curriculum. Specifically, the study recommends that university students keep on using the natural language processing way of learning to program but should not leave the school-based curriculum.

Lectures should adopt and teach students how to use natural language processing ethically and lawfully.

5.5 Recommendations for Further Research

Due to the constraints highlighted in the first chapter; exploring the use of natural language processing to learn programming, this study could not exhaust all the parameters needed to assess the use of natural language processing to learn programming for university students. This study could not capture the total perception of the students about using natural language processing to learn programming. Other factors like the mode of study, ethical guidance, and experience aspect, just to mention a few, require further investigation on learning programming

REFERENCES

A. Alam, "Should Robots Replace Teachers? Mobilization of AI and Learning Analytics in Education," 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3), Mumbai, India, 2021, pp. 1-12, doi: 10.1109/ICAC353642.2021.9697300.

Alexander, T., Stefanova, V., & Zahidi, S. (2018). The Future of Jobs Report 2018. Cologny/Geneva (Switzerland): World Economic Forum.

Blumberg, B., Cooper, D., & Schindler, P. (2014). *EBOOK: Business research methods*. McGraw Hill.

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. *Industrial Psychiatry Journal*, *19*(1), 60.

CastaÃ±o, C., & Webster, J. (2011). Understanding women's presence in ICT: The life course perspective. *International Journal of Gender, Science, and Technology*, *3*(2), 364-386.

Dúo-Terrón, P., Moreno-Guerrero, A. J., & Marín-Marín, J. A. (2022). ICT motivation in sixth-grade students in pandemic times—the influence of gender and age. *Education Sciences*, *12*(3), 183.

Jr Fowler, F. J., & Fowler Jr, F. J. (2014). *Survey research methods*.

Barenkamp, M., Rebstadt, J. & Thomas, O. Applications of AI in classical software engineering. *AI Perspect* **2**, 1 (2020). https://doi.org/10.1186/s42467-020-00005-4

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069-6104.

Khramova, M. V., Nesterov, M. V., & Kurkin, S. A. (2019, September). Problems of learning programming in the introductory course. In *2019 International Conference" Quality Management, Transport, and Information Security, Information Technologies"(IT&QM&IS)* (pp. 522-525). IEEE.

King, T. M., Arbon, J., Santiago, D., Adamo, D., Chin, W., & Shanmugam, R. (2019, April). AI for testing today and tomorrow: industry perspectives. In *2019 IEEE International Conference On Artificial Intelligence Testing (AITest)* (pp. 81-88). IEEE.

M. Ismail and A. Ade-Ibijola, "Lecturer's Apprentice: A Chatbot for Assisting Novice Programmers," 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Vanderbijlpark, South Africa, 2019, pp. 1-8, doi: 10.1109/IMITEC45504.2019.9015857.

Mäkelä, M. (2019). Utilizing artificial intelligence in software testing.

Maloney, J. H., Peppler, K., Kafai, Y., Resnick, M., & Rusk, N. (2008, March). Programming by choice: urban youth learning programming with scratch. In *Proceedings of the 39th SIGCSE technical symposium on Computer Science education* (pp. 367-371).

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

Perchik, J. D., Smith, A. D., Elkassem, A. A., Park, J. M., Rothenberg, S. A., Tanwar, M., ... & Sotoudeh, H. (2022). Artificial intelligence literacy: developing a multi-institutional infrastructure for AI education. *Academic radiology*.

Sarkar, A., Gordon, A. D., Negreanu, C., Poelitz, C., Raghavan, S. S., & Zorn, B. (2022). What is it like to program with artificial intelligence? *ArXiv*. https://doi.org/10.48550/arXiv.2208.06213

Talwar, R., & Hancock, T. (2010). The shape of jobs to come. Possible new careers emerge from science and technology advances (2010-2030). *Final report, January*.

Jenkins, T. (2002, August). On the difficulty of learning to program. In *Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences* (Vol. 4, No. 2002, pp. 53-58).

Kelan, E. K. (2007, November). 'I don't know why'—Accounting for the scarcity of women in ICT work. In *Women's Studies International Forum* (Vol. 30, No. 6, pp. 499-511). Pergamon.

Hourani, H., Hammad, A., & Lafi, M. (2019, April). The impact of artificial intelligence on software testing. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 565-570). IEEE.

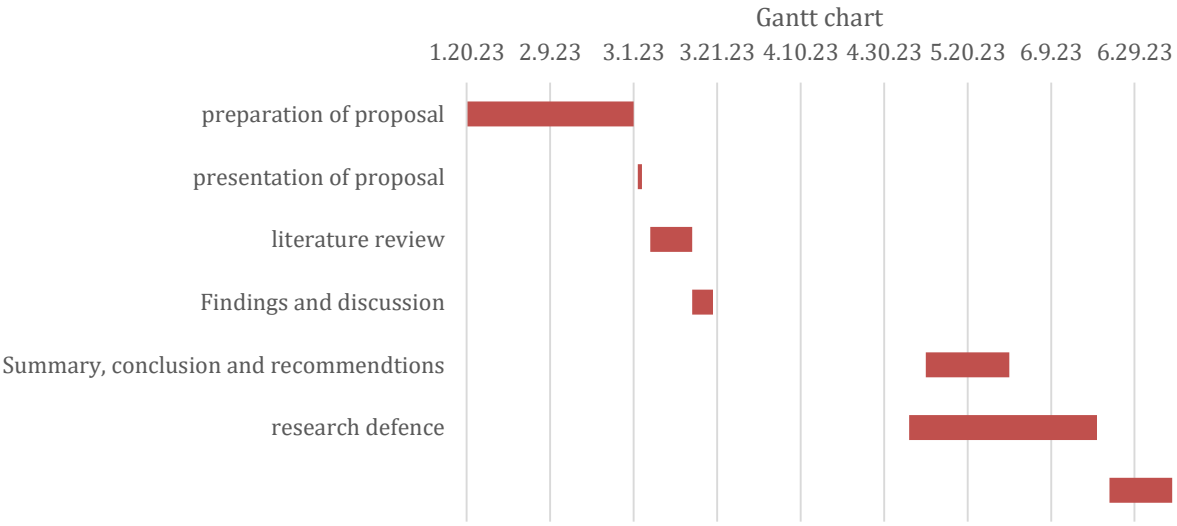
Jaakkola, H., Henno, J., Mäkelä, J., & Thalheim, B. (2019, May). Artificial intelligence yesterday, today, and tomorrow. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 860-867). IEEE.

Nigh, C., Bhargava, G., & Blanton, R. D. (2021, October). Aaa: Automated, on-ate ai debug of scan chain failures. In *2021 IEEE International Test Conference (ITC)* (pp. 314-318). IEEE.

Alam, A. (2022, April). A digital game based learning approach for effective curriculum transaction for teaching-learning of artificial intelligence and machine learning. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 69-74). IEEE.

APPENDICES

4.1 Project plan



	research defence	Summary, conclusion and recommendtions	Findings and discussion	literature review	presentation of proposal	preparation of proposal
start date	23.06.2023	06.05.2023	10.05.2023	15.03.2023	05.03.2023	02.03.2023	20.01.2023
duration	15	45	20	5	10	1	40

Figure 10: Project plan

4.2 Project budget

Budget items	Figure allocated
Modem and internet connection	2000Ksh
Printing of proposal	1000Ksh
Questionnaire printing	600Ksh
Printing	3000Ksh
Laptop	120000Ksh
Miscellaneous	700Ksh
Total	127,300Ksh

Table 12: Project budget

4.3 Research questions

Please take a few minutes to complete this survey. This survey aims to determine the use of Natural Language Processing to assist university students in learning programming. It would be of great value if you could share your wealth of knowledge by completing this questionnaire. The findings of this research are solely for academic purposes; hence, your responses will be handled with the highest anonymity and confidentiality.

SECTION A: RESPONDENTS DEMOGRAPHIC DETAILS

1. Gender
   [ ] Male
   [ ] Female

2. Education level
   [ ] Pursuing Bachelor's degree
   [ ] Pursuing Diploma
   [ ] Pursuing Certificate

3. Year of study
   [ ] First-year
   [ ] Second-year
   [ ] Third-year
   [ ] Fourth-year

4. Status
   [ ] Fulltime
   [ ] ODL

5. Age
   [ ] 18-22
   [ ] 23-25
   [ ] 26 and above

SECTION B: TOOLS OF NLP		SD	D	N	A	SA
i.	There are NLP tools specifically designed to learn to program					

ii.	It is easier to operate through NLP tools that are available					
iii.	The NLP tools integrated into IDE ease the learning process in a far better way					
iv.	Do the NLP tools give out the satisfaction needed					
v.	The NLP tools are easier to download and install					

SECTION C: INFRASTRUCTURE OF NLP		SD	D	N	A	SA
i.	There's a well-established infrastructure to assist in learning programming using NLP					
ii.	The infrastructure involved in the process of learning to program using NLP is affordable					
iii.	The infrastructure involved in learning to program using NLP is easily available					
iv.	The infrastructure used does not need many physical requirements					
v.	The infrastructure available considers all people with disability					

SECTION D: NLP CODE DEBUGGING FEATURES		SD	D	N	A	SA
i.	The use of NLP code debugging features eases the debugging process					
ii.	NLP gives users the ability to understand the code easily					
iii.	More programming experience is earned through learning using NLP					
iv.	It is fast and fun to learn to program using NLP					