Journal of Information Systems Applied Research and Analytics

Volume 19, No. 1
April 2026
ISSN: 1946-1836

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Journal of Information Systems Applied Research and Analytics** (JISARA) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is four issues a year. The first date of publication was December 1, 2008. The original name of the journal was Journal of Information Systems Applied Research (JISAR).

JISARA is published online (https://jisara.org) in connection with the ISCAP (Information Systems and Computing Academic Professionals) Conference, where submissions are also double-blind peer reviewed. Our sister publication, the Proceedings of the ISCAP Conference, features all papers, teaching cases and abstracts from the conference. (https://iscap.us/proceedings)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%) and other submitted works. The non-award winning papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in JISAR. Currently the acceptance rate for the journal is approximately 35%.

Questions should be addressed to the editor at editor@jisara.org or the publisher at publisher@jisara.org. Special thanks to members of ISCAP who perform the editorial and review processes for JISARA.

Journal of Information Systems Applied Research and Analytics

One Health: AI in Healthcare Summarization:

Chiazam Izuchukwu
ci02061@georgiasouthern.edu
Georgia Southern University
Atlanta, GA 30302

Hayden Wimmer
hwimmer@georgiasouthern.edu
Georgia Southern University
Atlanta, GA 30302

Loreen Marie Powell
lpowell@maryu.marywood.edu
Marywood University
Scranton, PA18509

Abstract

The research uses a One Health text document as our main dataset. Two summarization methods were employed in this study, with the first being recursive (summary of the summary), which was repeated twice to achieve a progressively concise word count, and the second a controlled and direct summary of the original document, which was repeated to accomplish the same set word count. Performance metrics like ROUGE, BLEU, and BERT scores were calculated to assess the effectiveness of both methods. This study goes beyond automated metrics by incorporating human evaluation and assessing readability and coherence with the original document to ensure a qualitative validation of the results. Additionally, the results from all summaries generated provide a comparative analysis between traditional LLM-based summarization and conversational AI establishing performance baselines for validating clarity and coherence for effective summarization of healthcare documents.

Keywords: Large language models, Clinical summarization, Healthcare AI, Natural language processing, Data preprocessing, Machine learning

Recommended Citation: Izuchukwu , C., Wimmer, H., Powell, L.M., (2026). One Health: AI in Healthcare Summarization. *Journal of Information Systems Applied Research and Analytics*, v19(n1) pp 4-14. DOI# https://doi.org/10.62273/TDBT3523

One Health: AI in Healthcare Summarization

Chiazam Izuchukwu, Hayden Wimmer and Loreen Marie Powell

1. INTRODUCTION

AI in healthcare is improving ways to organize and interpret qualitative data. Large language models (LLMs) have demonstrated remarkable skill in dealing with structured narrative information such as this. LLMs are designed to produce short and fluent summaries from large text corpora, a property that can be highly useful for health informatics. Healthcare documents often include elaborate medical protocols, reported research findings, and clinical narratives, all of which can greatly benefit from effective summarization to support both information retrieval and informed decisions. Such LLMs designed for summarization tasks include BART (Bidirectional and Auto-Regressive Transformers). Considering the robust performance of BART in text generation tasks, it might be a preferable choice for summarizing challenging healthcare documents such as the One Health document used in this study. One Health initiative emphasizes the interconnectedness of human, animal, and environmental health, resulting in detailed reports that require efficient summarization to support research and policymaking.

We used two different approaches to measure the performance of the summarization in this study. The first approach is BART-specific, which contrasts two different summarization techniques: (1) a recursive technique, where a further summary of summaries is made to compress and condense, and (2) a direct summarization technique, where only one single summary is generated within a set constraint specifying that length of the output. The second method is to compare BART against conversational AI models such as ChatGPT and NotebookLM. This analysis compares BART summarization performance against the conversational models with respect to quality and credibility. The evaluation takes into account typical evaluation metrics like ROUGE (Recall-Oriented Understudy for Gisting Evaluation), BLEU (Bilingual Evaluation Understudy), and BERT scores that capture the recall, precision, and other semantic similarity of the generated summaries.

This work directly addresses the usage and deployment of LLMs for a healthcare context by evaluating the performance of BART on various summarization methods and its performance relative to conversational artificial intelligent (AI) models. It demonstrates the potential of recursive summarization methods, explores the capabilities of traditional LLM versus conversational AI, and informs assimilation AI design to facilitate summarizing complex health data, elevate information retrieval, and support decision-making.

2. LITERATURE REVIEW

This is the first systematic review to explore the application of AI in healthcare and the challenges faced during its implementation. AI technologies like machine learning (ML) and natural language processing (NLP) are revolutionizing healthcare through the help of diagnostic tools, individualized treatment plans, patient monitoring devices and operational tweaking too. However, integrating AI into healthcare systems has many challenges such as data privacy, ethical and legal issues, interoperability to standards-compliant databases, scalability in training and operating conditions, and human-system interaction. A generation ago, scholars wrote the ethical guidelines and legal frameworks that allowed for interoperable systems, developing similar robust cybersecurity tools that should include universally shared standards, as well as, software to streamline the deployment of these practices by nonexperts. Secinaro et. al (2021) called for increased collaboration across disciplines and better education programs for healthcare providers to promote AI as an integrated tool alongside efforts with ongoing research development.

The DialoGPT, intended for chatbot purposes, is a neural conversational response generation model designed to generate replies similar to human conversation. Specifically, Y. Zhang et al. (2019), used the Hugging Face PyTorch transformer to work on 147 million conversation-like exchanges from Reddit spanning 2005–2017. It performs at near-human levels in single-turn dialogue for both automatic and human evaluations. DialoGPT obtains the number of related, context-relevant, and highest-quality replies as baseline systems.

They also released the pre-trained model and training pipeline to facilitate future research in neural response generation and open-domain dialogue systems.

Su, et al. (2022) investigated the use of automatic summarization with deep neural networks to lighten the load of physicians in writing documentation. The authors created an "extract-and-abstract" approach for crafting the History of Present Illness (HPI) section in a clinical note through BART. Similarly, a classifier was trained to predict whether each utterance belongs to the clinical section of interest, and a subset of correlated utterances was returned for summarization. They noted other types of filters to address the token limit of the BART model. Their approach can complement ROUGE scores with medical concept extraction and be scalable.

Radford et al. (2018) studied the practicality of a possible solution for better transfer learning using large unlabeled text corpora by suggesting some generative pre-training approach and task-specific fine-tuning. They repeated these experiments for multiple benchmarks. Results were consistent in that the proposed task-agnostic model outperformed models trained to solve a given problem. Among their results, these improvements included an 8.9% increase in accuracy for logical reasoning, a 5.7% improvement for question-answering tasks, and a 1.5% performance on textual entailment. Evaluating the method across multiple benchmarks demonstrates the model's ability to achieve state-of-the-art performance on a diverse set of NLP understanding tasks.

3. TECHNICAL BACKGROUND

LLMs are transformative technologies in NLP. BART is known as encoder-decoder model, primarily designed for NLP generation tasks. Typically, this kind of model utilizes vast datasets and advanced ML techniques to accurately understand and generate human language. The foundational architecture behind LLMs typically involves deep learning techniques, such as transformers, which allow them to process and produce text that closely mimics human linguistic abilities (Vaswani et al., 2017).

Brown et al. (2020) aimed to close this gap and investigate how scaling of Language Models affects broadly generalizable, few-shot performance. They took advantage of the language model GPT-3, one with 175 billion parameters, which is an order of magnitude larger than the predecessors, such as BERT and XLNet, for few-shot absence by gradient-based optimization or fine-tuning. The approach focused on showing tasks to GPT-3 using only text-based interaction, measuring performance across translation pairs and question answering, as well as cloze-style fill-in-the-blank tests that required reasoning and adapting models from different domains. The test revealed that GPT-3 was proficient at a wide range of NLP tasks and could generate human-like text, but it struggled on certain datasets or specific civil issues with training on large web corpora.

BART is a denoising autoencoder that is used to pre-train sequence-to-sequence models by leveraging the Transformer-based encoder-decoder architecture. It perturbs the text using a noising function and then trains a model to reconstruct the original text. Because it uses both left-to-right generation and a bidirectional encoder, BART generalizes previous models like BERT (due to its bidirectional encoder) and GPT (due to its decoder). It gets impressive results with only random shuffling of sentences and a simple in-filling procedure where spans of text are replaced by a single mask token. When fine-tuned for text generation tasks, Lewis et al. (2019) reported that BART showed remarkable comprehension and generation performance. On GLUE and SQuAD, it matches RoBERTa performance (and achieves new state-of-the-art results for abstractive dialogue, question answering, and summarization). BART also yields a 1.1 BLEU point improvement over the back-translation baseline with only pretraining in the target language.

The Transformer architecture by Vaswani et al. (2017) suggest that it is possible to implement a sequence-to-sequence model without using RNNs and replace them with attention mechanisms only. Such an approach also facilitates parallelization and validation of the model tuned in this way much faster. The model achieves 28.4 BLEU on the English-to-German and 41.0 BLEU on the English-to-French translation, outperforming all the previously reported single models, at a fraction of their size. On the English-to-German task, it improves over the best ensemble by more than 2 BLEU points. The model is more efficient, as shown by reaching these scores at a fraction of the training cost in 3.5 days on eight GPUs for the English-to-French task.

After it has been found that using NLP models directly on the biomedical domain text is far from optimal due to different expressions specific to the biomedical language. Biomedical

Text Mining has become an even more promising area as it runs in parallel with the huge amount of scientific data. In response, Lee et al. (2020) introduced BioBERT, a domain-specific pre-trained language model. BioBERT outperforms BERT and previous state-of-the-art models by a notable margin in biomedical named entity recognition, relation extraction, and question answering. It significantly improves F1 score and MRR on these tasks, showing the effectiveness of domain-specific pre-training. Their study argues that pre-training BERT on biomedical texts can give it a head start in understanding the intricate languages in this domain.

Clinical responses to big textual data, for example, in electronic health records (EHR), are time-consuming. Introduction to LLMs have proven effective in NLP. However, their ability to summarize clinical information across various tasks is unclear. Van Veen et al. (2023) explored the behavior of eight LLMs using domain adaptation strategies on six datasets, which include radiology reports, patient queries, and progress notes for model performance. By comprehensive quantitative analysis, the work points out model-adaptation trade-offs and cases where more modern LLMs may not be better than previous ones. The best LLMs returned summaries that human evaluators rated complete and more accurate than those written by humans in a clinical reader study with 10 MDs. A follow-up qualitative analysis identified pain points common to both LLMs and human summarizers, while traditional NLP metrics were mapped against physician preferences for better-grounded scores. This is the first study claiming that having LLMs available can allow one to do better on a large number of tasks in clinical summarization compared with human experts and possibly reduces documentation burdens, allowing clinicians greater capacity for providing personalized care.

Improvements in large pre-trained neural networks such as BERT have changed how NLP is performed, and researchers are continually exploring what kind of linguistic knowledge these models can pick up from unlabeled data. Most of the previous works have explored the analysis of model outputs and internal representations. At the same time, Clark et al. (2019) instead focuses more on BERT in terms of its attention mechanisms. The methodology includes assessing the behaviors of attention heads, such as delimiter-token-focused attention or concentrating on neighboring phrases and sentence-level broader attention. Some heads that correspond very well to linguistic notions such as syntax and coreference, for example, by properly recognizing the direct object or a reference of determiners are noteworthy. They further contribute with an attention-based probing classifier that shows, in addition to being improved by BERT features on SRL and constrained dependency parsing tasks, this transformer is capturing a lot of syntactic pressure. The results highlight BERT's capacity to learn complex syntactic structures via attention, providing strong support for tuning and adaptation in NLP.

4. METHODOLOGY

In this study, we performed our summarization in Google Collab using Python-based libraries namely hugging face transformer, NLTK and some other utility packages for text preprocessing and metric calculations. Because the model summarization and evaluation were carried out in a free cloud-based environment with limited GPU and RAM support, the text data was first converted from pdf to txt document to ensure efficient processing of the large data. Figure 3.1 displays the process flow.

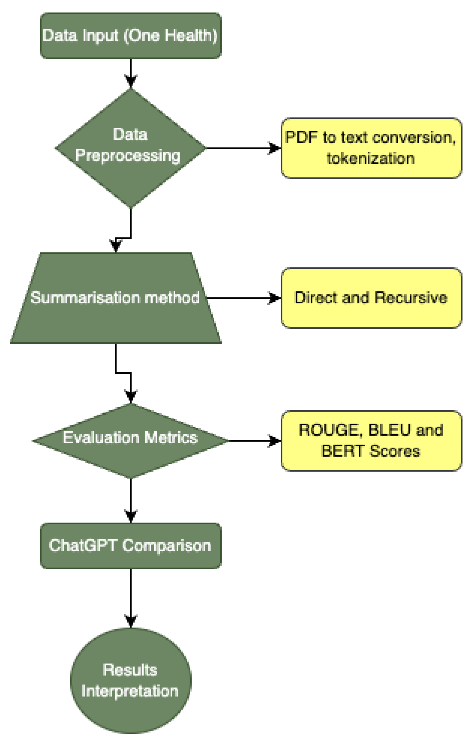

Figure 3.1: Process flowchart

Dataset
The dataset used in this study is a One Health Joint plan of action (Organization, Programme, & Health, 2022). The Quadripartite Organizations—the FAO, UNEP, WOAH, and WHO work around the globe with a single health approach that addresses threats to human health at the interface between people-animal-plant environment. In response to international calls for pandemic prevention and sustainable health promotion, they developed the One Health Joint Plan of Action 2022-2026 an 86-page document dedicated to this integrated approach to management.

The document was chosen because of its qualitative nature, relevance in health informatics and how much detail it contains. The document was first converted from PDF to text file to enable preprocessing using NLP tools. Preprocessing was then applied to the dataset to include tokenization, stop word removal and normalization to ensure compatibility with our preselected LLM.

BART
BART is a sophisticated LLM developed by Facebook AI which combines the positives of BERT bidirectional encoding with GPT's autoregressive decoding, making it very efficient in a range of NLP tasks (Lewis et al., 2019). The framework of this LLM allows it to autofill missing information and generate coherent information.
This phase involved the utilization of two different strategies, which are highlighted as follows.

- Direct Summarization: We obtained the direct summarization using the BART model. The goal for this method was to result in a shorter version of the document yet within bounded restrictions on how many words could be included as part of the summary and having some control over what information would stay or what we considered essential.
- Recursive Summarization (Summary of Summary): Here, the initial OG summary is made with the BART model. The resulting summary is then further summarized again to achieve a more concise output. This recursive approach aimed to maximize brevity while retaining essential information.

Evaluation Metrics
ROUGE, BLEU, and BERT score metrics are employed to assess the effectiveness of the summarization strategies. The evaluation of these models is crucial otherwise, we would have to depend solely on human evaluation which can be a more time consuming and costly path.

ROUGE Score
ROUGE scores were calculated to evaluate the overlap of n-grams and word sequence between the generated and referenced summaries and between the generated summaries and OG. Lin (2004) stated that ROUGE score is a set of metrics used to evaluate the quality of summaries by comparing them to reference summaries. ROUGE metrics are commonly used in NLP tasks to measure the similarity between generated and reference summaries.

BLEU Score
BLEU scores were calculated to ascertain the precision of word overlap, which sheds insight into the accuracy of the generated summary. According to Papineni et al. (2002), the BLEU score is a metric used to evaluate the quality of text that has been machine-translated from one natural language to another. The BLEU score compares the n-grams of the candidate translation's n-grams with the reference translations and counts the number of matches. These matches are then used to calculate precision for the candidate translation.

BERT Score
The BERT Score was calculated to evaluate the semantic relationship between the summary and the OG, offering more insights into how well the summary captured the main themes of the text dataset. BERT Score is a metric for evaluating text generation quality based on BERT embeddings. It calculates the similarity between the reference and generated text at the token level using contextual embeddings from a pre-trained BERT model (Zhang, et. al., 2019). BERTScore considers precision, recall, and F1 scores based on token similarity. Devlin et al. (2018) provided the formula that computes the cosine similarity between the generated and reference text.

ChatGPT & NotebookLM Comparison
This benchmark provides an evaluation of ChatGPT and Google NotebookLM versus a BART model looking at summarization in terms of quality, brevity, and fidelity. BART is a transformer model on which the training data is fine-tuned, and it can generate summaries of the tex. Still, due to its structured nature, where the input and output are sampled from similar distributions, which work great for structured datasets, it does not generalize well to many

other contexts. ChatGPT, a conversational AI, can solve more versatile tasks and summarize them in an engaging way that sounds natural to humans. Google NotebookLM is another conversational AI application for personal knowledge management that improves summarization thanks to context awareness and document amalgamation based on uploaded files. ChatGPT and NotebookLM are also as powerful but more versatile, allowing for more nuanced output compared to BART, which returns a clear and concise summary. We compare model performance on both frameworks, illustrating the strengths and trade-offs of traditional LLM vs conversational AI models to inform tool selection according to the summarization use case.

Human Evaluation
Human evaluation was provided as manual assessment of the generated summaries and their agreement with OG quality, beyond automated metrics. The ratings were done using a Likert scale of 1–7, where 1 strongly agreed that the summary was of high quality and 7 strongly disagreed that the document was of high quality. The evaluation was carried out by 24 one & public health professionals, via an IRB approved study, who were all recruited through a prominent one health conference via voluntary participation. The data obtained from the evaluation is used to perform various statistical analysis to gain more insights. The models evaluated are defined in the table 3.1, with recursive summarization referring to the summary of a previous summary in iteration until the desired length is achieved and direct summarization referring to summarizing the original text directly to the desired length without any iterations. ChatGPT and NotebookLM represent conversational AI-based summarization systems.

GROUP	SUMMARY TECHNIQUE
1	Recursive
2	Direct
3	ChatGPT (OpenAI)
4	NotebookLM

Table 3.1: Grouping of Summarization Methods

Descriptive Statistics
We provided descriptive statistics to summarize key characteristics of the responses generated by the human evaluation. We computed the following metrics:

- Mean (μ): The average score of the responses given to the group, where lower mean scores in the evaluation signified that reviewers found the summaries to be highly aligned with the original content, while higher mean scores suggested a weaker quality.
- Standard Deviation (SD) (σ): Quantifies the spread of summary performance across methods.
- Median: A value that represents a central point in the distribution of scores and is more resistant to extreme scores than the mean, thus making it a further alternative measure of central tendency.
- Minimum and Maximum Values: Identifies the score range, helping detect extreme values.

One-Way ANOVA
In this study, Analysis of Variance (ANOVA) was used to assess whether there are statistically significant differences in the performance of multiple summarization methods. The study compared the performance of multiple summarization techniques like recursive summarization, direct summarization, ChatGPT, and Google NotebookLM. Due to the nature of the comparison, one-way ANOVA was the suitable statistical test for analyzing variation in performance. Prior to running ANOVA we confirmed homogeneity of variance via Levene's test statistic. The null hypothesis (H_0) and the alternative hypothesis (H_1) was as follows:

- H_0: There is no statistically significant difference in mean performance between different summarization approaches.
- H_1: There is a statistically significant difference in mean performance between different summarization approaches.

Post-Hoc Analysis
Where ANOVA indicated a significant difference, Tukey's Honest Significant Difference (HSD) test was conducted to determine which summarization methods differed significantly from each other. This test is done to confirm the differences detected were not attributable to random variation but were real differences in summarization quality.

5. RESULTS

We utilized three evaluation metrics for each method, ROUGE (n-gram overlap), BLEU (word precision) and BERTScore (semantic similarity). All these metrics give a robust view of individual performance of each summarization approach with respect to content retention, brevity, and semantic similarity.

Table 4.1 investigates a recursive summarization process in which the text is summarized at successive stages and the quality at each stage is evaluated based on standard metrics. The comparison begins by evaluating the initial summary against the original (OG) text, followed by comparing the 2000-word summary to the initial (IN) summary (sum), and finally, comparing the 145-word sum to the 2000-word version.

Recursive Summarization					
Comparison	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	BERT-F1
IN sum vs OG	0.76	0.66	0.64	0.45	0.86
2000 vs IN	0.19	0.17	0.16	0.0	0.82
145 vs 2000	0.13	0.13	0.12	3.20 E-06	0.86

Table 4.1: Recursive Method

Table 4.2 is like table 4.1 except the comparison involves evaluating all the summaries 2000-word, and 145-word versions against the OG.

Recursive Summarization					
Comparison	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	BERT-F1
2000 vs OG	0.14	0.13	0.13	3.16 E-06	0.80
145 vs OG	0.01	0.01	0.01	1.53 E-79	0.79

Table 4.2: Recursive Summarization (Sum vs OG)

Table 4.3 investigates a direct summarization process in which the text is summarized within bounded restrictions and is evaluated based on standard metrics. The comparison involves evaluating all the summaries 2000-word, and 145-word versions against the OG.

Direct Summarization					
Comparison	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	BERT-F1
2000 vs OG	0.139338	0.130013	0.132797	4.16E-06	0.815139
145 vs OG	0.011799	0.011145	0.010853	1.51E-79	0.814662

Table 4.3: Direct Method

Fig 4.1 compares of direct and recursive summarization performance across ROUGE, BLEU, and BERT-F1 metrics at two levels of compression (2000 vs OG and 145 vs OG). Results show minimal differences in ROUGE and BLEU, with direct summarization slightly outperforming recursive summarization on BERT-F1.

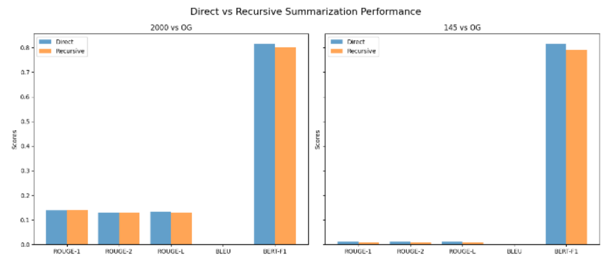

Fig 4.1: Performance Comparison of Recursive and Direct Summarization

Table 4.4 investigates a ChatGPT summarization process in which the text is summarized based on AI prompts and evaluated based on standard metrics. The comparison involves evaluating all the summaries IN, 2000-word, and 145-word versions against the OG.

ChatGPT Summarization					
Comparison	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	BERT-F1
IN vs OG	0.03	0.00	0.02	3.98 E-25	0.66
145 vs OG	0.02	0.00	0.01	8.58 E-25	0.66
2000 vs OG	0.03	0.00	0.02	4.53 E-25	0.66

Table 4.4: ChatGPT Method

Table 4.5 investigates a notebookLM summarization process in which the text is summarized based on AI prompts and evaluated based on standard metrics. The comparison involves evaluating all the summaries 145-word versions against the OG.

NotebookLM Summarization					
145 vs OG	0.01	0.01	0.01	1.22 E-85	0.81

Table 4.5: NotebookLM Method

Comparison of summarization approaches (Recursive vs Direct)

ROUGE score

Recursive summarization presents relatively higher ROUGE scores, especially in ROUGE-1, ROUGE-2, and ROUGE-L. It seems the recursive approach preserves word and phrase overlap better.

Direct summarization has comparatively lower ROUGE scores, especially when directly summarizing down to 145 words. This corresponds to the fact that with increasing direct summarization, the method accounts for fewer word overlaps with the original text, resulting in a loss of accuracy.

BLEU score

Minimal BLEU scores over recursive and direct methods allude to low precision and direct word overlap. On the other hand, when summarizing in smaller steps, recursive summarization achieves even slightly higher BLEU scores, which corresponds to better alignment in word choices. The BLEU score of the direct summarization approach is especially poor at the 145-word count, indicating that condensing such a large input corpus into one output token causes a loss in accuracy.

BERT score

Both methods resulted in a reduced BERT-F1 score after each iteration, however, recursive summarization yielded a consistently higher score at intermediate steps (145 vs 2000 at 0.857962), suggesting that the recursive method can preserve semantic meaning across iterations more effectively. While direct summarization retained a stable but lower BERT-F1 score compared to recursive summarization, this could suggest a trade-off between conciseness and semantic preservation.

Comparison with ChatGPT Summary (Bart vs ChatGPT)

ROUGE and BLEU score

The ROUGE and BLEU scores for ChatGPT summaries are the lowest across all our comparisons. The scores are particularly low for ROUGE-2 and BLEU, signifying that there is minimal overlap in phrasing and word choice with the OG. This implies that ChatGPT summaries could focus more on fluency than text fidelity. The differences in ROUGE and BLEU scores between ChatGPT and both BART-based summarization methods suggest that ChatGPT may not be as detail-focused when it comes to summarization, focusing more on synthesizing overall content.

BERT score

Across recursive and direct summarization BERT-F1 scores, ChatGPT's performance is below them with an average score of ~0.660550. This shows that ChatGPT holds high semantic similarity but cannot preserve the original meaning and context as well as BART-based models.

Comparison with NotebookLM Summary (Bart vs NotebookLM)

The summary from the NotebookLM yields lower BLEU and ROUGE scores compared to recursive and direct summarization approaches. Still, with a comparable BERT score, the Notebook approach does not replicate the original text very closely in terms of precision and fidelity but is somewhat aligned in terms of semantics. Like ChatGPT summaries, the NotebookLM method is not best suited when keeping the generated text very close to the OG. However, it is impactful when a general conceptual summary is acceptable.

Statistical Analysis

The descriptive statistics give some vital information on the relationship between summarization methods and OG. The mean ratings illustrated that Group 3 had the lowest average score (2.33), which implied that the summaries generated in this group were considered the highest quality with the document being summarized. In contrast, Group 2 had the lowest mean (5.5), indicating that reviewers considered the summaries in this group to be of lower quality. Table 4.6 provides the descriptive statistics.

Descriptive Statistics						
Data	N	Mean	Std. Dev	Std. Error	Min	Max
1	6	3.67	1.51	0.62	2	5
2	6	5.5	1.98	0.81	3	7
3	6	2.33	1.37	0.56	1	5
4	6	2.67	1.03	0.42	1	4
Total	24	3.54	1.89	0.39	1	7

Table 4.6: Descriptive Statistics

The SD was used to illustrate variability in the ratings. Group 2's scores were the most heterogeneous (SD = 1.98), indicating that participants' ratings were most spread. Meanwhile, Group 4 recorded the lowest SD

(1.03), implying that group's ratings were more homogeneous, as consensus was reached by the participators concerning quality level. The minimum and maximum values tell us some of the difference between the groups, with Group 3 reporting the lowest rated response (1), indicating strong alignment with the OG, and Group 2 provided a rating (7) indicating a weak quality.

An ANOVA test was conducted to determine if these ratings exhibited statistical significance. The p-value was obtained as 0.01, which is less than 0.05 significance value. This means that at least one group's ratings differed significantly from the others, which implies that some summarization methods produce summaries like the OG than others. Table 4.7 displays the results of ANOVA test.

ANOVA					
Data	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	36.46	3	12.15	5.34	0.01
Within Groups	45.50	20	2.28		
Total	81.96	23			

Table 4.7: ANOVA

To determine specific differences between groups a Tukey HSD post-hoc test was conducted. The results (Table 4.8) showed that Group 2 ratings were significantly higher than Groups 3 (p = 0.01) and 4 (p =0.02) as evidenced by a lower mean (where 1 = strongly agree in a 7 scale Likert format). Since a higher mean score signifies weaker agreement, this confirms that Group 2's summaries were significantly rated with less agreement than those from Group 3 and Group 4. Nonetheless, there were no significant differences among Groups 1, 3 and 4 indicating that the summarization methods employed in these cases led to summaries whose quality were similar.

(I) Group	(J) Group	Mean Diff (I-J)	Std. Error	Sig.
1	2	-1.83	0.87	0.19
	3	1.33	0.87	0.44
	4	1.00	0.87	0.67
2	3	3.17*	0.87	**0.01**
	4	2.83*	0.87	**0.02**
3	4	-0.33	0.87	0.98

Table 4.8: TUKEY

6. DISCUSSION

The results of this work demonstrate the empirical performance profiles of recursive summarization, direct summarization and ChatGPT summarization on complex qualitative health documents. While they all provide different advantages and disadvantages, each have trade-offs between content fidelity, semantic coherence and brevity of the summaries they produce. Semantic Fidelity: Adding to the BERT-F1 scores, recursive summarization proves to be the best method for maintaining the essential semantics of the OG whilst continuously shortening the text. This is useful for large documents where multiple passes of compression are needed to retain the overall context.

Word Repetition and Structure: direct summarization struggles to retain exact word usage and structure, even with short lengths. This is equally apparent in the drop in ROUGE and BLEU scores in the second stage of summarization. On the other hand, recursive summarization has a more gradual decrease in ROUGE scores, suggesting that content overlap is preserved better with iterative summarization. By contrast, the conversational AI (ChatGPT and NotebookLM) tends to preserve less focus on the text due to lower ROUGE and BLEU scores. This suggests that conversational AI models generally have wider applicability but may not be the best choice for high-fidelity summarization tasks compared to models such as BART that better specialize in structured forms of compression. These findings are further corroborated by the combination of statistical analysis and human observation. Group 3 (ChatGPT) received the most similar to original text, followed by Group 4 (NotebookLM) and finally Group 2 (Direct Summarization) obtained the highest mean, indicating lower quality. The results show that recursive summarization can maintain the main ideas of the source document much better than direct summarization, which tends to introduce more noise or information loss. While the AI models were able to summarize well enough in a general way, it may still not be precise enough for accurate technical summarization tasks.

Some recommendations to improve summarization in health informatics can be made based on the findings from this study. If the health document is extensive and lengthy, then recursive summarization, which iteratively shortens the content, can be performed to maintain key information. Another possibility comes from combining these approaches and using a hybrid model that includes recursive and direct summarization to achieve some balance between compression/absorption and length maintained within the summary, where summaries of different lengths but still potentially varying levels of detail about the source sentences are required. ChatGPT and NotebookLM is less accurate but useful for high-level summaries for non-technical audiences. To improve content fidelity, it is recommended that BART be fine-tuned on health-specific data.

7. CONCLUSIONS

Recursive summarization holds promise for creating ultra-short yet informative summaries of qualitative health data, so this study highlights its potential. The mechanism used in this method folds the content iteratively in a way that ensures shorter text while still retaining information, making it ideal for intricate and content-rich documents. Although direct summarization is more efficient with small compressed content, it can lose more information and coherence. On the other hand, ChatGPT summarization provides a general tool but may not offer the precision needed for clinical high-fidelity technical summarizations.

These results indicate that recursive summarization with LLMs such as BART produces concise and meaningful summaries for health-related documents outperforming direct summarization method in preserving meaning for complex content. With fine-tuning, models such as ChatGPT or Google's T5 may offer equally strong or superior performance. This paper is not without limitations. However, it could usefully guide future work in investigating the hybridization of recursive and direct summarization methods or domain-centric tuning to improve summary quality.

8. REFERENCES

Brown, T. B.; Mann, B., Ryder, B., Subbiah, M. Kaplan, J. Dhariwal, P. et al (2020). Language models are few-shot learners. *Proceedings of the 34th International Conference on Neural Information Processing System. 159, 1877 - 190. https://dl.acm.org/doi/ pdf/10.5555/3495724.3495883*

Clark, K., Khandelwai, U., Levy, O.,Manning, C. D. (2019). What does Bert look at? An analysis of Bert's attention. *https://arxiv.org/abs/1906.04341*

Devlin, J. Chang, M-W., Lee, K., & Toutanova, K. (2018). Pre-training of deep bidirectional transformers for language understanding. *https://arxiv.org/abs/ 1810.04805v1*

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics, 36*(4), 1234-1240. https://doi.org/ 10.1093/ bioinformatics/btz682

Lewis, M., Liu,Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871–7880, Online. Association for Computational Linguistics. *https://doi.org/10.18653/ v1/2020.acl-main.703*

Lin, C.-Y. (2004). *Rouge: A package for automatic evaluation of summaries.* Paper presented at the text summarization branches out. Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL, 74-81, Barcelona, Spain, 25 July 2004. https://www.aclweb.org/anthology/W04-1013.pdf

Organization, W. H., Programme, U. U. N. E., & Health, W. O. f. A. (2022). *One health joint plan of action (2022–2026): working together for the health of humans, animals, plants and the environment*. World Health Organization. https://www.who.int/publications/i/item/9789240059139

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *Bleu: a method for automatic evaluation of machine translation.* Paper presented at the Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 311 - 318. https://doi.org/10.3115/1073083.1073135

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018). Improving language understanding by generative pre-training. https://cdn.openai.com/ research-covers/language-unsupervised/language_understanding_paper.pdff

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. *BMC medical informatics and decision making, 21*, 1-23. https://doi.org/10.1186/s12911-021-01488-9

Su, J., Zhang, L., Hassanzadeh, H. R., & Schaaf, T. (2022). *Extract and Abstract with BART for Clinical Notes from Doctor-Patient Conversations.* Paper presented at the Interspeach, Incheon, Korea. https://www.isca-archive.org/interspeech_2022/su22b_interspeech.pdf.

Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., Polacin, M., Reis, E.P., Seehofnerova, A., Rohatgi, N. Hosamani, P., Collins, W. Ahuja, N., Langlotz, C., Hom, J. Gatidis, S. Pauly, J. Chaudhari, A. (2023). Clinical text summarization: adapting large language models can outperform human experts. *Research Square*. https://doi.org/10.21203/rs.3.rs-3483777/v1

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. 31st Conference on Neural Information Processing System Long Beach, California. https://dl.acm.org/doi/pdf/10.5555/3295222.3295349

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. *arxiv.org/abs/1904.09675*

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Dolan, B. (2019). Dialogpt: Large-scale generative pre-training for conversational response generation. *https://arxiv.org/pdf/1911.00536*

Architecting the Academic Assistant: A Review of Dual-Processing Architectures and Hybrid Chatbot Design in Higher Education

Sai Prakash Reddy Rao
srao11@students.kennesaw.edu

Lei Li
lli13@kennesaw.edu

Hisham Haddad
hhaddad@kennesaw.edu

Selena He
she4@kennesaw.edu

Kennesaw State University
Marietta, GA 30060

ABSTRACT

The integration of hybrid chatbots into academic websites is transforming institutional communication by fusing the structured precision of rule-based systems with the generative flexibility of Artificial Intelligence. This survey examines the core design mechanisms that drive this redefinition of student support, specifically focusing on dual processing architectures, contextual adaptation, and intent recognition. It further analyzes data integration strategies used to deliver personalized experiences, such as tailored course recommendations and dynamic event suggestions. Beyond technical architecture, this study evaluates the strengths, limitations, and practical applications of hybrid chatbots in enhancing accessibility, scalability, and user satisfaction within the university environment. These insights equip researchers and developers with a strategic guide for enhancing chatbot efficacy within higher education. Distinguishing itself from generalist reviews, this survey specifically addresses the challenge of integrating academic jargon, offering a concrete blueprint for building resilient, domain-aware student support systems.

Keywords: Hybrid Chatbots, AI in Education, Generative AI, Dual Processing Architecture, Contextual Adaptation, Intent Recognition.

Recommended Citation Rao, S., Li, L., Haddad, H., He, S., (2026). Architecting the Academic Assistant: A Review of Dual-Processing Architectures and Hybrid Chatbot Design in Higher Education. *Journal of Information Systems Applied Research and Analytics*, v19(n1) pp 15-25. DOI# https://doi.org/10.62273/XQBK2345

Architecting the Academic Assistant: A Review of Dual-Processing Architectures and Hybrid Chatbot Design in Higher Education

Sai Prakash Reddy Rao, Lei Li, Hisham Haddad and Selena He

1. INTRODUCTION

A Chatbot is a software program that mimics communication between a human and a machine to simulate human conversation with a user in text or voice. It uses Natural Language Processing (NLP) to provide appropriate responses based on the user query. From customer service, healthcare, academia to e-commerce, chatbots are used in a variety of industries due to their real-time response, customization features, and job automation.

Not all digital pathways are easy to navigate; however, some can get lost within mazes of menus, flashy visuals, and buried information. The research of (Pittsley and Memmott, 2012) has shown that smooth navigation is crucial, if users cannot quickly and easily find what they need, they are less likely to return. This is even more important for people with disabilities who need accessible design, not just design with accessibility considerations. Websites are now the digital gateways that connect us to the information we seek in a world where it is simply a click away. Like other organizations do, universities built their own online portals to help connect with their students, staff, and partners.

Educational institutions are now integrating Artificial Intelligence (AI) to better support and guide students on their university journey. AI-based chatbots have become major power tools to automate routine work, answer students' questions, and facilitate administration work. To cut through the noise and get engagement, chatbots have emerged and are becoming the go-to for websites looking to elevate user experience.

The advancement of chatbot technology has transformed educational settings, especially academic websites. These include bots providing information, increasing participation, and simplifying administrative tasks for students and staff. This review explores literature on different types of chatbots - Rule-based, AI-based, and Hybrid chatbots, focusing on their application within academic contexts.

Rule-based chatbots use predefined rules and scripts, which makes them a good fit for handling repetitive queries. Rule-based chatbots lack the capacities to handle nuanced queries, as they heavily rely on keyword matching or strict rule structures. Their inability to learn from interactions or adjust dynamically to changing scenarios is a persistent limitation in academic settings, where students frequently have context-specific questions. By interpreting all forms of user input and producing conversational contextual responses, AI-based chatbots can be more interactive and adaptable using machine learning and NLP. AI-based chatbots can be a resource intensive data used as training material and computational power needed. The lack of a quality dataset for correct answers is one of the major challenges to their use in academia. Hybrid chatbots, as the name suggests, use both rule-based and AI-assisted components to provide users with high flexibility, dependability, and availability for simple to complex questions. Simple questions will be answered by rule-based logic and AI will do the rest for more complex or context-specific questions.

Most surveys either address the limitations of rule-based systems or the potential of AI-based chatbots, while there is a noticeable gap in literature focused specifically on the design mechanisms of hybrid chatbots. The combination of these approaches, which can provide a more effective and flexible solution, is less explored, particularly focusing on the design approaches suggested for academic websites. The originality of this survey lies in its focus on integration strategies for hybrid chatbots in academia, particularly how rule-based and AI-driven methods can be adapted to address academic jargon and context-specific queries that are not adequately covered in existing reviews. This survey seeks to address this research gap by analyzing in detail design elements specific to hybrid chatbots for academic websites. This survey will contribute to outlining the key considerations for developing chatbots that can enhance their performance efficiently. This paper gives a comprehensive overview of the history of chatbots. In addition, this survey aims to answer the following questions:

RQ1: What are the key design mechanisms currently used in hybrid chatbots for academic websites?
RQ2: How effectively do hybrid chatbots switch between rule-based and AI-driven responses in real academic scenarios, and what are the common challenges encountered?
RQ3: To what extent do current hybrid chatbot implementations on academic websites support personalized responses, and what are the barriers to achieving full personalization?

The rest of the survey is organized as follows: The next subsection discusses the survey methodology and Section II covers the foundational years of chatbots, rule-based, AI-based, and hybrid chatbots. Section III discusses the design mechanisms of hybrid chatbots and delves into the inherent challenges of these chatbots and explores the outlook in this field. Finally, Section IV concludes the survey, summarizing its key findings and overall contributions.

1.1. Survey Methodology

This survey followed a structured methodology to ensure comprehensive coverage. Literature was sourced from IEEE Xplore, ACM Digital Library, Springer, and arXiv using keywords such as ‘Hybrid chatbot’, ‘Rule-based chatbot’, ‘AI chatbot’, and ‘Academic chatbot’. The search covered 2010–2025. Inclusion criteria were: (a) studies focusing on hybrid or academic chatbots, (b) peer-reviewed conference/journal papers, and (c) articles presenting design mechanisms or case studies in academia. Excluded were purely commercial chatbot studies or works without methodological detail. No new empirical data was collected; rather, this study synthesizes findings across existing literature to highlight design mechanisms relevant to academia.

2. PRELIMINARIES OF CHATBOTS

Chatbots have become an essential part of digital communication, changing how businesses reach customers and how users interact with technology. Chatbots have seen variants in forms and shapes ranging from plain, rule-based systems to advanced conversational AI over the years.

2.1. Rule Based Chatbots

(Weizenbaum, 1966) made one of the first breakthroughs with his pioneering effort ELIZA. It was based on pattern matching techniques, so it could provide limited conversation masking, but all the response came from a predefined ELIZA dataset only. The system was simple but showed that machines could replace humans in dialogue and laid the foundation for new chatbot technology.

Concerning the specific category, rule-based chatbots are very popular in distinct areas because of their effectiveness in the regulation of conventional, formalistic communication. These systems are most useful in cases where the interactions can be script-based, and therefore, are particularly structured in nature. For example, customer support and engagement could deploy rule-based chatbots to interact with users asking common questions in a knowledge base, reducing the load of human agents while ensuring faster assistance.

Rule-based chatbots come in handy for information seeking, where a chatbot takes a user's query and matches it with an existing database returning precise answers in a timely fashion without the need for forwarded assistance. In general, these bots are best used in environments where human–bot interaction is routine, and therefore conversations are normal and the same from one session to the other. However, where the dialogues are more complex, dynamic or challenging they lack the flexibility needed and have therefore given rise to more sophisticated AI-driven systems.

2.2. AI Based Chatbots

Current AI-based chatbots have improved a lot from rule-based systems that dominated the initial stage of their development. SmarterChild (2001) built on the base established by knowledge-based bots by incorporating real-time information retrieval into messaging systems, in preparation for today’s smart personal assistant technologies (Zemcik, 2019).

The evolution of chatbot development occurred with the introduction of intelligent agents. Assistants like Google Assistant, Siri and Alexa also pushed AI into the realm of voice recognition and the Internet of Things (IoT) to make AI assistants more ubiquitous for everyday use and establish them as the center of convenience and engagement (Tulshan & Dhage, 2019).

The use of transformer-based models defines a new generation of AI chatbots. Some recent developments like BERT (Devlin et al., 2019) and GPT-3 (2020) offered improvements in terms of Natural Language Processing models by

Model	Strengths	Limitations	Academic Use Cases
GPT-4 (Meyer et al., 2023)	Generates coherent, multi-turn conversations, excellent for detailed, open-ended queries and personalized advising.	High computational demands, risk of biased responses.	Student advising, personalized recommendations, complex multi-turn queries.
BERT (Wang et al., 2024)	High precision in intent recognition.	Not designed for open-ended response generation.	FAQ handling, standard academic assistance, intent interpretation.
RAG (Lewis et al., 2020)	Combines retrieval for factual accuracy with generative flexibility, ideal for knowledge grounded responses.	Rely on the quality of the retrieval system and knowledge base.	Research assistance, policy-based advising, knowledge grounded queries
MAMBA (Albert and Dao, 2023)	Modular design allows for domain-specific adaptability, enhances response accuracy and relevance.	Still experimental, it requires careful module integration.	Domain-specific tasks like enrollment, library support, research queries.

Table 1: Different Deep Learning Models and their Use Cases in Academic settings

incorporating self-attention mechanisms to achieve better contexts for the understanding of promotional verbal expressions and for making a consistent and smooth response. ChatGPT (2022) and Google Bard (2023) improved this strategy and could develop a spectrum that provided necessary and appropriate and accurate responses in various subject areas ranging from healthcare to business and education. The future trends in the AI chatbot will be its greater adaptability with ethical problems like privacy and data management and the right use of AI in the conversation.

2.2.1 Need for Deep Learning Based Models in Academic Websites

As chatbots began to require greater adaptability, AI based chatbots emerged as a promising solution. Machine learning algorithms are applied to these chatbots to respond dynamically, augmenting their application to more complicated, conversational queries that rule-based engines are incapable of handling. In academic settings, AI-based chatbots have shown potential to provide personalized advising, in-depth course guidance, and research support. Generative models process context and user intent in real time serving customized, coherent responses that help increase user engagement and satisfaction.

The AI based chatbots powered by models like GPT-4, BERT, and the future MAMBA architecture have changed academic support to dynamic, context aware interactions. The deep transformer-based architecture of GPT-4 makes it good at handling multiturn conversations, for example personalized student advising and complex queries, but its high resource requirements and possible biases need ethical monitoring (Huang and Marcus, 2021). Because BERT is bidirectional, it excels at intent recognition in structured queries, which is perfect for handling FAQs, and for typical academic assistance where exact interpretation is essential. Larger LLMs, such as PaLM (Chowdhery et al., 2023) and GPT–4, are capable of the deeper contextual understanding which is needed for cross-disciplinary research support, but at such scale, computational power and monitoring balance are needed to avoid generating gratuitous responses. Conversely, MAMBA enables modular adaptability by which academic chatbots can switch between specialized modules that support tasks such as enrollment or research support and enhanced accuracy in responses to various academic topics. Table 1 provides strengths, limitations, and use cases of different Deep Learning models in academic settings. This collection of models helps to create this comparative landscape in generative AI, balancing accuracy, adaptability, and ethics to improve the digital academic experience.

2.3. Hybrid Chatbots

The evolution of chatbot technology has resulted in huge development of hybrid chatbots. Hybrid, as the name suggests, combines rule-based and AI-based chatbot mechanisms that ensure the efficiency and experience of users. They overcome the limitations of rule-based systems that depend solely on pre-defined scripts, while also pushing toward AI-Based chatbots, which are generative. The combination of these two rule-based and AI-based chatbots is useful to provide both structured responses and complex responses.

The integration of hybrid chatbots in the educational environment is highly significant. (Sonderegger & Seufert, 2022), put forward a conceptual framework for chatbot use cases, in which chatbot can be tutor, learning analyst and support analyst. This framework highlights the educational benefits of chatbots, including personalized learning experiences and enhanced accessibility. Furthermore, chatbots have been employed for mental health support. (Nieva et al., 2020) studied a chatbot called Woebot, which was created to help students who were experiencing academic stress. This presented that these technologies could provide a secure and a stigma free platform for individuals to convey their concerns. They showed that hybrid models work by saying that chatbots can be used for information sharing, risk evaluation and health care management.

Design Mechanism	**Role in Hybrid Chatbots**	**Relevance to Academic Websites**
Dual Processing	Integrates rule-based and AI models.	Efficient handling of simple FAQs and complex academic queries
Contextual Adaptation	Maintains coherence in conversations.	Enables multi-turn advising and jargon interpretation.
Intent Recognition	Categorizes user queries.	Improves accuracy of academic service responses.
Data Integration	Personalizes responses via user data.	Recommends courses, events, and resources tailored to students.

Table 2: Overview of Hybrid Chatbot Design Mechanisms with Academic Use Contexts

Overall, hybrid chatbots display a technical advancement, combining rule-based with AI advancement. Their diverse applications in education, healthcare, and customer service underline their transformative evolution. However, ongoing research and development are crucial to address issues such as user trust, ethical considerations and the need for robust and scalable solutions. To provide an at-glance summary of the design mechanisms discussed in later sections, Table 2 outlines the core mechanisms of hybrid chatbots, their functional roles, and specific relevance to academic settings.

3. DESIGN MECHANISMS OF HYBRID CHATBOTS IN ACADEMIC SETTINGS

3.1. Dual Processing Mechanism

The development of hybrid chatbots has been growing especially in academia where there is the need to employ several models to achieve both efficiency and flexibility. Among the most used techniques in these chatbots, the dual processing architecture is widely adopted, integrating rule-based and generative models. This approach provides a good approach to managing both common and ambiguous questions, which in a way enhances the user experience. The architectural view of the dual processing model is that an API-based middleware layer is used to manage these two engines to balance their performance and manage real time requests.

3.1.1. Pattern Matching Algorithms

These algorithms constitute the core of the rule-based components in dual processing architectures. Some of the examples are Aho-Corasick and Boyer-Moore algorithms which enable the matching of a user query with a set of pre-determined replies. Thus, it answers the recurring questions that students may have, for example, about academic dates or course requirements. The use of the matching pattern guarantees that frequent and recurring queries are executed with great efficiency (Aho & Corasick, 1975).

Use Case: “What are the library hours?” - The pattern matching algorithm instantly retrieves this information from the database and provides a quick, consistent response.

3.1.2. Deep Learning Models

The generative part is based on state-of-the-art deep learning models, such as LSTM, as well as transformer-based architecture such as GPT, BERT. LSTMs are especially good at handling sequential data which makes them ideal for use in conversations that happen over time. Transformers can capture word-level relationships over the entire input and hence handle complex contexts. These models help the chatbot to respond appropriately to different, sometimes even ambiguous questions, thus increasing the level of system’s capabilities to address different student queries (Hochreiter & Schmidhuber, 1997).

Use Case: "How do I apply for a Ph.D. program" - The chatbot uses NLP models (e.g., BERT or GPT) to interpret the context and generate relevant, human-like responses.

3.1.3. Fallback Mechanisms

Dual processing architecture uses confidence-based fallback methods to provide reliability. If the ML model's prediction confidence is too low, the chatbot falls back to the rule-based system, which provides a predefined response. This method maintains a balance between adaptability and dependability, ensuring that the chatbot responds accurately even when presented with ambiguity (Yildirim et al., 2023).

Use Case: "Give me some information about the research opportunities I have" - If the chatbot faces an ambiguous or vague query, like the one above, and if the ML model's confidence score is low, it defaults to the rule-based response, such as "You can explore research opportunities by visiting the student research portal."

3.2. Contextual Adaptation

Contextual adaptation is an essential design technique for hybrid chatbots, particularly in academic settings in which conversations tend to incorporate technical language, long-term dependencies, and coherence across multi-turn interactions. In doing so, it guarantees that the chatbot can indeed interpret and respond to questions naturally creating a well-informed user experience that is more appropriate to the respective context.

3.2.1. Dialogue State Tracking (DST)

DST algorithms are designed to keep track of the evolving state of a conversation. These algorithms use slot-filling techniques to store and update important contextual information, such as the user's current academic focus, preferred courses, or prior concerns. For example, if a student asks about which courses are available and then asks them about course prerequisites those two topics can interweave seamlessly from previous conversation as the chatbot can recall. The underlying mechanism is enabled by DST and helps the chatbot maintain the conversational flow by understanding follow up reasonably well (Matthew et al., 2014).

Use Case: A student asking, "What courses can I take for my major?" followed by "What are the prerequisites for the AI course?" gets seamless context-aware responses where the chatbot dynamically links both queries.

3.2.2. BERT (Bidirectional Encoder Representations from Transformers)

BERT is a pre-trained transformer model known for its ability to understand context by processing text bidirectionally. This means that BERT can consider both the left and right context of a word simultaneously, which is essential for interpreting academic queries that may involve complex terminology or subtle nuances. By capturing the meaning of words in context, BERT enables the chatbot to handle subject-specific language more accurately. In academic settings it can really matter when you are trying to make sense of technical questions or when you are trying to distinguish between similar terms that only have different meaning in the context. Additionally, BERT's robust contextual understanding allows the chatbot to remember details from previous conversations, helping it generate responses that build on prior conversations (Wang et al., 2024).

Use Case: For complex academic queries like "Help me understand the differences between supervised and unsupervised learning?" BERT provides accurate and contextually rich answers tailored to the curriculum.

3.2.3. Recurrent Neural Networks (RNNs) and Variants

Recurrent Neural Networks, including their advanced variants like Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks, play a vital role in maintaining context memory over prolonged conversations. RNNs are specifically designed to handle sequential data and are employed to track and recall past interactions within a dialogue. RNNs let the system refer or ask an open-ended question from depth back in a conversation with a user. For something like walking a student through a multi-step process, like course registration or working out graduation requirements, this memory is crucial. It can remember what happened in earlier parts of the conversation so that conversations are consistent and improve the user experience (Cho et al., 2014).

Use Case: During a prolonged conversation about a learning path, RNNs allow the chatbot to recall previous interactions, such as suggested courses or topics of interest, providing coherent recommendations over multiple sessions.

3.3. Intent Recognition

The design of academic chatbots starts with intent recognition as a fundamental component. This helps the system understand and categorize user queries well so that it can provide relevant and timely responses. To improve chatbot ability to identify different intents, several machine learning techniques and models are used that have a different contribution to the overall efficiency of the system.

3.3.1. Support Vector Machines (SVMs)

SVMs are very powerful classification algorithms used for intent recognition on chatbots. They help in categorizing the user queries based on the extracted features, which makes them very effective for both binary and multi-class classification problems. SVMs obtain higher precision in intent recognition by using feature engineering techniques. SVMs can be used in academic chatbots to assist the users in the form of targeted actions, and it will also make the user interactions easy. The main strength of SVMs is in the ability to work with many features that makes it possible to classify queries correctly even in complicated conditions (Cortes & Vapnik, 1995).

Use Case: Queries like "What's the grading policy?" and "How do I enroll in summer classes?" are classified into "policy inquiries" and "enrollment assistance" categories for accurate responses.

3.3.2. Convolutional Neural Networks (CNNs)

CNNs, although traditionally used in image processing, have proven effective in text-based applications, particularly for hierarchical feature extraction. In the context of academic chatbots, they apply complex patterns and dependency capacity in understanding the text. Moreover, CNNs improve performance when processing word embeddings to differentiate between slightly dissimilar intentions such as regarding different academic services. By considering unique words and writing structures, a CNN-based model can randomly learn the difference between inquiries having almost the same category (Kalchbrenner et al., 2014).

Use Case: For the queries, "How do I apply for scholarships?" vs. "What's the scholarship deadline?" CNNs identify distinct intents, ensuring precise routing to relevant resources.

3.3.3. Word Embeddings (Word2Vec, GloVe)

Word embeddings like Word2Vec and GloVe are essential for understanding the semantic relationships between words in user queries. These models transform words into dense vector representations that capture meanings and relationships beyond simple keyword matching. In academic chatbots, word embeddings facilitate better understanding of user intents, even when the phrasing differs from the training data. By using pre-trained embeddings, chatbots can achieve a deeper understanding of language semantics, improving their ability to respond appropriately (Mikolov et al., 2013).

Use Case: For the phrases like "exam schedule" instead of "test dates," Word2Vec captures the semantic similarity, allowing the chatbot to respond accurately with the relevant examination timetable.

3.4 Data Integration for Personalization

Data integration and personalization were the keys to improving the effectiveness of academic chatbots. Chatbots provide custom support by utilizing advanced algorithms and data driven techniques to enhance user satisfaction. These techniques provide relevant, context-aware recommendations to users appropriate to their interests and academic needs.

3.4.1. Collaborative Filtering

Collaborative filtering is a recommendation algorithm that analyzes user behavior patterns to suggest content based on the preferences of similar users. For academic chatbot, this approach is useful to suggest the courses, or any relevant campus event to the user according to their interests. As an example, if a user interacts with content often on artificial intelligence, the chatbot might recommend them related workshops or study groups. The chatbot greatly improves the user experience by delivering personalized, contextual information using collaborative filtering, leading to a sense of support and engagement (Sarwar et al., 2001).

Use Case: The chatbot analyzes usage patterns and suggests study groups or peer networks for subjects the student has shown interest in, such as offering a coding workshop for frequent programming-related queries.

3.4.2. Content-Based Filtering

Content based filtering concentrates on features of content which a user has already shown interest in. It employs metadata of academic materials such as subjects, keywords, or research areas for producing suggestions. For instance, a chatbot may recommend a lecture or a paper in this field if the student asks about machine learning a relevant amount of time. This technique gives a guarantee that the suggestions are tailored to the customer since it is based on the customer's personal interest collection (Lops & Semeraro, 2010).

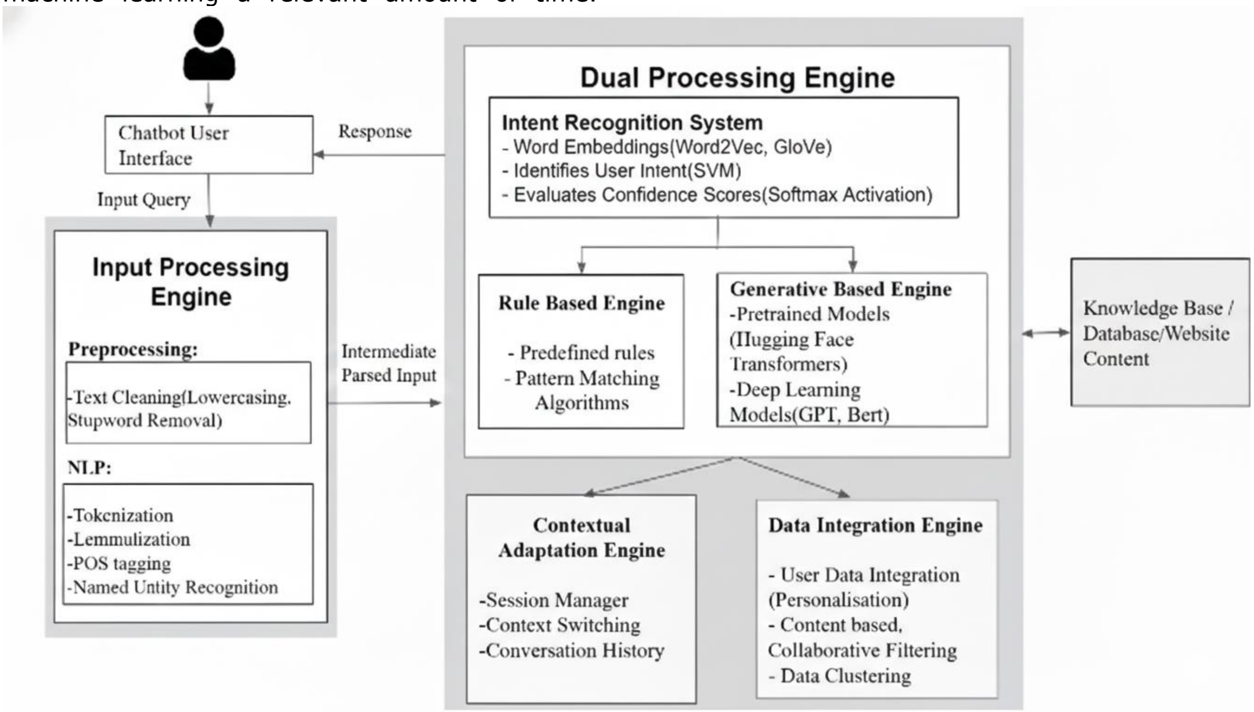

Figure 2: Conceptual Architecture of Hybrid Chatbot

Use Case: If a student expresses interest in courses related to data science, the chatbot recommends additional resources, like online tutorials or relevant academic clubs, based on metadata analysis.

3.4.3. Data Clustering Algorithms

K-means and Hierarchical Clustering are examples of data clustering algorithms that allow us to segment user data into groups according to common characteristics or interests. The groups of the clustered data make it possible for the chatbot to tailor more of the information provided in the form of responses and suggestions. For example, there are students who are at the undergraduate level, and therefore, it would be pertinent to provide information related to the principles level courses. On the other hand, students at a graduate level would probably be provided with information related to the advanced level of research resources or about dissertation writing. Thus, knowledge of user features and requirements is instrumental in data clustering since the chatbot will be able to give the users relevant, concise reports and respond to the needs of the entire academic community (Jain, 2010).

Use Case: Clustering students based on their field of study, such as engineering or humanities, allows the chatbot to tailor responses, like informing engineering students about technical fairs or humanities students about literary events.

3.5. Comparison and Discussion

Hybrid chatbots used in academic settings merge rule-based precision with AI-powered flexibility for design mechanisms, revolutionizing user satisfaction and quality of interaction. With dual processing architecture, the chatbot can elegantly manage predictable queries using rule-based responses and AI based models like transformers and LSTMs for complex context rich queries. Routine queries are responded using pattern matching algorithms, where responses are efficient, consistent and reliable, which enhances user satisfaction. However, AI models give nuanced responses, which feels natural and comprehensive in terms of speed and contextual understanding, this is necessary for academic environments where questions may be straightforward, or highly specific. Figure 2 presents the conceptual architecture of hybrid chatbot, that integrates rule-based and generative techniques for efficient and dynamic interactions. User queries are processed through the Input Processing Engine, where text is cleaned, tokenized, and analyzed using NLP techniques like lemmatization and Named Entity Recognition. The Dual Processing Engine identifies user intent using machine learning models (e.g., Word2Vec, SVM) and directs the query to either the Rule-Based Engine (for predefined rules) or the Generative-Based Engine (using advanced models like GPT or BERT for dynamic responses). Context is maintained via the Contextual Adaptation Engine, which tracks session history and manages conversation flow, while the Data Integration Engine personalizes responses using user data and recommendation techniques. Responses are sourced from a Knowledge Base or generated dynamically and delivered back to the user, ensuring accurate, context-aware, and personalized interactions.

The ability of chatbots to conduct multi-turn conversations is enabled by contextual adaptation mechanisms including DST and models like BERT. Providing the context of academic jargon, chatbots are helpful by giving meaningful step by step guidance in course selection or even academic advising. Using this approach not only increases interaction quality, but it also enables personalized and coherent interaction which is very important for user satisfaction.

SVMs, CNNs and word embeddings are used for intent recognition, so that the chatbot correctly understands the user queries thus ensuring students receive relevant responses.
Personalization is enabled through data integration mechanisms using collaborative as well as content-based filtering. The chatbot suggests relevant courses or research resources to the user by clustering users on matching academic interests. User satisfaction is increased by personalized, context aware responses which recognize the user's current journey through the academic spectrum. But there are still barriers, for example for the person concerned there are data privacy concerns and the risk of biased recommendations. Robust data protection and high-quality personalization continues to be both challenging problems.

3.5.1. Discussion on Switching in Chatbots

Hybrid chatbots use several methods to transition smoothly between rule-based and AI-generated content to provide the best results for user engagement. Among the approaches that are employed, the most basic one is confidence thresholds. For each query that reaches the system, the AI model generates a confidence level for the answer it generates. If this score is high enough, the response is provided by the AI generated system; if not, the chatbot reverts to its rule-based approach to guarantee the correctness of the response. This is especially helpful when working with academic (or other) contexts where the queries can be as basic as asking for library hours, to the deeper questions such as, "Can you provide me with articles for my work on neural networks?".

API integration provides an additional capability that makes it easier for the chatbot to transition between different systems, via an interface. APIs helps a chatbot to get real time data like course timings or event details and the same information can be used in its replies. Although not the only switching mechanism, APIs are crucial for managing when the chatbot requires external knowledge to answer a user's question. However, such mechanisms pose some difficulties which include the minimum latency and the consistent user experience when in use. These issues require constant optimization and evaluation to meet these challenges and enhance the performance of the chatbot as highlighted in studies concerned with real-time implementation of hybrid models.

3.5.2 Evaluation of Personalized Responses in Chatbots

Evaluating the effectiveness of personalized responses in academic chatbots is crucial to understanding their impact on user engagement and satisfaction. Personalization includes making interactions depending upon user's academic history, preferences and behavior using methods such as collaborative filtering and content-based filtering. Collaborative filtering studies user behavior to recommend courses and research papers, or more generally scientific resources, to users that have interests like those of other users.

Content-based filtering, on the other hand, uses the features of academic resources, such as keywords and subject areas, to personalize recommendations. This method ensures a higher degree of personalization by directly aligning suggestions with the user's academic interests. However, effective personalization requires access to high-quality and comprehensive datasets.

Data Privacy and Ethical Concerns are significant challenges in implementing personalized chatbots. Problems of data privacy and data protection are posed by the collection and use of the sensitive academic data. Emerging techniques, such as federated learning, attempt to solve this privacy problem by having data processed locally on user devices without sharing the data, only the model updates, with central servers.

Finally, the evaluation of personalized responses must involve metrics that assess both user satisfaction and the relevance of recommendations. Traditional techniques such as A/B test and User Feedback analysis are embraced to determine the effectiveness of personalization features. Appropriate measures such as monitoring and fine-tuning of algorithms must be exercised to prevent or resolve any biases and inaccuracies.

4. Future Work and Conclusion

Although considerable progress has been made with the hybrid chatbot design of academic sites, there are some areas which are still uncharted and provide a basis for future explorations and advancements. The current state of the art in hybrid chatbots use things like DST and transformer models to keep track of context. Future work could investigate embedding memory augmented neural networks or knowledge graphs to further extend long term context handling thereby building chatbots to provide even more personalized, and context aware responses across long running conversations.

Chatbots should seamlessly integrate across many platforms, including academic portals, mobile applications and voice-based systems, to increase accessibility. To begin with, research could target the definition of universal API frameworks, and lightweight models, which will satisfy various deployment environments. Furthermore, the future of Hybrid Chatbots in academia depends a lot on the continued improvements in machine learning, natural language processing and ethical AI frameworks.

In conclusion, hybrid chatbots present a positive avenue to make academic websites much more interactive, accessible, and user friendly than before. These systems close the gap between rule-based consistency and AI driven flexibility, and the implications of bridging that gap allow our digital experience to become smarter, allowing for smarter, more adaptive, educational technology.

5. REFERENCES

Aho, A. V., & Corasick, M. J. (1975). Efficient string matching. *Communications of the ACM*, *18*(6), 333–340. https://doi.org/10.1145/360825.360855

Cho, K., Bart, V. M., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014, June 3). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. arXiv.org. https://arxiv.org/abs/1406.1078

Chowdhery, A., Narang, S., Devlin, J., & Maarten Bosma. (2023). *PaLM: Scaling Language Modeling with Pathways* [Journal-article]. https://www.jmlr.org/papers/volume24/22-1144/22-1144.pdf

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. https://doi.org/10.18653/v1/n19-1423

De Nieva, J. O., Joaquin, J. A., Tan, C. B., Te, R. K. M., & Ong, E. (2020). Investigating Students' Use of a Mental Health Chatbot to Alleviate Academic Stress. *CHIuXiD '20: 6th International ACM In-Cooperation HCI and UX Conference*, 1–10. https://doi.org/10.1145/3431656.3431657

Gu, A., Dao, T., Machine Learning Department, Carnegie Mellon University, & Department of Computer Science, Princeton University. (n.d.). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv*. https://arxiv.org/pdf/2312.00752

Huang, X. (2021). CHATBOT: DESIGN, ARCHITECUTRE, AND APPLICATIONS. In University of Pennsylvania, School of Engineering and Applied Science, *ASCS CIS 498 Senior Capstone Thesis*. https://www.cis.upenn.edu/wp-content/uploads/2021/10/Xufei-Huang-thesis.pdf

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014, April 8). *A convolutional neural network for modelling sentences*. arXiv.org. https://arxiv.org/abs/1404.2188

Lewis, P. S. H., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. *Neural Information Processing Systems*, *33*, 9459–9474. https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Long Short-Term memory. (1997, November 15). MIT Press Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/6795963

Lops, P., De Gemmis, M., & Semeraro, G. (2010). Content-based recommender Systems: state of the art and trends. In *Springer eBooks* (pp. 73–105). https://doi.org/10.1007/978-0-387-85820-3_3

Meyer, J.G., Urbanowicz, R.J., Martin, P.C.N. *et al.* ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining* **16**, 20 (2023). https://doi.org/10.1186/s13040-023-00339-9

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, January 16). *Efficient estimation of word representations in vector space*. arXiv.org. https://arxiv.org/abs/1301.3781

Pittsley, K. A., & Memmott, S. (2012). Improving independent student navigation of complex educational web sites: An analysis of two navigation design changes in LibGuides. *Information Technology and Libraries*, *31*(3), 52–64. https://doi.org/10.6017/ital.v31i3.1880

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *https://dl.acm.org/doi/proceedings/10.1145/371920*, 285–295. https://doi.org/10.1145/371920.372071

Sonderegger, S., & Seufert, S. (2022). Chatbot-mediated Learning: Conceptual Framework for the Design of Chatbot Use Cases in Education. *CSEDU 2022 - 14th International Conference on Computer Supported Education*, 207–215. https://doi.org/10.5220/0010999200003182

The second dialog State tracking challenge. (2014). In *Proceedings of the SIGDIAL 2014 Conference* (pp. 263–272). Association for Computational Linguistics. https://aclanthology.org/W14-4337.pdf

Tulshan, A. S., & Dhage, S. N. (2019). Survey on virtual Assistant: Google Assistant, Siri, Cortana, Alexa. In *Communications in computer and information science* (pp. 190–201). https://doi.org/10.1007/978-981-13-5758-9_17

Wang, J., Huang, J. X., Tu, X., Wang, J., Huang, A. J., Laskar, M. T. R., & Bhuiyan, A. (2024). Utilizing BERT for information retrieval: survey, applications, resources, and challenges. *ACM Computing Surveys*, *56*(7), 1–33. https://doi.org/10.1145/3648471

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Yildirim-Erbasli, S. N., Bulut, O., Epp, C. D., & Cui, Y. (2023). Conversation-Based Assessments in Education: design, implementation, and cognitive walkthroughs for usability testing. *Journal of Educational Technology Systems*, *52*(1), 27–51. https://doi.org/10.1177/00472395231178943

Zemčík, M. T. (2019). A brief history of chatbots. *DEStech Transactions on Computer Science and Engineering*, *aicae*. https://doi.org/10.12783/dtcse/aicae2019/31439

.

A Framework for Addressing Research Gaps in AI-Driven Classrooms: Toward Equity, Ethics, and Sustainable Innovation

Mark Pisano
pisanom1@southernct.edu
Southern Connecticut State University
New Haven, CT 06515

Abstract

Artificial Intelligence (AI) is reshaping the landscape of educational environments, presenting new opportunities for personalized learning, increased efficiency, and improved equity. However, the swift integration of AI into classroom settings has exposed significant gaps in research and practice. This article synthesizes the literature to identify these persistent gaps and introduces a comprehensive five-pillar framework aimed at effectively addressing them. The framework emphasizes the importance of equity, ethics, teacher agency, and student empowerment in AI implementation within educational contexts. Central to this approach is a commitment to interdisciplinary collaboration that includes educators, policymakers, and community stakeholders, enabling the development and deployment of AI tools that are socially responsible and ethically sound. This article offers detailed recommendations for advancing the responsible integration of AI in education, ensuring that these technologies contribute positively to learning outcomes for all students, particularly those from marginalized backgrounds. By shifting the focus from mere identification of issues to actionable solutions, the proposed framework seeks to foster an environment of continuous improvement in educational practices and outcomes, thus maximizing the potential of AI to support diverse learners effectively.

Keywords: Artificial intelligence, Education, Framework, Ethical Technology, Professional Development

Recommended Citation: Pisano, M.A., (2026). A Framework for Addressing Research Gaps in AI-Driven Classrooms: Toward Equity, Ethics, and Sustainable Innovation. *Journal of Information Systems Applied Research and Analytics* v19(n1) pp 26-32. DOI# https://doi.org/10.62273/YNQL6973

A Framework for Addressing Research Gaps in AI-Driven Classrooms: Toward Equity, Ethics, and Sustainable Innovation

Mark Pisano

1. INTRODUCTION

AI-powered technologies are shifting from experimental projects to integral components of everyday classrooms. They influence instruction, assessment, and administration, raising critical questions about their pedagogical, ethical, and societal ramifications. Teachers and researchers must not only harness AI's potential but also proactively mitigate risks like bias, inequity, and diminished teacher agency. Despite abundant rhetoric about AI closing achievement gaps, evidence reveals ongoing risks, especially regarding bias and the creation of new exclusions (Holmes et al., 2023; Luckin, Holmes, 2022).

Amid these developments, a holistic framework is urgently needed to shift the educational AI research community from gap identification to effective problem-solving. Such a comprehensive approach would not only facilitate the recognition of existing challenges but also provide structured methodologies for addressing them. By transitioning from merely pinpointing what is lacking in current systems to actively generating and implementing concrete solutions, this framework would encourage collaboration, innovation, and measurable progress within the field. This shift is essential to ensure that educational AI research can effectively support learners, educators, and institutions in adapting to the rapidly changing landscape of technology-enhanced education.

2. IDENTIFICATION AND PERSISTENT RESEARCH GAPS

AI now automates assessment, adapts learning, allocates resources, and enhances student creativity. Adaptive programs and smart tutoring systems promise tailored support, especially for struggling students (Zawacki-Richter et al., 2019; Chen et al., 2022). Yet, analyses and policy reviews uncover persistent and interlinked gaps such as diversity and inclusion, professional development, ethical concerns, dynamic and responsive curricula, and AI literacy shown below:

Diversity and inclusion
While AI has broadened educational access and personalization, there is limited evidence that it narrows equity gaps for marginalized or under-resourced groups. Although artificial intelligence technologies are often hailed for their capacity to revolutionize learning by making high-quality resources and adaptive instruction more widely available, these benefits are not always equitably distributed. Most algorithms are not tuned to local cultural and linguistic diversity, risking new forms of exclusion and amplifying existing disparities (Schiff, 2021; Holmes et al., 2023). This means that many AI-driven educational tools can inadvertently reinforce structural inequalities, as they may overlook the specific needs, contexts, or backgrounds of the students they aim to serve. As a result, rather than serving as a great equalizer, AI could unintentionally widen gaps for learners from historically marginalized or underserved communities, unless deliberate efforts are made to address these biases and tailor solutions to diverse populations.

Professional development
The pace of adoption of educational technologies and innovative pedagogical tools far exceeds the current availability of robust, adaptive, and ongoing professional learning opportunities designed specifically for teachers. This growing gap places significant pressure on educators, who must rapidly adapt to new systems without sufficient structured support. Existing competency models, such as UNESCO's AI Competency Framework, while comprehensive in scope, remain significantly underutilized across many educational contexts. These frameworks often do not sufficiently address or reflect the diverse professional development requirements and backgrounds of teachers, making it challenging for educators from various disciplines and experience levels to reap their full benefits (Zhou et al., 2022; Dillenbourg, 2023). As a result, there is a pressing need to develop and disseminate more tailored, flexible, and accessible professional development solutions that can empower teachers to keep pace with technological advancements and the evolving demands of the educational landscape.

Ethical concerns
Persistent questions and unresolved debates continue in areas such as algorithmic bias within AI tools, the privacy and safeguarding of student data, and the establishment of clear, transparent frameworks for the governance of AI systems in education. These ethical considerations demand scrutiny and ongoing oversight to foster trust and accountability among stakeholders (Jobin et al., 2019; Selwyn, 2023).

Despite widespread consensus among policymakers, educators, and researchers on the crucial importance of transparency and explainability in the use of educational technologies, the majority of existing ethical guidelines still fail to evolve into concrete, actionable steps that can be practically implemented in everyday school environments. Furthermore, these guidelines often lack accessible frameworks or models that teachers, students, and families can easily understand and apply in real-world educational contexts, leaving a significant gap between high-level ethical intentions and their effective translation into practice (Jobin et al., 2019; Selwyn, 2023).

Dynamic and responsive curricula
AI-driven tools often focus on test scores and skills mastery over real-time, project-based, or creative learning. Many of these systems are designed to measure student performance through standardized assessments, focusing primarily on whether students can recall and apply specific knowledge or complete isolated skill-based tasks. As a result, these tools may overlook important aspects of learning, such as critical thinking, collaboration, innovation, and problem-solving in novel contexts. Few systems support dynamic curriculum adaptation linked to authentic real-world problems, such as using current events or community-based challenges to guide instructional content and learning strategies. This gap highlights a significant limitation in the ability of AI technologies to foster holistic educational experiences that mirror the complexities and demands of the world outside the classroom (Luckin & Holmes, 2022; Baker & Hawn, 2021).

AI literacy
Most initiatives emphasize technical proficiency while neglecting critical, interpretive, and participatory competencies (AI literacy, ethical reasoning, and student involvement in the design and evaluation of AI tools) (Holmes et al., 2023; Ng et al., 2022). While there is considerable attention given to teaching students how to use and understand the mechanics of artificial intelligence technologies, less focus is directed towards cultivating students' abilities to critically assess the broader social, ethical, and cultural implications of AI. In addition, many of these programs overlook the importance of engaging students as active contributors in both the design and evaluation stages of AI tool development, which is essential for ensuring that emerging technologies are relevant, inclusive, and ethically grounded. Therefore, a more holistic approach would encompass technical skills alongside explicit instruction and opportunities for students.

3. HOLISTIC FRAMEWORK FOR ADDRESSING RESEARCH GAPS

Presented below is a cohesive five-pillar framework aimed at systematically addressing the ongoing and complex gaps in both research and practice within AI-driven classrooms. This framework synthesizes foundational and contemporary insights from the literature and incorporates a range of perspectives to establish a solid base for advancing future research, informing policy development, and guiding practical implementation in educational settings that utilize artificial intelligence.

Primary Pillar	**Research Gap**	**Secondary Pillars**	**Notes**
Pillar I: AI Design	Diversity and Inclusion	Pillar III, Pillar V	Pillar I directly addresses equity, accessibility, and stakeholder engagement in design; Pillar III supports fairness and accountability; Pillar V contributes to AI literacy with bias awareness.
Pillar II: Professional Development	Professional Development	Pillar I, Pillar IV, Pillar V	Pillar II is the core PD driver; Pillar I involves teachers in design; Pillar IV aligns curricula; Pillar V

			embeds AI literacy into PD.
Pillar III: Ethical, Transparent, Accountable AI	Ethical Concerns	Pillar I, Pillar II, Pillar V	Pillar III leads ethics/governance; Pillar I ensures ethics-by-design; Pillar II and Pillar V reinforce ethics through practice and literacy.
Pillar IV: Dynamic and Responsive Curricula	Dynamic and Responsive Curricula	Pillar II, Pillar I, Pillar V	Pillar IV enables adaptable, inquiry-based learning; Pillar II trains teachers; Pillar I engages stakeholders; Pillar V embeds literacy.
Pillar V: AI Literacy	AI Literacy	Pillar II, Pillar IV, Pillar I	Pillar V centers on critical, participatory literacy; Pillar II integrates literacy into PD; Pillar IV weaves literacy into curricula; Pillar I frames literacy within equitable design.

Table 1: Pillar mapping to gaps

Pillar I: AI design
Actively engage a diverse group of stakeholders, including teachers, students, families, and members of the local community, in both the initial design and subsequent evaluation process. By involving these groups at each stage, the initiative can better address the unique needs, perspectives, and cultural nuances present within the local context, ensuring that solutions are not only relevant but also sustainable and widely accepted (Ng et al., 2022). Additionally, establish a routine and recurring process of conducting thorough equity audits. These audits should systematically assess, identify, and address any forms of inequitable impact or unintended negative consequences. By revisiting practices and outcomes regularly, interventions can be adjusted and improved to promote fairness and inclusivity, ensuring that no group is disproportionately disadvantaged (Selwyn, 2023). Proactively incorporate principles of accessibility and inclusiveness from the very start of development. This approach seeks to eliminate barriers to participation and learning by considering the full range of potential users' needs, regardless of ability, language, or background, thereby fostering an environment where learning opportunities are genuinely accessible and meaningful to everyone involved (Rose & Meyer, 2022).

Pillar II: Professional development
Provide flexible professional development that evolves with teachers' needs by offering a range of workshops, mentorship programs, and on-demand resources tailored for various experience levels and subject areas, ensuring continuous growth and adaptability in response to classroom challenges and educational trends (Zhou et al., 2022). Foster professional learning communities focused on shared practice by encouraging the formation of collaborative groups where educators regularly engage in reflective discussions, share resources, observe each other's teaching, and participate in joint problem-solving, thus building a culture of continuous improvement and mutual support among staff members (Luckin & Holmes, 2022). Provide real-time analytics for educators by integrating intelligent data-driven platforms that assess teaching strategies, monitor student progress, and deliver actionable recommendations, enabling teachers to refine their instructional approaches and maximize student outcomes through timely and personalized support (Baker & Hawn, 2021).

Pillar III: Ethical transparent accountable AI
Embed privacy, transparency, and fairness throughout every stage of the AI system lifecycle, from conception and design to deployment, use, and ongoing maintenance. Ensure that ethical principles are not simply a checklist but are consistently integrated into policies, algorithms, and oversight mechanisms, fostering an accountable and ethically grounded

technology culture (Jobin et al., 2019). Also, create review bodies that include the active participation of a wide range of stakeholders such as students, parents, educators, technology specialists, and community advocates. These boards should meet regularly to review AI applications, ensure alignment with institutional values, and adapt governance approaches in response to evolving challenges (Selwyn, 2023). All of this helps to support feedback loops that routinely collect, analyze, and publicly report on equity outcomes and learning metrics associated with AI implementations, transparently showcasing both successes and areas for improvement. Gather input not only from quantitative data but also from qualitative user feedback to refine practices, address disparities, and foster community trust (Schiff, 2021).

Pillar IV: Dynamic and responsive curricula
Implement AI tools that empower students to engage in inquiry-based, open-ended learning experiences. Such tools should encourage exploration, creativity, and critical thinking skills, guiding learners to investigate complex problems and generate innovative solutions (Baker & Hawn, 2021). These innovative educational tools continually support the evolution of learning materials, ensuring content remains relevant, engaging, and closely connected to real-world applications and scenarios (Luckin & Holmes, 2022). While actively leveraging and integrating diverse student interests in the learning process to foster much deeper engagement, motivation, and personalization of educational experiences. This approach not only makes learning more meaningful but also tailors the curriculum to resonate with individual learners (Ng et al., 2022).

Pillar V: AI literacy
Develop not only basic programming skills but also a broad spectrum of critical, ethical, and participatory abilities to analyze and understand the multifaceted societal impacts of artificial intelligence systems and technologies. This includes fostering awareness of biases, accountability, transparency, and the importance of informed decision-making about AI in everyday life (Holmes et al., 2023; Ng et al., 2022). This should actively engage students at every stage of AI-related projects and curricula, enabling them to participate in the design, development, and ongoing critical evaluation of artificial intelligence tools. This process empowers learners to shape AI applications and reflect on the broader implications of their choices and solutions (Luckin & Holmes, 2022).

Also, ensure that robust infrastructure, comprehensive digital resources, and targeted outreach strategies are prioritized to reach underserved and marginalized populations. This commitment facilitates fair opportunities for all learners to access, interact with, and benefit from quality AI education, regardless of their geographical, social, or economic context (Zhou et al., 2022).

4. RECOMMENDATIONS

This framework is a living one; it is designed to spark actionable progress toward equitable, ethical, and effective AI integration. Each of the five components represents a mutually reinforcing area; neglect in one risks undermining the whole structure. Future research should move from describing gaps and frameworks to actively bridging them using participatory, interdisciplinary, and agile research approaches. Longitudinal and mixed-methods studies, as well as rapid feedback mechanisms, are vital for keeping pace with ongoing change (Zawacki-Richter et al., 2019). To help meet the goals of the framework, the following are recommended:
Fund collaborative, inclusive projects that actively and meaningfully engage teachers, students, parents, and marginalized groups throughout every phase of AI development, from ideation to implementation and evaluation. Ensure that diverse voices are incorporated through open calls, collaborative workshops, and sustained partnerships, so that educational AI systems genuinely reflect the needs, aspirations, and cultural contexts of those most affected.

Mandate rigorous, transparent ethical review protocols and detailed public reporting requirements for all stages of educational AI deployment, including design, testing, and real-world use. Require ongoing oversight by independent ethics boards, create clear documentation of risks and benefits, and ensure mechanisms for addressing potential harms, privacy concerns, and algorithmic biases are in place and regularly updated.

Scale up accessible, differentiated, and ongoing professional development programs that integrate AI literacy, pedagogical strategies, and hands-on opportunities for teachers to experiment and give feedback. Prioritize building teacher agency by involving educators in the design of training materials and technology tools, supporting collaborative learning communities, and incentivizing peer-led mentoring around responsible AI use in classrooms.

Actively promote research projects and funding schemes that intentionally combine expertise from education, computer and data science, ethics, policy, psychology, and community organizations. Facilitate cross-sector collaboration through conferences, joint publications, and research networks to address the complex challenges of educational AI and share innovative solutions globally.

Establish robust, transparent mechanisms for public feedback, impact reporting, and continual improvement in all educational AI initiatives. Construct easily accessible platforms for students, families, teachers, and the wider community to report concerns, offer suggestions, and engage in monitoring impacts. Institute regular, independent evaluations, incorporate lessons learned into policy updates, and publicly share progress to ensure ethical, socially responsive AI in education.

Adopt Open-source AI tools and educational materials to significantly reduce barriers: by releasing datasets, models, and benchmarks under permissive licenses, developers and educators can adopt and adapt new technologies without prohibitive costs. Open-source frameworks also spur innovation by promoting standardized, widely vetted platforms that support diverse applications across varied educational contexts. AI can further address workload challenges by automating routine administrative tasks that consume a substantial share of teachers' time. By offloading these tasks, AI tools free valuable time that educators can devote to other areas mentioned within the framework.

5. CONCLUSION

AI's promise in classrooms remains conditional upon robust, equity-centered, and ethical frameworks realized in practice. To fully harness the transformative potential of artificial intelligence in education, it is essential that these frameworks are not merely theoretical constructs but are actively implemented and continuously evaluated for effectiveness and inclusivity. Drawing from global research and innovation, which reflect diverse educational needs and local contexts, the field must build collective capacity for inclusive, sustainable, and human-centered educational transformation. This involves fostering a culture of collaboration among educators, policymakers, researchers, and communities to ensure that AI technologies are developed and deployed to address equity gaps, support personalized learning, and promote student well-being while upholding ethical principles. In doing so, the education sector can move towards a future where AI empowers every learner, regardless of background, to achieve their full potential in a rapidly changing world.

6. REFERENCES

Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. In V. Kumar (Ed.), Proceedings of the International Conference on Artificial Intelligence in Education (pp. 123-138). Springer.

Chen, L., Chen, P., & Lin, Z. (2022). Artificial intelligence in education: A review. IEEE Access, 10, 21680-21695. https://doi.org/10.1109/ACCESS.2022.3141375

Dillenbourg, P. (2023). Orchestrating AI and teachers: Challenges and opportunities for hybrid classrooms. Learning Analytics Review, 1, 27-41.

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education: Promises and implications for teaching and learning. MIT Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Luckin, R., & Holmes, W. (2022). AI for school teachers. Routledge.

Ng, W., Leung, K., & Chiu, S. (2022). Equity and inclusion in AI-driven classrooms. International Journal of Inclusive Education, 26(10), 1019–1033. https://doi.org/10.1080/13603116.2021.1895014

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Rose, D. H., & Meyer, A. (2022). Universal design for learning: Theory and practice. CAST Professional Publishing.

Schiff, D. (2021). The impact of artificial intelligence on the practices of educational leadership. AI & Society, 36, 1119–1129. https://doi.org/10.1007/s00146-020-01090-1

Selwyn, N. (2023). Should robots replace teachers? AI and the future of education. Polity.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0

Zhou, W., Fong, J., & Yang, M. (2022). Teacher competency development for AI-driven classrooms. Computers & Education: Artificial Intelligence, 3, 100070. https://doi.org/10.1016/j.caeai.2022.10007

Does Dark Humor Work? An Inspection of Social Media-Based Marketing Strategies

Matthew Bagatta
mbagatta3@walsh.edu
Walsh University
North Canton, OH 44720

Philip Kim
pkim@walsh.edu
Walsh University
North Canton, OH 44720

Richard Metzer
rlmst26@mail.rmu.edu
Robert Morris University
Pittsburgh, PA 15108

Abstract

This study explores the effectiveness of dark humor in social media marketing by analyzing consumer sentiment and engagement across YouTube, TikTok, and Facebook. Advertisements from Mint Mobile, Doritos, and Dr. Squatch were examined to compare audience reactions to dark humor versus traditional humor. Sentiment analysis and engagement rate calculations reveal that dark humor generates higher engagement on visually driven platforms like YouTube and TikTok, while traditional humor is better received on Facebook. These findings underscore the importance of aligning humor styles with platform characteristics and audience expectations to enhance marketing impact.

Keywords: Dark humor, Social media marketing, Sentiment analysis, Engagement metrics

Recommended Citation: Bagatta, M., Kim, P., Metzer, R., (2026). Does Dark Humor Work? An Inspection of Social Media-Based Marketing Strategies. *Journal of Information Systems Applied Research and Analytics*, v19(n1) pp 33-45. DOI# https://doi.org/10.62273/XOCU1723

Does Dark Humor Work? An Inspection of Social Media-Based Marketing Strategies

Matthew Bagatta, Philip Kim and Richard Metzer

1. INTRODUCTION

Social media has fundamentally reshaped the digital marketing landscape, enabling businesses to communicate directly with consumers through tailored, interactive content. With billions of users across platforms such as TikTok, Facebook, and YouTube, organizations are increasingly shifting from traditional advertising channels to social media to build brand awareness and foster engagement (Atske, 2023; Statista, 2024). Unlike traditional advertising, social media content encourages active audience interaction—likes, shares, comments, and reposts—which in turn provides rich, real-time feedback for marketers. Humor has long played a central role in marketing by appealing to emotions, increasing message retention, and creating a sense of brand relatability (Cline et al., 2003; Isaza, 2022). However, a more provocative form—dark humor—has emerged as a notable tactic, particularly among younger demographics. Defined by its use of irony, cynicism, and taboo topics, dark humor reflects a cultural shift in how consumers engage with content that addresses uncomfortable or complex issues through levity (Splitters, 2023). Despite its prevalence in entertainment and viral media, dark humor remains understudied in the context of digital marketing—and virtually absent from information systems (IS) literature.

From an IS perspective, there is a growing need to understand how sentiment and user engagement data, derived from humorous content, can inform platform-specific marketing strategies. While sentiment analysis and natural language processing (NLP) techniques have been adopted to study customer satisfaction and brand perception (Barney, 2023; Poeczea et al., 2018), little attention has been paid to how dark humor affects these outcomes across diverse digital platforms.

This study examines how dark humor in advertising shapes user sentiment and engagement across YouTube, TikTok, and Facebook. By comparing advertisements from Mint Mobile, Doritos, and Dr. Squatch, the research helps address a gap in the IS literature around content analytics and offers insight into how humor functions strategically within social media environments. The findings provide IS professionals, digital marketers, and analytics researchers with a clearer understanding of consumer response patterns and content strategy optimization.

2. LITERATURE REVIEW

The role of humor in marketing has been well-documented across disciplines such as communications, psychology, and business, where it is frequently linked to higher message recall, increased likability, and stronger brand loyalty (Hasanova, 2019; Barry & Graça, 2018). Humor captures attention and fosters emotional engagement—two key variables in influencing consumer behavior. Existing scholarship often generalizes humor, overlooking the nuanced differences between light, satirical, or dark comedic styles.

Humor in Advertising

Advertising scholars have noted that humor can facilitate persuasion by reducing arguments, increasing ad memorability, and enhancing perceived credibility (Cline et al., 2003; Eisend, 2018). The Elaboration Likelihood Model (ELM) provides a theoretical framework to explain this dynamic. Humor, when processed peripherally, can enhance attitudes toward the advertisement or brand without requiring deep cognitive processing. Humor may increase involvement when aligned with message relevance (*Humorous Ads*, 2023; Petty & Cacioppo, 1986). Humor styles influence engagement differently depending on context (*Affiliative and Aggressive Humor*, 2024; Muntinga, Moorman, & Smit, 2011). Humor is versatile and can be impactful—but its effectiveness hinges on context, audience, and tone.

Light or affiliative humor is the most common style in advertising due to its broad appeal and lower risk of offense (Martin et al., 2003). Humor in social media marketing can boost ad performance significantly (*"Investigate the Impacts,"* 2023). Satirical humor, which critiques social norms or institutions, has gained traction in purpose-driven marketing. Dark humor is defined by its use of morbid, taboo, or ironic content for comedic effect. It has remained on the margins of advertising practice due to perceived risks of alienating audiences or damaging brand reputation (Napp, 2023; Splitters, 2023).

Dark Humor’s Cultural Function
Recent research suggests that dark humor resonates deeply with certain demographics, particularly younger audiences such as Generation Z. This cohort, raised amid global crises, political polarization, and hyper-digital culture, often uses dark humor as a mechanism for self-expression, identity formation, and psychological coping (Jacob, 2023). Studies in media psychology highlight that humor helps individuals regulate emotions and build resilience, especially during high-stress events such as the COVID-19 pandemic (McGraw & Warren, 2010). For example, Jordanian Facebook users during the pandemic adopted dark humor to satirize government policies and make light of existential anxieties (*Jordanian Social Media*, n.d.).

In this sense, dark humor may not merely be a provocative tool, but a culturally adaptive one. It offers catharsis, solidarity, and meaning in environments perceived as unstable or absurd. These insights challenge the assumption that dark humor is inherently alienating and suggest that it may foster authenticity and relatability—two increasingly valued brand traits in the digital age (Isaza, 2022).

Platform-Specific Norms
The platform itself is a crucial variable in determining how humor is received. Research shows that content norms vary significantly across social media sites (Thomas et al., 2019; Voorveld, 2019). TikTok, with its short-form, trend-driven culture, favors fast-paced, visual humor, often with absurd or unexpected elements. YouTube supports longer-form storytelling, allowing for more nuanced comedic arcs. Facebook, increasingly skewed toward older users, tends to reward more conventional, narrative-based humor styles (Taecharungroj & Nueangjamnong, 2015). These platform characteristics shape not only what content gets shared, but how users interpret and engage with humor.

As Barry and Graça (2018) argue, user engagement with humor varies significantly across video-based platforms, depending on audience expectations and content format. Yet, little research has explored how dark humor performs across these digital ecosystems, especially in terms of interaction stats like comments, shares, and sentiment polarity.

Gaps in IS Literature
While humor in advertising has been studied in communication and psychology, it remains underexplored in the field of information systems. Within IS, research on digital marketing tends to focus on technical optimization, performance metrics, and user behavior, rather than content tone or emotional strategies. Sentiment analyses have typically been applied to customer reviews, satisfaction studies, or general brand perception (Barney, 2023). These analyses often rely on polarity detection (positive, neutral, negative), without accounting for emotional nuance, sarcasm, or humor subtypes.

This creates a methodological blind spot. Humor, and especially dark humor, is deeply context-dependent and difficult for automated systems to interpret correctly. For example, sarcasm detection is still an emerging area in natural language processing (NLP), and even state-of-the-art models struggle with culturally coded humor or irony (Liu et al., 2019). So, applying sentiment analysis tools to humorous content—without accounting for these limitations—may yield misleading results.

Srivastava’s (2015) comparative study of humorous versus emotional advertising provides an exception within IS-related scholarship, showing that humor can outperform emotional appeals in generating favorable consumer attitudes. However, the study did not differentiate between humor styles or assess variation across digital platforms. Poeczea et al. (2018) examined influencer marketing performance using sentiment and interaction stats but did not consider tone as a variable in their analysis. This leaves open questions about how humor affects interaction dynamics, particularly in comment sections, where brand perception is actively negotiated by users.

3. METHODOLOGY

To evaluate the effectiveness of dark humor in social media advertising, this study adopted a mixed-methods research design integrating both qualitative and quantitative analysis. This approach was selected to provide a more holistic view of audience responses, combining sentiment insights derived from natural language processing (NLP) tools with platform interaction stats. The research approach was grounded in prior literature on humor theory, digital marketing strategy, and sentiment analysis practices within the information systems (IS) domain (Barney, 2023; Barry & Graça, 2018; Hasanova, 2019).

The research began with a comprehensive literature review to identify foundational theories related to humor in marketing and its potential effects on consumer engagement. Emphasis was placed on the emotional dimensions of advertising, the emergence of dark humor in generational subcultures, and the expanding use of sentiment analysis tools in brand perception studies (Thomas & Fowler 2021; Warren, Brannon, & Kopel, 2019). This foundational stage informed the design of the research questions and guided the selection of analytical methods, helping to align the study with both marketing communication theory and IS evaluation tools.

Three major social media platforms—YouTube, TikTok, and Facebook—were chosen as the sites of analysis. These platforms were selected for their diverse user bases, global reach, and unique content structures. YouTube supports long-form video content and often attracts engaged audiences seeking entertainment or information. TikTok, known for short-form, viral-oriented content, emphasizes trend-based participation and rapid engagement. Facebook, in contrast, offers a more traditional social media experience l humor relied more on relatability, celebrity appeal, or mild absurdity. The comparative structure ensured that the analysis remained balanced, avoiding bias toward a specific brand or content style.

Audience sentiment was analyzed through 100 top-level user comments per advertisement per platform, for a total of 1,800 comments across the six ads and three platforms. This sampling strategy aligns with recommendations from Google Cloud and previous sentiment research, which suggest that a minimum of 100 documents per category offers a reliable baseline for natural language analysis (Google Cloud, 2024; Rashid & Huang, 2018). To reduce sampling bias, comments were selected based on chronological order of appearance and filtered to ensure they were unique, relevant to the advertisement, and free from automated bot language.

The sentiment analysis was conducted using the Text2Data plugin for Google Sheets, a cloud-based NLP tool capable of scoring textual data based on polarity and emotional tone. Each comment was assigned a sentiment score on a continuous scale from -1 (strongly negative) to +1 (strongly positive), with values near zero representing neutral sentiment. The tool also categorized each comment as positive, neutral, or with mixed media formats and an older demographic skew (Sheikh, 2025). The use of these platforms enabled cross-comparative insights into how dark humor performs in varied digital environments.

The sample included advertisements posted between 2016 and 2024, a time frame representing the exponential rise of social media marketing and growing cultural acceptance of edgier humor styles. Ads were selected from three consumer brands—Mint Mobile, Doritos, and Dr. Squatch—based on their consistent use of humor in digital campaigns and their cultural relevance among online audiences. These brands operate in different industries (telecommunications, food/snacks, and personal hygiene), allowing for diverse content styles and humor executions.

Two advertisements per brand were selected, resulting in a total of six ads under study. Each brand contributed one advertisement categorized as using "dark humor" and another considered more traditional or light-hearted in tone. The categorization was based on definitions established in prior humor literature, with dark humor involving elements of morbidity, irony, taboo subjects, or socially transgressive content(Splitters, 2023; Napp, 2023). Traditiona negative based on its internal thresholds. By analyzing a large and diverse comment set, we aimed to average out anomalies and obtain a reliable sentiment profile for each ad. The system uses NLP algorithms that identify positive or negative sentiment based on keywords, phrase structure, and contextual cues.

One of the challenges of using such tools is a limited capacity to interpret tone or sarcasm. For example, a comment like "Wow, what a great ad " with an eye roll emoji may be interpreted as positive due to the words "wow" and "great," despite being clearly sarcastic. This underscores the need for sentiment classifiers that are trained on humorous and context-rich datasets—an area where IS research and machine learning innovation could help.

Descriptive statistics were applied to the sentiment data to calculate mean, median, and confidence intervals for each advertisement's sentiment distribution. These values helped quantify the overall emotional tone of audience reactions and allowed for direct comparison between humor styles and platforms.

In addition to sentiment analysis, behavioral engagement was assessed using an engagement rate ratio. This metric was calculated by dividing the sum of public interactions (likes, shares, comments) by the total number of followers or subscribers on the brand's account, then multiplying by 100 to yield a percentage. This rate offers a normalized measure of how effectively each advertisement stimulated audience participation, adjusting for differences in brand size and reach. Engagement data were extracted from public platform statistics and verified through manual sampling.

This methodology allows for an exploration of how humor tone influences sentiment and engagement across platform environments. It also provides a replicable framework for future IS research examining emotional content strategies, platform-user dynamics, or NLP applications in digital marketing contexts.

4. RESULTS

YouTube	Results1
Mean	0.15195
Standard Error	0.0525582
Median	0.1595
Mode	1
Standard Deviation	0.5255819
Sample Variance	0.2762363
Kurtosis	- 1.1749234
Skewness	- 0.2307057
Range	1.684
Minimum	-0.684
Maximum	1
Sum	15.195
Count	100
Confidence Level(95%)	0.1042868

Table 1: Mint Mobile Sentiment Analysis results for Dark Humor Ad on YouTube

Mint Mobile

The sentiment analysis for Mint Mobile revealed platform-specific variations in audience reception to humor styles. On YouTube, the dark humor advertisement yielded a mean sentiment score of 0.152, indicating a largely neutral reception. The non-dark humor ad performed slightly better at 0.167, suggesting a marginally more favorable response. This indicates that YouTube users appreciated both ads but leaned slightly toward the lighter tone. The modest gap may reflect YouTube's tolerance for offbeat or edgy humor, while still rewarding clearer messaging or nostalgic appeal.

YouTube	Results1
Mean	0.16716
Standard Error	0.0498242
Median	0.174
Mode	0.489
Standard Deviation	0.4982416
Sample Variance	0.2482447
Kurtosis	-1.007735
Skewness	-0.441853
Range	1.726
Minimum	-0.726
Maximum	1
Sum	16.716
Count	100
Confidence Level(95%)	0.0988619

Table 2: Mint Mobile Sentiment Analysis results for Normal Ad on YouTube

On TikTok, the dark humor ad also remained within the neutral range at 0.142, while the non-dark humor ad scored higher at 0.186. This difference suggests that TikTok users responded more positively to lighter humor styles, potentially due to the platform's emphasis on short, accessible entertainment and high visual tempo. TikTok's algorithm tends to favor content that is emotionally clear and quickly digestible, which may explain why dark humor—often requiring a buildup or contextual irony—receives a more subdued reaction despite being stylistically bold.

Facebook presented a notable reversal. The dark humor ad received a mean sentiment score of 0.118, while the non-dark humor ad scored just 0.017, indicating a weaker reception. While both scores remained in the neutral range, the significantly lower rating for the non-dark humor ad may suggest that Facebook users either found the content dull or failed to connect with the ad's style.

TikTok	Results2
Mean	0.14176
Standard Error	0.0487939
Median	0.184
Mode	0.7
Standard Deviation	0.4879386
Sample Variance	0.2380841
Kurtosis	- 0.8402292
Skewness	- 0.4536318
Range	1.873
Minimum	-0.873
Maximum	1
Sum	14.176
Count	100
Confidence Level(95%)	0.0968176

Table 3: Mint Mobile Sentiment Analysis results for Dark Humor Ad on TikTok

TikTok	Results2
Mean	0.18592
Standard Error	0.0488108
Median	0.207
Mode	0.249
Standard Deviation	0.4881076
Sample Variance	0.238249
Kurtosis	-0.740133
Skewness	-0.467313
Range	1.748
Minimum	-0.748
Maximum	1
Sum	18.592
Count	100
Confidence Level(95%)	0.0968511

Table 4: Mint Mobile Sentiment Analysis results for normal ad on TikTok

Facebook	Results3
Mean	0.1176139
Standard Error	0.0496397
Median	0.168
Mode	0.205
Standard Deviation	0.4988725
Sample Variance	0.2488738
Kurtosis	- 1.0050834
Skewness	- 0.2933773
Range	1.996
Minimum	-0.996
Maximum	1
Sum	11.879
Count	100
Confidence Level(95%)	0.0984837

Table 5: Mint Mobile Sentiment Analysis results for Dark Humor Ad on Facebook

Facebook	Results3
Mean	0.0171089
Standard Error	0.0508468
Median	0.128
Mode	-0.134
Standard Deviation	0.511004
Sample Variance	0.2611251
Kurtosis	-1.361985
Skewness	0.0636105
Range	1.742
Minimum	-0.748
Maximum	0.994
Sum	1.728
Count	100
Confidence Level(95%)	0.1008786

Table 6: Mint Mobile Sentiment Analysis results for normal ad on Facebook

The level of interaction supports these findings. The dark humor ad achieved engagement rates of 17% on YouTube and 9% on TikTok, while Facebook lagged significantly at 0.4%. In contrast, non-dark humor ads received lower engagement on all platforms, including 10% on

YouTube, 4% on TikTok, and 0.5% on Facebook. These results suggest that, while sentiment remained mostly neutral, dark humor drove stronger interaction on visually rich platforms, particularly when the content was attention-grabbing and well-aligned with brand identity.

Doritos

Doritos advertisements demonstrated stronger overall engagement and more pronounced differences in sentiment compared to Mint Mobile. On YouTube, the dark humor ad achieved a positive sentiment score of 0.301, the highest in the study. This strongly suggests that viewers appreciated the ad's bold narrative and unconventional punchline. The non-dark humor ad scored 0.168, falling within the neutral range. This contrast suggests that YouTube viewers were more responsive to the bolder comedic approach, which aligns with Doritos' history of producing provocative and memorable ads for high-visibility events like the Super Bowl.

YouTube	Results1
Mean	0.30098
Standard Error	0.046321
Median	0.3455
Mode	1
Standard Deviation	0.463214
Sample Variance	0.214567
Kurtosis	- 0.446541
Skewness	- 0.665477
Range	1.638
Minimum	-0.638
Maximum	1
Sum	30.098
Count	100
Confidence Level(95%)	0.091912

Table 7: Doritos Sentiment Analysis results for Dark Humor Ad on YouTube

YouTube	Results1
Mean	0.1682222
Standard Error	0.0520793
Median	0.171
Mode	1
Standard Deviation	0.5181824
Sample Variance	0.268513
Kurtosis	- 1.0985742
Skewness	-0.312324
Range	1.727
Minimum	-0.727
Maximum	1
Sum	16.654
Count	100
Confidence Level(95%)	0.1033497

Table 8: Doritos Sentiment Analysis results for normal ad on YouTube

TikTok	Results2
Mean	0.054686
Standard Error	0.042187
Median	0.182
Mode	0.247
Standard Deviation	0.426066
Sample Variance	0.181532
Kurtosis	- 0.771776
Skewness	- 0.578713
Range	1.659
Minimum	-0.873
Maximum	0.786
Sum	5.578
Count	100
Confidence Level(95%)	0.083687

Table 9: Doritos Sentiment Analysis results for Dark Humor Ad on TikTok

Conversely, on TikTok, the dark humor ad scored only 0.055—still neutral but significantly lower than the non-dark humor ad's 0.226. This reversal reinforces the idea that TikTok users, who favor lighthearted, quick humor and trend-based formats, may be less inclined to engage with content requiring setup or context. The structured nature of the dark humor ad, which involved a storyline and a punchline about premature birth, may have been too jarring or narratively complex for TikTok's browsing environment.

TikTok	Results2
Mean	0.2264653
Standard Error	0.0495067
Median	0.247
Mode	-0.582
Standard Deviation	0.497536
Sample Variance	0.2475421
Kurtosis	- 0.9557613
Skewness	- 0.5737663
Range	1.638
Minimum	-0.638
Maximum	1
Sum	22.873
Count	100
Confidence Level(95%)	0.0982198

Table 10: Doritos Sentiment Analysis results for normal ad on TikTok

Facebook	Results3
Mean	0.272216
Standard Error	0.050481
Median	0.513
Mode	1
Standard Deviation	0.509836
Sample Variance	0.259933
Kurtosis	- 0.821251
Skewness	- 0.646819
Range	1.676
Minimum	-0.676
Maximum	1
Sum	27.766
Count	100
Confidence Level(95.0%)	0.100141

Table 11: Doritos Sentiment Analysis results for Dark Humor Ad on Facebook

Facebook sentiment scores once again favored dark humor, with a score of 0.272 compared to the non-dark humor ad's 0.045. This result underscores a surprising trend: although Facebook users are generally older, they may still respond positively to dark humor if the narrative is clear and the brand is trusted. Doritos, as a longstanding and culturally embedded brand, may benefit from nostalgic recognition that softens the perceived risk of dark content.

Facebook	Results3
Mean	0.0447549
Standard Error	0.0400215
Median	0.197
Mode	0.197
Standard Deviation	0.404197
SampleVariance	0.1633752
Kurtosis	-0.717591
Skewness	- 0.5127395
Range	1.497
Minimum	-0.748
Maximum	0.749
Sum	4.565
Count	100
Confidence Level(95%)	0.0793918

Table 12: Doritos Sentiment Analysis results for normal ad on Facebook

Interestingly, this sentiment did not translate into behavioral engagement. While the dark humor ad achieved high engagement on YouTube (28%) and TikTok (44%), Facebook recorded a minimal rate of 0.04%, despite relatively positive sentiment. This points to a growing pattern across platforms: Facebook users may feel positively toward dark humor content but are less inclined to interact publicly with it. This divergence between sentiment and engagement could be explained by social visibility norms—users may not want to be seen endorsing content that others might interpret as offensive or inappropriate.

These results suggest that while sentiment and engagement can align—as on YouTube and TikTok—they do not always. Dark humor may trigger appreciation without prompting action, particularly on platforms where user behavior is more passive or where content is consumed silently without social interaction.

Dr. Squatch advertisements displayed the most nuanced and arguably most intriguing results. On YouTube, the dark humor ad scored -0.059, indicating a slightly negative response. In contrast, the non-dark humor ad scored 0.143, suggesting that viewers preferred the lighter, product-focused tone. Given that the dark humor ad relied heavily on visual shock and profanity, the slight dip into negative sentiment could indicate viewer discomfort or fatigue with exaggerated masculine branding tropes.

YouTube	Results1
Mean	-0.05947
Standard Error	0.0494906
Median	0.1075
Mode	-0.467
Standard Deviation	0.4949058
Sample Variance	0.2449318
Kurtosis	- 1.3409075
Skewness	0.2521005
Range	1.707
Minimum	-0.707
Maximum	1
Sum	-5.947
Count	100
Confidence Level(95%)	0.0982001

Table 13: Dr. Squatch Sentiment Analysis results for Dark Humor Ad on YouTube

TikTok sentiment scores for both ads were nearly identical—0.100 for dark humor and 0.097 for non-dark—implying neutrality regardless of humor style. These results suggest that TikTok users may be less influenced by humor and more by the entertainment value or relatability of the content itself. Dr. Squatch's aesthetic branding and novelty may have driven interaction more than the humor itself, especially on TikTok, where product-feature videos and "oddly satisfying" content often outperform narrative ads.

YOUTUBE	Results1
Mean	0.1434753
Standard Error	0.0439817
Median	0.158
Mode	0.653
Standard Deviation	0.442011
Sample Variance	0.1953738
Kurtosis	- 0.9440037
Skewness	- 0.4230946
Range	1.632
Minimum	-0.638
Maximum	0.994
Sum	14.491
Count	100
Confidence Level(95%)	0.0872585

Table 14: Dr. Squatch Sentiment Analysis results for normal Ad on YouTube

TikTok	Results2
Mean	0.09992
Standard Error	0.0414423
Median	0.197
Mode	0.247
Standard Deviation	0.4144226
Sample Variance	0.1717461
Kurtosis	- 0.8005902
Skewness	- 0.6102802
Range	1.322
Minimum	-0.638
Maximum	0.684
Sum	9.992
Count	100
Confidence Level(95%)	0.0822304

Table 15: Dr. Squatch Sentiment Analysis results for Dark Humor Ad on TikTok

TIKTOK	Results2
Mean	0.0973824
Standard Error	0.0402182
Median	0.1715
Mode	0.159
Standard Deviation	0.4061839
Sample Variance	0.1649854
Kurtosis	- 0.3164259
Skewness	- 0.6224305
Range	1.742
Minimum	-0.996
Maximum	0.746
Sum	9.933
Count	100
Confidence Level(95%)	0.0797821

Table 16: Dr. Squatch Sentiment Analysis results for normal Ad on TikTok

Facebook	Results3
Mean	- 0.0198039
Standard Error	0.0501431
Median	0.097
Mode	-0.638
Standard Deviation	0.5064208
Sample Variance	0.256462
Kurtosis	-1.341169
Skewness	0.1863887
Range	1.696
Minimum	-0.696
Maximum	1
Sum	-2.02
Count	100
Confidence Level(95%)	0.0994705

Table 17: Dr. Squatch Sentiment Analysis results for Dark Humor Ad on Facebook

This trend was echoed on Facebook, where sentiment scores were -0.020 (dark humor) and 0.026 (non-dark). Again, while both scores hovered near neutral, the preference for the milder content implies that Facebook audiences may be more interested in clear value propositions or aesthetics than comedic edge. This is particularly relevant for a product category like hygiene, where trust and perceived product quality are central to decision-making.

Facebook	Results3
Mean	0.0260294
Standard Error	0.0465995
Median	0.1465
Mode	-0.638
Standard Deviation	0.4706316
Sample Variance	0.2214941
Kurtosis	- 1.1821423
Skewness	- 0.1397316
Range	1.75
Minimum	-0.75
Maximum	1
Sum	2.655
Count	100
Confidence Level(95%)	0.0924408

Table 18: Dr. Squatch Sentiment Analysis results for normal Ad on Facebook

Engagement data, however, told a different story. The dark humor ad achieved a remarkably high engagement rate of 66% on YouTube and 30% on Facebook. TikTok, by contrast, showed only 0.8%. The engagement rates for the non-dark humor ads were significantly lower across all platforms: YouTube (11%), Facebook (0.5%), and TikTok (0.6%). These numbers illustrate an important divergence: even when sentiment is neutral or slightly negative, dark humor can still drive high engagement.

For Dr. Squatch, the strategy proved especially effective on platforms where users expect bold, irreverent content, like YouTube. The brand's identity, built on exaggerated masculinity, ruggedness, and anti-mainstream tone, aligns well with the type of humor that shocks or entertains, even if it divides opinion. This suggests that dark humor may function as an amplifier for brands seeking to disrupt established norms or differentiate from conventional messaging.

Overall, these findings highlight how brand identity, platform culture, and humor tone interact in complex ways. Engagement and sentiment do not always move together—and in the case of Dr. Squatch, they appear to diverge in ways that still benefit brand visibility. This reflects a broader theme in digital marketing: polarizing content may risk alienating some users, but it can also galvanize loyalists and spark conversation. In a crowded digital environment, that trade-off may be worthwhile.

Engagement Rate Formula
Engagement Rate (%) = (Likes + Shares + Comments) / Total Followers × 100

Score Range	Interpretation
+0.20 to +1.00	Positive Sentiment
-0.20 to +0.19	Neutral Sentiment
-1.00 to -0.21	Negative Sentiment

Table 19 – Sentiment Score Categories

Brand	YouTube (%)	TikTok (%)	Facebook (%)
Mint Mobile	17	9	0.4
Doritos	28	44	0.04
Dr. Squatch	66	0.8	30

Table 20 – Engagement Summary by Platform (Dark Humor Ads)

5. CONCLUSION

This study explores the role and effectiveness of dark humor in social media advertising by analyzing sentiment and engagement across three major platforms: TikTok, YouTube, and Facebook. Using a comparative framework involving three brands—Mint Mobile, Doritos, and Dr. Squatch—the research examined how different humor tones performed across varying digital environments and user demographics. We found that while audience sentiment towards dark humor ads often remains neutral or mildly positive, these ads consistently yield higher engagement levels than their non-dark counterparts, especially on visually driven platforms such as YouTube and TikTok.

In particular, brands with strong, irreverent identities—such as Doritos and Dr. Squatch—were able to leverage dark humor effectively to generate attention and interaction. These brands saw significantly higher engagement rates for their edgier content, even in cases where sentiment scores were flat or slightly negative. Mint Mobile presented more variable outcomes, illustrating that platform norms and audience preferences can mediate the effectiveness of a humor strategy. Facebook emerged as an outlier, with some of the highest sentiment scores for dark humor but the lowest levels of engagement. This suggests that appreciation does not always translate into action, possibly due to the older, more passive user base of the platform or changing norms around public expression on Facebook.

These findings underscore the importance of contextual alignment in digital advertising. Humor—particularly dark humor—is not a one-size-fits-all strategy. Brands should consider not only their own voice and identity but also the platform-specific culture and audience expectations. For marketers and information systems professionals alike, this implies a need to integrate content tone analytics into campaign planning, performance tracking, and platform selection. Dark humor holds considerable promise as a high-engagement marketing tactic. However, its success depends on strategic execution— grounded in cultural awareness, platform fluency, and a nuanced understanding of audience dynamics.

Opportunities for IS Research and Integration
The omission of humor typologies in IS research represents a missed opportunity. Integrating emotional tone analysis into platform-specific user analytics could offer deeper insights for marketers, brand strategists, and systems designers. For example, dashboards that distinguish between affiliative, aggressive, and dark humor could help tailor ad delivery in real time. Additionally, incorporating these distinctions into training datasets could improve the accuracy of AI-driven content moderation or sentiment interpretation systems.

Additionally, user-generated reactions to humorous content—especially those that go viral—could serve as high-value data points for understanding digital engagement. Analyzing how humor triggers conversation threads, resharing behavior, or meme generation can inform theories of online influence and virality, which are core concerns in the IS domain.
In sum, we believe that humor is not just entertainment; it is data. And dark humor—often dismissed as risky or fringe—may, in the right contexts, offer powerful advantages in brand differentiation and digital engagement. As IS continues to evolve toward more human-centered and emotionally intelligent systems, the need to understand, measure, and interpret humor is essential.

6. REFERENCES

Atske, S. (2023, October 24). *Social media use in 2021*. Pew Research Center. https://www.pewresearch.org

Barry, J. M., & Graça, S. S. (2018). Humor effectiveness in social video engagement. *Journal of Marketing Theory and Practice*, *26*(1–2), 158–180. https://doi.org/10.1080/10696679.2017.1389247

Barney, N. (2023, December 21). *Sentiment analysis (opinion mining)*. TechTarget. https://www.techtarget.com

Bradley, L. (2021). *The marketing power of Ryan Reynolds: Case study on Mint Mobile*. *Forbes*. https://www.forbes.com

Chan, A., & Lowe, B. (2024). How to cook a meme: Exploring content strategies in brand and user compliance. *Cogent Social Sciences*, *10*(1), 2366001. https://doi.org/10.1080/23311975.2024.2366001

Cline, T. W., Altsech, M. B., & Kellaris, J. J. (2003). When does humor enhance or inhibit ad responses? *Journal of Advertising*, *32*(3),31–45. https://doi.org/10.1080/00913367.2003.10639142

Eisend, M. (2018). Explaining the use and effects of humorin advertising:An evolutionary perspective. *International Journal of Advertising*, *37*(1), 3–24. https://doi.org/10.1080/02650487.2017.1349037

Isaza, A. (2022). Being funny pays off: Let’s bring humor back to advertising. *Forbes*. https://www.forbes.com

Jacob, R. (2023). Unveiling the dark humour and self-image of Generation Z in a polymedia context. *ResearchGate*. https://www.researchgate.net

Muntinga, D. G., Moorman, M., & Smit, E. G. (2011). Introducing COBRAs: Exploring motivations for brand-related social media use. *International Journal of Advertising*, *30*(1), 13–46. https://doi.org/10.2501/IJA-30-1-013-046Oikarinen, E.-L. (2022). The boundaries of a small company’s human voice: Insights into dark humour in Internet recruitment advertising. *Corporate Reputation Review*, *26*(3),223–241. https://doi.org/10.1057/s41299-022- 00138-5

Poeczea, I., Ebster, C., & Strauss, C. (2018). Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. *Journal of Retailing and Consumer Services*, *45*,149–157. https://doi.org/10.1016/j.jretconser.2018.08.014

Saavedra Torres, J. L., Bhattarai, A., Dang, A., & Rawal, M. (2023). Do you want to be roasted? The boundaries of using dark humor as a brand-to-brand communication strategy. *Journal of Research in Interactive Marketing*. https://doi.org/10.1108/JRIM-12-2022-0370

Sheikh, M. (2025, January 6). Social media demographics to inform your 2025 strategy. SproutSocial. https://sproutsocial.com/insights/new- social-media-demographics/

Splitters, S. (2023, March 13). Why we use dark humor: The psychology of comedy. *Sidesplitters*. https://www.sidesplitters.com

Srivastava, R. (2015). A comparative study of humour versus emotional advertisements on consumer behavior. *Asian Journal of Marketing*, *10*(1–3), 8–21.

Statista. (2024, April 29). *Global social networks ranked by number of users 2024*. https://www.statista.com

Taecharungroj, V., & Nueangjamnong, P. (2015). Humour 2.0: Styles and types of humour and virality of memes on Facebook. Journal of Creative Communications, 10(3), 288-302.

Thomas, R., & Fowler, H. (2021). Manipulative intent and consumer response to brand humor on social media. *Journal of Interactive Marketing*, *55*,75–88. https://doi.org/10.1016/j.intmar.2020.12.002

Voorveld, H.A. M. (2019). Brand communication in social media:An integrated framework. *International Journal of Advertising*, *38*(3),477–487. https://doi.org/10.1080/02650487.2019.1575101

Warren, J., Brannon, L., & Kopel, M. (2019). Developing brand coolness through disruptive social media interactions. *Journal of Consumer Behaviour*, *18*(3), 221–234. https://doi.org/10.1002/cb.1773

Designing An Artificial Intelligence Adoption Assistance Platform for Small and Medium-sized Enterprises: Identifying Opportunities and Concerns

Brian Alvarez
bga5780@uncw.edu
University of North Carolina Wilmington
Wilmington, NC 28403-5611

Lukas Rudolf Germut Fitz
fitz@b-tu.de
Brandenburg University of Technology
Cottbus-Senftenberg
03046 Cottbus, Germany

Judith Gebauer
gebauerj@uncw.edu

Yao Shi
shiy@uncw.edu

University of North Carolina Wilmington
Wilmington, NC 28403-5611

Abstract

Small and medium-sized enterprises (SMEs) often face significant barriers when adopting artificial intelligence (AI), including concerns related to cost, ease of use, and accessibility. This project introduces a web-based prototype designed to overcome these concerns by providing a secure, low-cost, and user-friendly platform for deploying AI tools such as sales forecasting and customizable chatbots. Following design science research methodology, the platform's initial features, layout, and included tools were informed by potential users, which identified the most desired AI functionalities and common adoption concerns among SMEs. To evaluate the AI adoption assistance platform, a demonstration and testing iteration was conducted with owners and employees of SMEs and university students in North Carolina, which led to insights on several design meta-requirements. The SME participants expressed strong interest in using the platform regularly and emphasized the importance of data privacy, intuitive design, and flexible pricing; the student evaluators provided additional insight into navigation and interface clarity, identifying areas for refinement such as clearer user feedback, improved discoverability of tools, and consistent formatting. Future work includes expanding the chatbot functionality, refining the onboarding experience, and improving design consistency, followed by gathering additional user feedback.

Keywords: Artificial Intelligence, Adoption, SME, Design Science, Chatbot, Sales Forecasting

Recommended Citation: Alvarez, B., Fitz, L.R., Gebauer, J., Shi, Y., (2026). Designing An Artificial Intelligence Adoption Assistance Platform for Small and Medium-sized Enterprises: Identifying Opportunities and Concerns. *Journal of Information Systems Applied Research and Analytics*, v19(n1) pp 46-63. DOI# https://doi.org/10.62273/ZQPS7795

Designing An Artificial Intelligence Adoption Assistance Platform for Small and Medium-sized Enterprises: Identifying Opportunities and Concerns

Brian Alvarez, Lukas Rudolf Germut Fitz, Judith Gebauer and Yao Shi

1. INTRODUCTION

Artificial Intelligence (AI) has swiftly impacted the business landscape, presenting an unprecedented opportunity to transform operations, improve decision-making, and drive efficiency in all sectors. While AI tools assist large corporations to gain an edge, numerous small and medium-sized enterprises (SMEs) encounter various obstacles when implementing these solutions. These barriers often prevent such smaller actors from accessing the full benefits of AI and thus limit their ability to compete in an increasingly data-driven marketplace.

To address these barriers, this study follows a two-fold approach: (1) Through a narrative literature review and surveys, we investigate the challenges of AI adoption by SMEs and identify key obstacles such as lack of technical knowledge or financial resources, concerns about AI bias, complexity in setup, privacy, fairness, and trust issues, particularly in relation to data access and management, and the need to comply with regulatory requirements, such as the European Union’s General Data Protection Regulation (GDPR) (Bhalerao et al., 2022; Jackson et al., 2022; OECD, 2021; Steven & Blake, 2024; Soudi & Bauters, 2024); (2) We conceptualize and instantiate a prototypical design artifact to address the challenges. Guided by the principles of Design Science Research (DSR), we thereby approach AI adoption as a complex problem that requires iterative development, grounded in both practical needs and established theoretical knowledge. The artifact, an AI adoption assistance platform, is conceptualized, prototypically built, and evaluated through a feedback loop that bridges the practical environment of SMEs and the prevalent knowledge base around AI solutions and best practices (Hevner et al., 2004).

In line with elicited user requirements and with the intent to provide a broad foundation for feedback and further evaluation, our prototype consists of two SME-oriented use cases for AI adoption assistance: (1) an AI-driven predictive sales forecasting tool to help businesses anticipate demand and manage inventory more effectively (Crenshaw, 2024); (2) an AI-driven tool to create custom AI chatbots that may be embedded in SME websites and enable automated customer support.

Altogether, this design-based and application-oriented study responds to the need for an SME-tailored AI adoption assistance solution. We combine research on the barriers to AI adoption with the development of a user-friendly AI platform to empower smaller businesses to harness the benefits of AI technology and to enhance our understanding of SME needs around AI adoption from a research perspective. The goal is to create a solution that not only supports the integration of AI but also builds trust and confidence in AI among SME stakeholders, which might help pave the way for wider AI adoption in the SME community.

2. BACKGROUND

The adoption of AI technologies by SMEs is an evolving area of research, driven by the potential benefits and challenges these businesses face. Understanding these factors is crucial for developing effective solutions tailored to the needs and constraints of smaller organizations.

Importance of AI in SMEs

Integrating AI into SMEs provides substantial benefits, promoting business growth, enhancing decision-making, and improving operational efficiency across various regions. According to the U.S. Chamber of Commerce (2023), SMEs leveraging advanced technology report higher growth rates in their sales, employment, and profits, with 87% of small business owners noting improved operational efficiency. Additionally, businesses using AI have a 12-point higher likelihood of profit growth than those without, with positive effects particularly evident in marketing and communications (U.S. Chamber of Commerce, 2023). Bhalerao et al. (2022) stress the role of AI in improving SMEs’ core functions, such as decision-making and customer engagement. AI allows SMEs to analyze data more effectively, aiding in customer behavior insights, inventory management, risk mitigation, and cybersecurity protections. It also streamlines human resource functions, simplifying talent acquisition, employee development, and performance management (Bhalerao, 2022). In summary, AI adoption in SMEs drives growth and efficiency while offering smaller businesses the tools to remain competitive in a globalized economy, maximizing the benefits of advanced technology.

Current Use of AI in SMEs

Businesses in North Carolina (this study's focal region) and across the U.S. that currently use or plan to use AI in the next six months demonstrate varying levels of adoption. 28% of businesses in North Carolina and 30% of U.S. businesses plan to use AI for data analytics. In terms of specific AI applications, 25% of North Carolina businesses use AI for text analytics, compared to 23% nationally, while 20% of North Carolina businesses utilize machine learning, versus 22% at the national level (Smith, 2024a). However, only 5.1% of North Carolina and 5% of U.S. enterprises currently employ AI in their operations. This data reveals a significant divide between current and planned usage, with data analytics being a major area of interest for AI adoption in these businesses (Smith, 2024b).

Additionally, 28% of both North Carolina and U.S. national businesses are interested in adopting virtual chatbots for customer interactions (Smith, 2024a). Marketing automation will be particularly impactful in sectors, such as real estate, retail, accommodation, food services, construction, education, and agriculture (Smith, 2024b). Artificial intelligence in customer engagement extends beyond marketing, with tools like chatbots and virtual assistants playing an increasing role in enhancing operational efficiency and customer satisfaction (Crenshaw, 2024).

In addition to enhancing customer service, AI automation can help SMEs improve productivity, reduce costs, and enhance workplace safety and conditions. By automating routine functions, businesses can minimize human errors, reduce administrative bottlenecks, and operate more efficiently, even offering 24/7 customer interaction at a lower cost (OECD, 2023). AI-driven solutions are particularly impactful in streamlining marketing, customer relationship management, and sales processes, as noted by Watney and Auer (2021), who highlight how third-party AI systems can transform operational workflows for SMEs.

Notably, based on a thorough literature review and empirical framework, Schwaeke et al. (2025) conclude that when "AI aligns with the specific needs and innovation objectives of SMEs, the likelihood of their enthusiastic adoption increases" (p. 1319).

Barriers to AI Adoption in SMEs

Regulatory burdens add complexity, particularly for SMEs. Firms in Europe, for instance, must comply with multiple data protection laws such as the General Data Protection Regulation (GDPR) and the Digital Services Act. Watney and Auer (2021) found that European businesses often struggle with legal uncertainty, citing liability for potential damages (33%), data standardization requirements (33%), and regulatory obstacles (29%) as major external challenges to AI adoption. For SMEs, maintaining compliance with overlapping and complex regulations can be a significant barrier to innovation, often reducing investment in new technologies due to the perceived risks.

Cost and resource limitations represent significant barriers for SMEs in adopting AI technologies, as these businesses often lack the financial capacity to invest in and maintain complex AI systems. The cost-related concerns are further intensified by the high perceived risk associated with adopting advanced AI solutions. Smaller firms are often cautious about experimenting with cutting-edge technology that could lead to financial losses if implementation does not deliver expected efficiencies or strategic benefits, according to Schwaeke et al. (2025). Compared to larger corporations, SMEs may not have the resilience to absorb short-term setbacks from costly technological initiatives, making them more reluctant to adopt innovations that might be unprofitable in the short run (Bhalerao et al., 2022).

AI solutions are often perceived as too complex, with user interfaces and operational requirements that are unintuitive for those without specialized knowledge (Bettoni et al., 2021). A study of Danish SMEs showed that a "shortage of workforce skilled in AI" was rated a significant obstacle, with respondents scoring this barrier an average of 9 out of 10 (Iftikhar & Nordbjerg, 2021). The intricate demands of AI, combined with a critical lack of expertise within SMEs, create a challenging landscape for AI adoption. SMEs face not only a shortage of skilled personnel but also a need for accessible, simplified AI systems to bridge the complexity gap (Bhalerao, 2022).

With SMEs using AI-driven tools for their customers, they may inadvertently promote biased recommendations or pricing practices, potentially eroding customer trust and damaging brand reputation (Steven & Blake, 2024). A key contributor to this issue is the lack of explainability in AI systems, especially those powered by machine learning. Unlike traditional software, which operates according to clearly defined rules, many AI models make decisions based on complex statistical correlations that are difficult to interpret (Hitchings, 2024). Unlike larger firms that may have teams to interpret, monitor, and adjust AI systems, SMEs typically lack the resources to oversee these models (Roa Baez & Igbekele, 2021). This gap creates an additional vulnerability for SMEs, as they may unknowingly deploy biased or opaque AI systems that inadvertently harm their business operations (Bettoni et al., 2021).

3. METHODOLOGY

Design Science for Applied Information Systems Research

Design science research (DSR) is a foundational paradigm in Information Systems research that emphasizes the creation of innovative artifacts aimed at solving real-world problems. Unlike behavioral science, which seeks to understand and predict phenomena, design science is inherently constructive. It advances knowledge through the conceptualization and/or instantiation and evaluation of design artifacts such as models, methods, constructs, and IT systems (Hevner et al., 2004). DSR thus enables research through tangible prototypes that address identified problems in between practical environment and knowledge base. Through iterative design activities, DSR may explore not only what works, but why and how it works in a specific context (Johannesson & Perjons, 2014).

DSR operates under a set of guiding principles that stress relevance, rigor, and utility. The process begins with identifying a relevant problem, followed by designing an effective artifact that addresses it, and then rigorously evaluating the artifact's performance (Hevner et al., 2004). This paradigm values both theoretical grounding and practical applicability: each artifact should be informed by existing knowledge and simultaneously offer contributions that are actionable. The cyclic interaction between design and behavioral research fosters both utility (here: in the form of functioning systems) and truth (here: in the form of both actionable and usable knowledge), forming a research approach well-suited for addressing complex socio-technical challenges within information systems (Hevner et al., 2004).

Requirements Collection

From a preliminary survey, we gathered insights into the current uses of AI within SMEs, identified potentially sought-for new applications, and understood the concerns and barriers faced by these businesses. We administered the preliminary survey with a structured questionnaire via Qualtrics (Appendix A), targeting SMEs owners, managers, and employees in North Carolina, USA. The questionnaire focuses on three primary areas: the current application areas of AI, the perceived benefits of AI tools, and the challenges and concerns that SMEs encounter in adopting AI.

Artifact Conceptualization

The conceptual design phase establishes a foundational blueprint for the AI adoption assistance platform, presenting a mock website that showcases the proposed features and functionality. In alignment with DSR guidelines, this phase serves as the design and development stage, where theoretical knowledge is transformed into a practical blueprint through iterative modeling and visualization. Therefore, we crafted user stories to capture the needs and actions of various user types, such as SME owners and general staff. These user stories guide design decisions by highlighting specific user goals and tasks, ensuring the platform functionality aligns with users' real-world needs and workflows (Möller et al., 2020).

Artifact Instantiation

The instantiating development phase brings the design to life by building the AI adoption assistance platform using Python, Django, and SQLite, with a strong focus on security, usability, and performance. The development process embodies the building activity within DSR, where the conceptual design is transformed into a functional platform prototype that addresses the identified SMEs needs of integrating robust security measures, user-friendly functionalities, and scalable design to create a secure environment for SMEs using the platform.

LLM and Machine Learning Integration. The AI and machine learning functionalities leverage Python libraries such as Pandas and NumPy to develop custom AI models tailored to meet SMEs needs. For the sales forecasting function, the system utilizes the Catboost, Statsmodels, scikit-learn, and Pmdarima libraries, which supports widely used forecasting models, including RandomForestRegressor, Exponential Smoothing, AutoArima, and RNNs/LSTMs (Zhao et al., 2024). These models analyze historical sales data to identify seasonal trends, recurring patterns, and future sales projections.

Additionally, a custom AI chatbot creator is integrated to assist businesses in automating customer support and operations. The chatbot is deployed using the Hugging Face platform with serverless architecture, ensuring scalability and ease of maintenance. For optimal performance, the Hugging Face Spaces chatbot, powered by the ChatGPT-3 API, is utilized, offering a fast, high-performing large language model (LLM) with a generous free-tier and compliance with EU privacy regulations (Zhao et al., 2024). Furthermore, the system generates insights into frequent customer concerns and enables intelligent support routing, escalating unresolved issues to human agents when necessary. To ensure data privacy and build user trust, the chatbot is trained with strict protocols to prevent the disclosure of personal identifiable information (PII), maintaining full compliance with data protection standards.

Web Application Security Enhancements. Hosted on an AWS EC2 Instance (Iaas), the application benefits from built-in Distributed Denial of Service (DDoS) protection and a Web Application Firewall (WAF), helping mitigate common threats and unauthorized access attempts. All data transmitted between users and the platform is encrypted using TLS/SSL, enabled through Cloudflare, which enforces HTTPS connections to safeguard information in transit (Vallabhaneni et al., 2024).

On the backend, Django's security framework provides strong safeguards by default. Sensitive information, such as passwords, is automatically hashed, ensuring that even in the event of a data breach, exposed data remains protected. The platform uses SQLite as its database solution, which is lightweight and well-suited for development and low- to moderate-traffic production environments. The access to sensitive data is tightly controlled through Django's adherence to the principles of least privilege and zero-trust security, ensuring that users and processes only have access to the resources necessary for their role. These security measures aim to ease concerns that SMEs may have when adopting AI applications, ensuring data privacy, compliance, and operational safety.

API Management. Effective API management is essential for integrating AI functionalities across the platform. Instead of FastAPI, the platform utilizes Gradio and Gradio_Client to manage interactions between the user interface and AI models, particularly Hugging Face Spaces. Gradio provides secure, rate-limited connections by default, simplifying deployment while ensuring that APIs remain protected from abuse and unauthorized access. (Vallabhaneni et al., 2024).

User Authentication and Authorization. The platform implements secure user authentication and role-based access control using Django's built-in authentication system. Django provides robust session management, password hashing, and protection against common web threats such as cross-site request forgery (CSRF) using CSRF tokens.

While multi-factor authentication (MFA) is not currently implemented, Django's authentication framework supports user login, registration, and permissions management, allowing for the definition of distinct user roles such as admin, SME employee, and guest. Each role is granted access only to the features and data relevant to their level of authorization, ensuring secure and appropriate access throughout the application.

Backend Optimization. To ensure a responsive and efficient user experience, the platform offloads heavy processing tasks—such as chatbot interactions—to external infrastructure. By deploying the chatbot on Hugging Face Spaces, the computational load is shifted away from the core server and handled by Hugging Face's high-performance infrastructure. This approach reduces strain on the backend, improves response times, and ensures the platform remains lightweight and scalable without requiring complex task queues or caching systems (Abdali et al., 2024).

Monitoring and Logging. The platform leverages AWS EC2's built-in logging and monitoring tools to track performance and maintain system health. AWS provides real-time activity logs, error reporting, and performance metrics, allowing for efficient issue detection and resolution (Casola et al., 2024). This streamlined approach simplifies infrastructure management while ensuring that the platform remains reliable and responsive.

User Experience Enhancements. To support usability and adoption among SMEs, the platform prioritizes clear, step-by-step instructional guidance designed with an "explain like I'm five" (ELI5) mindset. Comprehensive documentation will be created to walk users through each feature, helping those with little to no technical background navigate and effectively use the platform.
To build transparency and foster user trust, the documentation will include plain-language descriptions of the AI models powering the platform. These explanations will highlight each model's strengths and limitations, allowing users to better understand how outputs are generated and make informed decisions based on the results. The aim is to bridge the technical skills gap by offering simple, approachable resources, including video demonstrations, that show exactly how to interact with the platform's tools. This user-centered approach is designed to reduce uncertainty, increase confidence, resources, and empower SMEs to adopt AI solutions with clarity and purpose.

Deployment and Scaling Strategy. The platform utilizes a Continuous Integration/Continuous Deployment (CI/CD) pipeline through GitHub, automating testing and deployment to ensure smooth and efficient updates. The platform was developed on a secure web server hosted by an EC2 instance, due to its free tier, as one of the main goals of the project is to deliver a low-cost solution. AWS EC2 also offers auto-scaling capabilities to handle traffic surges and intensive computation tasks, ensuring reliable and scalable performance as user demand grows.

Demonstration and Testing

The final design phase of this study involves testing to evaluate the platform's effectiveness, user-friendliness, and overall suitability for SMEs with the goal of obtaining potential design meta-requirements. The testing was conducted with a sample of participants based in North Carolina, including both practitioners and non-technical students. After using the platform, participants will complete a questionnaire on Qualtrics to gather qualitative and quantitative feedback on their experience, including their trust in its security and whether they would consider implementing it in their business (Appendix C). We conducted interviews to gather additional information and feedback.

4. RESULTS OF REQUIREMENTS COLLECTION

The preliminary survey results collected from twelve owners or employees of SMEs (Appendix A) not only provide insights into AI adoption and interest among SMEs of various sizes (7 micro-businesses, 2 small and 3 medium-sized) and industries but also reinforce the key barriers, particularly cost, lack of expertise, and data security concerns, which continue to hinder AI adoption among SMEs. Despite the small sample size, the ex-ante outcome informs our solution design with input from the practical SME environment bordering our targeted design space.

Although none of the preliminary surveyed SMEs currently use AI solutions, a majority expressed interest in a no-code/low-code AI platform to support their business functions. Among them, especially micro-businesses (1-10 employees) see machine learning sales forecasting as the most valuable AI application, with additional interest in automated customer support/chatbots and voice-to-text automation. SMEs with 11-50 employees prioritized document summarization, data dashboard automation, and fraud detection tools, while SMEs with 51-500 employees favored support chatbots and data dashboard automation. These findings highlight a strong potential demand for tailored AI tools for SMEs.

The preliminary survey also reveals several barriers and concerns regarding AI adoption. In fact, 36.4% of respondents expressed concern about AI model bias and lack of in-house expertise, while complexity of setup (54.5% of respondents) and cost of implementation (72.7% of respondents) are also commonly considered obstacles (**Figure 2**). Micro-business respondents, on the other hand, all mentioned the cost of implementation as a main barrier. Data privacy and security stand out as universally critical concerns, with all respondents marking data privacy as essential to any potential AI adoption.

5. ARTIFACT DEVELOPMENT

Landing Page

The landing page of the website (**Figure 3**) establishes the first point of contact for users and reflects the core values of accessibility, trust, and functionality. A clean and intuitive navigation bar is placed just below the header, featuring links to Home, Testimonials, Services, FAQs, About, Contact, Login, and Register.

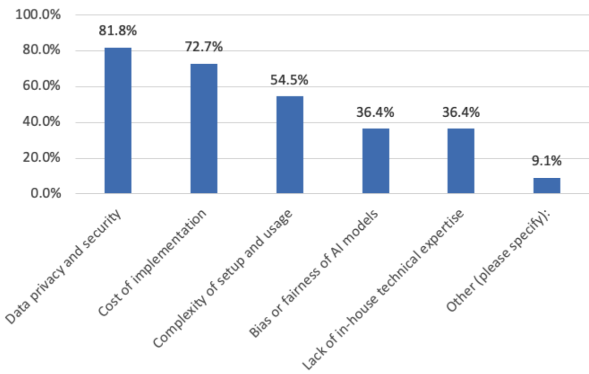

Figure 2: Distribution of main concerns about AI adoption among preliminary survey respondents

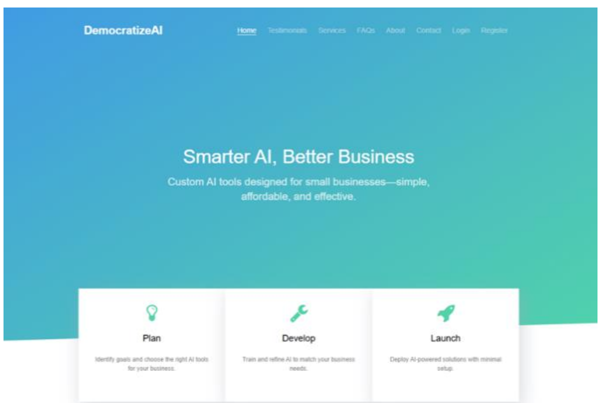

Figure 3 Landing Page

User Authentication

The Register and Login functionalities are essential components of the platform's secure access model and are presented as distinct pages accessible through the navigation bar (Appendix B). The Register form collects basic information such as first name, last name, phone number, email address, username, and password, while the Login form allows existing users to securely access their accounts with only their username and password. Both forms were built using responsive front-end design, with server-side validation and hashing for password storage to ensure secure data storage from the start. Once authenticated, users are redirected to a private, user-specific interface where they can begin utilizing the features of the platform. This gated structure not only supports secure access but also lays groundwork for future features.

Custom AI Chatbot Creator

This feature allows users to quickly configure and deploy a customer-facing chatbot capable of responding to frequently asked questions, providing service updates, or directing inquiries, all through a simple, form-based interface. Users are not required to write or understand any code; instead, they follow a guided setup process that walks them step-by-step through customizing their chatbot's responses and deployment settings.

After the initial setups, users are directed to the Manage Chatbot page, a centralized hub for controlling all aspects of their to-be-created chatbot (**Figure 4**). This page includes clearly written instructions on how to use and update the tool, ensuring that even first-time users can navigate the process confidently. Users can upload or modify an individual reference document that powers the chatbot's knowledge base, test the chatbot's behavior in real time, and delete it at any time. Allowing users to upload documents rather than filling out a structured form offers greater flexibility and mirrors existing business workflows, reducing friction for SMEs who may already have relevant materials prepared in their own formats.

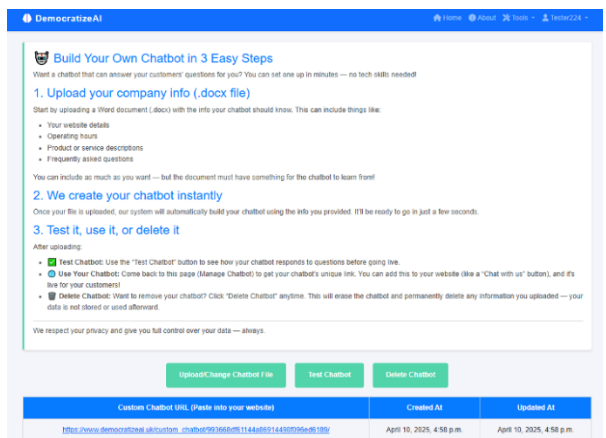

Figure 4 Manage Chatbot Page

After uploading their file, the Manage Chatbot page provides users with a unique, embeddable link that can be easily integrated into their own business website. This link enables their customers to interact directly with the AI chatbot, offering a seamless extension of their customer service capabilities without requiring complex installation or technical setup. When accessed through the embeddable link, the Chatbot Page (**Figure 5**) provides a clean, user-friendly interface where end users can interact with the AI in real time, receiving accurate, context-aware responses based on the SME's uploaded documentation.

In addition, the delete functionality reinforces the platform's commitment to user privacy and aligns with the principle of data ownership and the right to be forgotten. When a user deletes their chatbot, all associated data, including uploaded documents and stored interactions, is permanently removed from the system.

Sales Forecasting

The Sales Forecasting tool enables SMEs to upload sales data in a spreadsheet format, CSV or XLSX, and receive 30-days forward-looking predictions based on trends identified by models trained with custom SME data (**Figure 6**). The tool presents results in easy-to-understand visualizations, helping business owners identify sales patterns and potentially use the results to make informed decisions around staffing, inventory, or marketing strategy.

Users are guided through a clear, step-by-step process that outlines how to format their data and interpret the results. The platform includes brief explanations to ensure users can get started even if they have no prior experience with forecasting or data science. Once data is processed from the uploaded file, the tool displays a graph and chart comparing historical sales to projected future values: Last 30 Days and Future 30 Days (**Figure 6**). Users can hover over data points to view exact figures and trends to find key takeaways, such as predicted high-volume months or declining trends. This output is tailored for non-technical users and avoids jargon, providing real insights without requiring background knowledge in analytics or AI.

All sales data is processed securely at runtime and is never stored on the server or saved after the forecast is generated. This design ensures full data privacy by default, aligning with the principle that users retain complete control over their information without needing to manually delete it.

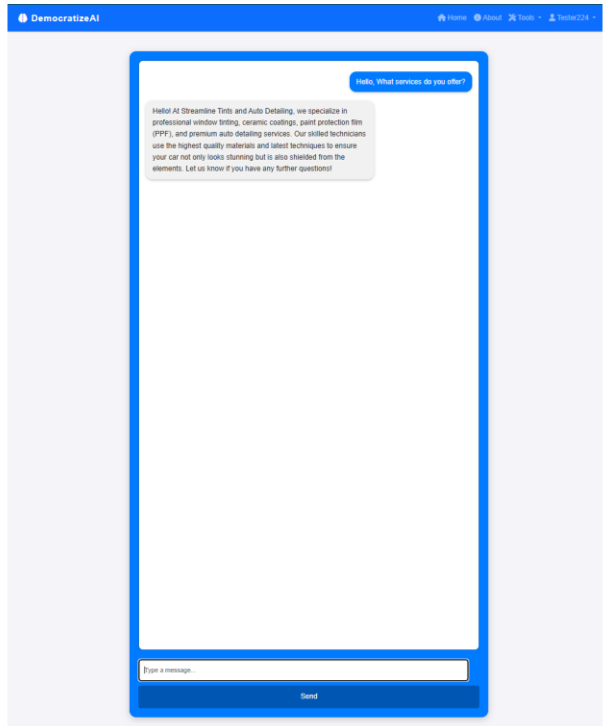

Figure 5 Chatbot Page

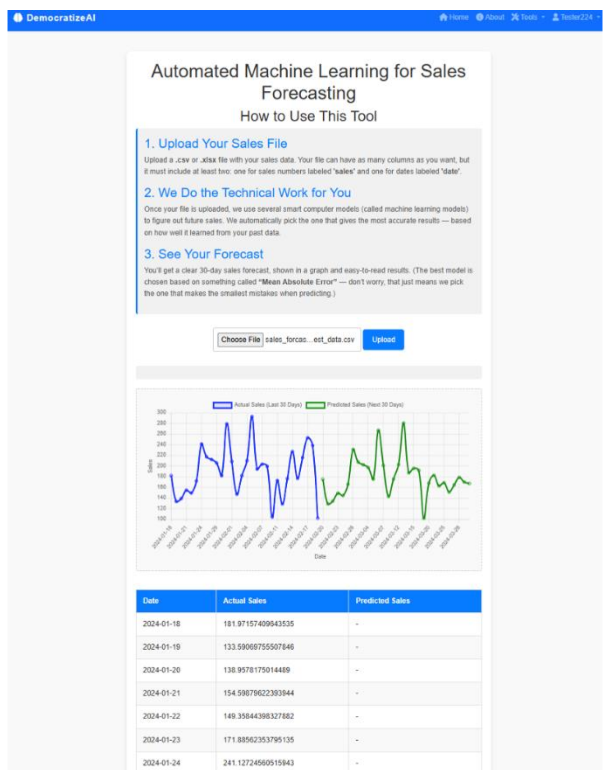

Figure 6 Sales Forecasting Page

Manage User

The Manage User page allows authenticated users to update basic account information, specifically their first name, last name, or email address (**Figure 11** in Appendix B). All changes are made through a simple form-based interface, designed for ease of use and clarity. Users also have the option to delete their account entirely, which triggers the permanent removal of all associated data across the platform. This functionality upholds the "right to be forgotten" and reinforces the platform's commitment to user data privacy and control.

6. RESULTS OF DEMO AND TESTING

To evaluate the effectiveness, clarity, and user-friendliness of the web-based AI platform, usability testing was conducted with owners or employees of SMEs and non-technical university students. Participants were asked to complete common tasks based on a predefined scenario description within the platform and provide feedback on their experience, including areas of confusion, interface design, and overall satisfaction (Appendix C). A total of 28 valid responses were obtained and merged in a way that preserves proportional contributions, with greater analytical emphasis on SME insights to align with the research objective of addressing AI adoption challenges specific to that context.

A recurring theme across both groups was the need for improved discoverability of key tools, particularly when users were not logged in. This issue was most evident in the Data Security and Privacy FAQs, which were only visible to unauthenticated users, a design oversight that will be corrected by making the FAQ content accessible regardless of authentication status. This change supports both transparency and ease of navigation, especially for new users evaluating the platform.

A key priority emerging from the SME participants' feedback is the expansion of AI tool functionality, particularly the development of a conversational chatbot tailored for internal business queries. While the current chatbot supports basic customization, future iterations will introduce a more dynamic interface that allows users to build and refine their chatbot directly within the platform, removing the need for external document uploads. In addition, options for updating chatbot content in response to evolving business needs, such as product offerings or internal FAQs, were requested.

Participants from both groups highlighted the importance of clearer tool guidance, especially around the AI Sales Forecasting feature. Several users either missed the output entirely or were unsure where to view the results after uploading a dataset. To address this, the interface will be redesigned to include an automated success message, clearer upload confirmation, ability to export results, and auto-scroll functionality that brings users directly to the forecast results. Additional tooltip guidance and a revised CSV upload prompt will ensure users understand file format requirements upfront.

Feedback also emphasized the need for onboarding support, particularly for less technical users. As such, guided video tutorials will be developed and embedded into the site to demonstrate tool usage, dataset preparation, and chatbot creation. These tutorials will be optional but easily accessible, aligning with DSR's goal of maximizing artifact adoption without imposing unnecessary friction on more confident users.

7. FUTURE WORK

The artifact testing revealed a range of valuable insights that guide the future development of AI adoption assistance solutions for SMEs.

From an infrastructure standpoint, the AI adoption assistance platform is currently hosted using a free-tier EC2 architecture on AWS. While this setup has enabled rapid prototyping, long-term hosting solutions will need to be evaluated. AWS can present unpredictability in terms of scaling and cost, especially after the one-year free tier expires. Future considerations include continued use of cloud services or transitioning to a self-hosted local server to maintain affordability and control as user traffic increases.

Potential business model development opportunities remain an ongoing consideration. Feedback around pricing model preferences was notably diverse, with participants expressing interest in freemium, ad-supported, and subscription-based models. A tiered pricing structure is being explored to accommodate a wide range of user preferences and financial capacities, though the final monetization strategy will depend on further user research. Ensuring accessibility for smaller businesses will remain a guiding principle, while premium tiers may offer added value for advanced users. In this light, future work might explore viable and desirable ways of deploying an AI adoption assistance platform for SMEs across various sectors.

Advanced features under consideration for premium services include greater flexibility in chatbot customization, enhanced forecasting options, and the introduction of a business intelligence dashboard that automates data visualization based on uploaded sales data. This would allow business owners to “talk to their business data,” bridging both the forecasting and conversational AI features into a cohesive, user-guided analytics experience.

8. CONCLUSION

This article has reported on a project that targets the design of an AI adoption assistance solution for SMEs. Following a DSR project cycle framework, relevant aspects and requirements from the application domain were gathered in a stakeholder-centric manner, alongside an unstructured review of related literature, to inform a first design iteration implementation. Moreover, this article has argued and described the conceptualization and technical development of an early artifact instantiation: a platform showcasing custom AI chatbot creation and AI-assisted sales forecasting. Tentative insights from demos and tests with experience-based inquiries were collected from both SMEs and non-technical student users. Towards design knowledge obtained so far, key insights on meta-requirements address the need to strengthen AI adoption assistance capabilities by enhancing the chatbot’s functionality for internal business use, enabling in-platform customization without external uploads, and ensuring users can easily access and interpret AI-generated outputs through clearer guidance, improved feedback mechanisms, and contextual tooltips.

9. ACKNOWLEDGEMENTS

This work was supported by a fellowship of the German Academic Exchange Service (DAAD).

10. REFERENCES

Abdali, S., Anarfi, R., Barberan, C. J., & He, J. (2024). Securing Large Language Models: Threats, Vulnerabilities and Responsible Practices. arXiv Preprint. https://doi.org/10.48550/arXiv.2403.12503

Bhalerao, K., Kumar, A., Kumar, A., & Pujari, P. (2022). A Study of Barriers and Benefits of Artificial Intelligence Adoption in Small and Medium Enterprises. *Academy of Marketing Studies Journal*, *26*(1), 1–6.

Bettoni, A., Matteri, D., Montini, E., Gładysz, B., & Carpanzano, E. (2021). An AI Adoption Model for SMEs: A Conceptual Framework. *IFAC-PapersOnLine, 54*(1), 702-708. https://doi.org/10.1016/j.ifacol.2021.08.08

Casola, V., De Benedictis, A., Mazzocca, C., & Orbinato, V. (2024). Secure Software Development and Testing: A Model-Based Methodology. *Computers & Security*, *137*, 103639. https://doi.org/10.1016/j.cose.2023.10363

Crenshaw, J. (2024). Enhancing Entrepreneurship: AI’s Big Impact on Small Business: Exploring Real-World Success Stories of Small Businesses Benefiting from Artificial Intelligence. U.S. Chamber of Commerce. Retrieved May 2, 2025 from https://www.uschamber.com/technology/enhancing-entrepreneurship-ais-big-impact-on-small-business

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75-105. https://doi.org/10.2307/25148625

Hitchings, B., Rouse, D., & Brown-Graham, A. (2024). AI Uses in North Carolina. UNC School of Government, ncImpact Initiative. Retrieved May 2, 2025 from https://ncimpact.sog.unc.edu/2024/04/ai-uses-in-north-carolina

Iftikhar, N., & Nordbjerg, F. E. (2021). Adopting Artificial Intelligence in Danish SMEs: Barriers to Become a Data Driven Company, Its Solutions and Benefits. *Proceedings of the 2nd International Conference on Innovative Intelligent Industrial Production and Logistics*, 1, 131-136. http://doi.org/10.5220/0010691800003062

Jackson, D., Allen, C., Michelson, G., & Munir, R. (2022). Strategies for Managing Barriers and Challenges to Adopting New Technologies. CPA Australia.

Johannesson, P., & Perjons, E. (2014). An Introduction to Design Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-10632-8

Möller, F., Guggenberger, T.M., & Otto, B. (2020). Towards a Method for Design Principle Development in Information Systems. In: S. Hofmann, O. Müller, M. Rossi (eds) *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry. DESRIST 2020. Lecture Notes in Computer Science* (Vol.

12388, pp. 208-220). Springer. https://doi.org/10.1007/978-3-030-64823-7_20

OECD (2021). The Digital Transformation of SMEs. OECD Publishing, Paris. https://doi.org/10.1787/01a4ae9d-en

OECD (2023). Generative AI for SMEs: Separating the Chit and the ChatGPT. Cogito OECD. Retrieved April 26, 2025 from https://oecdcogito.blog/2023/02/08/digital-smes-separating-the-chit-and-the-chat-gpt/

Schwaeke, J., Peters, A., Kanbach, D. K., Kraus, S., & Jones, P. (2025). The New Normal: The Status Quo of AI Adoption in SMEs. *Journal of Small Business Management*, *63*(3), 1297–1331. https://doi.org/10.1080/00472778.2024.2379999

Smith, M. (2024a). Are North Carolina Businesses Using AI? A Look at Current Use and Future Projections. NC Department of Commerce, Labor & Economic Analysis. Retrieved May 13, 2025 from https://www.commerce.nc.gov/news/the-lead-feed/are-nc-businesses-using-ai

Smith, M. (2024b). What Industries Are Using AI? Current Use and Future Expectations. NC Department of Commerce, Labor & Economic Analysis. Retrieved June 4, 2025 from https://www.commerce.nc.gov/news/the-lead-feed/what-industries-are-using-ai

Soudi, M. S., & Bauters, M. (2024). AI Guidelines and Ethical Readiness Inside SMEs: A Review and Recommendations. *Digital Society*, *3*(1). https://doi.org/10.1007/s44206-024-00087-1

Steven, A., & Blake, G. (2024). AI and Business Intelligence: Boosting SME Growth while Navigating Ethical and Security Challenges [Working Paper]. http://doi.org/10.13140/RG.2.2.17182.14408

U.S. Chamber of Commerce (2023). The Impact of Technology on Small Business Report 2023 Edition. U.S. Chamber of Commerce Technology Engagement Center. Retrieved September 14, 2025 from https://www.uschamber.com/assets/documents/The-Impact-of-Technology-on-Small-Business-Report-2023-Edition.pdf

Vallabhaneni, R., Vaddadi, S. A., Pillai, S. E. V. S., Addula, S. R., & Ananthan, B. (2024). MobileNet Based Secured Compliance through Open Web Application Security Projects in Cloud System. *Indonesian Journal of Electrical Engineering and Computer Science*, 35(3), 1661-1669. http://doi.org/10.11591/ijeecs.v35.i3.pp1661-1669

Watney, C., & Auer, D. (2021). Encouraging AI adoption by EU SMEs. Progressive Policy Institute. Retrieved May 13, 2025 from https://www.progressivepolicy.org/encouraging-ai-adoption-by-eu-smes/"

Zhao, J., Wang, S., Zhao, Y., Hou, X., Wang, K., Gao, P., Zhang, Y., Wei, C., & Wang, H. (2024). Models are Codes: Towards Measuring Malicious Code Poisoning Attacks on Pre-Trained Model Hubs. arXiv Preprint. https://doi.org/10.48550/arXiv.2409.09368

APPENDIX A
Requirement Collection: Preliminary Survey of AI Platform

Section	Question	Answer Options
Company Size and Industry	What is the size of your company?	(A) 1-10 employees (B) 11-50 employees (C) 51-200 employees (D) 201-500 employees (E) 500+ employees
	What industry does your company operate in?	(A) Retail (B) Manufacturing (C) Financial Services (D) Healthcare (E) Technology (F) Hospitality (G) Other (please specify): ____
	How familiar are you with AI tools and technologies?	(A) Not familiar at all (B) Slightly familiar (C) Moderately Familiar (D) Very Familiar (E) Extremely Familiar
Platform Feature Interest	Would your company benefit from a no-code/low-code platform to create AI models for tasks like sales forecasting or customer segmentation?	(A) Yes, definitely (B) Maybe (C) No, we already have AI solutions in place (D) No, we don’t have a need for AI right now
	Which of the following AI tools would you find most valuable? (Select all that apply)	(A) Automated customer support/chatbots (B) Sales forecasting tools (C) Image generation for marketing (D) Document summarization (E) Data Dashboard automation (F) Fraud detection tools (G) Voice-to-text automation (H) Sentiment Analysis (I) Recruitment & Resume Screening (J) “Talk to your data” chatbot
Concerns and Barriers	What is your main concern about using AI tools? (Select all that apply)	(A) Data privacy and security (B) Cost of implementation (C) Complexity of setup and usage (D) Bias or fairness of AI models (E) Lack of in-house technical expertise (F) Other (please specify): ____
	How important is data privacy and security in deciding whether to use AI tools?	(A) Extremely important (B) Important (C) Neutral (D) Not important
	Would your company be more likely to adopt AI solutions if they complied with privacy laws like GDPR and offered encryption and data protection features?	(A) Yes, definitely (B) Maybe (C) No, privacy compliance is not a major concern
	What would make you hesitate to use AI tools provided by a third-party service? (Select all that apply)	(A) Uncertainty over who owns the data (B) Concerns about hidden costs (C) Lack of confidence in the accuracy of AI (D) Poor support or training for the tools (E) Other (please specify): ____

Open-Ended Questions	Are there any specific AI tools or features not listed that you believe would benefit your business?	(A) Please Specify: _________
	What concerns or barriers do you think would prevent your company from adopting an AI-powered platform like this one?	(A) Please Specify: _________

Table 2 Preliminary Survey of AI Platform

APPENDIX B
Prototype of the AI Platform for SMEs

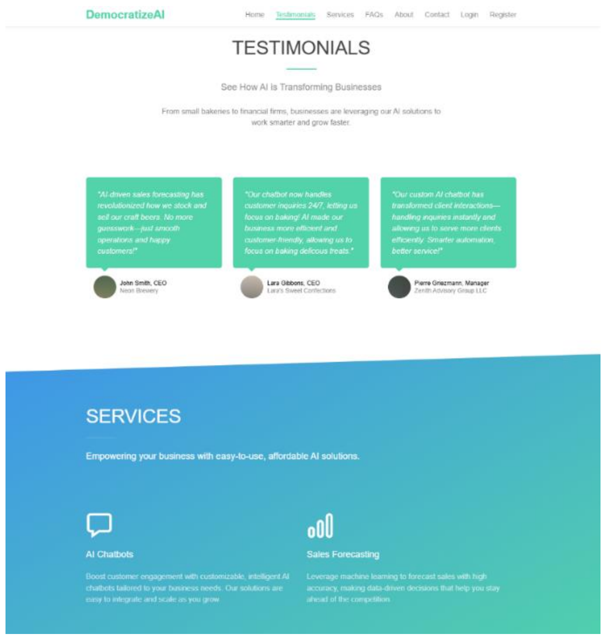

Figure 7 Testimonials and Services

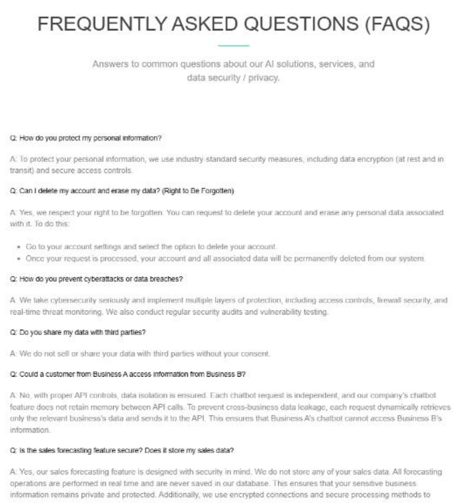

Figure 8 Frequently Asked Questions (FAQs)

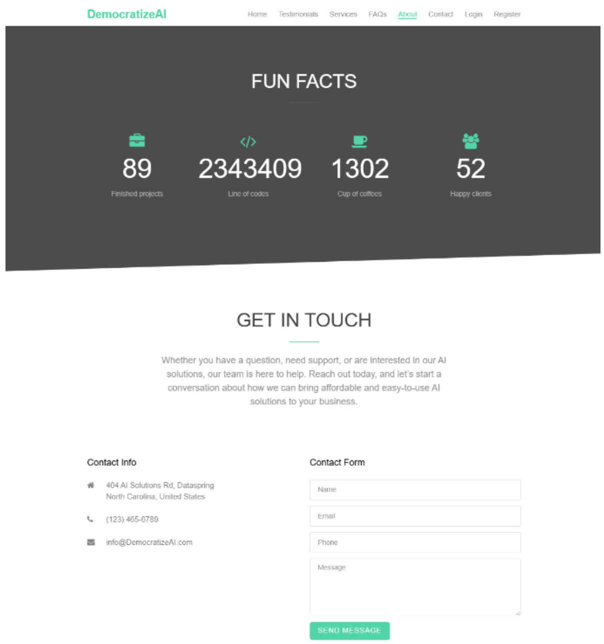

Figure 9 Contact Form

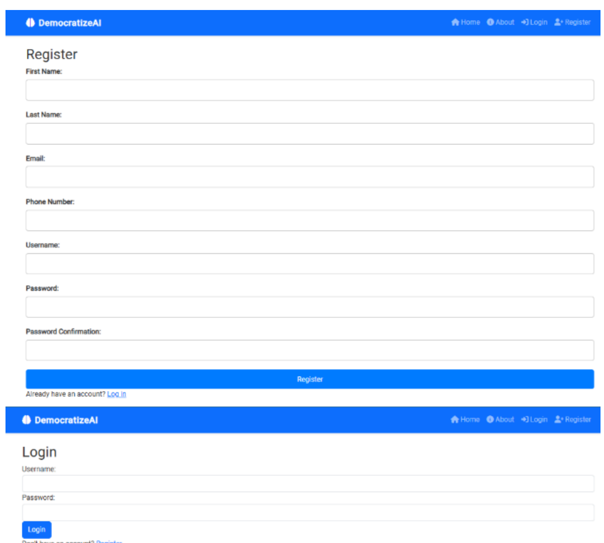

Figure 10 Register & Login Form

Figure 11 Manage & Update User Page

APPENDIX C
Usability and Feedback Survey

Questions	Answer Options
Pre-Test Questionnaire	
Q1: What industry does your business operate in?	Open-ended
Q2: How familiar are you with AI tools?	(A) Not familiar at all (B) Slightly familiar (C) Moderately familiar (D) Very familiar (E) Extremely familiar
Q3: How much of a concern are the following factors in preventing you from adopting or using more AI tools in your business?	0–5 scale (Affordability, Accessibility, Ease-of-use)
Q4: What AI-powered tools (if any) do you currently use in your business?	Open-ended
Task 1: Explore the website. Create an Account.	
Q1: How would you rate the landing page and navigation?	0–5 scale (Landing Page, Navigation)
Task 2: Navigate to the Data Security and Privacy FAQs section and read the FAQs.	
Q1: Did you easily find the information you were looking for?	(A) Yes; (B) No
Q2: Which security/privacy feature stood out to you the most?	(A) The ability to delete my account and permanently erase all personal data (Right to Be Forgotten). (B) The use of industry-standard encryption and secure access controls to protect personal information. (C) Strict policy of not sharing or selling my data to third parties without consent. (D) Data isolation between businesses to ensure one customer can’t access another’s information. (E) Secure, real-time sales forecasting that doesn’t store any business data.
Q3: How confident do you feel about the platform’s data privacy measures?	(A) Not confident at all (B) Slightly confident (C) Moderately confident (D) Very confident (E) Extremely confident
Q4: After reading the FAQs, do you feel more inclined to utilize these tools?	(A) Yes; (B) No

Task 3: Navigate to the AI Sales Forecasting Tool and forecast the sales for the next month. (Use example CSV in the folder, look over CSV if you'd like)	
Q1: How easy was this task?	(A) Not easy at all (B) Slightly easy (C) Moderately easy (D) Very easy (E) Extremely easy
Q2: How likely would you use this tool in your business?	(A) Not at all (B) Slightly likely (C) Somewhat likely (D) Very likely (E) Extremely likely
Q3: What, if anything, was unclear about the process?	Open-ended
Task 4: Navigate to the Custom Chatbot Feature create your personalized chatbot and utilize the personalized chatbot link to test your chatbot.	
Q1: How easy was this task?	(A) Not easy at all (B) Slightly easy (C) Moderately easy (D) Very easy (E) Extremely easy
Q2: How likely would you use this tool in your business?	(A) Not at all (B) Slightly likely (C) Somewhat likely (D) Very likely (E) Extremely likely
Q3: What, if anything, was unclear about the process?	Open-ended
Post-Test Questionnaire	
Q1: Overall, how easy was it to navigate and use the site?	(A) Not easy at all (B) Slightly easy (C) Moderately easy (D) Very easy (E) Extremely easy
Q2A: How likely are you to use this platform regularly?	(A) Not at all (B) Slightly likely (C) Somewhat likely (D) Very likely (E) Extremely likely
Q2B: What would encourage you to use this platform regularly?	Open-ended
Q3A: Are there any AI features you expected but didn't find?	(A) Yes; (B) No

Q3B: What additional AI tools would be valuable for your business?	Open-ended
Q4: Would guide video tutorials enhance your experience with the site?	(A) Yes; (B) Maybe; (C) No; (D) I’m not sure
Q5: Which pricing model would you find most reasonable for this service?	(A) Free with limited features (B) Free with ads (C) Monthly subscription (D) Only if free

Table 3 Usability and Feedback Survey

Developing Florida Digital Divide Index: A Comprehensive Analysis of Internet Accessibility and Socio-economic Characteristics

Tushya Vemuri
tushya09@gmail.com
Southeastern Grocers

Karthikeyan Umapathy
k.umapathy@unf.edu
University of North Florida
Jacksonville, FL 32224

Abstract

Along with the technological advancement and infiltration of Internet-based devices into our daily lives, Digital Divide research domain has evolved to focus on social development issues. Purdue researchers have developed the Digital Divide Index to measure digital access and use gaps for U.S. counties. Digital Divide Index goes beyond the access gap to focus on identifying communities disconnected from social ties and economic opportunities of the 21st century. However, the Digital Divide Index currently is calculated only at the county level. We have developed the Florida Digital Divide Index at the Zip Code level. We collected relevant datasets from the Census Bureau and the Ookla speed test. We applied Random Forest modeling to the index scores and gathered data variables to identify top importance features. The findings from the machine learning model were used to develop interactive dashboards to explore Florida zip codes with digital divide index scores.

Keywords: Digital Divide, broadband access, socio-economic, random forest, dashboards

Recommended Citation: Vemuri, T., Umapathy, K., (2026). Developing Florida Digital Divide Index: A Comprehensive Analysis of Internet Accessibility and Socio-economic Characteristics. *Journal of Information Systems Applied Research and Analytics*, v19(n1) pp 64-79. DOI# https://doi.org/10.62273/SXIC6488

Developing Florida Digital Divide Index: A Comprehensive Analysis of Internet Accessibility and Socio-economic Characteristics

Tushya Vemuri and Karthikeyan Umapathy

1. INTRODUCTION

President Biden remarked that "for today's economy to work for everyone, internet access is just as important as electricity, or water, or other basic services" during the Broadband Equity Access and Deployment (BEAD) program announcement at the White House (Mason & Renshaw, 2023, para. 4). Given the vitality of connectivity, digital accessibility casts a long shadow across communities and demographics (Powell, Bryne, & Dailey, 2010). Broadband access stands as the invisible barrier that separates the technologically privileged from the underprivileged, marking a distinction far deeper than just access to hardware. The digital divide encapsulates disparities in the ability to engage with the digital world, access information, and utilize technology for advancement.

The digital divide is more than a gap that can be overcome by providing equipment or access to a service. Rather, the digital divide is a social development issue that needs to be addressed by integrating information and communication technologies into impacted communities (Warschauer, 2004). In today's society, being connected and digitally literate means having access to education, job opportunities, healthcare information, and social networks. It's about managing your finances online, applying for jobs, completing your education, or even starting a business. For some, these opportunities are just a click away. But for others, barriers like lack of internet access, affordability, and digital skills stand in the way.

Despite the United States being a global hub for technological advancements and innovation, significant disparities exist in digital access across different regions and communities. Roberto Gallardo from the Purdue University Center for Regional Development has developed the Digital Divide Index (DDI) to rank and identify counties in the United States with the highest digital divide (Gallardo, 2016, 2024). The DDI paints a concerning picture of digital accessibility disparities, emphasizing the uneven distribution of digital resources and connectivity. This discrepancy underscores the importance of conducting focused studies on the digital divide in different parts of the USA. By identifying regions with high DDI scores, researchers and policymakers can better understand the underlying causes of digital disparities and implement targeted interventions. These studies are crucial for ensuring that the benefits of technology and digital access are equitably shared, supporting educational opportunities, economic development, and social inclusion across all American communities.

The DDI doesn't just highlight these gaps; it also points us toward solutions. Breaking down the digital divide into measurable components shows us where to direct resources and efforts. Whether providing internet access to remote areas, making technology more affordable, or offering digital literacy programs, the DDI guides policymakers, educators, and community leaders in making informed decisions to ensure everyone can benefit from the digital revolution.

However, DDI values are provided at the county level. Florida Philanthropic Network (FPN) is an organization focused on addressing issues with the 2020 Census and planning for the 2030 Census. The Census Bureau has revealed that the census count for Florida has an estimated net coverage error of -3.48%, which means around 749,529 people in Florida were undercounted in the 2020 census, as the recorded Florida population is 21,538,187 (America-Counts-Staff, 2022). FPN engaged the Florida Data Science for Social Good (FL-DSSG) team to develop data-driven strategies aimed at mitigating the undercount observed in the 2020 Census. The collaboration seeks to inform and enhance methods for ensuring that the 2030 Census data collection process avoids similar disparities in representation. As the Census Bureau predominantly utilizes online data collection methods, FPN sought digital divide measurement at the zip code level. Analyzing the digital divide at the zip code level provides a fine-grained understanding of areas at risk of digital access disparities that can directly impact Census response rates. As county-level data may obscure localized challenges, zip code level insights would allow researchers and FPN to identify communities at higher risk of undercount due to limited internet access.

Florida's diverse cultural and topographical landscape presents a unique case study for exploring the digital divide issue. With bustling metropolitan hubs and secluded rural locales, the state is a microcosm of the nation's wider digital disparities. We expand on Gallardo's Digital Divide Index (DDI) work and offer a quantifiable look into DDI at Zip Code levels in Florida, shining a light on the areas where the digital age is but a distant echo and those where it resonates clearly.

This paper delves into the fine-tuned DDI's findings to paint a comprehensive picture of Florida's digital landscape. By mapping out the contours of connectivity and access, we aim to provide a foundation to build more inclusive digital strategies, ensuring that all Floridians can confidently navigate the digital future.

2. BACKGROUND – DIGITAL DIVIDE INDEX

The digital divide is a sociotechnical phenomenon that has attracted public policy and information systems researchers' attention. Vassilakopoulou and Hustad (2023) conducted a systematic literature review of information systems research on the digital divide on articles published from 2010 to 2020. One of the findings identified by researchers is the lack of studies that focused on innovative approaches to address the grand challenge of multi-faceted dimensions of the digital divide and drawing insights into bridging the digital divide. The Digital Divide Index (DDI), a nationwide measurement of the digital divide, is one such effort. Roberto Gallardo developed the Digital Divide Index (DDI) from the Purdue University Center for Regional Development (Gallardo, 2024). DDI score ranges from 0 (low) to 100 (highest digital divide). The DDI measurement uses the Ookla speed test and Census data. The DDI score comprises infrastructure/adoption (INFA) and the socio-economic (SE) dimension scores. Infrastructure dimension score is calculated based on broadband infrastructure and adoption variables. Socio-economic dimension scores is calculated based on variables known to impact technology adoption. INFA and SE scores are combined to calculate the overall DDI score for each county in the United States. Figure 1 displays a map of DDI score across US.

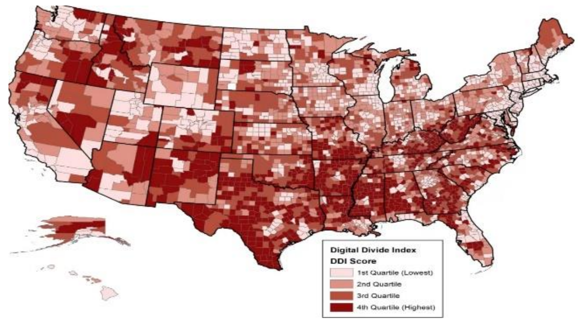

Figure 1. Digital Divide Index across the United States.

2.1. Literature Review

Literature on the digital divide has evolved through increasingly granular levels of analysis, starting with national-level comparisons and progressing into sub-national and behavioral dimensions, reflecting growing awareness of intra-country disparities. Initial research efforts focused on disparities within the United States (NTIA, 1999), while international comparisons emerged with efforts such as Carrocher and Ordanini (2002). The conceptualization of a "global digital divide" gained traction through works by Norris (2001), and Dasgupta, Lall, and Wheeler (2005), which lead to continent-wide analyses such as Fuchs and Horak's (2008) study of Africa.

Sub-national analyses have grown in prominence, reflecting a shift toward localized understandings of digital inequality. Pioneering studies in the U.S. (Atkinson & Coduri, 2002) and China (Jin & Xiong, 2002) paved the way for regional assessments across India, Europe, Australia, and beyond. These studies often rely on first-level metrics—such as access and infrastructure—though some, like Korovkin, Park, and Kaganer (2023), have begun to explore second-level divides in urban contexts. The increasing frequency of sub-national research post-2010 suggests a growing recognition of intra-national disparities and the need for targeted policy interventions.

Despite the proliferation of sub-national studies, there is a notable absence of research that operationalizes the digital divide at the ZIP code level. This gap is particularly consequential given the increasing policy emphasis on hyper-local interventions and the need to identify digital exclusion within neighborhoods and communities, limiting their utility for targeted resource allocation and community-based programming.

In sum, literature underscores a critical methodological and empirical gap: the absence of ZIP code-level digital divide measurement. Addressing this gap requires integrated models that combine multidimensional indicators with geospatial precision, enabling researchers and policymakers to more effectively diagnose and address digital inequities at the community level.

Carrocher and Ordanini proposed a model for measuring the digital divide for 10 countries. For calculating digital divide, authors utilized 36 indicator variables and applied principal components analysis for aggregating the variables into synthetic index of digitalization. The digital divide measurement framework was used to highlight opportunities and risks for business managers working in the digital economy environment. The research described in this paper differs in the geographic unit level of analysis and variables used for calculating digital divide index.

3. IMPLEMENTATION METHODOLOGY

After reviewing DDI the methodology outlined in (Gallardo, 2024), the Florida Digital Divide Index (FL-DDI) was calculated to incorporate the latest U.S. Census data. This approach allowed for a contemporary assessment of the digital divide, capturing nuances in internet access, digital literacy, and technology usage across different zip codes in the Florida region.

The methodology from (Gallardo, 2024) provides a robust framework for analyzing the digital landscape, considering variables such as Households with no computer, Households without an internet subscription, average download speed, average upload speed, population percentage of 65 years and above, population percent 25 years and above with Less Than High School (LTHS), Percentage of population with disability, Percentage of population below poverty rate. Using U.S. Census data enhances the DDI's reliability by grounding it in comprehensive and systematically collected information.

All the attributes listed in Table 1 in Appendix A have been sourced from the 2022 American Community Survey (ACS) 5-year census data sources. This comprehensive dataset allows for a detailed and scaled analysis, providing a broad yet nuanced snapshot of demographic, economic, and technological factors across various regions. The extended duration of the data collection ensures that the attributes reflect sustained trends and patterns, making them highly reliable for in-depth analysis in studies such as the Digital Divide Index (DDI). This approach ensures that the attributes encompass a wide range of variables, from socio-economic status to technology access, which is crucial for accurately assessing the scope and impact of the digital divide in different communities.

3.1. Data Collection

We explored the U.S. Census Bureau's website to identify data sources relevant to the Digital Divide Index (DDI). We explored a range of Census data profiles to identify relevant datasets and determine the availability of Florida-specific data. This approach was taken to gather a comprehensive dataset encompassing a range of attributes relevant to understanding the facets of the digital divide across Florida. Appendix A contains a Table that outlines the specific attributes selected for analysis and the corresponding profiles from which these data were extracted.

3.2. FL-DDI Calculation

The calculation of the Florida Digital Divide Index (FL-DDI) involved a detailed process of calculating scores for factors like infrastructure access (INFA) and social equity (SE). This methodological approach aims to comprehensively understand the digital divide in specific areas, incorporating both quantitative and qualitative aspects of digital access and literacy. Here's a step-by-step breakdown of how the FL-DDI is calculated and the additional process of standardizing and scaling data for further analysis.

Step 1: Standardizing Internet Speed Data: Initially, the average download (avg_d_mbps_wt) and upload (avg_u_mbps_wt) internet speeds are extracted from the Ookla speed test website (Ookla, 2024). Ookla Speed Test is a widely recognized tool for assessing the performance of internet connections globally. It measures download and upload speeds to provide users with a clear view of their internet service performance.

Download speed, measured by the Ookla Speed Test, refers to the rate at which data is transferred from the internet to a user's device. Broadband download speed is typically expressed in megabits per second (Mbps). Higher download speeds allow for smoother streaming of high-definition videos, faster

loading of webpages, and more efficient downloads of large files.
Upload speed measures how quickly data is sent from a user's device to the internet. The upload speed is also measured in Mbps. Upload speeds are crucial for sending large amounts of data, such as video conferencing, uploading large files to a server, or live streaming.

A comprehensive approach has been followed to retrieve and analyze broadband speed data across different geographical areas, specifically focusing on Florida's zip codes. The goal is to extract average upload and download speeds for each zip code, using weighted averages where the weights are the number of tests conducted. A Python script (see Appendix B) has been implemented to extract data that involves the following steps.

1. The script generates a URL to download speed test data for fixed broadband services in the second quarter of 2020. This dataset is read into a GeoDataFrame named tiles.
2. Downloads and reads a shapefile of U.S. state boundaries into a GeoDataFrame named States.
3. Downloads and reads a shapefile of U.S. zip code boundaries into a GeoDataFrame named ZipCodes.
4. The script filters the States GeoDataFrame for Florida (state FIPS code '12'), ensuring the coordinate reference system (CRS) matches that of the zip code data. An inner spatial join (sjoin) is then performed to extract zip codes that fall within Florida, resulting in a GeoDataFrame florida_zipcodes. Performs an inner spatial join between the broadband speed test tiles and Florida zip codes, resulting in tiles_in_florida_zipcodes containing broadband data specifically for areas within Florida zip codes.
5. The code computes weighted average download and upload speeds (avg_d_mbps_wt and avg_u_mbps_wt) for each zip code. Weights are based on the number of tests conducted, reflecting a more accurate measure of broadband speeds experienced by users. This is done by grouping the data by zip code and using the np.average function with tests as weights.
6. The weighted average speeds are then merged with the total number of tests conducted in each zip code, resulting in the zipcode_stats DataFrame. This DataFrame is saved to a CSV file, providing a ready-to-use dataset for analysis of broadband speeds by zip code in Florida.

After extracting the average download and upload speed, the values are standardized using Z-scores. This standardization process converts the raw speed data into a format that reflects how many standard deviations each value is from the mean, facilitating comparison across different scales and distributions. This ensures that the values are in sync with other features, all in percentile.

Step 2: Handling Missing Data and Data Status Tagging: In the process of analyzing broadband speeds across Florida's zip codes, we encounter an issue common to many datasets: not all zip codes have complete data for the attributes being studied. To address this and ensure the integrity and usability of our analysis, we implement a two-way approach to manage missing values and tag data completeness.

Replacing Missing Values with Column-Wise Medians: To maintain the statistical validity of our dataset without discarding incomplete records, we opt to replace missing values with the median value of the respective attribute across all zip codes. This method is chosen because the median is less sensitive to outliers than the mean, making it a robust measure for imputing missing data. For each attribute with missing values, we calculate its median based on available data and fill in the gaps accordingly. This ensures that every zip code has a value for each attribute, allowing for comprehensive statewide analysis.

Tagging Data Status: To maintain transparency in data analysis, we introduce a "Data Status" column to our dataset. This column categorizes each zip code based on the completeness of its data.

Complete Data: This tag is assigned to zip codes where all attributes have original, non-imputed values. It indicates that the data for these zip codes is complete and has not been subjected to imputation.

Partial Data Available: This tag is applied to zip codes that require imputation for one or more attributes. It signals to users of the dataset that while the zip code is included in the analysis, some of the values have been filled in using median imputation due to the absence of original values.

This approach enhances the dataset's usability by filling in missing information and maintains data transparency by clearly indicating which records have been altered. Users can easily identify which zip codes have complete data and which have been partially imputed, enabling informed decision-making and analysis. This meticulous attention to data quality and integrity is crucial for accurately assessing broadband access and performance across Florida, providing a solid foundation for further research and policy development.

Step 3: Calculating the Infrastructure Access Score: The Infrastructure Access (INFA) score is computed by incorporating the percentages of households without a computer and without an internet subscription, each weighted at 35%, as shown in Figure 2. The standardized (Z-scored) average download and upload speeds are subtracted from this total, each weighted but negatively at 15%. This calculation reflects the positive impact of having computer access and internet subscriptions while accounting for the quality of internet access as indicated by speeds. The weightings used for calculation is same as the original DDI used by Gallardo (2024).

```
data['INFA'] = data['No computer'] * 0.35 +
data['Without an Internet subscription'] *
0.35 - data['avg_d_mbps_wt_1'] * 0.15 -
data['avg_u_mbps_wt_1'] * 0.15
```

Figure 2. Code snippet for INFRA score calculation.

Step 4: Computing Socio-economic Score: The Socio-economic (SE) score aggregates factors that reflect demographic vulnerabilities or disparities affecting digital access. It includes the percentage of the population over 65 years, the *percentage with less than a high school education, the percentage with a disability, and the percentage below the poverty* line. These components are summed directly without explicit weights as shown in Figure 3, underlining their collective impact on digital equity.

```
data['SE'] = data['65 years and over'] +
data['percent population 25 and over with less
than high school(LTHS)'] + data['Total Civilian
Noninstitutionalized Population!!With a
disability'] + data['Percent below poverty
level!!Population for whom poverty status is
determined']
```

Figure 3. Code snippet for SE score calculation.

Step 5: Deriving FL-DDI Score: The FL-DDI is then calculated as the sum of the INFA and SE scores, as shown in Figure 4. This final metric captures a holistic view of the digital divide, integrating considerations of both the physical infrastructure and the broader socio-economic conditions that influence digital access and utilization.

```
data['DDI'] = data['INFA'] + data['SE']
```

Figure 4. Code snippet for DDI score calculation.

Step 6: Scaling for Analysis: For comparative analysis and visualization, an additional step of rescaling the Z-scores of each column (excluding non-numeric or identifier columns like "ZipCode") to a 0-100 range is performed. This is achieved by subtracting the minimum Z-score in each column from every Z-score and dividing the result by the range of Z-scores. The scaled values are then multiplied by 100. This transformation maintains the distribution of the original data while standardizing the scale for ease of interpretation and analysis.

Through this detailed computation and scaling process, FL-DDI scores can guide researchers in identifying areas most affected by the digital divide. It can enable the prioritization of interventions and resources for those most in need and bridge the gap in digital access across the population.

4. MODELING AND DASHBOARD

FL-DDI scores were plotted against zip codes in Florida, which is visually represented in the map as shown in Figure 5. Varying colors illustrate the extent of the digital divide with minimum represented with green and max represented with red. The map uses a color-coded system to indicate the severity of the digital divide across different areas:

- Red Areas: These regions exhibit higher DDI scores, suggesting a significant digital divide. Residents in these areas may face challenges due to limited internet connectivity, fewer households with computers or smart devices, and potentially lower digital literacy rates.
- Green Areas: In contrast, green areas indicate lower DDI scores. These regions will likely have better access to digital resources, including higher rates of internet subscriptions, greater availability of computers and smart

devices, and possibly a more digitally literate population.

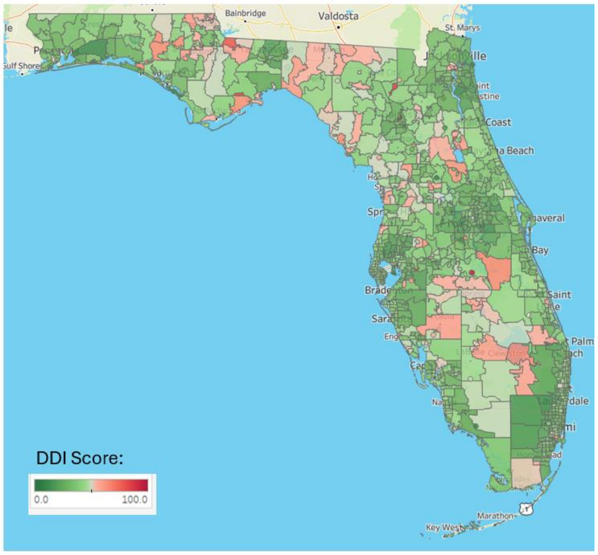

Figure 5. Digital Divide Index Scores Spread Across Florida Zip Codes.

The most severely high DDI scores are exhibited by zip codes in rural counties in Northwest and North Central regions of Florida. While, high scores are occurring for rural zip codes, each urban areas like Miami, Palm Beach, Tampa, and Jacksonville have one zip code with DDI score in range of 55 except for Orlando which have zip codes that can be considered as high DDI score. It is not surprising that zip codes with lowest DDI scores are in the urban areas with high populations. However, there were few zip codes in rural counties with low DDI scores which were further away from population hubs. These insights suggest that key variables like population and household density does not guarantee access to internet or lack off emphasizing the complexity of digital divide and the need for targeted policies to address the disparities.

4.1. Machine Learning Model Results

Leveraging machine learning for feature selection facilitates the identification of influential variables, thereby streamlining the design of interactive dashboards. This systematic prioritization of the most salient data points ensures that user attention remains focused on significant analytical drivers. In this study, a Random Forest model (Breiman, 2001) was employed to understand the feature importance of different attributes relevant to FL-DDI. The Random Forest model excelled in its predictions, boasting an impressive R-squared value of 0.9286, which means its predictions are closely aligned with the actual data. The feature importance scores are displayed in Figure 6.

- The “Total Civilian Noninstitutionalized Population with a disability” feature had the most significant impact, with an importance score of approximately 0.33.
- “Households without an Internet subscription” was the second most influential feature, with a score of around 0.15.
- The percentage of the “Population 65 years and over” closely followed, also with an importance score near 0.15.
- The “Percentage of Population below poverty level” had a notable contribution, with a score of around 0.13.

```
Model Performance: 38.138765651146315
R-squared: 0.9286161659623637
                                                          importance
Total Civilian Noninstitutionalized Population!...         0.329450
Without an Internet subscription                           0.154825
65 years and over                                          0.152115
Percent below poverty level!!Population for who...         0.129262
percent population 25 and over with less than h...         0.101639
No computer                                                0.063635
Under 18 years                                             0.020947
Smartphone with no other type of computing device          0.009893
Civilian veterans                                          0.009631
Households with Cellular data plan with no othe...         0.007591
Less than $20,000:!!Without an Internet subscri...         0.005105
avg_d_mbps_wt_1                                            0.004284
avg_d_mbps_wt                                              0.003875
Estimate!!Percent limited English-speaking hous...         0.003716
avg_u_mbps_wt                                              0.002372
avg_u_mbps_wt_1                                            0.001659
```

Figure 6. Random Forest Model feature importance scores.

With the above interpretation from machine learning model, we plotted the features with top importance against to FL-DDI score to understand the data distribution. Scatter plots in Figure 7 show the comparison between DDI score and two socio economic variables poverty status and Disability Status across various Zip Codes. Each point on the graph represents a different zip code, differentiated by color.

FL-DDI Score vs. Poverty Status (Top Plot): A trend appears to be that as the percentage of the population living below the poverty level increases, the FL-DDI score tends to increase as well. An increased FL-DDI score typically indicates a higher digital divide, suggesting that areas with more poverty might experience less digital inclusion. But there are outliers or exceptions where even though the poverty level is near 100, the FL-DDI score is at an average level between 50-60. This indicates the necessity of considering several factors to be stressed for areas with a high digital divide.

FL-DDI Score Vs Disability Status (Bottom Plot): There seems to be a positive correlation between the disability z-score and the FL-DDI score. This suggests that zip codes with a higher proportion of individuals with disabilities may also have a higher digital divide, facing greater challenges in accessing digital technologies.

Figure 7. FL-DDI Score vs. Poverty Status & Disability Score.

4.2. DDI Dashboard

A Tableau dashboard, depicted in Figure 8, was developed to present a multi-faceted view of the Florida Digital Divide Index (FL-DDI). By analyzing this dashboard, stakeholders can identify regions at risk, patterns that might inform policy, and the overall landscape of digital inclusion within the state, all of which are critical for designing effective interventions to bridge the digital divide.

DDI > 70 vs. Other Attributes (Bar Graph): This chart compares zip codes with a DDI score greater than 70 against other socio-economic attributes like the percentage of households without a computer, Household with income less than $20000 and without an internet subscription, Household without internet subscription, Percentage of population 25 and above without a high school diploma, percentage below the poverty line, and percentage under 18 years of age. The darker shaded bars represent the zip code with a high DDI score. This visualization suggests that zip codes focus on addressing digital divide issues, which are indicated by higher DDI scores for attributes correlated with providing conditions for digital equality.

FPN could use this visualization to focus census campaign outreach on high FL-DDI zip codes where households are more likely to lack internet access, live in poverty, have low educational attainment, and include a higher percentage of children—ensuring these digital and socioeconomically disadvantaged communities are accurately counted. This analysis reveals that zip codes with high FL-DDI scores also face compounded socio-economic challenges—such as low internet access, poverty, and limited educational attainment—offering digital divide researchers' critical insights into systemic inequities, while guiding public policy toward targeted investments in broadband access, digital literacy, and educational support to bridge the digital gap.

DDI Score vs. Top Importance Features (Time Series): This graph is represented to track the DDI score concerning top importance features across different zip codes over time. The features include factors identified in the previous section as part of the machine learning model.

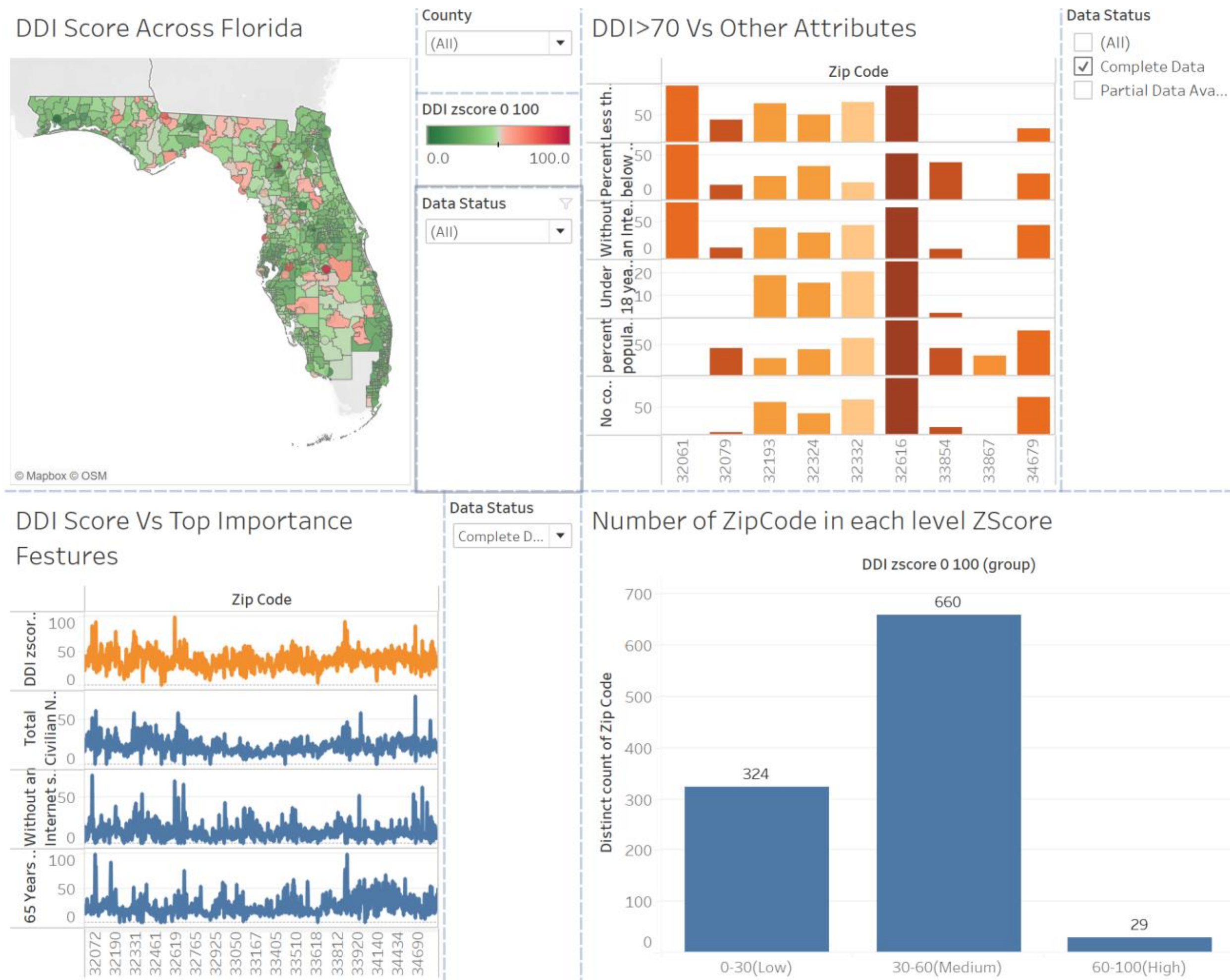

Figure 8. FL-DDI Dashboard.

Specifically attribute "percent of population with disability" shows almost a linear relationship with the DDI score, with both the peaks and lows matching. This graph can also help identify patterns or trends that warrant further investigation.

FPN could use this time series analysis to identify zip codes where rising digital divide risks align with higher percentages of residents with disabilities and seniors, enabling more timely and targeted interventions that address accessibility and connectivity barriers for these vulnerable populations to take part in the Census. Digital divide researchers could use evidence observed from time series visuals to investigate how digital exclusion disproportionately affects vulnerable populations, while policymakers could use this finding to determine zip codes to implement design-inclusive strategies—such as accessible technology programs, senior-focused digital training, and disability-friendly infrastructure—to reduce systemic barriers over time.

Number of Zip Codes in each level Z-Score: This histogram categorizes zip codes into groups based on their DDI z-scores, standardizing DDI values for comparison. It shows the number of zip codes within low, medium, and high DDI score ranges. Although the number of zip codes with a high DDI may be relatively small, it remains crucial to allocate resources effectively to these areas. Ensuring digital inclusivity and providing equitable technological access are essential steps toward integrating all communities into the rapidly evolving tech landscape.

Florida Philanthropic Network can use this histogram to identify and prioritize high-DDI zip codes—despite being fewer in number—as critical areas for targeted investment, ensuring that communities most at risk of digital exclusion receive the necessary support to achieve equitable access to technology and participation in the Census. This distribution of standardized DDI scores highlights a smaller cluster of zip codes facing the highest levels of digital exclusion, offering digital divide researchers a clear focus for studying concentrated digital inequities, while guiding public policy toward strategically allocating resources to these high-need areas—ensuring that no community is left behind in the transition to a digitally connected society.

5. CONCLUSIONS

The rapid development of information and communication technologies (ICT) has exacerbated inequalities between developed and underdeveloped communities. There seems to be limited research focused on the digital divide at the localized or zip code unit level of analysis. In this research paper, we extended existing Digital Divide Index scores that are calculated at the county level; we recalculated them at the zip code level. We calculated the Florida Digital Divide Index (FL-DDI) for Florida zip codes. The FL-DDI scores and related data sets were analyzed using the Random Forest model to identify key important features. Modeling results were utilized as key items for designing and developing visuals for interactive dashboards. We developed a Tableau Dashboard that the Florida Philanthropic Network will utilize as a part of a larger research addressing the Census undercount issue in Florida.

While this study provides valuable insights into Florida's digital divide at the zip code level, future data science and analytics research could explore the impact of targeted interventions in high FL-DDI areas and incorporate Florida-specific data and socio-economic factors such as access to affordable housing, healthcare disparities, and the prevalence of households receiving public assistance. The study's limitations include reliance on available socio-economic data, the inability of the Random Forest model to capture all aspects of digital exclusion, and a lack of data context to inform cultural barriers and local infrastructure variations. Additionally, constraints on using federal data to estimate and calculate the DDI at such a fine-grained level may limit the precision and granularity of certain inputs. The approach presented in this paper can be replicated for other states to identify regional disparities, and comparative studies could help develop best practices for addressing digital inequities nationwide.

6. REFERENCES

America-Counts-Staff. (2022, May 19, 2022). 2020 Census Undercounts in Six States, Overcounts in Eight. Retrieved from https://www.census.gov/library/stories/2022/05/2020-census-undercount-overcount-rates-by-state.html

Atkinson, R. D., & Coduri, R. (2002). *The 2002 State New Economy Index. Benchmarking economic transformation in the states*. Retrieved from Washington, DC:

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/A:1010933404324

Corrocher, N., & Ordanini, A. (2002). Measuring the Digital Divide: A Framework for the Analysis of Cross-Country Differences. *Journal of Information Technology, 17*(1), 9-19. doi:10.1080/02683960210132061

Dasgupta, S., Lall, S., & Wheeler, D. (2005). Policy Reform, Economic Growth and the Digital Divide. *Oxford Development Studies, 33*(2), 229-243. doi:10.1080/13600810500137889

Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and Informatics, 25*(2), 99-116. doi:https://doi.org/10.1016/j.tele.2006.06.004

Gallardo, R. (2016). *2014 County-level Digital Divide Index*. Retrieved from https://www.researchgate.net/publication/308514127_2014_County-level_Digital_Divide_Index

Gallardo, R. (2024). Digital Divide Index. Retrieved from http://pcrd.purdue.edu/ddi

Jin, J., & Xiong, C. (2002). *The digital divide in National Informatization Quotient: The perspective of mainland China*. Paper presented at the International Conference on Digital Divides: Technology and Politics in the Information Age, Hong Kong.

Korovkin, V., Park, A., & Kaganer, E. (2023). Towards conceptualization and quantification of the digital divide. *Information, Communication & Society, 26*(11), 2268-2303. doi:10.1080/1369118X.2022.2085612

Mason, J., & Renshaw, J. (2023, June 26). US to spend $42 billion to make internet access universal by 2030. Retrieved from https://www.reuters.com/world/us/biden-detail-plans-42-billion-investment-us-internet-access-2023-06-26/

Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge: Cambridge University Press.

NTIA. (1999). Falling Through the Net: Defining the Digital Divide. Retrieved from https://www.ntia.gov/sites/default/files/data/fttn99/index.html

Ookla. (2024). Speedtest by Ookla Global Fixed and Mobile Network Performance Maps. Retrieved from https://registry.opendata.aws/speedtest-global-performance/

Powell, A., Bryne, A., & Dailey, D. (2010). The Essential Internet: Digital Exclusion in Low-Income American Communities. *Policy & Internet, 2*(2), 161-192. doi:https://doi.org/10.2202/1944-2866.1058

Vassilakopoulou, P., & Hustad, E. (2023). Bridging Digital Divides: a Literature Review and Research Agenda for Information Systems Research. *Information Systems Frontiers, 25*(3), 955-969. doi:10.1007/s10796-020-10096-3

Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge, MA: MIT Press.

APPENDIX A – Data Sources

Table 1. Data Sources Used in FL-DDI Calculation

Attribute Name	Description	Profile Name	Source
Estimate!!Percent!!Total households!!TYPES OF COMPUTER!!No computer	The percentage of households without a computer	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total households!!TYPE OF INTERNET SUBSCRIPTIONS!!Without an Internet subscription	Percentage of households without internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total population!!SELECTED AGE CATEGORIES!!65 years and over	Percentage of population above 65 years of age	ACS – S0101	U.S. Census Bureau
Estimate!!Percent!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Less than 9th grade+Estimate!!Percent!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!9th to 12th grade, no diploma	Percentage of population above 25 years of age with Less Than High school	ACS – S1501	U.S. Census Bureau
Percent!!DISABILITY STATUS OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION!!Total Civilian Noninstitutionalized Population!!With a disability	Percentage of Population with a disability	ACS – DP02	U.S. Census Bureau
Estimate!!Percent below poverty level!!Population for whom poverty status is determined	Percentage of Population below poverty rate	ACS – S1701	U.S. Census Bureau
Estimate!!Total!!Total households!!TYPE OF INTERNET SUBSCRIPTIONS!!With an Internet subscription:!!Dial-up with no other type of Internet subscription	Household with only cellular data plan and no other type of internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Total!!Total households!!HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2022 INFLATION-ADJUSTED DOLLARS)!!Less than $20,000:!!Without an Internet subscription	Percentage of household with less than $20000 household income and no internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total households!!TYPES OF COMPUTER!!Has one or more types of computing devices:!!Smartphone!!Smartphone with no other type of computing device	Percentage of Households with only smartphone and no other computing device	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!VETERAN STATUS!!Civilian population 18 years and over!!Civilian veterans	Percentage of Civilian veterans	ACS – DP02	U.S. Census Bureau
Estimate!!Percent!!Total population!!SELECTED AGE	Percentage of population below 18 years	ACS – S0101	U.S. Census Bureau

Attribute Name	Description	Profile Name	Source
CATEGORIES!!Under 18 years			
Estimate!!Percent limited English-speaking households!!All households	Limited English-speaking households	ACS – S1602	U.S. Census Bureau
avg_d_mbps_wt	Average download speed	Shapefile & tl_2019_us_zcta510.zip	Ookla Speed Test
avg_u_mbps_wt	Average Upload speed	Shapefile & tl_2019_us_zcta510.zip	Ookla Speed test

APPENDIX B – Python Script for Standardizing Internet Speed Data

```
# %%
%matplotlib inline

from datetime import datetime

import geopandas as gp
import matplotlib
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

from shapely.geometry import Point
from adjustText import adjust_text

# %%
def quarter_start(year: int, q: int) -> datetime:
    if not 1 <= q <= 4:
        raise ValueError("Quarter must be within [1, 2, 3, 4]")

    month = [1, 4, 7, 10]
    return datetime(year, month[q - 1], 1)

def get_tile_url(service_type: str, year: int, q: int) -> str:
    dt = quarter_start(year, q)

    base_url = "https://ookla-open-data.s3-us-west-2.amazonaws.com/shapefiles/performance"
    url = f"{base_url}/type%3D{service_type}/year%3D{dt:%Y}/quarter%3D{q}/{dt:%Y-%m-
%d}_performance_{service_type}_tiles.zip"
    return url

# %%
tile_url = get_tile_url("fixed", 2020, 2)
tile_url

# %%
tiles = gp.read_file(tile_url)

# %%
print(tiles.head())

# %%
# zipfile of U.S. county boundaries
state_url = "https://www2.census.gov/geo/tiger/TIGER2019/STATE/tl_2019_us_state.zip"
States = gp.read_file(state_url)
```

```
# %%
print(States.head())

# %%
# zipfile of U.S. county boundaries
ZipCode_url = "https://www2.census.gov/geo/tiger/TIGER2019/ZCTA5/tl_2019_us_zcta510.zip"
ZipCodes = gp.read_file(ZipCode_url)

# %%
print(ZipCodes.head())

# %%
ky_States = States.loc[States['STATEFP'] == '12'].to_crs(4326)

# %%
ky_States.head()

# %%
ky_States = ky_States.to_crs(ZipCodes.crs)

# %%
florida_zipcodes = gp.sjoin(ZipCodes, ky_States, how="inner", op='intersects')

# %%
florida_zipcodes.head()

# %%
florida_zipcodes['ZCTA5CE10'].nunique()

# %%
florida_zipcodes.info()

# %%
if 'index_left' in tiles.columns:
    tiles = tiles.rename(columns={'index_left': 'index_left_old'})

if 'index_right' in tiles.columns:
    tiles = tiles.rename(columns={'index_right': 'index_right_old'})

# Similarly, check and rename for 'florida_zipcodes' if needed
if 'index_left' in florida_zipcodes.columns:
    florida_zipcodes = florida_zipcodes.rename(columns={'index_left': 'index_left_old'})

if 'index_right' in florida_zipcodes.columns:
    florida_zipcodes = florida_zipcodes.rename(columns={'index_right': 'index_right_old'})

# %%
tiles_in_florida_zipcodes = gp.sjoin(tiles, florida_zipcodes, how="inner", op='intersects')

# %%
# convert to Mbps for easier reading
tiles_in_florida_zipcodes['avg_d_mbps'] = tiles_in_florida_zipcodes['avg_d_kbps'] / 1000
tiles_in_florida_zipcodes['avg_u_mbps'] = tiles_in_florida_zipcodes['avg_u_kbps'] / 1000

# %%
tiles_in_florida_zipcodes.head()

# %%
```

```
tiles_in_florida_zipcodes.info()

# %%
zipcode_stats = (
    tiles_in_florida_zipcodes.groupby(["GEOID10", "ZCTA5CE10"])
    .apply(
        lambda x: pd.Series({
            "avg_d_mbps_wt": np.average(x["avg_d_mbps"], weights=x["tests"]),
            "avg_u_mbps_wt": np.average(x["avg_u_mbps"], weights=x["tests"])
        })
    )
    .reset_index()
    .merge(
        # Aggregate total tests for each group
        tiles_in_florida_zipcodes.groupby(["GEOID10", "ZCTA5CE10"])
        .agg(tests=("tests", "sum"))
        .reset_index(),
        on=["GEOID10", "ZCTA5CE10"],
    )
)

# %%
zipcode_stats.head()

# %%
zipcode_stats.to_csv(r'zipcode_stats_output.csv', index=False)

# %%
table_stats = (
    zipcode_stats.loc[zipcode_stats["tests"] >= 50]
    .nlargest(20, "avg_d_mbps_wt")
    .append(
        zipcode_stats.loc[zipcode_stats["tests"] >= 50].nsmallest(20, "avg_d_mbps_wt")
    )
    .sort_values("avg_d_mbps_wt", ascending=False)
    .round(2) # round to 2 decimal places for easier reading
)

# %%
header = ["GEOID10", "ZCTA5CE10", "Avg download speed (Mbps)", "Tests"]

table_stats.rename(columns=dict(zip(table_stats.columns, header)))

# %%
zipcode_data_map = tiles_in_florida_zipcodes[['GEOID10', 'ZCTA5CE10']].merge(zipcode_stats,
on='GEOID10')

# %%
labels = ["0 to 25 Mbps", "25 to 50 Mbps", "50 to 100 Mbps", "100 to 150 Mbps", "150 to 200 Mbps"]

zipcode_data_map['group'] = pd.cut(
    zipcode_data_map.avg_d_mbps_wt,
    (0, 25, 50, 100, 150, 200),
    right=False,
    labels = labels
)

# %%
zipcode_data_map.head()
```

```
# %%
ky_places = gp.read_file("ftp://ftp2.census.gov/geo/tiger/TIGER2019/PLACE/tl_2019_12_place.zip")

# %%
ky_places = ky_places.loc[ky_places['PCICBSA'] >= "Y"].sample(15, random_state=1).to_crs(26916)
ky_places["centroid"] = ky_places["geometry"].centroid
ky_places.set_geometry("centroid", inplace = True)

# %%
ky_places.head()

# %%
zipcode_data=tiles_in_florida_zipcodes[['ZCTA5CE10','REGION','DIVISION','STATEFP','NAME','avg_d_
mbps','avg_u_mbps']]
print(zipcode_data.head())

# %%
zipcode=tiles_in_florida_zipcodes['ZCTA5CE10'].nunique()
print(zipcode)
```