Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations, Fall 2023 to Present

Graduate Studies

5-2024

Generative AI in Education From the Perspective of Students, Educators, and Administrators

Aashish Ghimire
Utah State University, aashish.ghimire@usu.edu

Follow this and additional works at: https://digitalcommons.usu.edu/etd2023

Part of the Computer Sciences Commons

Recommended Citation

Ghimire, Aashish, "Generative AI in Education From the Perspective of Students, Educators, and Administrators" (2024). *All Graduate Theses and Dissertations, Fall 2023 to Present*. 124.
https://digitalcommons.usu.edu/etd2023/124

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations, Fall 2023 to Present by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

UtahStateUniversity
MERRILL-CAZIER LIBRARY

GENERATIVE AI IN EDUCATION FROM THE PERSPECTIVE OF STUDENTS, EDUCATORS, AND ADMINISTRATORS

by

Aashish Ghimire

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

John Edwards, Ph.D. Major Professor	Soukaina Filali Boubrahimi, Ph.D. Committee Member
Shuhan Yuan, Ph.D. Committee Member	Steve Petruzza, Ph.D. Committee Member
Kevin Moon, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

Copyright © Aashish Ghimire 2024

All Rights Reserved

ABSTRACT

Generative AI in Education from the Perspective of Students, Educators, and Administrators

by

Aashish Ghimire, Doctor of Philosophy
Utah State University, 2024

Major Professor: John Edwards, Ph.D.
Department: Computer Science

This dissertation delves into the integration of generative artificial intelligence (AI) in educational settings, examining its potential to revolutionize teaching and learning processes across various disciplines. Through a series of studies, the research addresses critical aspects of AI in education, including the effectiveness of natural language processing (NLP) for legal text summarization, educators' perceptions and attitudes towards AI tools, the existing policy landscape for AI use in educational institutions, the impact of AI on student engagement and learning outcomes in foundational programming courses, and the factors influencing educators' acceptance and adaptation of AI technologies. This dissertation is composed of five distinct investigations, each exploring a different facet of generative artificial intelligence (AI) application in educational environments. These studies include: the utilization of AI for the summarization of legal court opinions, exploring educators' perceptions and attitudes towards generative AI tools within educational settings, examining administrators' views on the policy frameworks governing generative AI in education, assessing students' experiences and outcomes when using generative AI in introductory programming courses, and evaluating the adaptation of generative AI in classrooms through the lenses of the Technology Acceptance Model (TAM) and the Innovation Diffusion Theory (IDT). The findings highlight the transformative potential of AI in enhancing access to information, streamlining educational processes, and fostering pedagogical innovation. However, they also underscore the challenges of ensuring equitable access to AI tools, safeguarding data privacy, and maintaining academic integrity. This dissertation contributes to the broader discourse on the role of AI in education by offering evidence-based recommendations for policymakers, educators, and institutions to navigate the complexities of AI integration. It calls for ongoing collaboration and research to develop strategies that leverage AI's capabilities while addressing ethical and pedagogical concerns, ultimately aiming to enrich the educational experience and prepare students for a rapidly evolving technological landscape.

(131 pages)

PUBLIC ABSTRACT

Generative AI in Education from the Perspective of Students, Educators, and Administrators

Aashish Ghimire

This research explores how advanced artificial intelligence (AI), like the technology that powers tools such as ChatGPT, is changing the way we teach and learn in schools and universities. Imagine AI helping to summarize thick legal documents into something you can read over a coffee break or helping students learn how to code by offering personalized guidance. We looked into how teachers feel about using these AI tools in their classrooms, what kind of rules schools have about them, and how they can make learning programming easier for students. We found that most teachers are excited about the possibilities but also a bit cautious because they want to make sure these tools are used fairly and safely. There's also a lot that schools need to figure out in terms of setting up the right rules to make the best use of AI. Our study suggests that if we can address these challenges, AI could make education more engaging, accessible, and effective for everyone. It's a call to educators, policymakers, and tech developers to work together to ensure AI tools are used in ways that benefit all students and help prepare them for a future where technology plays an even bigger role in our lives.

"For my family, who reminded me 'patience is a virtue.' I found patience, lost it, and found it again in this process. Your patience with me was the real virtue. You kept saying 'take it one day at a time.' I did, and somehow, those days turned into years, and the journey was all the more special because of you."

ACKNOWLEDGMENTS

I am incredibly grateful to many people whose support and encouragement have been invaluable throughout this journey.

To my loving parents, Thakur and Krishna: thank you for instilling in me a love for learning. You always encouraged me to measure myself by what I learned, rather than what I earned. To my wonderful wife, Ritu: your unwavering belief in me, along with your patience and love, made this journey possible. To my sister, Asmita—you are still the hardest working in the family; I am merely being inspired by you. To my brother, Ashim—I dream big because you give me the courage to do so. To Subash and Sharada, welcome to the family and thank you for your constant encouragement. To Nitesh & Sumitra Rijal and Aadarsha & Sony Basnet—thank you for listening to me for hours on end. In a world of change, you are the constants I have anchored to for decades.

Next, my deepest gratitude goes to my advisor, Dr. John Edwards. His profound knowledge, insightful critiques, and unwavering faith in my capabilities have guided me throughout this process. I am also immensely thankful to my committee members: Dr. Soukaina Filali Boubrahimi, Dr. Shuhan Yuan, Dr. Steve Petruzza, and Dr. Kevin Moon. Special thanks go to the administrative staffs - Cora, Caitlin, and Genie - who have been extremely supportive and efficient in dealing with all my queries and requests. Shout out to Erik Falor, whom I TA'd for throughout my entire time here at USU.

Similarly, Mr. Bishnu Prasad Bastola, the entire VJHSS family, and my undergraduate co-advisors, the late Dr. Sisir Ray and Dr. Nicholas Eugene, helped set the right foundation for me. Dr. Sikharini Ray, Dr. Paramjit Sahdev, Dr. Ron Collins, DeChelle Forbes, Dr. Paul Bass, and the Coppin State University Honors College family were also instrumental in inspiring me. Dr. Tim Chung (Microsoft), Karnika Shah (Meta), and Don Hong (Esri), Dr. Jamal Uddin (Coppin Center for Nanotechnology), Ishan Srivastava (Purdue) as well as many others, mentored me, helping me to grow both personally and professionally during my internships and work.

Last but not least, to my friends—Abiral, Ashesh, Bhaskar, Biplav, Bishav, Dipen, Hrishiv, Ijan, Kavin, Prakriti, Prerana, Raj, Rajan, Ritesh, Santosh, Swastik, and Sujan—as well as the extended Nepalese family in Utah: you made my seven years in Utah a blast. You all have been my constant cheerleaders and have provided me with endless laughter and companionship during the most stressful times—Thank you!

Aashish Ghimire

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AI	artificial intelligence
GenAI	generative artificial intelligence
GPT	generative pre-training transformer
LLM	large language model
DOF	degree of freedom
NN	neural network
NLP	natural language processing
TF-IDF	term frequency – inverse document frequency
TF	term frequency
DF	document frequency
POS	parts of speech
NER	named entity recognition
LSTM	long short term memory
LCS	longest common subsequence
ROUGE	recall-oriented understudy for gisting evaluation
IDE	integrated development environment

CHAPTER 1

INTRODUCTION

1.1 Background

The dawn of the 21st century has been marked by unprecedented advancements in artificial intelligence (AI), with generative AI and Large Language Models (LLMs) standing at the forefront of this technological revolution. These innovations have not only transformed industries, commerce, and social interactions but have also begun to profoundly impact the educational sector. The application of generative AI tools, such as natural language processing (NLP) models and AI-driven educational aids, offers the promise of revolutionizing teaching methodologies, enhancing learning experiences, and democratizing access to education. This dissertation focuses on exploring the multifaceted implications of integrating generative AI technologies in educational settings, spanning legal education, policy formulation, programming courses, and broader educational practices.

The significance of generative AI in education cannot be overstated. In legal education, for instance, the vast and ever-expanding corpus of legal documents presents a formidable challenge. NLP-based summarization tools offer a potential solution by enabling the efficient distillation of lengthy court opinions into concise summaries, thus facilitating easier access to and understanding of legal precedents for both students and professionals. This application of AI not only streamlines legal research but also enhances the educational experience by making complex legal texts more accessible. The first part of this dissertation, chapter 2 addresses use of generative in summarization legal document – specifically court opinions.

Furthermore, the integration of generative AI tools like ChatGPT into classroom settings raises important questions about educators' awareness, attitudes, and the factors influencing their acceptance of these technologies. The rapid advancement of AI prompts a reevaluation of pedagogical strategies and the development of new frameworks for integrating technology into teaching and learning processes. As educators navigate this evolving landscape, understanding their perspectives is crucial for maximizing the benefits of AI in education while mitigating potential drawbacks. Chapter 3 explored the use of generative AI in classroom from the teachers' prospective.

Policy formulation around the use of AI in educational settings is another critical area of concern. The lack of comprehensive policies and guidelines for the ethical deployment of AI tools poses significant challenges, including issues related to student privacy, data security, and academic integrity. This dissertation examines the current policy landscape, highlighting the gaps and the urgent need for robust, flexible policy frameworks that can adapt to the fast-paced evolution of AI technologies in chapter 4.

In the realm of computer science education, specifically in foundational programming courses, AI tools present both opportunities and challenges. The use of AI in assisting students with coding assignments has the potential to enhance learning outcomes, foster engagement, and make programming more accessible to beginners. However, it also necessitates careful consideration of how these tools are integrated into the curriculum to ensure they complement rather than replace fundamental learning processes. In chapter 5, I address the use of AI as assistive tool in CS1 class.

Finally, this dissertation explores the broader acceptance and adaptation of generative AI tools in educational settings through the lenses of the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) in chapter 6. By examining educators' perceptions and attitudes towards AI, this research aims to identify the facilitators and barriers to the effective integration of AI technologies in education. Understanding these dynamics is essential for developing strategies that leverage the potential of AI to enrich teaching and learning experiences while addressing ethical and practical concerns.

In sum, this dissertation establishes the context for a comprehensive investigation into the application, implications, and integration of generative AI in education. Through a series of focused studies, this research seeks to contribute to the ongoing discourse on how best to harness the potential of AI technologies to advance educational goals, enhance learning outcomes, and shape the future of education in the digital age.

1.2 Objectives and Motivation

The overarching objectives of this dissertation are to critically examine the integration of generative artificial intelligence (AI) in educational settings, assess its implications, and develop insights that can guide effective, ethical, and sustainable AI adoption in education. These objectives are detailed below, reflecting the scope of the research across its various chapters. As AI technologies, particularly Large Language Models (LLMs) like GPT and NLP tools, become increasingly sophisticated, their integration into educational practices offers unprecedented opportunities for enhancing teaching and learning. However, this rapid technological evolution also introduces complex ethical, pedagogical, and policy challenges that necessitate thorough investigation and thoughtful consideration. This dissertation is motivated by the critical need to bridge the gap in existing research on the responsible integration of generative AI tools in education, focusing on the effective implementation, ethical considerations, and development of comprehensive policy frameworks to guide their use.

A. Bridging the Gap in Legal Education Through NLP

The first area of study, which examines the use of NLP for summarizing court opinions, underscores the need to make legal education and practice more accessible and efficient. Traditionally, humans have manually summarized court opinions and made them available for attorneys and clerks for a fee. This dissertation explores the possibility of generative summary using the generative AI. Even though this work is not directly related to AI in education, this served as a foundation in understanding the Large Language Models. This work used both traditional natural language processing (NLP) with LLM, and helped set the baseline for other studies. Furthermore, traditional methods of legal research and education struggle to keep pace with the sheer volume of legal texts generated annually. The motivation here is to leverage AI to distill complex legal information into manageable summaries, thereby democratizing access to legal knowledge and supporting the foundational principle of justice for all.

Objective 1: Get the Understanding of LLMs by Exploring the Potential of NLP-Based Legal Text Summarization

- The biggest objective for this project was to try both traditional NLP approaches and LLMs, build foundation on LLMs, and to understand the intricacies of generative AI and LLMs.
- To assess the efficacy of natural language processing (NLP) technologies in summarizing legal documents and court opinions.
- To evaluate the impact of NLP-based summarization tools on enhancing accessibility to legal information for both legal professionals and students.
- To explore how such tools can contribute to more efficient legal education and potentially broader access to justice.

B. Exploring Educators' Perspectives on AI

The second chapter delves into educators' awareness, sentiments, and the influencing factors towards generative AI in education. The motivation stems from understanding the pivotal role educators play in the integration of new technologies into teaching and learning processes. Identifying educators' attitudes and the variables that affect their acceptance of AI tools is crucial for designing pedagogical strategies that effectively incorporate AI into educational curricula, thereby enhancing the educational experience for students.

Objective 2: Understand Educators' Awareness, Attitudes, and Influencing Factors

- To investigate the level of awareness among educators regarding generative AI tools and their potential applications in education.
- To assess educators' attitudes towards the integration of AI tools in teaching and learning processes.

- To identify the key factors influencing educators' perceptions and acceptance of generative AI technologies in educational settings.

C. Addressing the Policy Vacuum

The examination of AI policies in educational settings highlights a significant policy vacuum. As AI tools like ChatGPT find their way into classrooms, there's an urgent need for policies that address ethical concerns, including student privacy and data security. This study is motivated by the pressing need for educational institutions to adopt flexible, robust policy frameworks that not only address current ethical challenges but are also adaptable to future technological developments.

Objective 3: Examine the Existing Policy Landscape Around AI in Education

- To analyze the current policy frameworks governing the use of AI tools in educational institutions.
- To identify gaps and challenges in the existing policy landscape related to the ethical deployment of AI technologies in education.
- To recommend strategies for developing comprehensive, adaptable policy frameworks that address ethical considerations, data privacy, and academic integrity in the context of AI usage in education.

D. Enhancing Programming Education with AI

The investigation into the use of AI tools in foundational programming courses is driven by the potential of these technologies to transform the way programming is taught and learned. The motivation here is to explore how AI can support students in overcoming the challenges of learning programming, thereby making computer science education more accessible and engaging for a broader audience.

Objective 4: Investigate the Impact and Usage Patterns of AI Tools in Foundational Programming Courses

- To explore how AI tools, particularly those akin to ChatGPT, are being utilized by students in foundational programming courses.

- To examine the impact of such tools on student learning outcomes, engagement, and interest in computer science education.

- To assess the potential of AI tools to make programming education more accessible and effective for students with diverse learning needs.

E. Understanding the Acceptance of AI Tools

Finally, the study on the adoption of generative AI tools in educational settings through the TAM and IDT lenses seeks to understand the factors influencing educators' acceptance of these technologies. The motivation is to identify barriers and facilitators to the effective use of AI in education, thereby informing strategies that encourage the responsible and beneficial integration of AI tools into teaching and learning practices.

Objective 5: Analyze Educators' Acceptance and Adaptation of Generative AI Tools

- To apply the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) in analyzing educators' acceptance and adaptation of generative AI tools in their teaching practices.

- To explore the relationship between perceived usefulness, ease of use, and the broader acceptance of AI technologies among educators.

- To identify targeted strategies that can facilitate the broader integration of AI tools in education, ensuring they align with educators' needs and teaching objectives.

In essence, the motivation behind this dissertation is to contribute to the responsible and effective integration of AI in education. By exploring these diverse yet interconnected areas, this research aims to provide insights into how generative AI can be harnessed to enhance educational outcomes, address ethical and policy challenges, and ultimately shape the future of education in an increasingly digital world.

Through these objectives, this dissertation aims to contribute valuable insights into the effective, ethical, and pedagogically sound integration of AI technologies in education. By addressing these objectives, the research seeks to inform policy, practice, and future research directions in the burgeoning field of AI in education.

1.3 Dissertation Structure

The dissertation is organized into seven chapters, each serving a distinct purpose within the overarching investigation of generative AI applications and policy considerations in education. After the introduction, which lays the foundation by presenting the background, motivations, and objectives of the study, the subsequent chapters delve into specific areas of research. Chapter 2 through 6 each focus on a unique aspect of generative AI in education, ranging from NLP-based legal text summarization [1] and educators' attitudes towards AI [2], to the examination of AI policies in education [3], the impact of AI tools on programming education [4], and the analysis of educators' acceptance of generative AI through the lenses of the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) [5]. Each of these chapter are papers either published or submitted to be published as a peer-reviewed article. These chapters collectively explore the multifaceted implications of AI integration in educational contexts, offering insights into the potential benefits, challenges, and policy needs associated with these technologies. Prior work such as [6], [7] and [8] helped gain vital research experiences as well as refine the processes.

The final chapter synthesizes the findings from the individual studies, providing a comprehensive analysis of the research questions and objectives outlined in the introduction. It discusses the implications of the findings for educational practice, policy formulation, and future research, concluding with recommendations for the effective, ethical, and pedagogically sound integration of AI technologies in education. This structure ensures a coherent narrative flow throughout the dissertation, guiding the reader through a detailed exploration of generative AI's role in transforming educational landscapes.

CHAPTER 2

Too Legal; Didn't Read (TLDR):Summarization of Court Opinions

2.1 Abstract

Access to justice remains one of the fundamental principles of the rule of law. The original US constitution was four pages and a few thousand words long [9]. But with new additions to laws and bills every year, understanding legal texts or navigating through them in itself requires specialized training and skills. Most of the legal processes and arguments rely on precedents from the past and the previous interpretation of laws. Thus, having access to the last case documents is really important and convenient. Unfortunately, these case documents are often very long, and parsing through them is time-consuming. Case summaries are meant to be of help but are written by experienced professionals and are expensive and labor-intensive. In this article we propose (Natural Language Processing) NLP based legal text summarization approach that can help professionals in writing summaries quickly with a minimum effort or create summaries automatically.

2.2 Introduction

Access to justice remains one of the fundamental principles of the rule of law. United States Institute of Peace declares, "Access to justice consists of the ability of individuals to seek and obtain a remedy through formal or informal institutions of justice for grievances" [10]. The original US constitution was four pages and a few thousand words long [9]. But with new additions to laws and bills every year, understanding legal texts or navigating through them in itself requires specialized training, skills and education. Moreover, most legal processes and arguments rely on precedents from the past and the previous interpretation of laws. Thus, having access to the last case documents is important and convenient for many legal professionals. Unfortunately, these case documents are often very long, and parsing through them is time-consuming. Case summaries are written to aid people, mainly professionals in legal services, to quickly parse through many legal documents by highlighting essential information in court opinions.

Creating a case summary is an expensive, labor-intensive task performed by trained humans [11]. The legal fee is expensive in the United States because parsing through past case histories and filings is the most costly part of access to the justice system [12]. A Natural Language Processing (NLP) approach to summarize a legal text can help trained professionals write summaries more quickly at a minimum and ideally would write the summaries automatically. Consequently, this can lower the cost of the barrier to seeking legal help and increase access to the legal system for people of lower-income brackets.

This paper has two contributions. First, we used different machine learning techniques for labeling sentences or paragraphs in a court opinion as having information important for a summary or not. These labels can be helpful for directing legal professionals to important information in the opinion. We also compared these approaches and found that LSTM-based classifier performs best among the four techniques that we tested. Second, is a domain-adapted, transformer-based model called $PEGASUS_{CourtOp}$ that outperforms all other legal text summary generators in both recall and f1 score.

2.3 Related Works

Summarization tasks, in general, can be divided into two broad categories: extractive and abstractive. Most of the works in the legal text have been focused on extractive summarization.

2.3.1 Extractive Summarization

Extractive Summarization is the process of identifying important phrases or sentences from the original text and extracting only these phrases from the text as the summary. Most of the prior work in legal text summarization until the last few years has been extractive summarization. The work done in this field can be further classified into two categories:

NLP-Based Latent Semantic Analysis and Exploration of the Thematic Structures and Argumentative Roles (rhetorical role-based approach). In 2003, Grover et al. [13] showed a primary annotation scheme of seven rhetorical roles — fact, proceedings, background, proximation, distancing, framing, and disposal, assigning a label specifying the argumentative role of each sentence in a fragment of the corpus. They used various Parts of Speech (POS) and grammar-based rules, manually defined. In 2004, Farzindar et al. [14] introduced LetSum (Legal text Summarizer), a prototype system, which determines the thematic structure of a judgment in four themes Introduction, Context, Juridical Analysis, and Conclusion. Then it identifies the relevant sentences for each theme. In 2012, Galgani et. al [15] proposed an ensemble model that used a wide range of techniques from Term Frequency – Inverse Document Frequency (TFIDF), Term Frequency (TF), Document Frequency (DF), CatchPhrase Occurance, POS, Named Entity Recognition (NER), etc., to create 23 rules. These rules described the selection of important sentences as candidate catchphrases and these rules are applied to get the summary. Later in 2016, Polsley et al. [16] proposed a tool for automated text summarization of legal documents which uses standard summary methods based on word frequency augmented with additional domain-specific knowledge. Summaries are then provided through an informative interface with abbreviations, significance heat maps, and other flexible controls. Marchent and Pande [17] published work on NLP-Based Latent Semantic Analysis for Legal Text Summarization in 2018 and this was also a fully extractive approach, based on sentence ranking. In 2019, Anand and Wagh [18] introduced a new deep learning approach to summarizing legal documents to generate the extractive summary. In addition, there have been surveying works to compare and highlight the work in legal text summarization by Kanapala et al. [19] and Jain et al. [20].

2.3.2 Abstractive Summarization

Abstractive summarization, on the other hand, is a technique in which the summary is generated by generating novel sentences by either rephrasing or using the new words, instead of simply extracting the important sentences. The complexities underlying the natural language text make abstractive summarization a difficult and challenging task. There has been research in abstractive summarization since the early 2000s, but one of the important works came in 2010 by Ganesen et al. [21] named A graph-based approach to abstractive summarization of highly redundant opinions. With the rise of deep learning and transformer-based architecture, a lot of work has been done in recent years. Paulus et al. [22] proposed a deep reinforced model for abstractive summarization in 2017. Gehrmann et al. [23] proposed a bottom-up attention step with neural networks for abstractive summarization. Later in 2020, Zhang et al. [24] published a paper on pre-training with Extracted Gap-sentences for Abstractive Summarization (PEGASUS) which was a massive language model trained for general-purpose summarization tasks. This model also included BillSum corpus [25] — 23,000 Congressional bills and human-written reference summaries for training. While there have been a lot of works in abstractive text summarization, very little has been done to adapt it to the legal domain. Huang et al. [26] in 2020 published work using the attention-based network but it was trained in public opinion data in the legal domain collected from several micro-blog sites (e.g., Peng Mei news, The Beijing News) and not an official court ruling. In our work, we present the domain-adapted abstractive summarizer trained in court opinions from various US State supreme courts and summaries created by legal professionals. Feijo [27] proposed splitting the text into smaller chunks according to predefined rules and using a BERT-based model to generate the summary. In doing so, they were able to compare the different strategies for creating those chunks and keep the best performing. They further used entailment to check the relatedness of the text and summaries. However, in this study, the dataset was somewhat labeled - the text was sectioned into the report, vote and judgment as well as contained the court-provided summary. In our study, we create summaries just from the court opinions - a blob of text with no sections by training them in summaries prepared by a separate entity.

2.4 Data

2.4.1 Data Acquisition and Cleaning

A court opinion is a statement the court announces in cases in which the court has heard oral arguments. Each sets out the Court's judgment and its reasoning. The Justice who authors the majority or principal opinion summarizes the opinion from the bench during a regularly scheduled session of the Court [28]. For our study, we use the opinion of the supreme court of Utah, Idaho, Arizona, New Mexico, Nevada, and Colorado. Table 2.1 gives the number of opinions for these states in our dataset.

Court	Opinion Count
Arizona Supreme Court	379
Colorado Supreme Court	925
Idaho Supreme Court	1411
Nevada Supreme Court	997
New Mexico Supreme Court	322
Utah Supreme Court	780
Total	4814

Table 2.1: Table with courts and count of their opinion in the dataset.

Each of these court opinions has a human-generated summary created by legal professionals for a legal information hub, Justia. Justia provided data to our research team under a data-sharing agreement. We performed basic data clean-ups including case conversion and punctuation removal, stop-words removal, tokenization, lemmatization, and vectorization (using word embeddings). In addition, we also tokenized the original text to words and sentences using state-of-the-art pre-trained models from the Natural Language Toolkit (NLTK) [18] and Spacy [29]. Gensim Word2Vec model [30] is a pre-trained word embedding representation where each word are represented with a unique vector representing the meaning of that word. This preserves the similarities and the distance representation between words. We vectorized the data with Gensim pre-trained word embedding vectors for LSTM-based models and indices for other binary classifiers.

2.4.2 Data Exploration

We also did some data exploration, including descriptive statistics of the opinion and summary text. The length of the summary for the original text can vary depending on the type of opinion. This information gave us an overview of the distributions of the opinion text and helped us to get an overall idea of the size of the generated summary for automatic summarization. Figures 2.1 and 2.2 are the histograms of the number of words in the opinion and the summary.

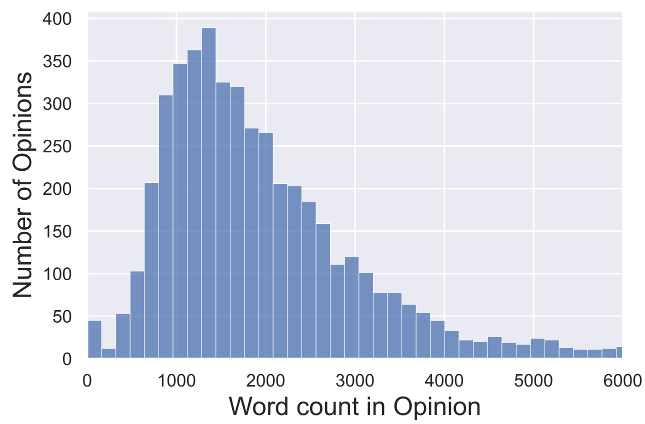


Fig. 2.1: Histogram of word count in Opinion
*Some opinions have 6000 words

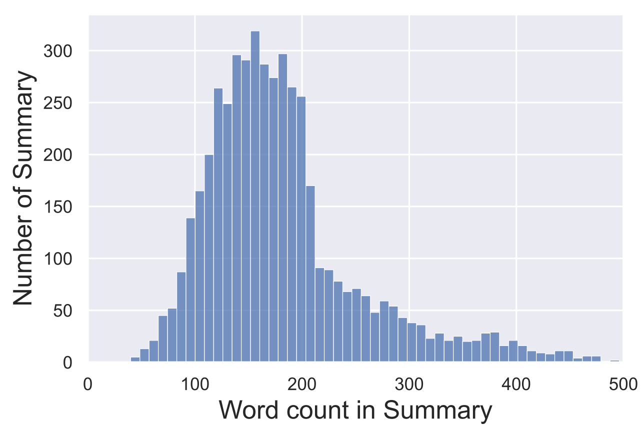


Fig. 2.2: Histogram of word count in Summary

2.4.3 Labeling the Opinion

Extractive Summarization is a widely used summarization method for text documents. This approach uses the portions, typically sentences, of the input text/documents to create a generated summary. We used sentences and paragraphs from the original text for the summarization. We created a classifier that tags sentences and paragraphs of the court opinion based on their relevance to a human-generated summary and uses these parts to synthesize a summary. Labeling the sentences and paragraphs of the court opinion was our first step. The most straightforward process for tagging would be to manually label the opinion parts (sentences and paragraphs) as relevant or not to the summary using domain expertise. To our knowledge, no such dataset exists, so we used different algorithms to automatically tag the relevant parts of the opinion. We discuss four approaches: N-Grams [31], Longest Common Subsequence (LCS) [32], Semantic Similarity using word2vec [33], and ROUGE score [34]. The main idea behind these algorithms is to find the relevant parts of the opinion that are most similar to the human-generated summary of the opinion document.

N-Gram

N-gram is a contiguous sequence of words from a given sample of text. N-Gram-based tagging looks for the n-grams from the sentence and paragraph of the document in the human-generated summary.

LCS-Score (Longest Common Subsequence)

The LCS-Score method uses the longest common subsequence between the sentence and paragraph of the opinion and the summary.

Semantic Similarity

A semantic similarity between text corpus can be determined by using word embeddings for sentences and paragraphs using a python NLP library like Spacy.

ROUGE Score

ROUGE [34] is a metric used for evaluating automated summarization text with the reference summary. Section 2.5.3 describes the ROUGE score in more detail. This approach can also be used to identify or determine the parts of the original text that are most similar to the original summary. The sentences and paragraphs can be compared with the original summary to calculate the ROUGE score, selecting the most relevant parts that exceed a certain threshold from the opinion document.

2.5 Method

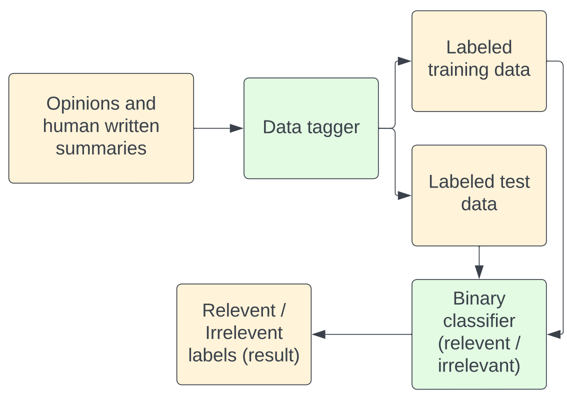


Fig. 2.3: Method A: Binary Classification of Text

2.5.1 Binary Classification for Extractive Summarization

We used the data labeling methods described in Section 2.4.3 to transform our dataset into a labeled dataset. The sentences and paragraphs of the original text are the input features, and their relevance to the summary is the labels. This partially casts our summarization problem as a classification problem. Using the labeled training data, we create a model and then use the model to tag sentences and paragraphs of a new opinion as relevant (label 1) or not relevant (label 0). These classified sentences and paragraphs are then used to create an extractive summary by joining them in order. Our approach for classification is summarized in figure 2.3. For classification, we use Scikitlearn [35] extensively.

Multinomial Naive Bayes Classifier

The Bayesian classifier is based on Bayes' theorem. Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes [36]. Naive Bayes is a learning algorithm that is commonly applied to text classification. When the assumption of independence holds, a Naive Bayes classifier performs better compared to other models like logistic regression and with less training data. The probability that a given document D contains all the words w_i, given a class C, is:

$$P(D \mid C) = \sum_i p(w_i \mid C)$$

where $p(w_i \mid C)$ is the conditional probability of term w_i to be of class C. We interpret $p(w_i \mid C)$ as a measure of how much evidence w_i contributes that C is the correct class, C is represented as 1 or 0, 1 being relevant and 0 not. The Naive Bayes Classifier acts as a benchmark baseline for comparison with other classifiers.

Decision Trees

A decision tree is a simpler and more interpretable classifier [37]. We trained a decision tree with the labeled dataset and compare the results with other classifiers. We tested with the tree depths from 2 to 50.

Random Forest

Random forest is an ensemble learning method for classification, regression, and other tasks. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression [38]. We experimented with various sizes of estimators and ultimately, used 250 estimators for the test.

Neural Network (LSTM)

We trained a feed-forward neural network to classify the text. This very simple network has an embedding layer, one Long Short Term Memory (LSTM) [39] layer, one dropout, and one dense layer with sigmoid activation.

The dropout layer will randomly drop the connection for 30% of the networks (tuned hyper-parameters) to prevent overfitting the network. The embedding layer is not a pre-trained model - it is a simple matrix, randomly initialized.

The labeled sentences and paragraphs are used to generate an extractive summary. We calculated the ROUGE score for the generated summary from each classifier when compared with the human-generated summary, described in the section 2.6.

2.5.2 Abstractive Summarization using Pre-Trained Language Models

We explored the use of a pre-trained language model for generating the abstractive summary of a court opinion, which not only has sentences from the opinions but also has paraphrasing and a human-like sentence structure. With the rise of very large language models with millions of parameters, it is now possible to start with those models as the base and fine-tune them to domain-adapt along with training with more domain-specific training sets. Our approach for abstractive summarization is shown in figure 2.4

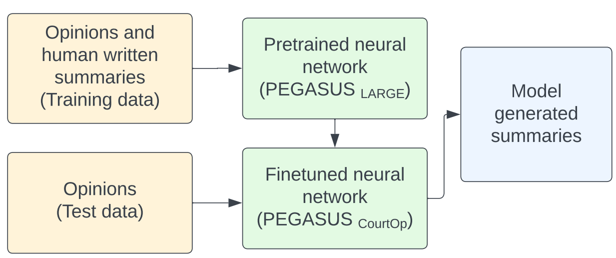


Fig. 2.4: Method B: Summary Generation using Pre-trained Models

The language model PEGASUS (Pre-training with Extracted Gap-sentences for Abstractive Summarization) is pre-trained for Gap Sentences Generation objective i.e. some portions of texts are selected to be masked (using a few different selections techniques) and the model is trained to fill in the masks. This is done together with the mask from the Mask language model (MLM) from something like BERT (Bidirectional Encoder Representations from Transformers) model. The majority of data in the PEGASUS project come from web common crawl, social media, and news. It however also includes the BillSum dataset [25]. BillSum (Kornilova & Eidelman, 2019) contains 23k US Congressional bills and human-written reference summaries from the 103rd-115th (1993-2018) sessions of Congress.

There are different versions of the PEGASUS language model depending on their size. We use the $PEGASUS_{LARGE}$ as our base language model.

On top of the $PEGASUS_{LARGE}$, we re-train the model fine-tuning it for the legal opinion domain. We used 3661 pairs of legal opinions and summaries (75 percent of our data, the rest 25 percent held for validation and benchmarking). We froze the weight of encoder layers and trained the decoders for our objective. We fine-tuned with following parameters as shown in the table 2.2:

Parameter	Value
Additional Retraining Examples	3661
Retraining Epochs	20
Encoder Layers	gradient forzen
Decoder Layers	gradient updated
Rate of Weight decay	0.01
evaluation strategy	*steps*

Table 2.2: Table showing the fine-tuning parameters $PEGASUS_{CourtOp}$ model

2.5.3 Benchmarking and Performance Metrics

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a widely used performance metric for a summarization task. It includes measures to automatically determine the quality of a summary by comparing it to other (ideal) summaries created by humans [34]. An n-gram is a contiguous sequence of n items from a given sample of text or speech. Formally, ROUGE-N is an n-gram recall between a candidate summary and a set of reference summaries. The ROUGE-N score of a candidate text (candidate) and a reference text (ref) is computed as the equation 2.1:

$$\text{Rouge-N} = \frac{\sum_{ref}\sum_{candidate} match(gram_n)}{\sum_{ref}\sum_{candidate} Count(gram_n)} \tag{2.1}$$

Here, $\sum_{candidate} match(gram_n)$ represent the number of common n-grams between candidate and reference text. The notation $\sum_{ref}\sum_{candidate} Count(gram_n)$ is the total number of n-grams in the text themselves. Since we have the human-written summary for each opinion, we can use them to get the ROUGE-N score for our model-generated summary. We calculated ROUGE-1, ROUGE-2, and ROUGE-L, but we are using ROUGE-1 for comparison because most of the prior literature reports the ROUGE-1 score.

In addition to the ROUGE-1 score, we also have the result of the classification for the binary classification task used for extractive summarization.

2.6 Results and Discussion

2.6.1 Extractive Summarization

In this section, we show the results obtained from the Binary Classifiers for Extractive Summarization of the legal document. Table 2.3 shows the classification report for the Paragraph Level and Sentence Level classification of legal documents using 5-fold cross-validation. From the table 2.3, we can see that the F1-Score and Recall decrease for the sentence level classification as compared to paragraph level classification. The original text document has a relatively larger number of sentences than paragraphs. Similarly, the number of irrelevant sentences is larger than irrelevant paragraphs. This makes the dataset highly imbalanced and introduces bias in the classifier. Random Forest classifiers provide better results for our classification when compared to other classifiers.

After the classification of the relevant paragraphs and sentences, we can create a summary by concatenating them. The ROUGE score between the generated summary and the human-generated summary is shown in Table 2.4. The LSTM-based summarization seems to have better ROUGE scores than other classifier-based summaries. LSTM is capable of

Classifier	Paragraph Level		Sentence Level	
	Recall	F1-Score	Recall	F1-Score
Naive Bayes	0.76	0.705	0.61	0.58
Decision Tree	0.70	0.69	0.59	0.55
Random Forest	0.84	**0.8**	0.69	**0.63**
LSTM NN	**0.85**	0.73	**0.7**	0.59

Table 2.3: Classification Results

learning order dependence in a sequential dataset. This order dependence plays a role in interpreting the sentences and paragraphs of the original text. This might be the main reason for the better performance of LSTM-based neural networks in contrast to other classifiers.

Classifier	ROUGE-1 F1	ROUGE-1 Recall
Naive Bayes	0.11	0.2
Decision Tree	0.26	0.48
Random Forest	0.29	0.5
LSTM NN	**0.34**	**0.55**

Table 2.4: Rouge Score for Binary Classifiers

The summary generated from extractive summarization doesn't seem to be natural, as it is just concatenating the relevant parts from the original text. The abstractive summarization, however, creates a more natural summary of the original text. The extractive summarization approach can be integrated with the abstractive approach in creating a more robust and human-like summarization approach in our future work.

2.6.2 Abstractive Summarization

For the abstractive summarization, three different tasks were performed. For the baseline comparison, we evaluated the pre-trained $PEGASUS_{LARGE}$ model in our test data. After that, we performed a test of our data in a pretrained **LEGAL PEGASUS** model [40]. This model was trained on a sec-litigation-releases dataset consisting of more than 2700 litigation releases and complaints. Finally, our domain-adapted and fine-tuned model $PEGASUS_{CourtOp}$ was tested for the same data set. The result is shown in table 2.5.

Classifier	ROUGE-1 F1	ROUGE-1 Recall
$PEGASUS_{LARGE}$	0.31	0.39
Legal Pegasus	0.41	0.55
$PEGASUS_{CourtOp}$	**0.53**	**0.66**

Table 2.5: Rouge Score for Binary Classifiers

The PEGASUS model was specifically designed to be fine-tuned and domain adapted with a relatively smaller number of examples for objective defined tuning. The word encoding side of the model was already well trained, and when we redefined the objective of the model to generate the opinion summary by fine-tuning with our example, the model performed better.

2.6.3 Summary Examples

Examples of summaries generated using different models are in table 2.6.

A perfect rouge score is almost impossible in an abstractive summary as it involved text generation by the human and model which are not necessarily present in the corpus. Our ROUGE-1 score is better than any out-of-the-box models for both extractive and abstractive summaries.

2.7 Conclusion

In this paper, we have presented work on the automatic summarization of legal texts. We created our own labeled corpus from a legal information hub, Justia, and discussed different summarization approaches and their results.

Human Generated Summary:	$PEGASUS_{CourtOp}$ **Summary (Ours), Rouge1-F1= 0.79**
CDA Dairy Queen, Inc. and Discovery Care Centre, LLC of Salmon (collectively, Dairy Queen) filed a class action against the Idaho State Insurance Fund (SIF) seeking a declaratory judgment that SIF violated Idaho Code section 72-915 by failing to distribute premium rate readjustments on a pro rata basis. The district court granted SIF's motion for summary judgment, finding that the Idaho Legislature's retroactive repeal of section 72-915 was constitutional and that Dairy Queen's action was thereby barred. Dairy Queen appealed and argued that the retroactive repeal violated article I, section 16 of the Idaho Constitution. Dairy Queen the Supreme Court to reverse the decision of the district court and remand for further proceedings consistent with the determination that the retroactive repeal was unconstitutional. Upon review, the Supreme Court agreed with Dairy Queen and reversed.	The Supreme Court reversed the decision of the district court granting summary judgment to the Idaho State Insurance Fund (SIF) seeking a declaratory judgment that SIF violated Idaho Code 72-915 by failing to distribute premium rate readjustments on a pro rata basis, holding that the retroactive repeal of Idaho Code 72-915 was constitutional and that Dairy Queen's action was thereby barred. Dairy Queen filed a class action against SIF seeking a declaratory judgment that SIF violated Idaho Code 72-915 by failing to distribute premium rate readjustments on a pro rata basis. The district court granted summary judgment for SIF. Dairy Queen timely appealed and argued that the retroactive repeal violated article I, 16 of the Idaho Constitution. The Supreme Court reversed and remanded for further proceedings consistent with the determination that the retroactive repeal is unconstitutional.
$PEGASUS_{LARGE}$ **Summary Rouge1-F1= 0.58**	**Legal PEGASUS Summary Rouge1-F1= 0.61**
6 Filed: January 23, 2013 Stephen Kenyon, Clerk Appeal from the District Court of the Third Judicial District of the State of Idaho, Canyon County. and Discovery Care Centre, LLC of Salmon (collectively, Dairy Queen) filed a class action against the Idaho State Insurance Fund (SIF) seeking a declaratory judgment that SIF violated Idaho Code 72-915 by failing to distribute premium rate readjustments on a pro rata basis. The district court granted SIF's motion for summary judgment, 1 finding that the Idaho Legislature's retroactive repeal of Idaho Code 72-915 was constitutional and that Dairy Queen's action was thereby barred. Dairy Queen asks this Court to reverse the decision of the district court and remand for further proceedings consistent with the determination that the retroactive repeal is unconstitutional.	On January 23, 2013, the U.S. Court of Appeals for the Third Judicial District of the State of Idaho ruled that the Idaho Legislature's retroactive repeal of a law requiring the state's insurance fund to distribute premium rate readjustments on a pro rata basis is unconstitutional. The court held that the law violated article I, 16 of the Idaho Constitution, and that Dairy Queen, Inc. and Discovery Care Centre, LLC's declaratory judgment against the Idaho State Insurance Fund was barred. Dairy Queen and Discovery Care Centre filed a class action against the State Insurance Fund for failing to distribute premium rate readjustments.

Table 2.6: Example of summary generated by humans and different models.

We presented different extractive approaches to extract relevant parts from the original legal text. This approach can be useful in identifying the relevant parts and reducing the time taken by a legal advisor on creating human-generated summaries. Furthermore, we also created a domain-adapted, fine-tuned summarizer model based on Google's

$PEGASUS_{LARGE}$ language model.

Our model improved on state-of-the-art models in both recall and F1 score for the specific task of summarizing the legal opinions. This can be used as an assistive tool to speed up the summarization by the human at the minimum, and furthermore, to generate an opinion summary automatically with a relatively good performance.

2.8 Future Works

Since our work on this topic, more powerful language models like GPT 3 and GPT 3.5 have been released. While these models are not open source yet and could not be used for comparison in this paper, fine-tuning such models with more parameters in legal summarization could yield better results. A legal-text-specific model Named Entity Recognition would be another important step in increasing the accuracy and performance of the summarization task. While there have been works in Named Entity Recognition in general, court opinion-specific work seems to be lacking. Furthermore, while our work can help humans to narrow down and focus on a specific part of the document, more work needs to be done to generate a language that is not present in the opinions.

CHAPTER 3

Generative AI in Education: A Study of Educators' Awareness, Sentiments, and Influencing Factors

3.1 Abstract

The rapid advancement of artificial intelligence (AI) and the expanding integration of large language models (LLMs) have ignited a debate about their application in education. This study delves into university instructors' experiences and attitudes toward AI language models, filling a gap in the literature by analyzing educators' perspectives on AI's role in the classroom and its potential impacts on teaching and learning. The objective of this research is to investigate the level of awareness, overall sentiment towards adoption, and the factors influencing these attitudes for LLMs and generative AI-based tools in higher education. Data was collected through a survey using a Likert scale, which was complemented by follow-up interviews to gain a more nuanced understanding of the instructors' viewpoints. The collected data was processed using statistical and thematic analysis techniques. Our findings reveal that educators are increasingly aware of and generally positive towards these tools. We find no correlation between teaching style and attitude toward generative AI. Finally, while CS educators show far more confidence in their technical understanding of generative AI tools and more positivity towards them than educators in other fields, they show no more confidence in their ability to detect AI-generated work.

keywords: LLM, Chatbot, ChatGPT, AI in Education, Teachers' attitude

3.2 Introduction

The rapid advancement of generative artificial intelligence (AI) and the increasing integration of large language models (LLMs) in various domains have sparked a debate surrounding their implementation within the educational sector [41–43]. This study aims to investigate instructors' experiences and attitudes toward harnessing AI language models in education, focusing on understanding the underlying factors that shapes these opinions. The study addresses a gap in the existing literature by comprehensively analyzing educators' perspectives on integrating AI technologies in the classroom and their implications for teaching and learning.

To achieve the research objectives, the study explores the following research questions:

RQ1 How aware are educators of Generative AI-based tools across various departments?

RQ2 What are educators' perceptions and sentiments about these AI tools?

RQ3 What factors contribute to variations in teachers' attitudes toward generative AI based tools?

RQ4 How do the attitudes and perceptions of CS educators differ from those of educators in different departments?

RQ5 What are the biggest opportunities and concerns identified by the educators?

The study employed a mixed-methods research design, incorporating both quantitative and qualitative data collection and analysis techniques. A survey was conducted to collect data on instructors' experiences and attitudes using a Likert scale, which was supplemented by free-form text entries and optional interviews to gain a more nuanced understanding of the factors shaping their perspectives. The data were analyzed using statistical and thematic analysis techniques to generate insights into the research questions.

By understanding educators' experiences and attitudes concerning the harnessing of AI language models in education, this study aims to contribute to the ongoing discourse on the role of AI technologies in shaping the future of education. The findings can inform policymakers, educators, and researchers about the potential benefits and challenges of integrating AI language models into the classroom and guide the development of strategies and practices that enhance teaching and learning outcomes.

3.3 Related Work

3.3.1 Generative AI tools in Computer Science Education

Recent advances in generative AI and natural language processing have enabled the development of large language models (LLMs) that show impressive capabilities in generating and reasoning about code [44]. Major LLM-based products like Generative Pre-trained Transformer (GPT-4), CodeX, GitHub Copilot, Bard and ChatGPT have significant implications for computing education research and practice [45].

A growing body of work has begun empirically evaluating how these LLMs perform on tasks and assessments commonly used in programming courses [46,47]. For instance, Chen et al. found that GPT-3, after generating 100 samples and selecting the sample that passed the unit tests, scored around 78% on CS1 exam questions, outperforming most students [48]. In more advanced CS2 assessments, Codex performed comparably to the students in top quartile [49]. GitHub Copilot was also shown to generate passing solutions for typical introductory programming assignments [50]. These studies clearly demonstrate the need to reconsider curriculum design and assessment in light of LLM capabilities.

Researchers have proposed adaptations such as focusing less on basic coding skills and more on higher-level thinking and analysis when LLMs can automate generation [48]. New forms of assessment may be required to prevent plagiarism and ensure students have true mastery [47, 51, 52]. There are also calls to explicitly teach the productive use of LLMs as aids rather than relying excessively on them [52–54].

Beyond assessment, researchers have identified opportunities for using LLMs in pedagogy. They can automatically generate solutions, explanations, and examples to scaffold learning and reduce instructor effort [41, 55–58]. LLMs may enable novel active learning approaches through personalized help, peer code reviews, and interactive coding activities integrated with LLMs [41, 59–61]. New programming problem types that utilize LLMs, such as Prompt Problems, are also beginning to emerge [62]. However, risks include the propagation of incorrect solutions or explanations if not vetted [41, 48].

The literature also highlights threats posed by LLMs regarding over-reliance impeding learning [63] and circumventing assessments [48, 51]. Challenges around plagiarism detection [49, 50], bias [64], and the greater socio-economical consequences [65] must also be addressed. Further research is critically needed to develop evidence-based practices for effectively leveraging LLMs in computing courses while mitigating their potential harms.

3.3.2 Teachers' attitudes towards AI tools in education

The attitudes and perceptions of instructors and educators are paramount in the adoption, rejection, success, or failure of these tools. Bii et al. investigated the attitude of teachers towards the use of chatbots in routine teaching by surveying teachers in Kenya, and the results showed that teachers have a positive attitude towards the use of chatbots [66]. The study found that teachers have some reservations about using chatbots, such as concerns about the accuracy of the information provided by chatbots and the potential for chatbots to replace teachers. However, overall, the study found that teachers are open to using chatbots in their teaching. Guillén-Gámez and Mayorga-Fernández investigated the factors that predict teachers' attitudes towards information and communication technologies (ICT) in higher education for teaching and research [67]. The results of the study showed that the professors' attitudes towards ICT were positively predicted by their age, gender, and participation in ICT-related projects. The professors' attitudes were also positively predicted by their teaching experience and their perception of the usefulness of ICT for teaching and research. Nazaretsky et al. investigated the factors that influence teachers' attitudes towards AI-based educational technology [68].The study found that teachers' attitudes were influenced by two human factors: confirmation bias and trust. Teachers who were more likely to engage in confirmation bias were more likely to ignore information about AI-based educational technology that contradicted their existing beliefs, thus becoming less likely to have positive attitudes towards AI-based educational technology.

Akgun and Greenhow provided an in-depth exploration of the ethical challenges inherent in the deployment of artificial intelligence (AI) within K-12 educational settings [69]. The authors highlight the importance of transparency, accountability, sustainability, privacy, security, inclusiveness, and human-centered design in the development and use of AI in education. Celik et al. explores the roles of teachers in AI research, the advantages of

AI for teachers, and the challenges they face in using AI [70]. They found that teachers have seven roles in AI research, including providing data to train AI algorithms and offering input on students' characteristics for AI-based implementation. The advantages of AI for teachers were identified in planning, implementation, and assessment, with AI providing timely monitoring of learning processes and assisting in decision-making on student performance. However, the study also highlighted challenges such as the limited technical capacity of AI, the lack of technological knowledge among teachers, and the context-dependency of AI systems. Kim and Kim investigated the perceptions of STEM teachers towards the use of an AI-enhanced scaffolding system developed to support students' scientific writing [71]. The results of the study showed that the teachers had a generally positive perception of the AI-enhanced scaffolding system. The teachers felt that the system could be used to provide personalized instruction, automate tasks, and provide feedback to students. Despite the positive expectations, the study noted that before AI can be effectively adopted in classrooms, teachers first need to learn how to use this technology and understand its benefits.

Chocarro et al. recently examined the factors that influence teachers' attitudes towards chatbots in education [72]. They used the dimensions of the Technology Acceptance Model (TAM), specifically perceived usefulness and perceived ease of use, to understand this acceptance. The study takes into account the conversational design of the chatbot, including its use of social language and proactiveness, as well as characteristics of the users, such as the teachers' age and digital skills. They found that formal language used by a chatbot increased teachers' intention to use them, and teachers' age and digital skills were related to their attitudes towards chatbots. Khong et al. aimed to construct a model that predicts teachers' extensive technology acceptance by examining the factors that influence their behavioral intention to use technology for online teaching by extending Technology Acceptance Model (TAM) [73]. The study suggested that cognitive attitude had a much larger impact on teachers' behavioral intention to teach online, and perceived usefulness of online learning platforms had greater influence on teachers' online teaching attitude than perceived ease of use, particularly on cognitive attitude.

The 2023 study by Iqbal et al. explored the attitudes of faculty members towards using ChatGPT [74]. The study used the TAM to investigate the factors that influence faculty members' attitudes towards using ChatGPT. The study found that faculty members had a generally negative perception and attitude towards using ChatGPT. Potential risks such as cheating and plagiarism were cited as major concerns, while potential benefits such as ease in lesson planning and assessment were also noted. Finally, Lau and Guo present the perspectives of 20 university instructors who teach introductory programming courses on how they plan to adapt to the growing presence of AI code generation and explanation tools such as ChatGPT and GitHub Copilot [42]. They report that instructors have different opinions on whether to resist or embrace these tools in their courses and propose a set of open research questions for the computing education community.

3.4 Methodology

3.4.1 Survey Design

To investigate teachers' attitudes toward AI tools and Language Learning Models (LLMs) in education, we conducted a quantitative study using a survey. The survey was designed to explore educators' perceptions of AI language models and their integration into pedagogical practices. It included questions that assessed participants' awareness of AI and LLMs, their beliefs about the potential benefits and challenges of these technologies, and their attitudes toward using them in the classroom.

The survey questions were developed based on relevant literature and the research questions listed above. The Likert scale was used for most questions, allowing participants to indicate their level of agreement or disagreement with specific statements. Additionally, the survey included open-ended questions to capture qualitative insights and gather more in-depth responses as well as some basic anonymous demographic data like age-group and tenure length. The survey was IRB approved at [anonymous].

3.4.2 Data Collection

We distributed the survey to faculty members at a mid-sized research university in the western United States via email. Each faculty member received the survey link only once to avoid duplicate responses. The email provided a brief introduction to the research study, assured confidentiality, and encouraged participation. Participants were informed about the voluntary nature of the survey and were given the option to opt-in for a follow-up interview.

3.4.3 Survey Responses

We received a total of 116 survey responses from email requests, representing a diverse sample from 8 colleges and 23 out of 39 departments at the university. The wide-ranging representation ensures a comprehensive understanding of educators' attitudes from various academic disciplines.

3.4.4 Interviews

To gain deeper insights into teachers' experiences and attitudes, we conducted semi-structured interviews with a subset of participants who opted-in for the follow-up interview. The interviews were approximately 25-30 minutes long and used open-ended questions to encourage participants to share their perspectives freely. The interview responses were recorded and later transcribed for analysis. Interviews were also conducted with IRB oversight.

3.4.5 Data Analysis

Quantitative Study

The quantitative survey data were analyzed using descriptive statistics and inferential methods. We calculated the mean, standard deviation, and frequency distributions to summarize participants' responses to Likert scale questions. We also performed hypothesis tests, confidence intervals, and regression analysis to identify potential correlations and trends in the data.

Grounded Theory and Qualitative Analysis

For the qualitative analysis, we adopted a grounded theory [75] approach to identify themes and patterns emerging from the interview data. Three independent evaluators coded two transcribed interviews each, and inter-rater reliability was evaluated using Cohen's Kappa coefficient. The evaluation resulted in an inter-rater reliability of over 85%, ensuring the consistency of the coding process.

Integration of Data

The coded interview data were integrated with the quantitative survey results to triangulate findings and provide a comprehensive understanding of teachers' attitudes toward AI tools and LLMs. The grounded theory approach allowed us to generate inductive insights from the interview data, which were then co-analyzed with the quantitative study's results.

3.4.6 Participation

The survey received a total of 116 responses from faculty members across various school/colleges and departments at the university. The colleges with the highest number of respondents were the College of Arts and Sciences (Science), the College of Education and Human Services (Education), and the School of Business (Business). In terms of follow-up interviews, 36 faculty members opted for further discussions, with a notable interest from the College of Science (Science) and the College of Engineering (Engineering). This diverse representation of faculty members provides a broad perspective on the attitudes and opinions towards AI tools and LLM-based technologies in education. Figure 3.1 shows the participants from each school and department.

3.5 Results and discussion

3.5.1 RQ1: How aware are educators of Generative AI-based tools across various departments?

To answer this question, we asked each survey participant about their familiarity and

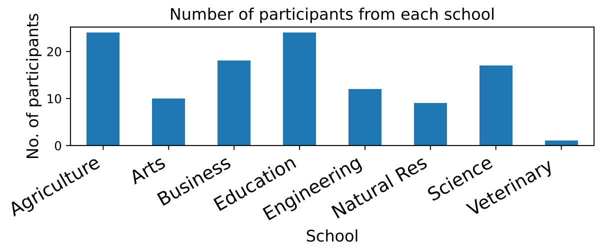


Fig. 3.1: Participants across various schools.

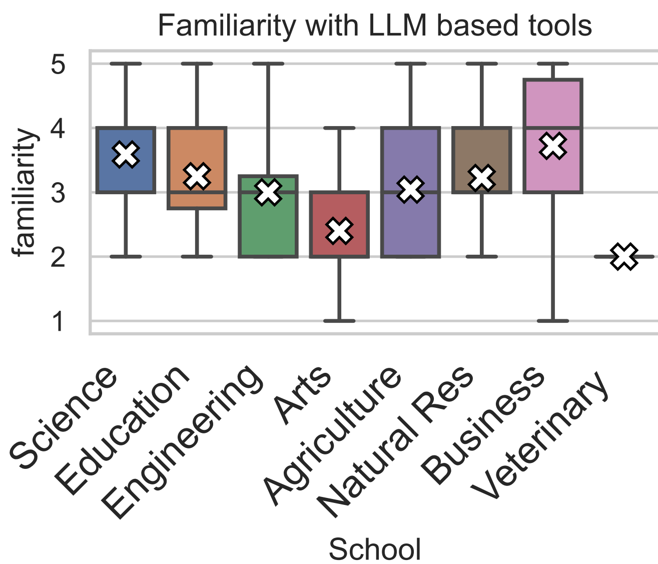


Fig. 3.2: Familiarity with LLMs by school

usage habits of these tools and followed up in an interview with questions about their usage habits, sources of introduction, etc.

Our survey revealed that most educators have at least heard of these tools or tried them. More than 40% of the faculty members said they use them at least periodically or regularly. While no significant difference was found across various age brackets and tenure lengths, the familiarity varied by school. The College of Science and School of Business have the highest familiarity overall, while the College of Arts affiliated educators were the least familiar. Figure 6.2 shows the familiarity by school and figure 3.3 shows familiarity by age-group.

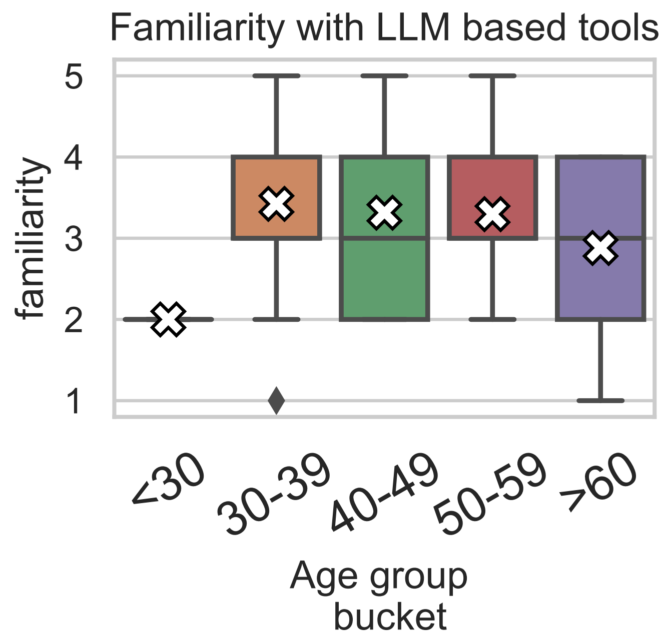
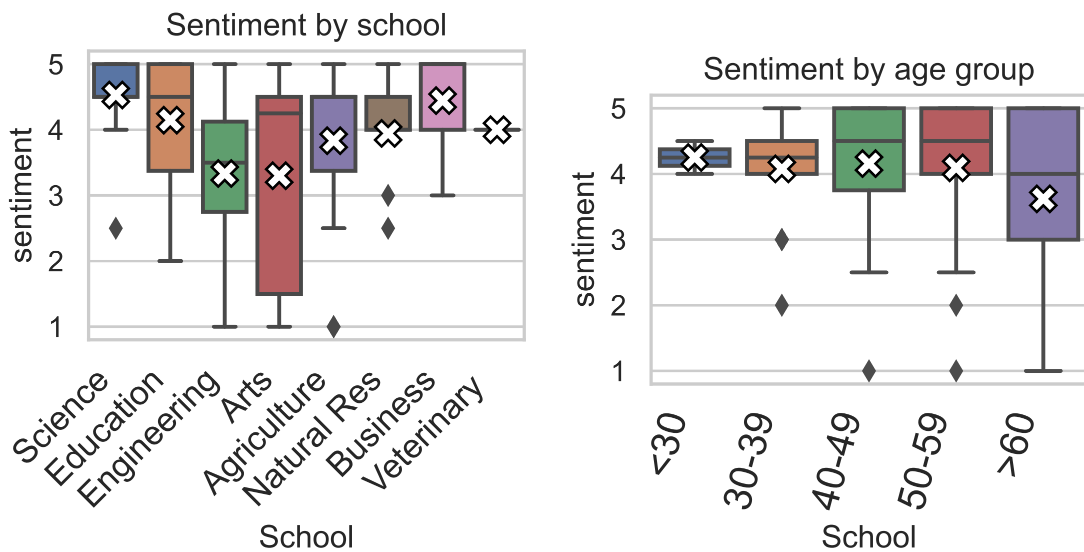


Fig. 3.3: Familiarity with LLMs by age-group

We followed up in the interview on how or in what context the educators were introduced to these tools. Table 3.1 shows the discovery source of these Generative AI-based tools among the educators. Through the interviews, we discovered multiple instances where faculty members who follow the development of these tools more closely held formal or informal workshops to inform their colleagues of these developments. Among those interviewed, 19% had a technical understanding of Generative AI and LLMs, while others had only a basic understanding. 38% of the interviewees were very aware that they lacked technical understanding of the tool.

Discovery Source	Proportion
News	33.33 %
Peers	16.67 %
Work/Training	16.67 %
Family Member	11.00 %
Social Media	11.00%
Others or unsure	11.33 %

Table 3.1: Discovery sources of Generative AI tools.



(a) Sentiment by school   (b) Sentiment by age group

Fig. 3.4: Sentiment by (a) school and (b) age-group.

3.5.2 RQ2: What are educators' perceptions and sentiments about these AI tools?

Next, we explored the educators' attitudes and sentiments towards these AI tools. We used the answers to the following questions for this:

1. AI tools like ChatGPT and Bard should be allowed and integrated into education. (beIntegrated)

2. I think the AI tools like ChatGPT and Bard should be banned in all academic settings. (beBanned)

We used the following equation to calculate the sentiment and ensure the value is between 1 and 5:

$$Sentiment = \frac{\text{beIntegrated} + (6 - \text{beBanned})}{2}$$

The overall sentiment towards these tools is positive, with a mean of 3.99. The median sentiment is 4.5, and the third quartile is 5. Only 12% of educators had worse than average sentiment (sentiment $<$ 3). Figure 3.4a shows the distribution of sentiment by school.

Following the familiarity trend, the College of Science and School of Business have the most positive sentiment, while the College of Arts has the lowest.

Figure 3.4b shows the distribution of sentiment by age group. While the overall mean sentiment is not different across age groups, the inter-quartile range becomes larger for the older age group.

We also asked about their initial impression as well as the change in impression since the first encounter. Most of the respondents, especially from outside the computer science department, used words like "amazed" or "mind-blown" to describe the initial impression. We saw more than 56% of interviewees grow more positive, 38% stayed the same and only 6% grew more negative.

3.5.3 RQ3 : What factors contribute to variations in teachers' attitudes toward AI language model?

Pedagogical Practices

In the survey, we asked the instructors questions about their teaching methodologies like lectures, labs and hands on experiments, discussions etc. as well as the testing methodologies they employ. There was no significant difference discovered between the perception about these AI tools in relation to their pedagogical practices. Comparing both the kinds of questions teachers use for their assignments and test as well as their teaching style, there were no statistically significant differences. We also followed up in the interview about how they see the need to adapt their pedagogical practices to address these new developments.

One of the most-repeated themes was that educators were more receptive of using a tool in advanced classes where students have already acquired the fundamentals of their discipline.

> *I have no problem with students using it in my advanced class. In fact, I don't mind that at all, might even encourage it. However, if they use it in the [Intro CS class], they are not going to learn anything.*

This quote from a computer science professor is one of many who indicated they are more positive toward adapting these technologies in higher level classes.

Identifying contributing features

Next, we delved into the process of identifying the key contributing features that influence educators' attitudes toward Generative AI and LLMs. To accomplish this, we employed regression analysis and utilized the LASSO (Least Absolute Shrinkage and Selection Operator) technique for feature selection. Through this analysis, we aimed to uncover the most significant factors that play a role in shaping educators' attitudes in order of importance.

The analysis yielded a list of features along with their corresponding coefficients, shedding light on the relative impact of each feature. Table 3.2 shows the most important factors that influence teacher's sentiment about Generative AI, listed in ranked order.

Rank	**Factor**	**Effect**
1	Benefit outweigh risks	Positive
2	Enhances the quality of education	Positive
3	Can easily be integrated	Positive
4	Decreases critical thinking	Negative
5	Increase cheating and dishonesty	Negative

Table 3.2: Top 5 Factors affecting sentiments ranked

It is evident that factors related to risk-to-rewards ratio, quality enhancement, and ease of integrating AI tools are among the most influential in shaping positive attitudes. Conversely, concerns about loss of creativity and potential for cheating and dishonesty impact attitudes negatively.

Moreover, we extended our analysis to different regression techniques, including Linear Regression, Random Forest, Gradient Boost, and XGBoost. The mean squared errors (MSE) obtained ranged between 0.4 and 0.5, indicating a reasonable level of predictive accuracy using these features.

Overall, our analysis unveils a hierarchy of factors that significantly contribute to educators' attitudes toward Generative AI and LLMs. These insights can guide educational practitioners, policymakers, and researchers in understanding the intricate interplay of factors that shape attitudes, thereby facilitating informed decision-making and effective implementation strategies.

3.5.4 RQ4: How do the attitudes and perceptions of CS educators differ from those from different departments?

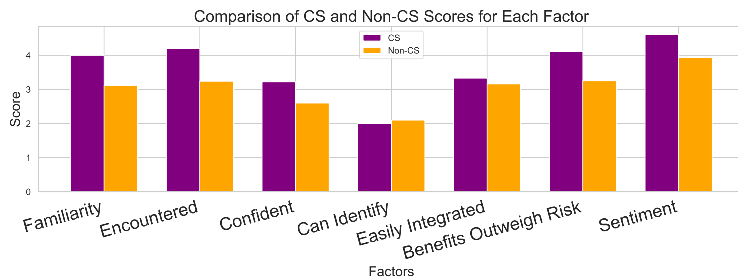


Fig. 3.5: Mean response between CS and non-CS instructors.

The survey included 9 Computer Science participants out of 116 total, and 6 out of 36 interviewees were from the Computer Science department. In terms of understanding, 83% of Computer Science instructors had technical understanding compared to only 10% of Non-CS. In terms of familiarity, as shown in Figure 3.5, the CS faculty members reported higher mean familiarity($M = 4, SD = .71$) as compared to non-CS($M = 3.15, SD = 1.08$); $t(113) = 3.28, p = .007$. The majority of CS respondents were confident that their students have used the tools ($M = 4.22, SD = 1.09$) while most non-CS ($M = 3.23, SD - 1.19$) respondents were unsure, $t(113) = 2.57, p = .028$

While we see that the Computer Science instructors had more technical understanding, they have a very similar level of confidence as to whether these new tools can be integrated into education. Similarly, the Computer Science faculty members were even less confident than other faculty members in identifying content generated by AI. This could be because of the nature of assignments (coding in CS vs. more creative writing), or simply because non-CS instructors are over-estimating their confidence. Additionally, some non-CS instructors who haven't used these tools were very surprised when shown an AI generated answer to their questions at the end of the interview. In terms of overall sentiment, computer science instructors had higher mean sentiment ($M = 4.61, SD = 0.41$) as compared to other instructors ($M = 3.95, SD = 1.11$) with $t(113) = 3.71, p < .001$.

During the interviews, it was observed that CS (Computer Science) instructors were less caught off guard and not as mesmerized by the capabilities of these tools as many Non-CS instructors were. This may be attributed to their gradual exposure to such technologies. Many CS instructors mentioned tools such as GPT, GPT-2, Codex, and Github Copilot, but the most common exposure for Non-CS instructors was to the ChatGPT (GPT-3.5 Turbo) model. CS instructors expressed their view of these tools as a change in approach rather than the end of a topic. One CS professor said:

> *"Think how many jobs [no-code solutions like] SquareSpace or Wix.com killed, but our web development class is thriving. I am not worried about it"*

On the other hand, some non-CS instructors expressed concern about their work or expertise being valued less.

3.5.5 RQ5: What are the biggest opportunities and concerns identified by the educators?

Opportunities	**Percentage**	**Concerns**	**Percentage**
Boosts efficiency	73%	Potential for cheating	38%
Thought Starter or Ideas generator	68%	Potential to stifle creativity	38%
Information at fingertips	53%	Concern about focus on product over process	36%
Automate mundane tasks	53%	Incorrect or fabricated results	27%
Personalized teaching & 24 hours TA access	31%	Equity and access	38%

Table 3.3: Biggest opportunities and concerns identified by instructors

In the followup interview, we asked the educators to discuss the biggest opportunities and challenges they see regarding adaptations of these tools in education. Table 3.3 shows the biggest opportunities and challenges identified by educators regarding these generative AI based tools.

Where there were a number of positives and negatives discussed, even the most frequent concern is discussed only 38% of time while there are four opportunities discussed over half the time, supporting the idea that educator attitudes are generally more oriented toward opportunities than concerns.

3.5.6 Additional observations

The survey results revealed a notable level of enthusiasm and optimism among faculty members concerning the integration of generative AI tools and Language Learning Models (LLMs) in education. A Business instructor said, *"This is just like the internet in 90s. This is going to change everything."* One of the significant advantages recognized by respondents is the potential for automating mundane and repetitive tasks, freeing up valuable time for educators to focus on more meaningful aspects of teaching. Some educators reported their own creative use of AI for help in grading assignments, generating personalized feedback, creating test questions and even finding flaws and biases in students' arguments. The notion of using AI as a personal tutor, catering to individualized learning needs, also garnered enthusiasm among respondents. One Arts instructor stated:

> *"I require them to use AI to complete their assignment and submit their prompts, as well as all outputs."*

AI tools are perceived as valuable aids in the creative process. Faculty members acknowledged the utility of AI in generating innovative ideas and offering fresh perspectives on complex concepts. By serving as a tool to bounce ideas off of, AI can challenge conventional approaches, encouraging educators to explore novel teaching methods and content delivery strategies. An instructor in Education said:

> *"[Generative AI] has been a lifeline for for people with learning disorders, or those who need little extra help."*

However, amidst the excitement, several unresolved questions and concerns were highlighted by the faculty members. One major concern pertains to the effective testing of students when AI tools are employed. Traditional testing methods may not adequately assess students' critical thinking and problem-solving skills when assisted by AI, as expressed by an Engineering instructor:

> *"I don't know what to test on anymore. I have no idea how to distinguish a genuine assignment with AI generated."*

Additionally, detecting plagiarism and ensuring academic integrity in an AI-driven learning environment poses a challenge. Moreover, the potential loss of creativity in a heavily AI-driven learning environment sparked debates among faculty members. Another prominent issue is the challenge of combating misinformation or fabricated information. Balancing the use of AI tools while preserving and nurturing students' creativity and originality is an ongoing concern.

3.5.7 Limitations

This survey was self reported, so it has the inherent self-reporting bias. Additionally, the survey was done in one university across different departments, so there could be variation in different institutions. This is also a fast-moving subject, and all the data reflects responses during May and June of 2023.

3.6 Conclusions

While some recent work has cast doubt on whether AI-based tools will or even should become integrated within classrooms [42,43], our findings reveal that educators are already seeing more positives than negatives. There is a general consensus from the survey and interview that these generative AI-based tools are going to be part of our education system, and being able to quickly adapt to this new reality sets the direction. While it may not be surprising that educators are aware of AI tools and are becoming more positive, this study's contribution primarily lies in identifying the factors that affect such an environment. Such information can help develop the right policies, conduct necessary training, and provide necessary resources so that we can take advantage of these tools while minimizing risks.

While the potential benefits are promising, it is crucial to navigate the complexities carefully and thoughtfully to ensure an inclusive, equitable, and effective learning experience for all students in the AI era. A larger study, encompassing a bigger sample-size, can help generalize these finding. This study also shed light on numerous big-picture philosophical questions that merit further exploration. Fundamental questions about the nature of teaching and learning in the context of AI tools need to be addressed. Existing uncertainty surrounding AI tools and LLM-based technologies in education calls for open dialogues and collaborations between researchers, educators, education policymakers, and technology developers. Together, they can address the emerging challenges, assess ethical considerations, and collectively shape the responsible integration of AI in education.

CHAPTER 4

From Guidelines to Governance: A Study of AI Policies in Education

4.1 Abstract

Emerging technologies like generative AI tools, including ChatGPT, are increasingly utilized in educational settings, offering innovative approaches to learning while simultaneously posing new challenges. This study employs a survey methodology to examine the policy landscape concerning these technologies, drawing insights from 102 high school principals and higher education provosts. Our results reveal a prominent policy gap: the majority of institutions lack specialized guidelines for the ethical deployment of AI tools such as ChatGPT. Where such policies do exist, they often overlook crucial issues, including student privacy and algorithmic transparency. Administrators overwhelmingly recognize the necessity of these policies, primarily to safeguard student safety and mitigate plagiarism risks. Our findings underscore the urgent need for flexible and iterative policy frameworks in educational contexts.

Keywords:LLM, Chatbot, ChatGPT, AI in Education, Administrator's attitude, Ethical AI Policy, Generative AI

4.2 Introduction

With the rapid advancement of technology, generative artificial intelligence (AI) tools, particularly Large Language Models (LLMs) like ChatGPT, are increasingly being adopted in various sectors, including education. These technologies offer promising avenues for pedagogical innovation, personalized learning, and administrative efficiency. However, their integration into educational settings is not without challenges, particularly concerning ethical considerations. Issues related to student privacy, data security, algorithmic transparency, and accountability are growing areas of concern.

While the application of these tools offers numerous advantages, the absence of comprehensive policy frameworks governing their ethical use in education can lead to unintended negative consequences. Inadequate policies may expose students to risks such as data misuse, algorithmic bias, and academic dishonesty. Educational institutions, thus, find themselves at a crossroads, balancing the potential benefits of emerging technologies against ethical and legal ramifications. Artificial Intelligence (AI) in education has garnered significant attention, leading to an increase in scholarly inquiries. The focus of these studies predominantly revolves around the implementation and efficacy of AI-powered educational tools, often sidelining essential discourses on policy, ethics, and administrative perspectives.

Given the escalating integration of AI tools like ChatGPT in educational settings, there is an imperative need to understand the current landscape of ethical policies, or the lack thereof, governing their use. Understanding administrators' attitudes and perceptions towards these ethical considerations is crucial for formulating effective policies that can guide responsible AI adoption in education.

The research focuses on addressing the following questions:

RQ1 What is the current landscape of policies related to Generative AI in educational settings and what do these policies cover?

RQ2 What are the perceived needs for future policy formulation in relation to Generative AI, and what recommendations can be made for an effective ethical framework?

To answer these questions, this study adopts a mixed-methods research design, incorporating both quantitative and qualitative data collected via a survey of over 100 educational administrators in the United States.

The remainder of this paper is organized as follows: Section 2 outlines the methodology, Section 3 presents the findings, Section 4 offers a discussion, and Section 5 concludes with recommendations for policy formulation.

4.3 Related Work

The integration of artificial intelligence (AI) in education is evolving rapidly, necessitating a multidimensional understanding of its applications, ethical considerations, governance frameworks, and pedagogical impacts. This section synthesizes key contributions across these areas, providing a coherent overview of the current research landscape.

4.3.1 Applications and Trends in AI in Education

Recent studies highlight significant advancements and trends in AI's educational applications. Zhai et al. [76] and Chen et al. [77] have identified critical research areas, including the Internet of Things, swarm intelligence, deep learning, and the application of natural language processing and neural networks in education. Works by Pradana et al. [78], Lo [79], and Choi et al. [80] emphasize the diverse applications of AI tools, notably ChatGPT, and the importance of addressing gaps in ethical and social considerations. Flogie and Krabonja [81] discuss the challenges and models for integrating AI into teaching, underscoring the field's evolving nature and the need for comprehensive research covering technological, ethical, and administrative aspects.

4.3.2 Ethical Challenges and Frameworks

The ethical implications of AI in education are complex, involving considerations of fairness, transparency, and privacy. Holmes et al. [82], Akgun and Greenhow [83], and Adams et al. [84] discuss the ethical challenges in deploying AI in educational settings. Halaweh et al. [85] and Sullivan et al. [86] propose frameworks for responsible implementation, emphasizing the need for policies that ensure student safety and academic integrity. Chiu [87] and Kooli [88] highlight the lack of policy considerations, calling for a balanced approach to leveraging AI's benefits while mitigating its risks.

4.3.3 Accountability, Fairness, and Governance

The governance of AI in education involves balancing technological benefits with ethical risks. Garshi et al. [89], Berendt et al. [90], and Filgueiras [91] explore frameworks for accountability and human rights in smart classrooms. Li and Gu [92] present a risk framework for Human-Centered AI, emphasizing accountability and bias. Memarian and Doleck [93], Nigam et al [94],Sahlgren [95], and Gillani et al. [96] discuss the challenges of fairness and transparency, necessity of security and privacy, ethical concerns, advocating for human-centered and politically aware governance models. Uunona and Goosen [97] explore ethical values in online education.

4.3.4 Policy Guidelines and Implications

The development of AI-specific policy guidelines is critical for ethical integration into educational systems. Miao et al. [98] and Chan [99] have contributed to guiding policymakers, though existing technology policies [100–103] often fall short in addressing AI's unique challenges. This underscores the need for more detailed and AI-focused educational policies.

4.3.5 Pedagogical Approaches and Curriculum Design

Pedagogical innovation is essential for integrating AI into education effectively. Ali et al. [104] advocate for AI literacy in curricula, while Sattelmaier and Pawlowski [105] propose a competence framework for incorporating generative AI into school curricula. Ouyang et al. [106] present a framework for understanding AI's role in learning, highlighting the shift towards learner-centric models.

4.3.6 Multidisciplinary Perspectives

A multidisciplinary approach is vital for understanding AI's impact on education. Dwivedi et al. [107] and Baidoo-Anu and Owusu Ansah [108] combine insights from various fields, addressing the capabilities and challenges of AI. Whalen and Mouza [109] emphasize the need for ethical uses.

4.4 Methodology

To gain insights into the current policy landscape regulating the use of AI tools such as ChatGPT in educational settings, as well as to understand the attitudes of educational administrators toward these policies, this study employed a survey. This survey, administered across a diverse array of educational institutions, consists of a mix of multiple-choice questions, Likert-scale questions, and free-form text entries. The survey was specifically designed to discover the current landscape of policies related to Generative AI in educational settings and the perceived needs for future policy formulation in relation to Generative AI. Influenced by prior research such as Nguyen et al. [110] and Adams et al. [84], the survey covered commonly identified policy areas and offered respondents the opportunity to express additional concerns and policy suggestions through free-form text. Section 4.4.1 outlines the questions included in the survey. Some options and language of questions were slightly changed to tailor the survey to high school and higher education administrators.

The primary focus of this study was on two groups of educational administrators: high school principals and academic officers or provosts in higher education institutions. These individuals were selected based on their pivotal roles in policy formulation and implementation within their respective organizations. The study garnered responses from over 100 administrators.

4.4.1 Survey Questionnaire

Demography

- How many years of experience do you have in education administration? — [Free Entry]
- What is the size of your student population? — [Free Entry]
- What is the size of your faculty (teaching and research) population? — [Free Entry]
- What is your school's type? — [Multiple Choice - Private/Public]

Current landscape of policies (RQ1)

- Policy on emerging technologies in place? [*Have policy/Working on a policy/No policy and not working on one/Don't know*] **M**

- How necessary is it to have a policy? [*Not/Somewhat/Very necessary*]

  The following questions are only shown if they have an AI policy :

- Current policies adequate? [*Likert scale: Strongly disagree to Strongly agree*]

- Policy specifically mentions LLMs such as ChatGPT? [*Yes/No/Unsure*]

- Which of the following elements are covered in your policy? [*Student privacy/Algorithmic transparency/Bias mitigation/Accountability mechanisms/Plagiarism/Other - Free entry*]**M**

- Primary motivations for implementing or revising policy governing use of these AI tools in the classroom? [*Stopping/Plagiarism/Ensuring student safety/Compliance with regulations/Ethical considerations/Research integrity**/Parental demand/Teachers' demand/Other - Free entry*]**M**

Perceived needs and recommendations (RQ2)

- Who should be primarily responsible for formulating policy? [*School administration/School board*/Teachers/Parent-Teacher Association*/Higher Education board**/Faculty Senate**/Independent body/Students/Other (free entry)*]**M**

- How much autonomy should individual schools have in setting or implementing policies? [*None/Some/Moderate/Most/All*]

- How much autonomy should individual teachers have in setting or implementing policies? [*None/Some/Moderate/Most/All*]

- In which areas should policies focus? [*Stopping Plagiarism/Ensuring student safety/Compliance with regulations/Pedagogical innovation/Research purposes**/Ethical* *considerations/Student engagement/Using these tools to help reduce the teacher's workload/Other (free entry)*] $^{\mathbf{M}}$

- What kind of support or resources would be helpful for your institution to create and implement policies? [*Professional development/Consultation with tech companies/Consultation with legal or ethics experts/Funding or resources/Model policies or guidelines from successful schools or districts/Other (free entry)*] $^{\mathbf{M}}$

- Are there any specific policy components that you believe should be included in guidelines? [*Free entry*]

Other Questions (RQ4)

- Overall opinion of LLMs? [*Likert scale : Dislike a great deal to Like a great deal*]

- Do you have a policy that allows for punishing students based on results from AI-detection tools?[*Such tools are banned/Such tools are used to narrow down but not as only factor to decide/Student can be punished based on the result of such tool-detected AI content. [Tool name]*]

- Additional comments. [*Free entry*]

- Interested in a follow-up interview?

Options: $^{\mathbf{M}}$ **Multiple selection allowed; * only to high school administrators, ** only to higher ed administrators**

4.4.2 Data Collection Instrument

The survey, structured to align with the four primary objectives of the study, was hosted on the Qualtrics platform. It featured both closed-ended questions, aimed at capturing quantifiable metrics, and open-ended questions designed to explore the subjective viewpoints and rationales of administrators.

For distribution, we utilized a publicly available directory to identify and reach out to high school principals. We downloaded the mailing list of school administrators from the state education board's website. Conversely, for higher education institutions, we employed a manually curated mailing list. To do this, we first obtained a list of all higher education institutes in the states, went to their websites, and looked up their provost's or chief academic officer's email. The survey was distributed across diverse geographic locations within the United States across Arkansas, Massachusetts, New Mexico, Utah and Washington to capture a wide range of perspectives. Survey responses were collected between June 19, 2023 and September 26, 2023.

4.4.3 Data Analysis

We performed χ^2 tests for each response against each of institution size, geographic location, and governance model (public or private). We also ran Pearson correlation tests for relation between need for policy, sentiment about AI tools, autonomy preference against administrators' experience length and student population. These tests were not significant.

4.5 Results

State	High Schools	Higher Education	Total
Arkansas	15	6	21
Massachusetts	13	3	16
New Mexico	19	4	23
Utah	18	5	23
Washington	16	3	19
Total	81	21	102

Table 4.1: Survey responses by institution type and states

We received over 126 survey responses from across five states, some of which were partially completed. We had 102 complete surveys that we use for analysis for this study. Table 4.1 shows the number of responses from each state and type of educational institution.

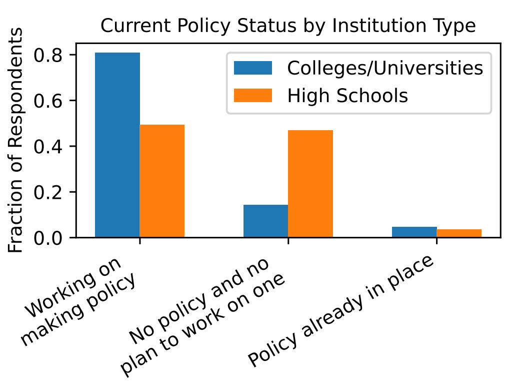
4.5.1 RQ1: What is the current landscape of policies related to Generative AI in educational settings and what do these policies cover?

The first research question investigates the presence and key components of policies or guidelines governing the use of emerging technologies such as Large Language Models (LLMs) and ChatGPT in educational environments.

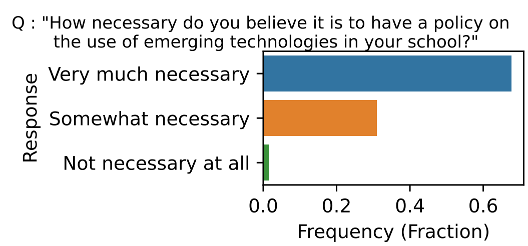
Existence of current policies

A majority of respondents indicated either ongoing efforts to formulate generative AI-related policies or the existence of established policies. Specifically, over 80% of higher education institutions reported active policy development, 5% already have a policy, and 15% have no plans to enact one. In contrast, only 50% of high schools are in the process of policy formulation, while approximately 45% neither have a policy nor plans to develop one. Figure 4.1a depicts these data. A statistically significant difference in policy status between high school and college was observed $\chi^2(2, N = 102) = 0.7.44, p = .0.024$ indicating that high schools are less inclined to work on policies than higher educational institutions.

Having a very small sample size in each category doesn't allow us to analyze and understand differences between the categories, but we can still understand a lot with the holistic review of the data.

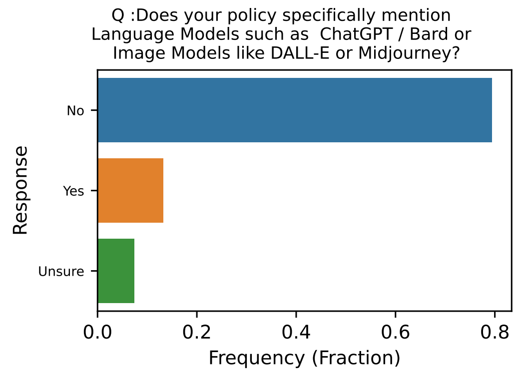


(a) Current policy status by institution type

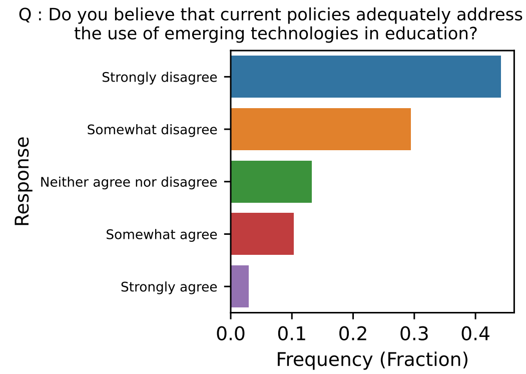


(b) Necessity of AI policies

Fig. 4.1: Administrators' responses on policy status and necessity



(a) Specificity of in-place or in-progress policies in covering AI models



(b) Adequacy of in-place or in-progress policies

Fig. 4.2: Administrators' responses on policy availability and adequacy

When asked if they need to have AI related policy, the prevailing sentiment among administrators was a critical need for these policies. Figure 4.1b shows the response on necessity of such policies. It can be seen that the necessity of AI related policy is almost universally agreed upon.

Adequacy of Current Policies

Administrators who reported the existence or development of policies were subsequently asked about what is covered on their AI policies and their adequacy. The majority expressed that current or in-progress policies inadequately address the integration of emerging technologies. Figure 4.2b shows the administrators' perceptions on the adequacy of existing or in-development policies. Even for many policies currently in development, administrators think these policies are not adequate.

Notably, only a small minority of these policies specifically mention LLMs like ChatGPT or Bard or image models like DALL-E. Figure 4.2a illustrates these findings, suggesting an awareness gap in tailoring policies to specific technological challenges.

We also asked administrators what their current or in-progress policies covered. Existing policies most commonly address issues like plagiarism, while elements like bias mitigation and algorithmic transparency are less frequently covered. Ethical considerations' emerged

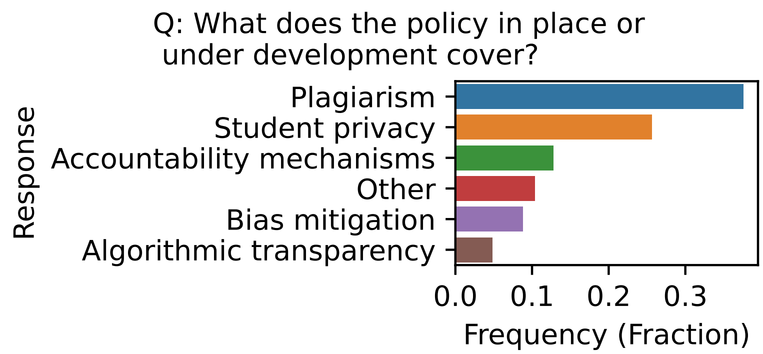


Fig. 4.3: Components included in existing or in-development policies (multiple selection allowed)

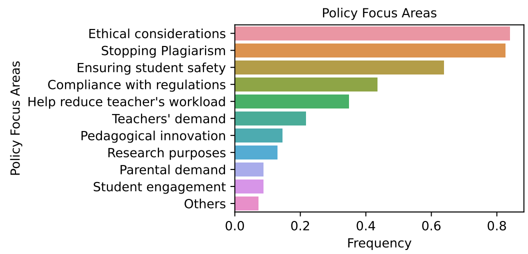
as the most frequently cited motivation (25.6%) for policy development or revision. This was followed by 'Ensuring student safety' (16.4%). Least cited were 'Parental demand' and 'Teachers' demand', both under 5%. Figure 4.3 indicates areas covered by current or in-progress policies. This indicates a perceived gap between existing governance mechanisms and the requirements for ethical and effective technology integration. Statistical tests revealed no significant associations between policy aspects and institution type, size, or location.

4.5.2 RQ2 : What are the perceived needs for future policy formulation in relation to Generative AI, and what recommendations can be made for an effective ethical framework?

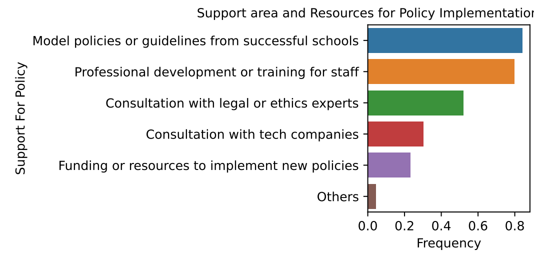
Our second goal of this study was to understand the key elements that educational administrators believe should be included in a policy framework for the ethical use of emerging technologies like ChatGPT in education as well as their overall sentiment on the policy and gather any additional insight and recommendation from the administrators.

Quantitative Analysis

Quantitatively, the focus was on the areas that respondents believe policies should primarily target and the kinds of support or resources they consider would be helpful for their



(a) Focus area for policy identified by the administrators (Multiple selection allowed)



(b) Important support resources identified by the administrators (Multiple selection allowed)

Fig. 4.4: Administrators' response in policy focus area and resources needed

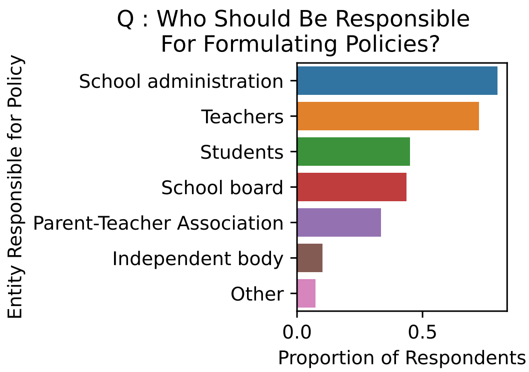
institutions. Question "In which areas should policies for the use of emerging technologies in education primarily focus?" allowed multiple selections as well as free form text entry to capture administrators' focus area for policy making. Figure 4.4a shows the policy focus area identified by school administers. The majority of respondents highlighted 'Ethical Considerations' and 'Stopping Plagiarism' as the top two areas, with over 80% of responses, followed by ensuring students' safety and compliance with regulations.

We also asked the administrators about the support resources that would help them make or update generative AI related policies. The administers' answers are shown in Figure 4.4b. A model guidelines from successful school or district was the most commonly deemed useful resources, followed by professional development and staff training and legal/ethical consultations. Need for funding and resources and consultation with tech companies were also identified.

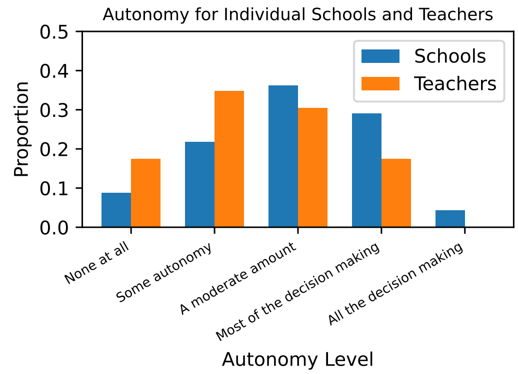
Centralized Oversight vs. Decentralized Autonomy

The responses indicate a diverse perspective on who should be responsible and involved for formulating the policies governing the use of emerging technologies like ChatGPT in education. School administrators are seen as the most responsible entities, followed by teachers and students, along with school board and parent-teacher association. Figure 4.5a shows the responsible entities identified for policy making purposes.

As for the autonomy and decision making given to schools and teachers, the respondents widely varied. For schools, the responses ranged from 'none' to 'all,' while the responses for



(a) Responsible entity for policy-making (Multiple selection allowed)



(b) Autonomy (decision making) for individual school and individual teachers

Fig. 4.5: Administrators' responses on responsible entity and autonomy

teachers' decision-making ranged from 'none' to 'most.' Figure 4.5b shows the response for the question about autonomy and decision making power for schools and teachers respectively. Interestingly, none of the administrators responded that teacher should have all the decision making power (total autonomy).

Overall, the data suggests a preference for a collaborative approach to policy formulation and implementation that includes various stakeholders at different levels of governance.

Qualitative Analysis

The qualitative analysis was based on free-form text entries. While the number of responses was too limited to be able to perform a qualitative coding and analysis, they provided valuable insights. Respondents expressed concerns about the rapid advancements in technology and the need for policies to be flexible and adaptive, offering some explanation for why so few policies are currently in place. For example:

- *"I believe that any policy should be reviewed and updated annually to keep up with advancements in technology."*

- *"The emerging AI platform will continue to grow and policies need to be flexible enough to adapt."*

- *"This area of technology is moving so quickly that it's hard for policy to keep up."*

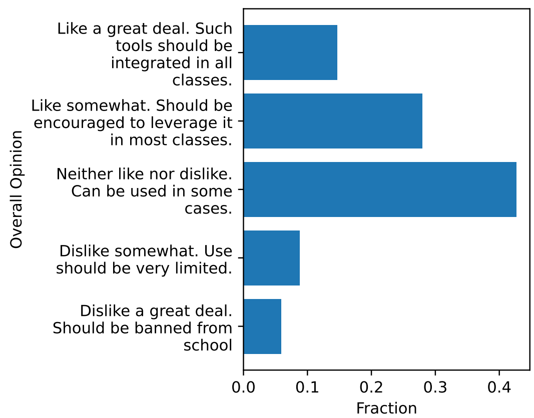
They also emphasized the importance of considering ethical implications, including potential biases in AI algorithms. One respondent noted, *"I am concerned about the potential for bias in AI and think this should be addressed in any policy."* Others emphasized the ethical and privacy aspects, stating, *"The policy must take into consideration the ethical implications of using AI in an educational setting.", and "I think privacy and data protection should be at the forefront of any policy concerning the use of AI technologies."* These quotes reflect the overarching sentiment that while technology is advancing rapidly, policies need to be robust yet flexible to adapt to these changes. Even when administrators are not clear what should be in the policy, they are quick to point out we have to be very careful on whatever policy we make:

- *"I'm not sure what the policy should contain, but I know it needs to be created carefully and with a lot of thought."*

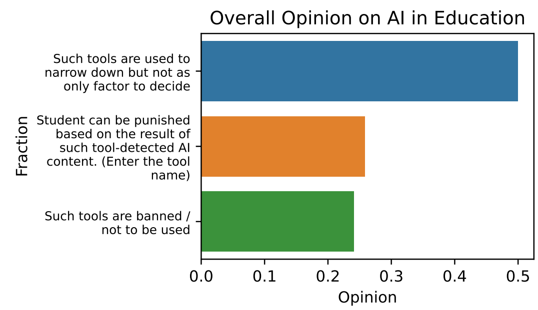
- *"We are observing how AI impacts student learning and will be formulating a policy based on these findings. We are deliberately being very careful"*

Additional Observations

Additionally, we asked a couple of questions to understand the overall sentiment about AI tool as well as sentiment about existing detection tools. Figure 4.6a shows the overall opinion from these administrators. Most of the administrators are either indifferent or positive, and very few are not in favor of the technology. When asked about the use of existing tool that claim to detect AI-generated content, about half of the respondent were in favor of using such tools to narrow down, but not as a final arbiter of truth. The remaining respondents are almost evenly split between banning such tools and using such AI-detection tools. Figure 4.6b shows the response for that question. We hypothesize that the high unreliability of these detection tools, their black-box nature and high cost of catching false positive are making the administrators take cautious approach towards detection tools.



(a) Overall opinion about AI in Education among school administrators



(b) Overall opinion on existing AI generated content detection tool

Fig. 4.6: Overall opinion on AI and AI detection tools

4.6 Conclusions and Discussion

This study aimed to address two primary research questions (RQs) regarding the policy landscape for AI and LLM-based tools like ChatGPT in education. RQ1 explored the current state of policies and their coverage, revealing a significant push, especially in higher education, to develop guidelines. Yet, these policies often fall short of addressing the unique challenges of technologies like LLMs. The necessity of policy development was universally recognized among administrators, driven by ethical considerations and student safety, though areas like algorithmic transparency and bias mitigation were less emphasized, indicating gaps in existing frameworks.

RQ2 investigated the perceived needs for future policy formulation and proposed recommendations for an ethical framework. A preference for a collaborative, multi-stakeholder approach was evident, alongside the recognition that policies must be iterative and adaptable to keep pace with technological advances.

The findings indicate an active acknowledgment of AI and LLM's potential in education, alongside a nascent governance stage for their ethical and practical integration. Notably, the disparity in policy development between higher education and high schools—where about 40% lack any policy efforts—points to potential resource or awareness discrepancies. This study underscores the critical gaps in policy adequacy and the necessity for policies to evolve alongside educational technologies. It emphasizes the importance of multi-stakeholder dialogues for creating governance mechanisms that are robust yet flexible enough to accommodate rapid technological changes.

The study concludes that the ethical and responsible integration of AI in education demands the continuous evolution of policies, practices, and attitudes. The findings of this study suggest for strategic, ethical, and collaborative governance, highlighting the imperative for developing comprehensive, adaptable policies to navigate the advancing landscape of AI technologies in educational settings.

4.6.1 Future Work

This study has laid important groundwork in understanding the state and direction of policies related to AI and LLMs in educational settings. However, several avenues for future research remain. The disparity in policy development between higher education and high schools warrants a more granular investigation. Future studies could focus on identifying the barriers and facilitators that influence policy-making at these disparate educational levels, possibly extending the research to include primary schools. Additionally, the evolving nature of AI and LLM technology itself calls for longitudinal studies that can track changes in administrative attitudes, policy adequacy, and implementation efficacy over time.

Another fruitful avenue for future work would be the exploration of multi-stakeholder perspectives, incorporating not just administrators but also teachers, students, and parents. Understanding these groups' attitudes and requirements could offer a more holistic view of what effective, comprehensive policies should entail. Investigations into the actual impact of AI and LLM-based tools on educational outcomes, based on these inclusive policies, could also provide valuable data for administrators and policy-makers.

4.6.2 Threats to validity

Our survey was not validated and no evaluation of reliability was made. Furthermore, all respondents were from institutions based in the United States, limiting external validity internationally. We did not collect any demographic information of participants. Finally, generative AI is a fast-moving technology and attitudes and policies are likely also changing quickly. This work represents a shapshot of policies and attitudes in mid-2023.

CHAPTER 5

Coding With AI: How Are Tools Like ChatGPT Being Used By Students In Foundational Programming Courses

5.1 Abstract

Tools based on generative artificial intelligence (AI), such as ChatGPT, have quickly become commonplace in education, particularly in tasks like programming. We report on a study exploring how students use a tool similar to ChatGPT, powered by GPT-4, while working on Introductory Computer Programming (CS1) assignments, addressing a gap in empirical research on AI tools in education. Utilizing participants from two CS1 class sections, our research employed a custom GPT-4 tool for assignment assistance and the ShowYourWork plugin for keystroke logging. Prompts, AI replies, and keystrokes during assignment completion were analyzed to understand the state of students' programs when they prompt the AI, the types of prompts they create, and whether and how students incorporate the AI responses into their code. The results indicate distinct usage patterns of ChatGPT among students, including the finding that students ask the AI for help on debugging and conceptual questions more often than they ask the AI to write code snippets or complete solutions for them. We hypothesized that students ask conceptual questions near the beginning and debugging help near the end of program development do not find statistical evidence to support it. We find that large numbers of AI responses are immediately followed by the student copying and pasting the response into their code. The study also showed that tools like these are widely accepted and appreciated by students and deemed useful according to a post-usage student survey. Furthermore, the findings suggest that the integration of AI tools can enhance learning outcomes and positively impact student engagement and interest in programming assignments.

Keywords: LLM, Chatbot, ChatGPT, BARD, AI in Education, AI Usage in Programming, Keystrokes

5.2 Introduction

The advent of generative artificial intelligence (AI) has ushered in a new era across various sectors, including education. Among these AI advancements, tools like ChatGPT, particularly those powered by Generative Pre-trained Transformers (GPT), have garnered increasing attention. Their integration into educational practices, particularly in programming and computer science education, signifies a notable shift in instructional methodologies. This shift raises questions about the role and effectiveness of these tools in enhancing student learning outcomes, particularly in foundational courses such as Introductory Computer Science.

This research aims to delve into the burgeoning field of AI application in education, focusing on the usage and impact of a ChatGPT-like tool in introductory programming class(CS1) coding assignments. Specifically, the study addresses these three research questions:

RQ1 *How do students employ generative AI-based tools, such as ChatGPT, while completing their CS1 coding assignments?* This question seeks to uncover the manner in which students utilize these tools, focusing on their prompts and responses.

RQ2 *What discernible patterns emerge from students' usage of this tool during assignments?* By analyzing students' keystrokes before and after engaging with the AI tool, this question aims to elucidate the nature of engagement and the type of support provided by the AI tool.

RQ3 *Does a tool like ChatGPT make programming classes more accessible, improve students' efficiency, or help new programmers learn programming?* This question investigates the broader impact of AI tools on the accessibility and efficacy of programming education.

To investigate these questions, the study utilized participants from two sections of a CS1 class, incorporating a custom GPT-4 tool designed for assignment assistance along with the ShowYourWork [111] plugin to the PyCharm integrated development environment(IDE) for recording keystrokes. Additionally, a post usage survey was conducted to collect students' feedback. This approach allowed for a comprehensive analysis of student interactions with the AI tool and a comparison of their performance in assignments completed with and without the aid of AI. The subsequent sections of this paper will detail the methodology, present the findings, and discuss the implications of these results in the context of modern computer science education.

5.3 Related Works

Recent advances in generative AI and natural language processing have enabled the development of sophisticated large language models (LLMs) like GPT-4, Codex, GitHub Copilot, and ChatGPT. These models are not just technical marvels but have profound implications for computing education research and practice [112].

Empirical evaluations of these LLMs in programming courses reveal their robust performance in tasks and assessments typical of such environments [46,47]. GPT-3, for instance, achieved about a 78% score on CS1 exam questions, surpassing many students, when the best out of 100 generated samples was chosen [48]. In more complex CS2 assessments, Codex's performance was on par with top-quartile students [49]. Similarly, GitHub Copilot demonstrated its efficacy by generating solutions that met the requirements of introductory programming assignments [50]. This notion is supported by Phung et al., who benchmark ChatGPT and GPT-4 against human tutors, demonstrating the near-human capabilities of these models in programming education [113].

These findings indicate the necessity of rethinking curriculum design and assessment strategies in the era of LLMs. There's a growing consensus on shifting focus from basic coding skills to higher-order thinking and problem-solving abilities [48]. Additionally, the advent of LLMs necessitates new forms of assessment to deter plagiarism and ensure genuine understanding [47,51,52].

In terms of pedagogy, LLMs offer promising avenues for automating the generation of solutions, explanations, and examples, potentially reducing instructor workload and enhancing learning [41,55,56]. They also enable innovative active learning strategies, including personalized assistance, peer reviews, and interactive coding activities [59,61,114]. However, caution must be exercised due to the risk of propagating incorrect information [41].

Sarsa et al. explore the use of Codex for generating programming exercises and code explanations, highlighting its potential for reducing instructor workload and enhancing learning, albeit with the need for quality oversight [56]. Shin and Nam survey automatic code generation from natural language, suggesting future research directions for improving this paradigm [115]. Watermeyer et al. examine the impact of generative AI on academia, discussing the balance between potential benefits and the reinforcement of existing challenges in the academic landscape [116]. Chiu's study investigates the effects of generative AI on school education, emphasizing teachers' perspectives on adapting to these technologies [87].

In the context of computing education, Zastudil et al. report on interviews with students and instructors, highlighting their perspectives on the use of generative AI tools and the emerging concerns and preferences for their integration [117]. Hedberg Segeholm and Gustafsson evaluate the use of generative language models for automated programming feedback, underscoring their potential in easing instructors' burden [118].

Kazemitabaar et al. delve into how novices use LLM-based code generators, revealing various approaches and the implications for self-regulated learning and curriculum development [119]. Zhang et al. focus on students' perceptions of AI-generated feedback in programming, emphasizing the need for specific, corrective feedback [120]. Carr et al.'s experiment with ChatGPT in database education shows its efficacy in generating SQL queries, suggesting new avenues for teaching and assessment [121]. Yilmaz and Karaoglan Yilmaz examine students' views on using ChatGPT for programming learning, revealing its advantages and limitations [122].

Surameery and Shakor explore Chat GPT's use in debugging, highlighting its potential as part of a comprehensive debugging toolkit [123]. Popovici assesses ChatGPT's potential in a Functional Programming course, discussing its strengths in generating code reviews [124]. Husain provides insights into programming instructors' perceptions of ChatGPT, contributing to the discourse on AI integration in programming education [125]. Speth et al. investigate the use of AI-generated exercises in programming courses, sharing insights on their quality and the time-saving aspect of using ChatGPT [126]. Wieser et al. explore ChatGPT's role in text-based programming education, underscoring its utility in supporting both students and teachers [127].

Finally, the literature highlights several challenges posed by LLMs, such as the potential for over-reliance, which may hinder learning [63], and issues surrounding assessment integrity [51]. Concerns about plagiarism detection [50], inherent biases in AI systems [64], and broader socio-economic impacts [65] also warrant attention. This underscores the urgent need for further research to develop evidence-based methodologies for integrating LLMs effectively in computing education while addressing their potential drawbacks.

5.4 Methodology

5.4.1 Participants

Our study was done in compliance with a protocol approved by our university's institutional review board (IRB). The participants of this study were students enrolled in two sections of a Computer Science 1 (CS1) course at our institution, a mid-sized research university in the United States. The study commenced during the final two weeks of the Fall 2023 semester. Students were given two programming assignments and given the option of using a tool based on LLMs for assistance. Students did not need to participate in the study in order to use the AI tool.

Programming assignments were to be completed in Python. The first programming assignment involves writing a graphical car racing game. Starter code provides the graphical structure and render loop. Students are asked to be creative in designing the game play. The second programming assignment provides starter code that provides a menu to allow the user to sort a deck of cards and search for specific cards. There are logic errors in

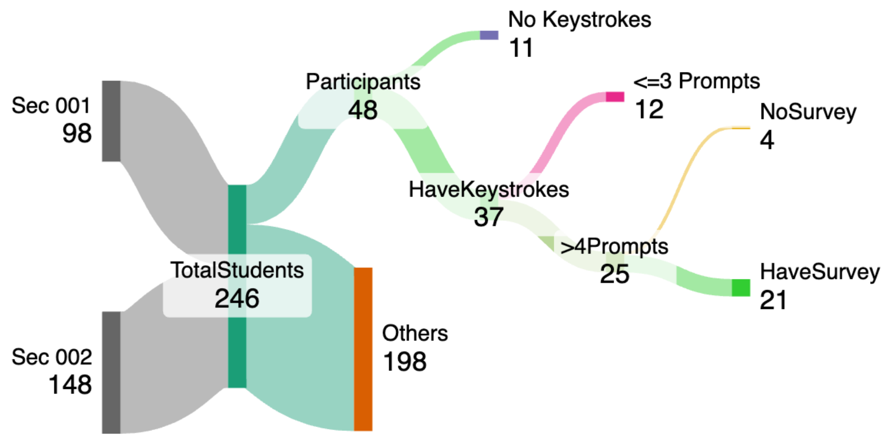


Fig. 5.1: Breakdown of the study participants

the starter code which make the program give the wrong results. The student is asked to identify and fix the errors. The assignments were designed to reinforce learning objectives related to methods, classes, objects, and operator overloading, involving work with multiple files and starter code.

A total of 48 students from both sections, out of 246, participated in the study. However, not all participants contributed to the dataset equally; some did not submit their keystroke data, and others did not engage with the LLM-based tools sufficiently to be included in the full data analysis. Ultimately, the keystroke data and AI tool usage data from 25 students were used for in-depth analysis. No demographic data were collected from the participants. The only background information gathered was regarding their prior programming experience, through a single question on the subject. Of the participants, 21 completed the post-assignment survey, providing valuable insights into their experiences and perceptions of using AI tools in their assignments. This selective participation and data contribution highlights the varied engagement levels with both the study and the LLM-based tool, underscoring the need for further investigation into factors influencing students' willingness to utilize such technologies in educational settings.

5.4.2 Tools

We primarily used the following three tools for data collection purpose:

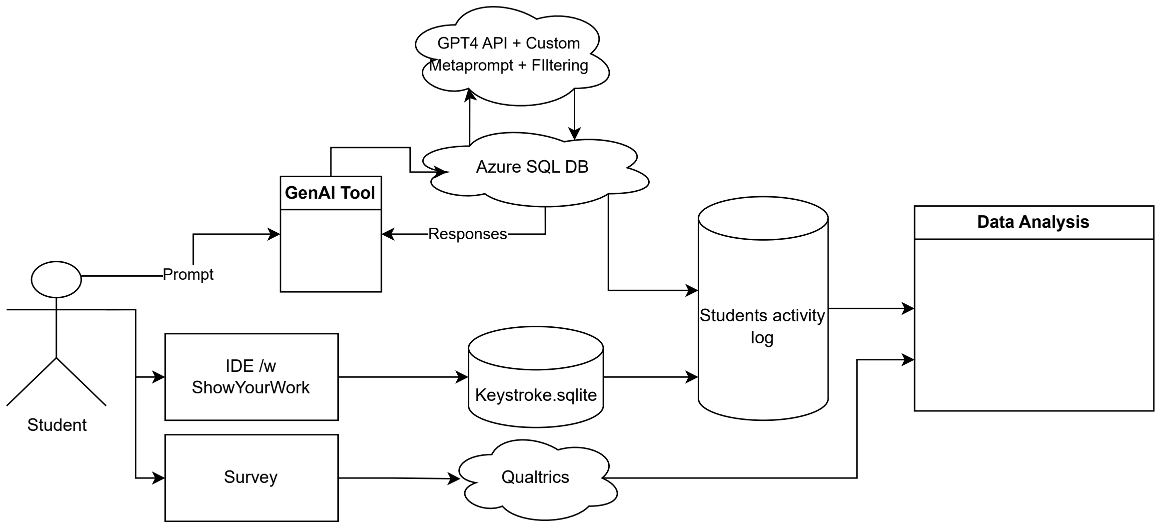


Fig. 5.2: Architecture of Custom GPT-4 tool

- **Custom GPT-4 Powered Tool for Assignment Assistance**: This tool, a wrapper around GPT-4, was specially designed for the study. It enabled the logging of prompts submitted by students, responses generated by GPT-4, and additional data such as timestamps, context, and follow-up counts. This setup allowed for a detailed analysis of the interactions between students and the AI, providing insights into how students leverage AI assistance in solving programming tasks. There was an extensive guardrail in the tool on top of OpenAI's guardrail to not let AI tool respond to questions that were not related to programming. Figure 5.2 shows the simplified architecture of the tool.

- **ShowYourWork Plugin for Keystroke Recording**: All students in the course were required to install the ShowYourWork plugin into the PyCharm IDE. ShowYourWork logs all keystrokes made within the PyCharm IDE during assignment completion. This data was crucial for understanding the coding process and habits of the students, offering a granular view of their programming workflow.

- **Post-Assignment Survey**: After completing the assignments, students were asked to fill out a survey. This survey gathered information about their prior programming experience and their perceptions of the usefulness of the AI tool in the assignment, helping to contextualize the quantitative data with qualitative insights.

5.4.3 Data Collection and Analysis

In collaboration with the course instructor (not an investigator in this study), students were instructed to complete their coding assignments with the option of freely using the custom GPT-4 powered tool. During this process, the ShowYourWork plugin continuously recorded their keystrokes, while the AI tool archived all student prompts and corresponding LLM responses in a structured database. This comprehensive dataset was pivotal for our analysis.

The analysis of the collected data focused on several key aspects:

- Identifying Patterns in AI Tool Usage: This involved examining the types of prompts given by students, the nature of GPT-4 responses, and how these interactions correlated with different stages of the assignment. The aim was to uncover how students navigate the problem-solving process with AI assistance.

- Analyzing the keystrokes data submitted alongside the assignments: Analyzing the keystrokes data provided insights on what was happening before, during and after the student use AI tool to get help with the assignment.

- Analyzing the survey : Analyzing the post-completion survey provided the direct feedback from students who were using the tool for their last two assignment.

This approach allowed for a multifaceted evaluation of AI tool integration in educational settings, illuminating both the the usage pattern as well as students' opinion and attitude towards such tool.

5.5 Results

In this section, we attempt to answer our research questions by analyzing the data collected from prompts, keystrokes, and surveys.

		GPT-4 rating				
		Complete	Part	Debug	Conceptual	total
Human rating	Complete	1	0	0	0	1
	Part	0	2	1	1	4
	Debugging	0	0	10	0	10
	Conceptual	0	1	0	4	5
	Total	1	3	11	5	20

Table 5.1: Inter-rater reliability between human and GPT-4

5.5.1 RQ1: How do students employ generative AI-based tools, such as ChatGPT, while completing their CS1 coding assignments?

A. Prompt type

The custom tool was programmed to not answer any questions that were not related to these topics, using meta prompting and system instructions. Students could ask any questions about computer science and mathematics to the AI tool. We first categorized the prompt types into 4 categories according to the following definitions:

1. **Debugging Help**: Prompts that seek help to identify or fix errors in the provided code snippet.

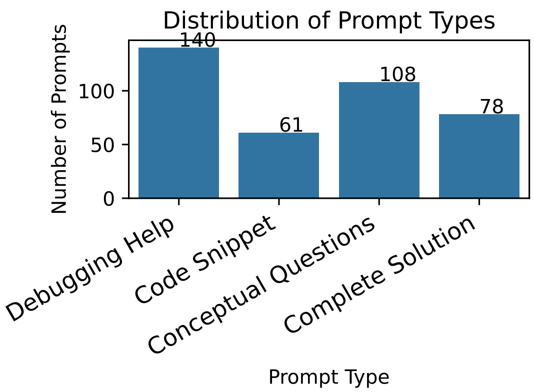
2. **Code Snippet**: Prompts that ask for a specific part of the code, like a function or a segment.

3. **Conceptual Questions**: Prompts that are more about understanding concepts or algorithms rather than specific code.

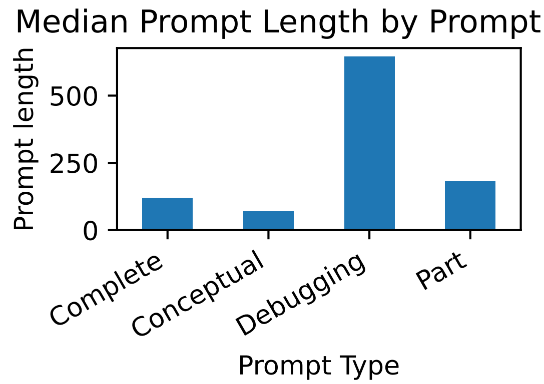
4. **Complete Solution**: Prompts that request an entire solution or a complete code snippet.

We leveraged OpenAI GPT-4 Turbo for categorizing the prompts using meta-prompting. Table 5.1 shows the agreement between human raters and GPT-4 on categorizing the prompts. We then calculated the inter-rater reliability between human raters and GPT-4.

For percentage agreement metrics, we observed a percentage agreement of $\frac{17}{20} = 85\%$ with Cohen's Kappa $\kappa = 0.76$.



(a) Type of prompts



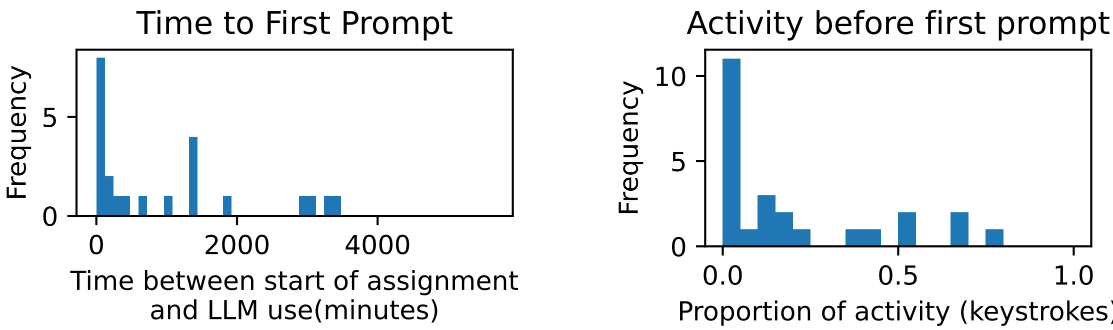
(b) Bar chart showing median prompt length to AI tool in no. of characters

Fig. 5.3: Prompt type and prompt length

Figure 5.3a shows the count of various types of prompts. Asking for help with debugging code and asking conceptual questions were the most common types of prompts, as opposed to asking for full or partial code directly. Figure 5.3b shows the bar chart of the median length of prompts sent to the AI tool. The median prompt length for Debugging prompts was over 500, whereas the median prompt length for each of the other three prompt types was under 250. This makes sense, since when asking for help debugging the student will include the code in the prompt. The plot indicates that most of the prompts are under 200 characters, meaning the most common prompts did not contain starter code but were rather more conceptual in nature or had a smaller code snippet.

5.5.2 RQ2: What discernible patterns can be identified from the prompts and responses exchanged between students and the LLM during the assignment?

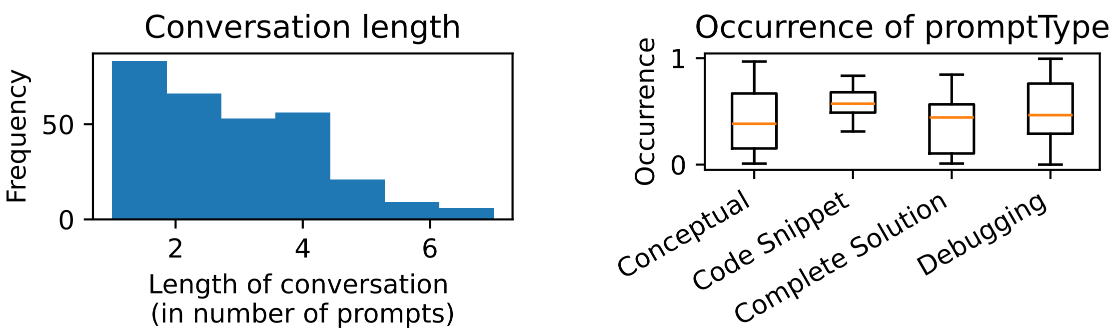
By examining the students' keystrokes before and after their engagement with the AI tool, this question seeks to understand the nature of engagement and the kind of support



(a)

(b)

Fig. 5.4: Time and activity before the first LLM call. (a) Time in minutes between the start of the assignment and the first LLM call. (b) Histogram of how the percentage of file edit events that were completed prior to the first AI prompt.



(a)

(b)

Fig. 5.5: (a) Histogram showing conversation length. (b) Occurrence of prompts.

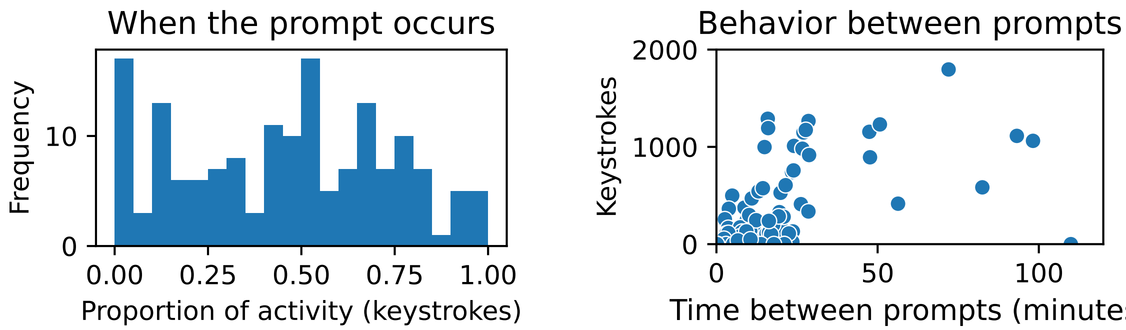
provided by the AI tool. We first examined when in the coding process students use the AI assistance. Figure 5.4a shows the time elapsed between the start of the assignment and the first use of AI, and Figure 5.4b shows the proportion of activity between the start of the assignment and the first AI prompt. Since the time graph is mostly spread over four days and the activity proportion occurs within the first one-fifth of the time of activity, it shows that students start slow on their assignments and don't immediately use the AI tool. However, students who use the AI tool use it at least once before they complete one-fifth of their assignment.

We hypothesized that the type of prompts students made (e.g., Debugging Help) would change as students progressed toward completion of the assignment. For example, we expected that students would ask more conceptual and/or complete solution types of questions near the beginning and debugging questions in the middle and at the end of development. We found no statistical support for this hypothesis. Median percentages of assignment completed for each query type were relatively close to each other (Debugging Help: 46%, Code Snippet: 57%, Conceptual Questions: 39%, Complete Solution: 44%). See Figure 5.5b. We performed two Mann-Whitney U tests between Conceptual Questions and Complete Solution with no statistical significance ($U = 407, p = 0.55$) and between Conceptual Questions and Debugging Help ($U = 1037, p = 0.19$). This could mean that, indeed, students ask varied types of questions throughout development, or that our sample size is insufficient to statistically detect the patterns.

We define a conversation chain as a series of prompts and responses with the AI tool that are uninterrupted by closing the webpage, a webpage refresh, or by a refresh of the context. The length of a conversation chain is the number prompts in the chain and is limited to seven, after which the context is refreshed. Most of the questions students asked for CS1 assignments were conceptual or debugging questions; therefore, the conversation chains were usually short. This means most of the student queries were solved in one or two responses. Figure 5.5a shows the histogram of conversation chain length.

Figure 5.6a shows proportion of work done when each prompt is made in terms of keystrokes. As expected, there is an initial burst of AI prompt activity as students are starting their assignments. Interestingly, usage appears to continue throughout program development. In addition, there appear to be roughly 10 prompts right near the end of development.

Figure 5.6b shows the time and keystrokes between two consecutive prompts by the same student. Most of the consecutive prompts occur within the first 5 to 15 minutes. This indicates that students are not spending a lot of time between prompts but rather trying the solution, varying their prompt, and asking again within a short time period. As expected, and as we see in the discussion below, many prompts are separated by few keystrokes but a

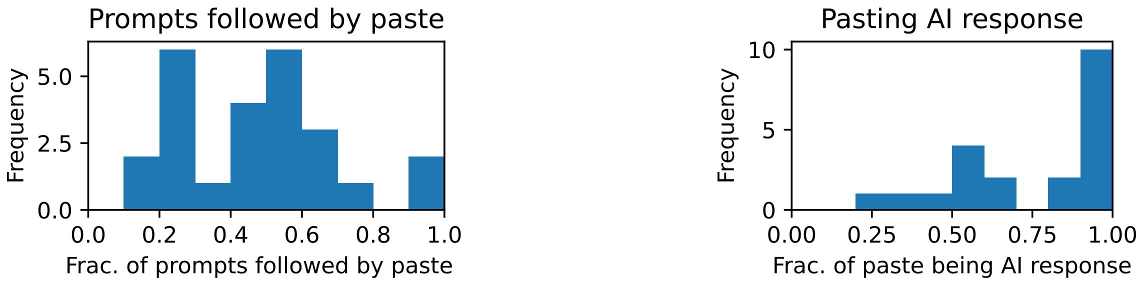


(a) (b)

Fig. 5.6: (a) Histogram of proportion of activity when AI is prompted. (b) Scatter plot of the number of keystrokes vs time (in minutes) between prompts. Prompt pairs with greater than 120 minutes between them (there are 21 such pairs) are not shown.

big paste, indicating students are copying the AI response into their code (this was allowed). However, a surprise is the number of cases where, in the short time between prompts (1-30 minutes), students typed 500 or even 1000 characters. These students engage in a flurry of programming activity between prompts, possibly trying out ideas from the AI response, or possibly typing in the AI-generated code instead of pasting it.

Next, we explored the activity that occurs immediately following a prompt to the AI. Figure 5.7a shows the histogram of the proportion of LLM calls that were followed by a large paste event of more than 20 characters for each student. For example, the figure shows that six students pasted the output from the AI tool into their code about half the time. All students pasted response text at least once. For this analysis, we only looked at prompts that were not classified as asking conceptual questions (which is the second most common type of prompt). We confirmed that the paste text came from the last response of the AI by testing if the pasted text was a substring of the AI response. Figure 5.7b shows that half of the pastes were exactly the AI response. For prompts asking for code or help with debugging, students often end up copying and pasting the response.



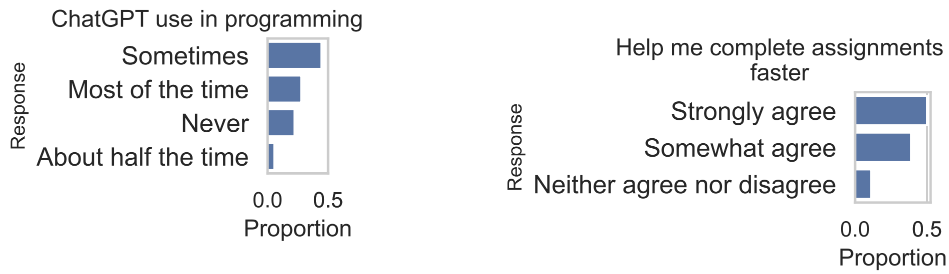
(a) (b)

Fig. 5.7: Paste activity following prompts. (a) Histogram of percentage of AI calls followed by a big 'paste' event (over 20 characters) for each student. (b) Percentage of paste events that are direct substrings of the AI response.

5.5.3 RQ3: Does a tool like ChatGPT make programming classes more accessible, improve students' efficiency, or help new programmers learn programming? How do students feel about such a tool?

We conducted a post-assignment survey, and students expressed that the tool was useful and helped them complete assignments more quickly. Figure 5.8a shows responses to the statement, "How often do you use tools like ChatGPT for help in your programming assignments?" It reveals that students are already using similar tools in their programming classes. Less than a third of the students said they never use it, hence the remaining two-thirds are using it in at least some capacity. Figure 5.8b shows responses to the statement, "The provided AI tool helped me complete the assignment faster." The vast majority (90%) of respondents agreed, and only 10% were neutral.

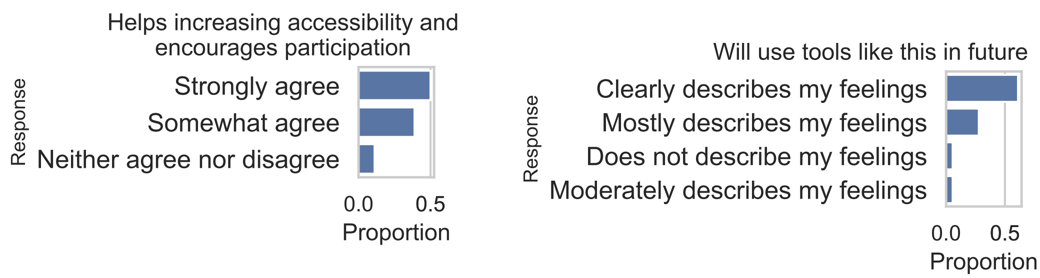
Programming can be an intimidating subject for some students, and tools like these have been touted for their potential as a personalized tutor. Figure 5.9a shows students' responses to the statement, "Tools like these help increase the accessibility of programming classes or encourage me to take programming classes." Again, over 90% of respondents agreed, with only 10% neutral and no one disagreeing. Finally, Figure 5.9b shows responses to the statement, "If offered, I would use tools like this one in future classes or assignments." Over 85% of students mostly agreed with the statement.



(a) Response to: "How often do you use tools like ChatGPT for help in your programming assignments?"

(b) Response to "The provided AI tool helped me complete the assignment faster"

Fig. 5.8: Survey Responses



(a) Response to: "Tools like these help increase the accessibility of programming classes or encourage me to take programming classes"

(b) Response to "If offered, I would use tools like this one in future classes or assignments"

Fig. 5.9: Survey Responses

5.6 Conclusion and Discussion

This study aimed to explore the impact and usage patterns of generative AI-based tools, like ChatGPT, on student performance and engagement in CS1 coding assignments. Through detailed analysis of interactions between students and the AI tool, as well as students' keystrokes and survey responses, several key findings emerged.

Firstly, the integration of a custom GPT-4 powered tool in programming assignments revealed significant usage among students, particularly for debugging and conceptual understanding. This suggests that such tools can serve as effective aids in the learning process, potentially reducing the time students spend stuck on particular problems and enhancing their overall learning experience.

The analysis of keystroke data and AI tool interactions indicated that students primarily used the AI tool for assistance with debugging and conceptual questions, with most interactions resulting in short conversation chains. This finding points to the efficiency of AI tools in providing targeted, immediate assistance, which, in turn, may contribute to improved problem-solving skills and deeper conceptual understanding.

Survey responses further supported the utility of the AI tool, with a vast majority of students reporting that it helped them complete assignments faster and made programming classes more accessible. These perceptions highlight the potential of AI tools to lower the barriers to entry for novice programmers and to support diverse learning needs in computer science education.

However, it is essential to discuss the implications of these findings in the context of pedagogy and ethical considerations. As pointed in the studies in related work section, while AI tools can enhance learning and engagement, they also raise questions about dependency, the development of critical thinking skills, and academic integrity. Educators must carefully integrate these tools into curricula, ensuring they complement traditional teaching methods and foster a balanced development of programming competencies.

5.6.1 Threats to Validity and Future Works

Our study was conducted at a single institution with a relatively small sample size, limiting generalizability. A threat to internal validity is the fact that we did not control which assignment the student was working on (due to small sample size) and behavior may have been different for different assignments.

Future studies could explore the long-term impact of AI tool usage on learning outcomes, investigate its effects across diverse educational contexts, and examine strategies to mitigate potential drawbacks. As AI technology continues to evolve, ongoing research and dialogue among educators, researchers, and policymakers will be crucial in harnessing its potential to enrich learning experiences while maintaining academic integrity and fostering comprehensive skill development.

CHAPTER 6

Generative AI Adoption in Classroom in Context of Technology Acceptance Model and the Innovation Diffusion Theory

6.1 Abstract

The burgeoning development of generative artificial intelligence (GenAI) and the widespread adoption of large language models (LLMs) in educational settings have sparked considerable debate regarding their efficacy and acceptability. Despite the potential benefits, the assimilation of these cutting-edge technologies among educators exhibits a broad spectrum of attitudes, from enthusiastic advocacy to profound skepticism. This study aims to dissect the underlying factors influencing educators' perceptions and acceptance of GenAI and LLMs. We conducted a survey among educators and analyzed the data through the frameworks of the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT). Our investigation reveals a strong positive correlation between the perceived usefulness of GenAI tools and their acceptance, underscoring the importance of demonstrating tangible benefits to educators. Additionally, the perceived ease of use emerged as a significant factor, though to a lesser extent, influencing acceptance. Our findings also shows that the knowledge and acceptance of these tool is not uniform, suggesting that targeted strategies are required to address the specific needs and concerns of each adopter category to facilitate broader integration of AI tools in education.

Keywords: Generative Artificial Intelligence (GenAI), Large Language Models (LLMs), Technology Acceptance Model (TAM), Innovation Diffusion Theory (IDT)

6.2 Introduction

The advent of generative artificial intelligence (GenAI) has heralded a new era in the technological landscape, offering unprecedented capabilities in creating text, images, code, and more from simple prompts. Among its various applications, the potential use of GenAI in educational settings is particularly compelling. Large language models (LLMs), a subset of GenAI, are poised to revolutionize teaching and learning practices by providing personalized learning experiences, automating content generation, and facilitating a more interactive and engaging learning environment. These technologies can augment the educational process, from crafting tailored educational materials to supporting diverse learning strategies, thereby enhancing the efficacy and accessibility of education. Furthermore, GenAI's ability to analyze and generate complex data can significantly contribute to research methodologies, enabling educators and students alike to explore new frontiers of knowledge and learning.

However, the integration of GenAI and LLMs into classroom settings is not without challenges. The adoption of new technologies in education is influenced by a multitude of factors, including but not limited to, perceived usefulness, ease of use, and the technological infrastructure available. To understand the dynamics of these technologies' acceptance and integration, it is crucial to delve into established theoretical frameworks that explain the adoption of technological innovations. The Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) offer robust lenses through which to examine these phenomena.

The Technology Acceptance Model (TAM) [128] posits that the perceived usefulness and perceived ease of use are fundamental determinants of the acceptance and usage of new technology. According to TAM, if users believe a technology will enhance their job performance (usefulness) and will be free of effort (ease of use), they are more likely to embrace and utilize the technology. On the other hand, Innovation Diffusion Theory (IDT) proposed by Rogers, explores how, why, and at what rate new ideas and technology spread through cultures [129, 130]. IDT suggests that innovation adoption is influenced by factors such as the innovation's relative advantage, compatibility with existing values and practices, complexity or ease of use, trialability, and observable results. Together, these frameworks provide a comprehensive understanding of the multifaceted process of technological adoption, enabling a nuanced analysis of the barriers and drivers behind GenAI's integration into the educational sphere. This paper has one primary research question:

RQ **What facilitators and barriers to the adoption of generative AI technologies exist in educational settings?**

In this paper, we aim to answer our research question by examining educators' perceptions and acceptance of GenAI and LLMs through the TAM and IDT frameworks. Understanding these factors is crucial for developing strategies to encourage the effective integration of GenAI tools in classrooms, thereby maximizing their potential benefits for teaching and learning. The following sections will delve into the methodology of our study, present our findings, and discuss their implications for the future of GenAI in education, setting the context for a comprehensive exploration of GenAI's role in reshaping educational paradigms. This inquiry not only contributes to the academic discourse on educational technology adoption but also provides practical insights for educators, policymakers, and technology developers aiming to foster an environment conducive to the innovative use of GenAI in education.

6.3 Related works

6.3.1 Teachers' perspectives on AI in education

Understanding the attitudes and perceptions of educators towards AI in education is crucial for its acceptance and integration into teaching practices. A survey of Kenyan teachers by Bii et al. revealed a generally positive outlook towards chatbot usage in education, despite concerns regarding their accuracy and potential to replace human teachers [66]. Similarly, Zhai et al.'s content analysis highlighted key research areas in AI education over a decade, including development and application [76]. Chen et al. noted an increased academic focus on AI, particularly in natural language processing and neural networks for educational purposes [77].

Research by Guillén-Gámez and Mayorga-Fernández found that factors such as age, gender, and ICT project involvement positively influence educators' attitudes towards ICT use in higher education [67]. Conversely, Nazaretsky et al. identified confirmation bias and trust as significant influencers on teachers' attitudes towards AI-based technologies, suggesting that pre-existing beliefs could hinder the adoption of such tools [68].

Akgun and Greenhow emphasized the ethical considerations necessary for AI deployment in K-12 settings, advocating for principles like transparency and inclusiveness [69]. Celik et al. explored the multifaceted roles of teachers in AI research and the challenges faced, including technical limitations and lack of technological knowledge [70]. Kim and Kim's study on STEM teachers' perceptions of an AI-enhanced scaffolding system for scientific writing indicated positive expectations, yet highlighted the need for teacher training on AI technologies [71]. Lastly, Lau and Guo's investigation into university instructors' views on AI tools like ChatGPT in programming education uncovered diverse strategies for adaptation, raising important questions for future research in computing education [42].

6.3.2 Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) to Explore the Adoption of Technology

The TAM, proposed by Fred Davis in 1989, posits that Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) play a critical role in user acceptance of information systems [128]. Masrom investigated the learning acceptance in terms of TAM and found that TAM could largely explain it's acceptance [131]. L Ritter performed a meta-analysis employing meta-analytic structural equation modeling (MASEM) to quantitatively synthesize studies that investigates college students' acceptance of online learning managements systems and got mixed results on how well it fits TAM [132]. Scherer et al. performed a meta-analytic structural equation modeling to investigate the Technology Acceptance Model's (TAM) validity in explaining teachers' adoption of digital technology in education [133]. Through robust statistical techniques, the study provided a comprehensive understanding of the factors influencing teachers' acceptance and use of technology, highlighting the role of perceived usefulness and ease of use. The results demonstrated the strong predictive power of TAM in teachers' technology adoption, offering a valuable framework for future research and technology integration strategies in the educational context. The role of certain key constructs and the importance of external variables contrast some existing beliefs about the TAM. Granic and Marangunic in their meta studey of 71 related papers found that TAM and its many different versions represent a credible model for facilitating assessment of diverse learning technologies and TAM's core variables, perceived ease of use and perceived usefulness, have been proven to be antecedent factors affecting acceptance of learning with technology. Zaineldeen et. al studied the TAM's concepts, contribution, limitation, and adoption in education [134].

Chocarro et al.'s application of the Technology Acceptance Model (TAM) to teachers' attitudes towards chatbots showed a preference for formal language and indicated that age and digital skills play roles in acceptance [72]. Khong et al. extended TAM to understand factors affecting teachers' acceptance of technology for online teaching, finding cognitive attitudes and perceived usefulness to be significant predictors [73]. A 2023 study by Iqbal et al. on faculty attitudes towards ChatGPT using TAM revealed mixed perceptions, with concerns about cheating balanced against the tool's benefits for lesson planning [74].

Similarly, innovation diffusion theory (IDT) have been used to study the acceptence and spreading of technology in education. Pinho et al.'s study on Moodle's use in higher education identified positive influences of Moodle's characteristics and personal innovativeness on its adoption, highlighting the importance of student-centered Learning Management Systems (LMS) [135]. Sahin provides a comprehensive overview of Rogers' Diffusion of Innovations theory, elaborating on its four main elements, the innovation-decision process, attributes of innovations, adopter categories, and its application in educational technology studies [136]. Menzli et al. examined the adoption of Open Educational Resources (OER) in higher education, finding that attributes such as relative advantage and observability positively impact faculty adoption, while also emphasizing the role of trialability, complexity, and compatibility in increasing OER adoption rates [137]. Frei-Landau et al. explored the mobile learning (ML) adoption process among teachers during the COVID-19 pandemic, uncovering 12 themes that denote the ML adoption process through Rogers' IDT, providing insights into promoting ML in teacher education under both routine and emergency conditions [138]. Finally, Al-Rahmi et al. combined the Technology Acceptance Model (TAM) with IDT to investigate students' intentions to use e-learning systems, demonstrating that innovation characteristics significantly influence students' behavioral intentions towards e-learning systems [139]. Ghimire et. al. explored the educators attitude towards these generative AI baded tools and found them to be generally positive [2].

6.4 Methodology - Evaluation Framework

6.4.1 Survey and Data

We conducted a quantitative study using a survey to gather educators' perspectives on AI tools in the classroom. We distributed the survey via email to faculty members at Utah State University (USU), a mid-sized research university in the western United States. Each faculty member received the survey link only once to avoid duplicate responses. The email provided a brief introduction to the research study, assured confidentiality, and encouraged participation. Participants were informed about the voluntary nature of the survey.

We received a total of 116 survey responses from email requests, representing a diverse sample from 8 colleges and 23 out of 39 departments at the university. The wide-ranging representation ensures a comprehensive understanding of educators' attitudes from various academic disciplines. For this study, we selected six survey questions that directly support our analysis using the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT) frameworks. Responses were captured using a Likert scale, allowing participants to express their agreement or disagreement with specific statements. The survey approved by the USU ethics review board (IRB). Since TAM identifies Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) as key determinants of technology adoption, the following questions were designed to represent these constructs:

1. AI tools like ChatGPT and Bard should be allowed and integrated into education. — **This question measures acceptance of technology, aligning with TAM's focus on the behavioral intention to use technology.**

2. I believe that AI tools like ChatGPT and Bard enhance the quality of education. — ($\mathbf{Q^{PU}}$)

3. I believe the benefits of incorporating large language models in education outweigh the potential risks and ethical concerns. — ($\mathbf{Q^{PU}}$)

4. I believe that the tools like ChatGPT and Bard are easy to use. —- ($\mathbf{Q^{PEOU}}$)

5. I believe that these AI tools like ChatGPT and Bard could be easily integrated into my current teaching methodology. — ($\mathbf{Q^{PEOU}}$)

6. Are you familiar with AI tool such as ChatGPT or Google Bard? – ($\mathbf{QIDT^{FM}}$)

Questions tagged with ($\mathbf{Q^{PU}}$) measure Perceived Usefulness (PU), and those with ($\mathbf{Q^{PEOU}}$) assess Perceived Ease of Use (PEOU). The question marked ($\mathbf{QIDT^{FM}}$) gauges familiarity, an important aspect of IDT.

6.4.2 Technology Acceptance Model (TAM) as an Evaluation Framework

Applying the Technology Acceptance Model (TAM) to understand teachers' attitudes and perceptions towards AI tools and Large Language Models (LLMs) such as ChatGPT and Bard can offer valuable insights. In this study's context, PU encompasses teachers' belief that specific AI tools or LLMs will enhance their teaching effectiveness and student learning outcomes. Conversely, PEOU refers to the ease with which educators can utilize these tools. Factors influencing PEOU include the user interface design, learning curve, and availability of technical support, which can significantly impact teachers' willingness to adopt AI technologies. Additionally, TAM helps identify potential barriers to technology adoption, such as perceived lack of IT skills or negative attitudes towards technology, guiding the development of professional training programs to mitigate these challenges and promote positive engagement with AI and LLMs in educational settings.

6.4.3 The Innovation Diffusion Theory (IDT) as an Evaluation Framework

The Innovation Diffusion Theory (IDT), proposed by Everett Rogers in 1962, offers a comprehensive framework for understanding the mechanisms through which new ideas and technologies are adopted within social systems. IDT delineates four key elements that influence the dissemination of an innovation: the characteristics of the innovation itself, the communication channels used to spread information about the innovation, the passage of time, and the nature of the social system. The theory categorizes the adoption process into five sequential stages:

1. Knowledge: This initial phase involves becoming aware of the innovation, albeit without detailed information about its functionality or application.
2. Persuasion: At this stage, interest in the innovation grows, prompting an active search for more information and a better understanding of its benefits and drawbacks.
3. Decision: Here, individuals or organizations critically assess the innovation, considering the pros and cons before making a decision to adopt or reject it.
4. Implementation: During implementation, the innovation is actively integrated into use, with adjustments and adaptations often made to fit specific needs.
5. Confirmation: In this final stage, the effectiveness and utility of the innovation are evaluated, influencing the decision to continue its use based on observed outcomes.

Moreover, IDT classifies adopters into five groups according to their propensity to embrace new technologies: Innovators, Early Adopters, Early Majority, Late Majority, and Laggards. This categorization helps in understanding the adoption timeline within a social system.

6.5 Results

6.5.1 Using TAM as a Framework

As explained in the methodology section, we utilized five survey questions to align with the TAM framework. Since the responses were on a Likert scale, they could be directly converted to numeric values. The response to the statement *"AI tools like ChatGPT and Bard should be allowed and integrated into education"* serves as a direct substitute for the dependent variable 'acceptance', as integrating it into coursework signifies full acceptance of the tool. For perceived usefulness (PU), we averaged the responses to the statements *"I believe that AI tools like ChatGPT and Bard enhance the quality of education"* and *"I believe the benefits of incorporating large language models in education outweigh the potential risks and ethical concerns"*. For perceived ease of use (PEOU), we averaged the responses to the statements *"I believe that the tools like ChatGPT and Bard are easy to use"* and *"I believe that these AI tools like ChatGPT and Bard could be easily integrated into my current teaching methodology."* This approach was adopted because the ease of use by educators should not only consider their own ease of use but also the ease of integrating it into their courses. Figure 6.1 shows the numeric Likert scale responses to these statements.

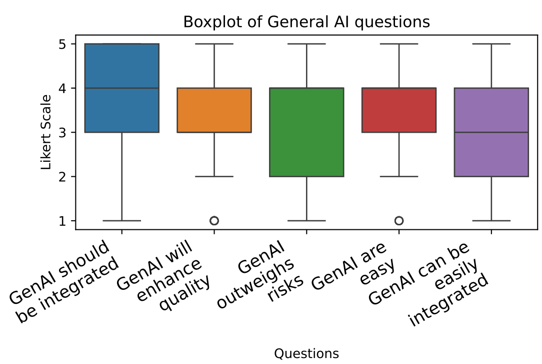


Fig. 6.1: Raw answers to the TAM-related questions

Next, we examine the correlation between acceptance, PU, and PEOU using the Pearson correlation coefficient.

As shown in Table 6.1, a strong positive correlation ($r = 0.734$) was found, indicating that as perceived usefulness increases, acceptance also tends to increase. A moderate positive correlation ($r = 0.542$) between acceptance and perceived ease of use was also observed.

	Perceived Usefulness	Perceived Ease of Use
Corr. with Acceptance	0.734	0.542
p-value	3.57^{-20}	8.33^{-10}

Table 6.1: Correlation table with acceptance

Regression analysis was performed to quantify how well acceptance is explained by perceived ease of use and perceived usefulness, including the significance of these predictors. It yielded an R-squared value of 0.566, indicating a moderate to strong fit. This suggests that perceived ease of use and perceived usefulness together explain a significant portion of the variance in acceptance. The coefficient for perceived usefulness was 0.678 with a p-value of 7.2^{-13}, showing a highly significant and strong positive effect on acceptance. Perceived ease of use had a coefficient of 0.227 with a p-value of 0.026, indicating a statistically significant positive effect on acceptance. This confirms that perceived usefulness is a significant and strong predictor of acceptance. The overall model is statistically significant, as indicated by an F-statistic p-value of 4.23×10^{-20}, meaning that the predictors together significantly explain the variability in acceptance.

	Coefficient	p-value
Perceived Usefulness	0.678	7.2^{-13}
Perceived Ease of Use	0.227	0.026

Table 6.2: Regression analysis results

6.5.2 The Innovation Diffusion Theory (IDT) to Explain the GenAI Use in Classrooms

The Innovation Diffusion Theory (IDT) offers a comprehensive framework for understanding the factors that facilitate the adoption of new technological ideas or systems within society. Unlike the Technology Acceptance Model (TAM), which provides a quantitative and concise explanation of innovation adoption, IDT offers insights into the adoption phase an individual or group might be in. IDT categorizes the population into five segments based on their adoption behavior:

1. Innovators: Individuals who embrace risks and are the first to experiment with new ideas.

2. Early Adopters: Those keen on exploring new technologies and affirming their usefulness within the community.

3. Early Majority: Individuals who contribute to mainstreaming an innovation within society, representing a significant portion of the population.

4. Late Majority: People who adopt an innovation following its acceptance by the early majority, integrating it into their daily lives as part of the wider community.

5. Laggards: Individuals who are slow to adopt innovative products and ideas, trailing behind the broader societal adoption curve.

While it is challenging to clearly categorize educators into these groups, such distinctions do exist. The range of familiarity with GenAI and LLM-based tools varies significantly across different departments and colleges. Figure 6.2 shows the familiarity with these tools in various schools.

In the context of education, particularly concerning teachers' attitudes and perceptions towards AI tools and Large Language Models (LLMs) like ChatGPT and Bard, IDT provides valuable insights:

1. Knowledge: Assessing teachers' awareness of AI tools and LLMs is crucial. Initiatives such as awareness campaigns, professional development sessions, and targeted marketing can significantly enhance this knowledge base.

2. Persuasion: Understanding teachers' interest in and attitudes towards these technologies is essential. Factors like perceived usefulness and ease of use play a critical role in shaping these attitudes.

3. Decision: The choice by teachers to incorporate AI tools into their classrooms is influenced by both individual preferences and institutional support structures.

4. Implementation: Effective integration of AI tools into teaching practices requires adequate support, training, and resources to ensure success.

5. Confirmation: Teachers' decisions to persist with the use of AI tools are influenced by the tangible benefits observed, feedback from students, and the availability of ongoing support.

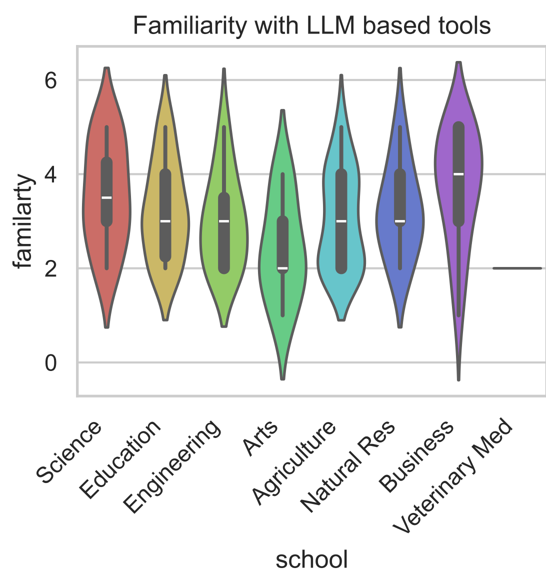


Fig. 6.2: Violin Plot showing familiarity with LLM-based tools among educators in various colleges.

By identifying where teachers stand in the diffusion process and recognizing their adopter category, strategies can be customized to facilitate the adoption of AI technologies. For example, while Innovators and Early Adopters may readily experiment with new tools, the Late Majority and Laggards might need more substantial evidence of the tools' effectiveness and comprehensive support systems to be persuaded.

ChatGPT became the fastest technology product to ever reach 100 million active users [140]. The spread of the technology is so rapid that it is challenging to gauge the sense of spread or adoption in the general public. Even in education, AI tools like these are rapidly becoming commonplace. Among the five steps of innovation diffusion outlined by IDT - knowledge, persuasion, decision, implementation, and confirmation - we could use the survey responses as proxies for some of the steps. For example, the knowledge step can be directly analogous to the question asking about familiarity with the AI tools. Similarly, the implementation and confirmation could be the execution of integrating the AI tool in class and its result, which are out of the scope of this paper.

6.6 Discussion and Conclusions

This paper explored the adoption and integration of generative artificial intelligence (GenAI) and large language models (LLMs) in educational settings, using the Technology Acceptance Model (TAM) and the Innovation Diffusion Theory (IDT) as theoretical frameworks. Our survey, conducted among educators at a medium-sized public research university in the United States, provided insights into their attitudes towards the use of AI tools like ChatGPT and Bard in the classroom. The findings indicate a generally positive perception towards these technologies, underscored by the perceived usefulness (PU) and perceived ease of use (PEOU) as significant predictors of their acceptance and integration into teaching methodologies.

The analysis revealed a strong positive correlation between the perceived usefulness of AI tools and their acceptance among educators, emphasizing the importance of demonstrating tangible benefits to enhance the adoption rate. Similarly, the perceived ease of use was found to have a significant, albeit moderate, positive effect on acceptance, highlighting the need for user-friendly and accessible AI tools in educational environments. TAM is a well-established theory that has been used to study the acceptance of new technologies in a variety of contexts. However, it is important to note that TAM is not a perfect theory. It has been criticized for being too simplistic and for not taking into account the full range of factors that influence users' intention to use a technology. TAM does not take into account the full range of factors that may influence teachers' attitudes and perceptions towards these technologies, such as their beliefs about the potential benefits and risks of AI, their level of comfort with technology, and their personal experiences with AI. The model does not account for the social and cultural factors that may influence teachers' acceptance of these tools.

Applying IDT, we categorized educators based on their adoption behavior and identified varied levels of familiarity with GenAI and LLMs across different departments. This diversity suggests the necessity for targeted strategies to address the specific needs and concerns of each adopter category, from Innovators to Laggards, to facilitate broader and more effective integration of AI tools in education. Based on interviews with educators, as detailed separately in [2], it was noted that early adopters are actively employing and incorporating these AI tools in their classes, expressing a need for clear policy guidelines. Meanwhile, laggards require training and education on the operation, advantages, and disadvantages of these tools, with many needing a combination of both approaches.

The rapid advancement of GenAI and LLMs presents a transformative opportunity for education. By embracing these technologies, educators can enhance the quality of education and foster a more engaging and personalized learning experience. Nevertheless, the successful integration of AI tools in education requires not only technological innovation but also a comprehensive understanding of the human factors influencing their adoption. Future research should therefore focus on longitudinal studies to track the evolution of educators' attitudes and the impact of AI tools on educational outcomes, as well as on the development of frameworks to address the ethical implications of AI in education.

6.6.1 Threats to validity

Our survey was not validated and no evaluation of reliability was made. Furthermore, all respondents were from a single institution based in the United States, limiting external validity regionally and internationally. Finally, generative AI is a fast-moving technology and attitudes and policies are likely also changing quickly. This work represents a shapshot of opinions, states and attitudes between May and June 2023.

6.7 Future Work

Future research should aim to extend the findings of this study by examining the long-term impact of GenAI and LLMs on educational outcomes and student engagement. Investigating the evolving attitudes of educators as they gain more experience with these technologies will also provide deeper insights into the barriers and facilitators of AI tool integration in educational settings. Additionally, exploring the ethical considerations and potential biases in AI applications in education will be crucial to ensure equitable and inclusive learning environments.

CHAPTER 7

Conclusion and Discussion

In this concluding chapter, I present a brief review of the pivotal discoveries made across the five studies, detailed in Section 7.1. We then explore the ramifications of our research for a range of stakeholders, outlined in Section 7.2. An analysis addressing the core research questions is provided in Section 7.3, which sets the stage for a candid examination of the study's limitations in Section 7.4 and recomentation for future research in Section 7.5. The chapter culminates in Section 7.6, where we encapsulate the essence of our findings and their broader impact on the field of educational technology and AI integration.

7.1 Summary of Key Findings

This dissertation explored the integration of generative AI in educational settings, examining its implications through various lenses including legal text summarization, educators' perceptions, policy landscapes, AI's role in programming education, and the acceptance of generative AI tools. Below, we summarize the key findings from each chapter:

- **Chapter 2: Summarization of Court Opinions Using NLP** revealed that NLP-based tools can significantly enhance the efficiency of legal document processing by automating the summarization process. This has profound implications for legal education and practice, offering a means to democratize access to legal information and facilitate more informed legal decision-making. In fact, at the time of publication of this dissertation, Justia, a major contributor of opinion summaries and with whom we worked to obtain data for the paper, has transitioned from manual summary to using generative AI.

- **Chapter 3: Generative AI in Education: Educators' Perspectives** found a generally positive attitude among educators towards the integration of AI tools in teaching and learning processes. However, the study also highlighted a need for further training and resources to fully leverage AI's potential in educational settings.

- **Chapter 4: From Guidelines to Governance: A Study of AI Policies in Education** identified a notable gap in existing policies governing the use of AI tools within educational institutions. The findings underscore the necessity for comprehensive, adaptable policy frameworks that address ethical considerations and promote responsible AI use.

- **Chapter 5: Coding With AI: Impact and Usage Patterns in Foundational Programming Courses** demonstrated that AI tools like ChatGPT can positively influence student engagement and learning outcomes in programming courses. The study suggests that these tools can make programming education more accessible and engaging for students.

- **Chapter 6: Generative AI Adaptation in Classroom in Context of Technology Acceptance Model and the Innovation Diffusion Theory** revealed that educators' acceptance of generative AI tools is significantly influenced by perceived usefulness and ease of use. The study emphasizes the importance of demonstrating tangible benefits to educators to facilitate the broader integration of AI tools in education.

These findings collectively highlight the transformative potential of generative AI in education, while also pointing to the challenges and considerations that must be addressed to realize this potential fully. The insights gained from this research contribute to a deeper understanding of how generative AI can be effectively, ethically, and sustainably integrated into educational practices.

7.2 Implications of the Research

The findings of this dissertation have several implications for educators, policymakers, and educational institutions regarding the integration of generative AI in educational settings. These implications span pedagogical practices, policy development, and ethical considerations, underscoring the need for a nuanced approach to leveraging AI technologies in education.

7.2.1 For Educators

The positive attitudes of educators towards AI tools, as highlighted in Chapter 3, suggest a readiness to integrate these technologies into teaching and learning processes. However, the necessity for additional training and resources indicates that professional development programs should be designed to enhance educators' AI literacy. This would enable them to effectively incorporate AI tools into their pedagogy, thereby enriching the learning experience and fostering a more engaging and personalized education environment.

7.2.2 For Policymakers

The policy gap identified in Chapter 4 emphasizes the urgent need for comprehensive, flexible policy frameworks that can adapt to the rapid advancements in AI technology. Policymakers should focus on developing guidelines that address ethical considerations, such as data privacy and academic integrity, while promoting the responsible use of AI in educational contexts. Collaboration with educational institutions, technology experts, and legal advisors will be crucial in formulating policies that balance innovation with ethical considerations.

7.2.3 For Educational Institutions

The dissertation's findings suggest that educational institutions should be proactive in adopting AI technologies, as demonstrated by the potential benefits in programming education and legal text summarization. Institutions should invest in the necessary technological infrastructure and support services to facilitate the integration of AI tools. Furthermore, establishing partnerships with AI technology providers could offer opportunities for co-developing educational applications that are tailored to the specific needs of students and educators.

7.2.4 Ethical and Pedagogical Considerations

The research underscores the importance of considering the ethical implications of AI integration in education. Institutions must ensure that the use of AI tools aligns with principles of fairness, transparency, and accountability, particularly concerning student data privacy and the prevention of algorithmic bias. Additionally, pedagogical strategies should be developed to complement AI tools with traditional teaching methods, ensuring that the technology serves as a support rather than a replacement for human interaction and critical thinking skills.

7.3 Discussion of the Research Questions

This section revisits the research questions introduced in the first chapter, discussing how the findings from each chapter contribute to answering these questions and extending the current understanding of generative AI's role in education.

7.3.1 Effectiveness of NLP in Legal Text Summarization

The first research question addressed the potential of NLP-based legal text summarization to enhance access to justice and legal education. Findings from Chapter 2 demonstrate that NLP tools can significantly reduce the time required to process and understand complex legal documents, thereby making legal information more accessible to professionals and students alike. This aligns with existing research on the efficiency of AI in legal contexts but extends it by providing empirical evidence of its application in educational settings.

7.3.2 Educators' Awareness and Attitudes Towards Generative AI

The second question explored educators' awareness and attitudes towards generative AI tools and the factors influencing these perceptions. Chapter 3's findings reveal a generally positive attitude but also highlight the need for further education and resources to fully leverage AI's potential. This suggests that while there is a growing interest in AI among educators, effective integration into pedagogy requires addressing the identified gaps in knowledge and resources.

7.3.3 Policy Landscape for AI in Education

Concerning the policy landscape around AI in education, the third question sought to identify existing gaps and needs for future policy development. The analysis in Chapter 4 underscores a significant policy vacuum, pointing towards the necessity for robust, adaptable policies that address ethical concerns and promote responsible AI use. This contributes to the discourse on AI governance in education by emphasizing the importance of proactive policy formulation.

7.3.4 Impact of AI Tools in Programming Education

The fourth question investigated the impact and usage patterns of AI tools like ChatGPT in foundational programming courses. As detailed in Chapter 5, the use of AI tools was associated with improved engagement and learning outcomes, suggesting that these technologies can serve as valuable aids in programming education. This finding enriches the debate on AI's educational utility by providing concrete examples of its positive effects on student learning.

7.3.5 Educators' Acceptance and Adaptation of AI Tools

Finally, the dissertation examined the extent to which educators' attitudes towards using generative AI tools in education could be explained by the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT). Chapter 6's findings indicate a strong correlation between perceived usefulness and educators' willingness to adopt AI tools, affirming the relevance of TAM and IDT in understanding technology adoption in educational contexts.

7.4 Limitations of the Study

This dissertation, while comprehensive in its scope and findings, is subject to several limitations that warrant consideration. These limitations not only highlight the challenges encountered during the research but also outline potential avenues for future investigations.

7.4.1 Scope of the Study

Firstly, the generalizability of the findings may be limited by the scope of the study. The research predominantly focused on specific educational contexts and applications of generative AI, such as legal text summarization and foundational programming courses. Consequently, the insights may not be directly applicable to other disciplines or educational settings without further investigation.

7.4.2 Sample Size and Diversity

Another limitation pertains to the sample size and diversity of the participants involved in the studies. While efforts were made to include a broad range of educators and institutions, the variability in AI adoption and attitudes across different educational landscapes could affect the representativeness of the findings. Future studies could benefit from a more diverse and larger sample to enhance the generalizability of the results.

7.4.3 Methodological Constraints

The methodologies employed in the research, including surveys and qualitative interviews, while effective in capturing a snapshot of educators' perceptions and policies around AI, may not fully encapsulate the dynamic and evolving nature of AI integration in education. Longitudinal studies could provide deeper insights into how these perceptions and policies change over time as educators and institutions gain more experience with AI tools.

7.4.4 Rapid Advancements in AI Technology

The fast-paced advancements in AI technology present another limitation. The tools and applications studied, such as NLP models and ChatGPT, are continually evolving, with new capabilities being developed at a rapid pace. As such, the findings must be contextualized within the technological landscape at the time of the study, acknowledging that future developments may alter the applicability and relevance of the results.

7.4.5 Potential Biases

Finally, potential biases in data collection and analysis must be acknowledged. Despite rigorous methodologies, the researchers' perspectives and the participants' self-reporting may introduce biases that could influence the interpretation of the findings. Future research should aim to mitigate these biases through diversified data collection methods and analytical approaches.

7.5 Recommendations for Future Research

Building upon the findings and acknowledging the limitations of this dissertation, several recommendations for future research emerge. These recommendations aim to extend the understanding of generative AI's integration in educational settings and address the gaps identified in the current study.

7.5.1 Expanding the Scope of AI Applications in Education

Future research should explore the integration of generative AI across a wider range of disciplines and educational contexts. Investigations could focus on subjects beyond legal education and computer science, such as the arts, humanities, and social sciences, to understand the broader applicability and impact of AI in education.

7.5.2 Longitudinal Studies on AI Adoption and Outcomes

To capture the evolving nature of AI integration in education, longitudinal studies are recommended. Such research would provide insights into how educators' perceptions, pedagogical strategies, and policy frameworks adapt over time, offering a dynamic view of the challenges and opportunities presented by AI technologies.

7.5.3 Investigating the Impact of AI on Diverse Learning Populations

Further studies should aim to understand the impact of generative AI tools on diverse student populations, including those with different learning needs and backgrounds. Research in this area could inform the development of inclusive AI-enhanced teaching practices that cater to a broad spectrum of learners.

7.5.4 Developing and Evaluating AI Literacy Programs for Educators

Given the importance of educators' awareness and understanding of AI, future research should focus on the development and evaluation of AI literacy programs. These programs would aim to equip educators with the knowledge and skills necessary to effectively integrate AI tools into their teaching practices.

7.5.5 Formulating and Assessing Ethical Guidelines for AI in Education

The ethical considerations surrounding the use of AI in education warrant further exploration. Research should be directed towards formulating comprehensive ethical guidelines for AI integration and assessing their implementation in educational institutions. This would contribute to the responsible and ethical use of AI technologies in educational settings.

7.5.6 Exploring the Technological Advancements and Their Educational Implications

As AI technology continues to advance, research should keep pace with these developments, exploring the implications of new AI capabilities for education. Studies could examine the pedagogical, ethical, and policy implications of emerging AI technologies, ensuring that educational practices remain aligned with the latest advancements.

7.6 Conclusion

This dissertation has embarked on a comprehensive exploration of the integration of generative artificial intelligence (AI) in educational settings, spanning from legal text summarization to educators' perceptions, policy considerations, the impact on programming education, and the adoption of AI tools through theoretical frameworks. The research presented has shed light on the transformative potential of AI in education, while also delineating the challenges, ethical considerations, and policy gaps that accompany its integration.

7.6.1 Reflecting on the Dissertation's Contributions

The contributions of this dissertation extend beyond the empirical findings of each chapter. Collectively, the research underscores the importance of a nuanced approach to integrating AI in education — one that balances technological innovation with ethical considerations, pedagogical effectiveness, and policy robustness. This work has provided valuable insights into how educators, policymakers, and educational institutions can navigate the complexities of adopting AI technologies, aiming to enhance learning outcomes while safeguarding ethical standards and fostering an inclusive educational environment.

7.6.2 The Future of AI in Education

Looking ahead, the potential of generative AI in education appears boundless, with advancements in AI technology continually opening new avenues for pedagogical innovation. However, the journey toward fully realizing this potential will require ongoing collaboration between educators, technologists, policymakers, and learners. As AI technologies evolve, so too must our strategies for their integration, ensuring that education remains a human-centric endeavor that leverages AI to enrich, rather than replace, the human elements of teaching and learning.

7.6.3 Final Words

This dissertation contributes to the foundational understanding of generative AI's role in education, offering a stepping stone for future research and practice. As we stand on the brink of a new era in educational technology, it is our collective responsibility to steer the integration of AI towards outcomes that are equitable, ethical, and aligned with the broader goals of education. The journey is just beginning, and the insights gleaned from this research illuminate the path forward, towards an educational landscape that harnesses the power of AI to unlock new potentials in teaching and learning.

REFERENCES

[1] A. Ghimire, R. Shrestha, and J. Edwards, "Too legal; didn't read (tldr): Summarization of court opinions," in *2023 Intermountain Engineering, Technology and Computing (IETC)*. IEEE, 2023, pp. 164–169.

[2] A. Ghimire, J. Prather, and J. Edwards, "Generative ai in education: A study of educators' awareness, sentiments, and influencing factors." Under Review, 2024.

[3] A. Ghimire and J. Edwards, "From guidelines to governance: A study of ai policies in education." Under Review, 2024.

[4] A. Ghimire and J. Edwrds, "Coding with ai: How are tools like chatgpt being used by students in foundational programming courses." Under Review, 2024.

[5] A. Ghimire and J. Edwards, "Generative ai adaptation in classroom in context of technology acceptance model and the innovation diffusion theory." Under Review, 2024.

[6] A. Ghimire, "Data-driven recommendation of academic options based on personality traits," in *Masters' Thesis, Utah State University*. USU, 2021.

[7] A. Ghimire, T. Dorsch, and J. Edwards, "Introspection with data: recommendation of academic majors based on personality traits," in *2022 Intermountain Engineering, Technology and Computing (IETC)*. IEEE, 2022, pp. 1–6.

[8] A. Ghimire, R. Ghimire, and J. Edwards, "Metadata in tweets: Broadcasting a lot more than what you tweet," in *2023 Intermountain Engineering, Technology and Computing (IETC)*, 2023, pp. 170–175.

[9] "Constitution of the united states, article i," 1787.

[10] U. S. I. of Peace, "Guiding principles for stabilization and reconstruction: Rule of law," Sep 2018. [Online]. Available: https://www.usip.org/guiding-principles-stabilization-and-reconstruction-the-web-version/rule-law

[11] N. C. for State Courts, "Case summaries and high-profile cases," 2021, accessed: 2022-4-20. [Online]. Available: https://www.ncsc.org/services-and-experts/areas-of-expertise/technology/web-best-practices/case-summaries-and-high-profile-cases

[12] S. Shavell, "The fundamental divergence between the private and the social motive to use the legal system," *The Journal of Legal Studies*, vol. 26, no. S2, pp. 575–612, 1997.

[13] C. Grover, B. Hachey, I. Hughson, and C. Korycinski, "Automatic summarisation of legal documents," in *Proceedings of the 9th international conference on Artificial intelligence and law*, 2003, pp. 243–251.

[14] A. Farzindar and G. Lapalme, “Letsum, an automatic legal text summarizing,” in *Legal Knowledge and Information Systems: JURIX 2004, the Seventeenth Annual Conference*, vol. 120. IOS Press, 2004, p. 11.

[15] F. Galgani, P. Compton, and A. Hoffmann, “Combining different summarization techniques for legal text,” in *Proceedings of the workshop on innovative hybrid approaches to the processing of textual data*, 2012, pp. 115–123.

[16] S. Polsley, P. Jhunjhunwala, and R. Huang, “Casesummarizer: a system for automated summarization of legal texts,” in *Proceedings of COLING 2016, the 26th international conference on Computational Linguistics: System Demonstrations*, 2016, pp. 258–262.

[17] K. Merchant and Y. Pande, “Nlp based latent semantic analysis for legal text summarization,” in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2018, pp. 1803–1807.

[18] D. Anand and R. Wagh, “Effective deep learning approaches for summarization of legal texts,” *Journal of King Saud University-Computer and Information Sciences*, 2019.

[19] A. Kanapala, S. Pal, and R. Pamula, “Text summarization from legal documents: a survey,” *Artificial Intelligence Review*, vol. 51, no. 3, pp. 371–402, 2019.

[20] D. Jain, M. D. Borah, and A. Biswas, “Summarization of legal documents: Where are we now and the way forward,” *Computer Science Review*, vol. 40, p. 100388, 2021.

[21] K. Ganesan, C. Zhai, and J. Han, “Opinosis: A graph based approach to abstractive summarization of highly redundant opinions,” 2010.

[22] R. Paulus, C. Xiong, and R. Socher, “A deep reinforced model for abstractive summarization,” *arXiv preprint arXiv:1705.04304*, 2017.

[23] S. Gehrmann, Y. Deng, and A. M. Rush, “Bottom-up abstractive summarization,” *arXiv preprint arXiv:1808.10792*, 2018.

[24] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, “Pegasus: Pre-training with extracted gap-sentences for abstractive summarization,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 11 328–11 339.

[25] A. Kornilova and V. Eidelman, “Billsum: A corpus for automatic summarization of us legislation,” *arXiv preprint arXiv:1910.00523*, 2019.

[26] Y. Huang, Z. Yu, J. Guo, Z. Yu, and Y. Xian, “Legal public opinion news abstractive summarization by incorporating topic information,” *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 9, pp. 2039–2050, 2020.

[27] D. de Vargas Feijo and V. P. Moreira, “Improving abstractive summarization of legal rulings through textual entailment,” *Artificial Intelligence and Law*.

[28] U. S. Supreme Court, “Supreme court of the united states : Opinions,” *Supreme Court of the United States*, 2022. [Online]. Available: https://www.supremecourt.gov/opinions/opinions.aspx

[29] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," 2017, to appear.

[30] R. Rehurek and P. Sojka, "Gensim–python framework for vector space modelling," *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, vol. 3, no. 2, 2011.

[31] M. Damashek, "Gauging similarity with n-grams: Language-independent categorization of text," *Science*, vol. 267, no. 5199, pp. 843–848, 1995.

[32] M. Paterson and V. Dančík, "Longest common subsequences," in *International symposium on mathematical foundations of computer science.* Springer, 1994, pp. 127–142.

[33] K. W. Church, "Word2vec," *Natural Language Engineering*, vol. 23, no. 1, pp. 155–162, 2017.

[34] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out*, 2004, pp. 74–81.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[36] K. M. Leung, "Naive bayesian classifier," *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, vol. 2007, pp. 123–156, 2007.

[37] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.

[38] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

[39] Y. Luan, Y. Ji, and M. Ostendorf, "Lstm based conversation models," *arXiv preprint arXiv:1603.09457*, 2016.

[40] "Legal pegasus pretrained model," 2021, accessed: 2022-4-20. [Online]. Available: https://huggingface.co/nsi319/legal-pegasus

[41] B. A. Becker, P. Denny, J. Finnie-Ansley, A. Luxton-Reilly, J. Prather, and E. A. Santos, "Programming is hard-or at least it used to be: Educational opportunities and challenges of ai code generation," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 2023, pp. 500–506.

[42] S. Lau and P. J. Guo, "From "ban it till we understand it" to" resistance is futile": How university programming instructors plan to adapt as more students use ai code generation and explanation tools such as chatgpt and github copilot," 2023.

[43] A. J. Ko, "More than calculators: Why large language models threaten learning, teaching, and education," https://medium.com/bits-and-behavior/more-than-calculators-why-large-language-models-threaten-public-education-480dd5300939, accessed: 2024-01-20.

[44] P. Denny, J. Prather, B. A. Becker, J. Finnie-Ansley, A. Hellas, J. Leinonen, A. Luxton-Reilly, B. N. Reeves, E. A. Santos, and S. Sarsa, "Computing education in the era of generative ai," *Commun. ACM*, jan 2024, online First. [Online]. Available: https://doi.org/10.1145/3624720

[45] J. Prather, P. Denny, J. Leinonen, B. A. Becker, I. Albluwi, M. Craig, H. Keuning, N. Kiesler, T. Kohn, A. Luxton-Reilly, S. MacNeil, A. Petersen, R. Pettit, B. N. Reeves, and J. Savelka, "The robots are here: Navigating the generative ai revolution in computing education," in *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education*, ser. ITiCSE-WGR '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 108–159. [Online]. Available: https://doi.org/10.1145/3623762.3633499

[46] J. Finnie-Ansley, P. Denny, B. A. Becker, A. Luxton-Reilly, and J. Prather, "The robots are coming: Exploring the implications of openai codex on introductory programming," in *Proceedings of the 24th Australasian Computing Education Conference*, ser. ACE '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 10–19. [Online]. Available: https://doi.org/10.1145/3511861.3511863

[47] J. Finnie-Ansley, P. Denny, A. Luxton-Reilly, E. A. Santos, J. Prather, and B. A. Becker, "My ai wants to know if this will be on the exam: Testing openai's codex on cs2 programming exercises," in *Proceedings of the 25th Australasian Computing Education Conference*, ser. ACE '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 97–104. [Online]. Available: https://doi.org/10.1145/3576123.3576134

[48] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. d. O. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman *et al.*, "Evaluating large language models trained on code," *arXiv preprint arXiv:2107.03374*, 2021.

[49] J. Savelka, A. Agarwal, C. Bogart, and M. Sakr, "Large language models (gpt) struggle to answer multiple-choice questions about code," *arXiv preprint arXiv:2303.08033*, 2023.

[50] B. Puryear and G. Sprint, "Github copilot in the classroom: learning to code with ai assistance," *Journal of Computing Sciences in Colleges*, vol. 38, no. 1, pp. 37–47, 2022.

[51] P. Denny, V. Kumar, and N. Giacaman, "Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, ser. SIGCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 1136–1142. [Online]. Available: https://doi.org/10.1145/3545945.3569823

[52] S. I. Ross, M. Muller, F. Martinez, S. Houde, and J. D. Weisz, “A case study in engineering a conversational programming assistant’s persona,” *arXiv preprint arXiv:2301.10016*, 2023.

[53] N. A. Ernst and G. Bavota, “Ai-driven development is here: Should you worry?” *IEEE Software*, vol. 39, no. 2, pp. 106–110, 2022.

[54] P. Denny, B. A. Becker, J. Leinonen, and J. Prather, “Chat overflow: Artificially intelligent models for computing education - renaissance or apocaiypse?” in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, ser. ITiCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 3–4. [Online]. Available: https://doi.org/10.1145/3587102.3588773

[55] S. MacNeil, A. Tran, D. Mogil, S. Bernstein, E. Ross, and Z. Huang, “Generating diverse code explanations using the gpt-3 large language model,” in *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 2*, ser. ICER ’22. New York, NY, USA: Association for Computing Machinery, 2022, p. 37–39. [Online]. Available: https://doi.org/10.1145/3501709.3544280

[56] S. Sarsa, P. Denny, A. Hellas, and J. Leinonen, “Automatic generation of programming exercises and code explanations using large language models,” in *Proceedings of the 2022 ACM Conference on International Computing Education Research - Volume 1*, ser. ICER ’22. New York, NY, USA: Association for Computing Machinery, 2022, p. 27–43. [Online]. Available: https://doi.org/10.1145/3501385.3543957

[57] P. Denny, S. Sarsa, A. Hellas, and J. Leinonen, “Robosourcing educational resources–leveraging large language models for learnersourcing,” *arXiv preprint arXiv:2211.04715*, 2022.

[58] S. MacNeil, A. Tran, A. Hellas, J. Kim, S. Sarsa, P. Denny, S. Bernstein, and J. Leinonen, “Experiences from using code explanations generated by large language models in a web software development e-book,” in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, ser. SIGCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 931–937. [Online]. Available: https://doi.org/10.1145/3545945.3569785

[59] J. Leinonen, A. Hellas, S. Sarsa, B. Reeves, P. Denny, J. Prather, and B. A. Becker, “Using large language models to enhance programming error messages,” in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 2023, pp. 563–569.

[60] J. Prather, B. N. Reeves, P. Denny, B. A. Becker, J. Leinonen, A. Luxton-Reilly, G. Powell, J. Finnie-Ansley, and E. A. Santos, ““it’s weird that it knows what i want”: Usability and interactions with copilot for novice programmers,” *ACM Trans. Comput.-Hum. Interact.*, vol. 31, no. 1, nov 2023. [Online]. Available: https://doi.org/10.1145/3617367

[61] B. Reeves, S. Sarsa, J. Prather, P. Denny, B. A. Becker, A. Hellas, B. Kimmel, G. Powell, and J. Leinonen, "Evaluating the performance of code generation models for solving parsons problems with small prompt variations," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, ser. ITiCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 299–305. [Online]. Available: https://doi.org/10.1145/3587102.3588805

[62] P. Denny, J. Leinonen, J. Prather, A. Luxton-Reilly, T. Amarouche, B. A. Becker, and B. Reeves, "Prompt problems: A new programming exercise for the generative ai era," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, ser. SIGCSE 2024. NY, USA: ACM, 2024.

[63] P. Vaithilingam, T. Zhang, and E. L. Glassman, "Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models," in *Chi conference on human factors in computing systems extended abstracts*, 2022, pp. 1–7.

[64] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu *et al.*, "Summary of chatgpt/gpt-4 research and perspective towards the future of large language models," *arXiv preprint arXiv:2304.01852*, 2023.

[65] D. Luitse and W. Denkena, "The great transformer: Examining the role of large language models in the political economy of ai," *Big Data & Society*, vol. 8, no. 2, p. 20539517211047734, 2021.

[66] P. Bii, J. Too, and C. Mukwa, "Teacher attitude towards use of chatbots in routine teaching." *Universal Journal of Educational Research*, vol. 6, no. 7, pp. 1586–1597, 2018.

[67] F. D. Guillén-Gámez and M. J. Mayorga-Fernández, "Identification of variables that predict teachers' attitudes toward ict in higher education for teaching and research: A study with regression," *Sustainability*, vol. 12, no. 44, p. 1312, Jan 2020.

[68] T. Nazaretsky, M. Cukurova, M. Ariely, and G. Alexandron, *Confirmation bias and trust: Human factors that influence teachers' attitudes towards AI-based educational technology*, Oct 2021.

[69] S. Akgun and C. Greenhow, "Artificial intelligence in education: Addressing ethical challenges in k-12 settings," *AI and Ethics*, vol. 2, no. 3, p. 431–440, Aug 2022.

[70] I. Celik, M. Dindar, H. Muukkonen, and S. Järvelä, "The promises and challenges of artificial intelligence for teachers: a systematic review of research," *TechTrends*, vol. 66, no. 4, p. 616–630, Jul 2022.

[71] N. J. Kim and M. K. Kim, "Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing," *Frontiers in Education*, vol. 7, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feduc.2022.755914

[72] R. Chocarro, M. Cortiñas, and G. Marcos-Matás, "Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of

social language, bot proactiveness, and users' characteristics," *Educational Studies*, vol. 49, no. 2, p. 295–313, Mar 2023.

[73] H. Khong, I. Celik, T. T. T. Le, V. T. T. Lai, A. Nguyen, and H. Bui, "Examining teachers' behavioural intention for online teaching after covid-19 pandemic: A large-scale survey," *Education and Information Technologies*, vol. 28, no. 5, p. 5999–6026, May 2023.

[74] N. Iqbal, H. Ahmed, and K. Azhar, "Exploring teachers' attitudes towards using chat gpt," *Global Journal for Management and Administrative Sciences*, vol. 3, Feb 2023.

[75] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction.* Morgan Kaufmann, 2017.

[76] X. Zhai, X. Chu, C. S. Chai, M. S. Y. Jong, A. Istenic, M. Spector, J.-B. Liu, J. Yuan, and Y. Li, "A review of artificial intelligence (ai) in education from 2010 to 2020," *Complexity*, vol. 2021, pp. 1–18, 2021.

[77] X. Chen, D. Zou, H. Xie, G. Cheng, and C. Liu, "Two decades of artificial intelligence in education," *Educational Technology & Society*, vol. 25, no. 1, pp. 28–47, 2022.

[78] M. Pradana, H. P. Elisa, and S. Syarifuddin, "Discussing chatgpt in education: A literature review and bibliometric analysis," *Cogent Education*, vol. 10, no. 2, p. 2243134, 2023.

[79] C. K. Lo, "What is the impact of chatgpt on education? a rapid review of the literature," *Education Sciences*, vol. 13, no. 4, p. 410, 2023.

[80] J. H. Choi, K. E. Hickman, A. Monahan, and D. Schwarcz, "Chatgpt goes to law school," *Available at SSRN*, 2023.

[81] A. Flogie and M. V. Krabonja, "Artificial intelligence in education: developing competencies and supporting teachers in implementing ai in school learning environments," in *2023 12th Mediterranean Conference on Embedded Computing (MECO).* IEEE, 2023, pp. 1–6.

[82] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt *et al.*, "Ethics of ai in education: Towards a community-wide framework," *International Journal of Artificial Intelligence in Education*, pp. 1–23, 2021.

[83] S. Akgun and C. Greenhow, "Artificial intelligence in education: Addressing ethical challenges in k-12 settings," *AI and Ethics*, pp. 1–10, 2021.

[84] C. Adams, P. Pente, G. Lemermeyer, and G. Rockwell, "Ethical principles for artificial intelligence in k-12 education," *Computers and Education: Artificial Intelligence*, vol. 4, p. 100131, 2023.

[85] M. Halaweh, "Chatgpt in education: Strategies for responsible implementation," 2023.

[86] M. Sullivan, A. Kelly, and P. McLaughlan, "Chatgpt in higher education: Considerations for academic integrity and student learning," 2023.

[87] T. K. Chiu, "The impact of generative ai (genai) on practices, policies and research direction in education: a case of chatgpt and midjourney," *Interactive Learning Environments*, pp. 1–17, 2023.

[88] C. Kooli, "Chatbots in education and research: A critical examination of ethical implications and solutions," *Sustainability*, vol. 15, no. 7, p. 5614, 2023.

[89] A. Garshi, M. W. Jakobsen, J. Nyborg-Christensen, D. Ostnes, and M. Ovchinnikova, "Smart technology in the classroom: a systematic review. prospects for algorithmic accountability," *arXiv preprint arXiv:2007.06374*, 2020.

[90] B. Berendt, A. Littlejohn, and M. Blakemore, "Ai in education: Learner choice and fundamental rights," *Learning, Media and Technology*, vol. 45, no. 3, pp. 312–324, 2020.

[91] F. Filgueiras, "Artificial intelligence and education governance," *Education, Citizenship and Social Justice*, p. 17461979231160674, 2023.

[92] S. Li and X. Gu, "A risk framework for human-centered artificial intelligence in education," *Educational Technology & Society*, vol. 26, no. 1, pp. 187–202, 2023.

[93] B. Memarian and T. Doleck, "Fairness, accountability, transparency, and ethics (fate) in artificial intelligence (ai), and higher education: A systematic review," *Computers and Education: Artificial Intelligence*, p. 100152, 2023.

[94] A. Nigam, R. Pasricha, T. Singh, and P. Churi, "A systematic review on ai-based proctoring systems: Past, present and future," *Education and Information Technologies*, vol. 26, no. 5, pp. 6421–6445, 2021.

[95] O. Sahlgren, "The politics and reciprocal (re) configuration of accountability and fairness in data-driven education," *Learning, Media and Technology*, vol. 48, no. 1, pp. 95–108, 2023.

[96] N. Gillani, R. Eynon, C. Chiabaut, and K. Finkel, "Unpacking the "black box" of ai in education," *Educational Technology & Society*, vol. 26, no. 1, pp. 99–111, 2023.

[97] G. N. Uunona and L. Goosen, "Leveraging ethical standards in artificial intelligence technologies: A guideline for responsible teaching and learning applications," in *Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines.* IGI Global, 2023, pp. 310–330.

[98] F. Miao, W. Holmes, R. Huang, H. Zhang *et al.*, *AI and education: A guidance for policymakers.* UNESCO Publishing, 2021.

[99] C. K. Y. Chan, "A comprehensive ai policy education framework for university teaching and learning," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, pp. 1–25, 2023.

[100] C. for School Networking, "Cosn strategic plan 2019-2022," 2019, [Accessed 06-10-2023].

[101] R. Chatila, K. Firth-Butterfield, and J. C. Havens, "Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems version 2," *University of southern California Los Angeles*, 2018.

[102] H. AI, "High-level expert group on artificial intelligence," p. 6, 2019.

[103] C. for School Networking, "CoSN Issues Guidance on AI in the Classroom — CoSN — cosn.org," 2023, [Accessed 06-10-2023].

[104] S. Ali, D. DiPaola, R. Williams, P. Ravi, and C. Breazeal, "Constructing dreams using generative ai," *arXiv preprint arXiv:2305.12013*, 2023.

[105] L. Sattelmaier and J. M. Pawlowski, "Towards a generative artificial intelligence competence framework for schools," in *Proceedings of the International Conference on Enterprise and Industrial Systems (ICOEINS 2023)*, vol. 270. Springer Nature, 2023, p. 291.

[106] F. Ouyang and P. Jiao, "Artificial intelligence in education: The three paradigms," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100020, 2021.

[107] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja *et al.*, ""so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *International Journal of Information Management*, vol. 71, p. 102642, 2023.

[108] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52–62, 2023.

[109] J. Whalen, C. Mouza *et al.*, "Chatgpt: Challenges, opportunities, and implications for teacher education," *Contemporary Issues in Technology and Teacher Education*, vol. 23, no. 1, pp. 1–23, 2023.

[110] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, "Ethical principles for artificial intelligence in education," *Education and Information Technologies*, vol. 28, no. 4, pp. 4221–4241, 2023.

[111] J. Edwards, K. Hart, R. Shrestha *et al.*, "Review of csedm data and introduction of two public cs1 keystroke datasets," *Journal of Educational Data Mining*, vol. 15, no. 1, pp. 1–31, 2023.

[112] P. Denny, J. Prather, B. A. Becker, J. Finnie-Ansley, A. Hellas, J. Leinonen, A. Luxton-Reilly, B. N. Reeves, E. A. Santos, and S. Sarsa, "Computing education in the era of generative ai," *arXiv preprint arXiv:2306.02608*, 2023.

[113] T. Phung, V.-A. Pădurean, J. Cambronero, S. Gulwani, T. Kohn, R. Majumdar, A. Singla, and G. Soares, "Generative ai for programming education: Benchmarking chatgpt, gpt-4, and human tutors," *International Journal of Management*, vol. 21, no. 2, p. 100790, 2023.

[114] J. Prather, B. N. Reeves, P. Denny, B. A. Becker, J. Leinonen, A. Luxton-Reilly, G. Powell, J. Finnie-Ansley, and E. A. Santos, ""it's weird that it knows what i want": Usability and interactions with copilot for novice programmers," *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2023.

[115] J. Shin and J. Nam, "A survey of automatic code generation from natural language," *Journal of Information Processing Systems*, vol. 17, no. 3, pp. 537–555, 2021.

[116] R. Watermeyer, L. Phipps, D. Lanclos, and C. Knight, "Generative ai and the automating of academia," *Postdigital Science and Education*, pp. 1–21, 2023.

[117] C. Zastudil, M. Rogalska, C. Kapp, J. Vaughn, and S. MacNeil, "Generative ai in computing education: Perspectives of students and instructors," in *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2023, pp. 1–9.

[118] L. Hedberg Segeholm and E. Gustafsson, "Generative language models for automated programming feedback," 2023.

[119] M. Kazemitabaar, X. Hou, A. Henley, B. J. Ericson, D. Weintrop, and T. Grossman, "How novices use llm-based code generators to solve cs1 coding tasks in a self-paced learning environment," *arXiv preprint arXiv:2309.14049*, 2023.

[120] Z. Zhang, Z. Dong, Y. Shi, N. Matsuda, T. Price, and D. Xu, "Students' perceptions and preferences of generative artificial intelligence feedback for programming," *arXiv preprint arXiv:2312.11567*, 2023.

[121] N. Carr, F. R. Shawon, and H. M. Jamil, "An experiment on leveraging chatgpt for online teaching and assessment of database students," in *2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. IEEE, 2023, pp. 1–8.

[122] R. Yilmaz and F. G. K. Yilmaz, "Augmented intelligence in programming learning: Examining student views on the use of chatgpt for programming learning," *Computers in Human Behavior: Artificial Humans*, vol. 1, no. 2, p. 100005, 2023.

[123] N. M. S. Surameery and M. Y. Shakor, "Use chat gpt to solve programming bugs," *International Journal of Information Technology & Computer Engineering (IJITC) ISSN: 2455-5290*, vol. 3, no. 01, pp. 17–22, 2023.

[124] M.-D. Popovici, "Chatgpt in the classroom. exploring its potential and limitations in a functional programming course," *International Journal of Human–Computer Interaction*, pp. 1–12, 2023.

[125] A. Husain, "Potentials of chatgpt in computer programming: Insights from programming instructors," *Journal of Information Technology Education: Research*, vol. 23, p. 002, 2024.

[126] S. Speth, N. Meißner, and S. Becker, "Investigating the use of ai-generated exercises for beginner and intermediate programming courses: A chatgpt case study," in *2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, 2023, pp. 142–146.

[127] M. Wieser, K. Schöffmann, D. Stefanics, A. Bollin, and S. Pasterk, "Investigating the role of chatgpt in supporting text-based programming education for students and teachers," in *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*. Springer Nature Switzerland Cham, 2023, pp. 40–53.

[128] F. D. Davis, "User acceptance of information systems: the technology acceptance model (tam)," 1987.

[129] E. M. Rogers, "Bibliography on the diffusion of innovations," 1961.

[130] E. M. Rogers and D. Williams, "Diffusion of," *Innovations (Glencoe, IL: The Free Press, 1962)*, 1983.

[131] M. Masrom, "Technology acceptance model and e-learning," *Technology*, vol. 21, no. 24, p. 81, 2007.

[132] N. L Ritter, "Technology acceptance model of online learning management systems in higher education: A meta-analytic structural equation model," *International Journal of Learning Management Systems*, vol. 5, no. 1, pp. 1–16, 2017.

[133] R. Scherer, F. Siddiq, and J. Tondeur, "The technology acceptance model (tam): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education," *Computers & Education*, vol. 128, p. 13–35, Jan 2019.

[134] S. Zaineldeen, L. Hongbo, A. Koffi, and B. Hassan, "Technology acceptance model'concepts, contribution, limitation, and adoption in education," *Universal Journal of Educational Research*, vol. 8, no. 11, pp. 5061–5071, 2020.

[135] C. Pinho, M. Franco, and L. Mendes, "Application of innovation diffusion theory to the e-learning process: higher education context," *Education and Information Technologies*, vol. 26, pp. 421–440, 2021.

[136] I. Sahin, "Detailed review of rogers' diffusion of innovations theory and educational technology-related studies based on rogers' theory." *Turkish Online Journal of Educational Technology-TOJET*, vol. 5, no. 2, pp. 14–23, 2006.

[137] L. J. Menzli, L. K. Smirani, J. A. Boulahia, and M. Hadjouni, "Investigation of open educational resources adoption in higher education using rogers' diffusion of innovation theory," *Heliyon*, vol. 8, no. 7, 2022.

[138] R. Frei-Landau, Y. Muchnik-Rozanov, and O. Avidov-Ungar, "Using rogers' diffusion of innovation theory to conceptualize the mobile-learning adoption process in teacher education in the covid-19 era," *Education and information technologies*, vol. 27, no. 9, pp. 12 811–12 838, 2022.

[139] W. M. Al-Rahmi, N. Yahaya, A. A. Aldraiweesh, M. M. Alamri, N. A. Aljarboa, U. Alturki, and A. A. Aljeraiwi, "Integrating technology acceptance model with innovation diffusion theory: An empirical investigation on students' intention to use e-learning systems," *Ieee Access*, vol. 7, pp. 26 797–26 809, 2019.

[140] H. Krystal, "Chatgpt sets record for fastest-growing user base," 2023. [Online]. Available: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

APPENDICES

APPENDIX A

Curriculum Vitae (CV)

Aashish Ghimire

Website: aashishg.com
Email: mail@aashishg.com
LinkedIn: aashishghimire
GitHub: github.com/GitAashishG

Education

Utah State University Logan, Utah
Ph.D. in Computer Science, Advisor: Dr. John Edwards 2024 (expected)

Utah State University Logan, Utah
M.S. in Computer Science 2019–2020

- Minor: Military Science and Leadership
- Thesis: "Data-Driven Recommendation of Academic Options Based on Personality Traits"

Coppin State University Baltimore, Maryland
B.S. in Computer Science 2011–2015

- Minor: Mathematics

Publications

Published

[1] **A. Ghimire**, R. Shrestha, and J. Edwards, "Too legal; didn't read (tldr): Summarization of court opinions", presented at the IEEE Intermountain Engineering, Technology, and Computing Conference (i-ETC), 2023.

[2] **A. Ghimire**, R. Ghimire, and J. Edwards, "Metadata in tweets: Broadcasting a lot more than what you tweet", presented at the IEEE Intermountain Engineering, Technology, and Computing Conference (i-ETC), 2023.

[3] **A. Ghimire** and J. Edwards., "Introspection with data : Recommendation of academic majors based on personality traits", presented at the IEEE Intermountain Engineering, Technology, and Computing Conference (i-ETC). Orem, UT, 2022.

[4] **A. Ghimire**, I. Srivastava, and T. S. Fisher, "Granular matter: Microstructural evolution and mechanical response", Citeseer, 2014.

Under Review

[5] **A. Ghimire**, J. Prather, and J. Edwards, "Generative ai in education: A study of educators' awareness, sentiments, and influencing factors", presented at the Innovation and Technology in Computer Science Education, 2024.

[6] **A. Ghimire** and J. Edwards, "Generative ai adaptation in classroom in context of technology acceptance model and the innovation diffusion theory", presented at the IEEE Intermountain Engineering, Technology, and Computing Conference (i-ETC), 2024.

[7] **A. Ghimire** and J. Edwards, "From guidelines to governance: A study of ai policies in education", presented at the Artificial Intillegence in Data Mining, 2024.

[8] **A. Ghimire** and J. Edwards, "Coding with ai: How are tools like chatgpt being used by students in foundational programming courses", presented at the Artificial Intillegence in Data Mining, 2024.

Work Experience

Microsoft Corp Redmond, Washington
Research Software Engineer, Office Of CTO (OCTO) Team 2023–Now

- Work in a smaller agile team within office of CTO on early tech prototyping and proof-of-concept of new and emerging technology to facilitate rapic technology transition within whole division
- Tech stack: Azure AI as a service, AutoGen, LangChain, Semantic Kernel

Research Intern, Advanced Autonomy and Applied Robotics (A3R) Team Aug 2022–Nov 2022

- Project: Project: Creating transformer-based Natural Language Processing model for code generation from English text for Robot Operating System (ROS).
- Tech stack: Pytorch, GPT, ROS, Gazebo, Python, Jupyter notebooks, CloudSim, Azure, CodeX

US Army Reserve Joint Base Lewis-McChord, Washington
Cyber Warfare Officer, 301 ME HHC April 2016 –now

- Supervise Cyber section within Brigade as a staff officer and lead the team of over 15 cyber-trained soldiers
- Previously, led platoon-sized element (about 40 soldiers) for conducting monthly battle assembly, and maintaining mission readiness as a platoon leader.

Meta Inc (Facebook) Menlo Park, California
Research Intern, Data and AI Platform Team, Reality Labs May 2022–Aug 2022

- Project: Created data pipelines and metrics anomaly detection system for large-scale log data into production.
- Tech stack: SQL and Presto, Python, Jupyter notebooks, Prophet for Machine Learning

Esri Inc. D.C. Regional Office (Remote)
SWE / Machine Learning Intern, Advanced Spatial Analytics Team May 2021–Aug 2021

- Created a machine learning solution pipeline for analyzing satellite imagery in geo-special context, including data cleanup and augmentation, machine learning model selection, image processing and training in a 3-person team.

Charles Schwab Corporation Phoenix, AZ (Remote))
SWE Intern, Tools, Audit and Automation Team Jan 2021–April 2021

- Created a machine learning solution pipeline for analyzing satellite imagery in geo-special context, including data cleanup and augmentation, machine learning model selection, image processing and training in a 3-person team.

Utah State University Logan, Utah
Research Assistant, EdwardsLab, Department of Computer Science April 2020–now
Project: Smart Career Recommendation System

- Researched, designed and implemented 'The Pocket Jeanie' - smart recommender system to find the optimal path for achieving a career goal for a high-school student with more than 1000 active users.
- Utilized different machine learning techniques to harvest large-scale web data, build a machine learning based recommendation model and integrated with user dataset for the recommendation system.

Utah State University Logan, Utah
Teaching Assistant, Department of Computer Science August 2019–April 2020

- Assisted the faculty member with classroom instruction material, preparing tests and exams, grading the assignments and projects and record-keepings.
- Conducted office hours and tutoring for the class "Developing dynamic, database-driven, web applications"

Microsoft IT via Unisys Corp Salt Lake City, Utah
Support and Escalation Engineer 2016 –2017

- Facilitated Microsoft's internal transition to Secure Admin Workstation (SAW) - locked down device that only runs pre-approved application to access any production servers.
- Supported Microsoft employee's active directory account, group policy, access control.

Purdue University West Lafayette, Indiana
Undergraduate Research Fellow Summer 2014

- Designed, coded and debugged the web based software tool to simulate the process of jamming and stress perturbation of granular matter and nanoparticles.
- Utilized Visualization Toolkit (VTK) rendering library to render the 3D modeling of initial and final configuration of the system.

Coppin State University Baltimore, Maryland
Research Assistant, Department of Natural Science 2012 –2015

- Studied and identified different semiconductor materials for making Multi-junction solar cell
- Simulated the output and efficiency of different semiconductor material in making solar cell

Teaching

- **Teaching Assistant** at Utah State University Fall 2019, Spring 2020, Fall 2021, Fall 2022
  Developing dynamic, database-driven, web applications (CS 2610, Undergraduate Class)
- **Head Teaching Assistant** at Coppin State University Fall 2014, Spring 2015
  Fundamentals of Programming (CS 131, Undergraduate Class)

Scholarships and Awards

- USU Graduate Research and Creative Opportunities (GRCO) award ($1000) 2023
- School of Graduate Study Student Travel Grant ($400) 2023
- Gold Award Intermountain Engineering, Technology, and Computing Conference (i-ETC) ($300) 2023
- Full Tuition awards and stipend (GTA/GRA), Utah State University 2019 –now
- Graduating Senior of the Year, Computer Science, Coppin State University 2015
- Best senior research thesis award, Coppin State University 2015
- Golden Eagle Honors Scholarship, Coppin State University (4 yr, full ride) 2011–2015
- Thurgood Marshall College Fund scholar (1 week retreat, conference) 2013, 2014
- STEM Grant (USD 1200 / semester) 2013, 2014
- Dean's List (All Semesters) 2011–2015
- Rising Scholar Award 2012
- Freshmen Male Initiative Award (iPad) 2013–2013
- Asia-Pacific Private Donor Award, CSU (USD 500/sem) 2012 –2014