HARNESSING CHATGPT-4 TURBO: OPTIMIZING REAL-TIME AND HISTORICAL DATA INTEGRATION FOR ACCURATE NFL QUIZ QUESTION GENERATION

Mark Karels

A Capstone Project Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Department of Computer Science
Department of Information Systems and Operations Management

University of North Carolina Wilmington

2024

Approved by

Advisory Committee

_____	_____
Dr. Lucas Layman	Dr. Kevin Matthews

_____
Dr. Curry Guinn, Chair

Accepted By

_____
Dean, Graduate School

TABLE OF CONTENTS
(Insert Automatic Table of Contents)

Page
Chapter 1: Introduction 1

Chapter 2: Literature Review and Analysis 4

Chapter 3: Methodology 12
Research Design 12
Prompt Engineering 13
Data Collection and Analysis 16
Quantitative Data Collection 16
Prompt Variation and Analysis 17
Qualitative Data Collection 18
Ethical Considerations and Data Integrity 18
Technological Architecture 19
Variables and Hypotheses 20
Independent Variables 21
Dependent Variables 21
Hypotheses 21
Limitations 22
Conclusion 23

Chapter 4: Completed Project and Experiment Results 25
Trial and Error 25
Observations During the Experiment 27
Experiment Data Collection and Analysis 29
Analyzing Results 35
Conclusion 39

Chapter 5: Future Work 42

References 46
Image References 49

Appendixes
A. Every Initial Prompt Used in Each Trial Run 51
B. Data Analysis Tables for Versions 2 and 3 61

ABSTRACT

Karels, Mark 2024. Capstone Paper, University of North Carolina Wilmington.

"Harnessing ChatGPT-4 Turbo: Optimizing Real-time and Historical Data Integration for Accurate NFL Quiz Question Generation" aims to advance educational technology by leveraging the sophisticated capabilities of ChatGPT-4 Turbo. This research focuses on enhancing prompt engineering techniques and implementing a rigorous method for ensuring the accuracy and uniqueness of quiz questions generated from both live and historical National Football League (NFL) data. The study endeavors to discover the most effective prompting strategies that enable ChatGPT-4 Turbo to produce content that is both accurate and responsive to the evolving nature of real-time sports events. It employs a feedback mechanism that utilizes ChatGPT-4 Turbo to evaluate the clarity, precision, and formatting of questions, thereby optimizing the generation process.

A key innovation of this project is the use of a structured finite state machine, built upon the Large Language Model (LLM) that powers ChatGPT-4 Turbo, to verify the historical accuracy of generated questions. This mechanism provides an integral data validation step. For generating questions based on live data, such as recent game statistics, the system necessitates the input of real-time information into the model. This approach highlights the challenges and potential costs associated with real-time data verification compared to historical data checking.

The anticipated outcomes of this research include gaining insights into the efficacy of ChatGPT-4 Turbo in processing and integrating real-time data, establishing best practices for AI-prompt interaction, and developing a versatile framework that can be applied to other domains requiring real-time data processing and educational content creation. This project aims to narrow the gap between the theoretical potential of AI and its practical application in the context of sports analytics and quiz question generation, thereby enriching the educational technology landscape.

LIST OF TABLES

Table Page

1. Verification Prompt Tracking Table 14

2. Generated Question Examples 30

LIST OF FIGURES

Figure Page

1. X-Ray Insight....5
2. AI Hallucinations....7
3. AI Tools....9
4. Prompt Engineering Example....15
5. Results of Version 2 Data Collection T-Test....32
6. Results of Version 3 Data Collection T-Test....34
7. Incorrect Question/Answer Categories....35
8. Results of Version 2 Live Data Collection....38

CHAPTER 1: INTRODUCTION

In the rapidly evolving field of artificial intelligence, the deployment of advanced language models such as ChatGPT-4 Turbo heralds a new era filled with both opportunities and obstacles. The use of AI to generate content, particularly within the dynamic and information-intensive realm of the National Football League (NFL), prompts essential inquiries regarding the ability of these models to deliver material that is not only precise and distinctive but also compelling. The core issue being explored is the adept utilization of ChatGPT-4 Turbo for the interpretation and application of both live and historical NFL data. The primary challenge encompasses striking an equilibrium between the continual updates of NFL games and the unchanging character of historical records, ensuring that the quiz questions generated are both accurate and relevant.

This investigation delves into the intricate process of prompt engineering and the integration of a feedback loop with the AI model, aiming to refine the creation of quiz content that reflects the latest developments in the NFL, as well as its rich historical context. The endeavor to harness ChatGPT-4 Turbo for this purpose underscores a broader quest to navigate the complexities of real-time data assimilation alongside the preservation of historical accuracy in educational content generation. By addressing these challenges, the research contributes to the broader discourse on the benefits and limitations of AI in transforming educational technologies, particularly in the context of sports analytics and interactive learning.

The hypothesis posited is that through strategic prompting and iterative refinement, ChatGPT-4 can be guided to produce quiz questions that are not only accurate and tailored to real-time events, as well as historical data, but also retain uniqueness, accuracy, and specificity. This project seeks to address questions surrounding

the model's ability to generate content that aligns with these parameters and the potential for these prompts to be adapted for generating quizzes in various academic, gaming, or training settings. To this end, the research will explore the intricacies of prompt design, the impact of data integration on content generation, and the methods for validating the generated questions against a backdrop of constantly evolving live sports data.

The methodological framework of this study employs a systematic analysis of prompt engineering, utilizing the vast repository of historical NFL data embedded within the ChatGPT-4 Turbo Large Language Model, in conjunction with data from ongoing live games. The objective is to develop an array of meticulously crafted prompts and systematic checks designed to enable ChatGPT-4 Turbo to accurately recognize and incorporate real-time data. Consequently, this facilitates the generation of quiz questions that not only test the user's knowledge but also capture the essence and volatility of live sporting events. An essential component of this approach is a feedback loop, meticulously devised to evaluate and enhance the accuracy and originality of the AI-generated questions. This process of continuous improvement aims to refine the AI's outputs, ensuring the production of questions that are both pedagogically beneficial and diverse, thereby fitting a broad spectrum of educational settings.

Through this research, the aim is to pioneer advanced AI applications in educational content creation, highlighting the capacity of AI to revolutionize learning environments that demand real-time, data-driven content. This endeavor seeks to demonstrate the potential of AI to adapt and thrive in dynamic learning scenarios, offering insights that could broaden its application across various data-intensive educational fields. By pushing the boundaries of AI in education, this study aims to contribute to the development of innovative teaching tools that are engaging, effective,

and capable of preparing learners for the fast-paced information age. By unlocking new capabilities in AI, the outcomes of this research may lay the groundwork for innovations that could transform the educational landscape, making learning more engaging, responsive, and attuned to the digital age.

CHAPTER 2: LITERATURE REVIEW AND ANALYSIS

The aim of this chapter is to conduct a thorough examination of the convergence between sophisticated language models like ChatGPT-4 and the dynamic field of sports analytics, with a particular focus on the creation of educational content that integrates both real-time and historical data. The study reviews a breadth of literature to extract insights into the practices of prompt engineering, assess the impact of AI applications in education and the medical field, as well as explore the potential of real-time data processing within AI frameworks. These insights are critical for the development of an innovative quiz interface that delivers content characterized by its accuracy, uniqueness, and pedagogical relevance.

The literature was selected through strict criteria to ensure the utmost relevance to the objectives of this research. Sources were chosen for their direct connection to artificial intelligence in technology, specifically those that detail prompt engineering and the use of large language models (LLMs) for content creation. Additional literature that delves into the real-time processing of dynamic data sets, particularly within the sports analytics domain, was deemed essential. This curation includes both theoretical framework and empirical studies, providing a comprehensive view of the current state of technology and its educational implications.

The thematic structure of the literature review organizes the selected references into three distinct areas, each reflecting a crucial aspect of the research project:

1) AI in Educational Technology: This theme groups together literature that discusses the transformative potential of AI in education. The capabilities of AI in crafting educational content that is engaging and instructive are highlighted, pointing to a more individualized learning experience. See Figure 1 for how AWS

is used to highlight key statistics live during an NFL game. The literature underscores the significance of a balanced approach between automated systems and human oversight to establish reliable educational platforms [2, 4, 8, 9, 14].

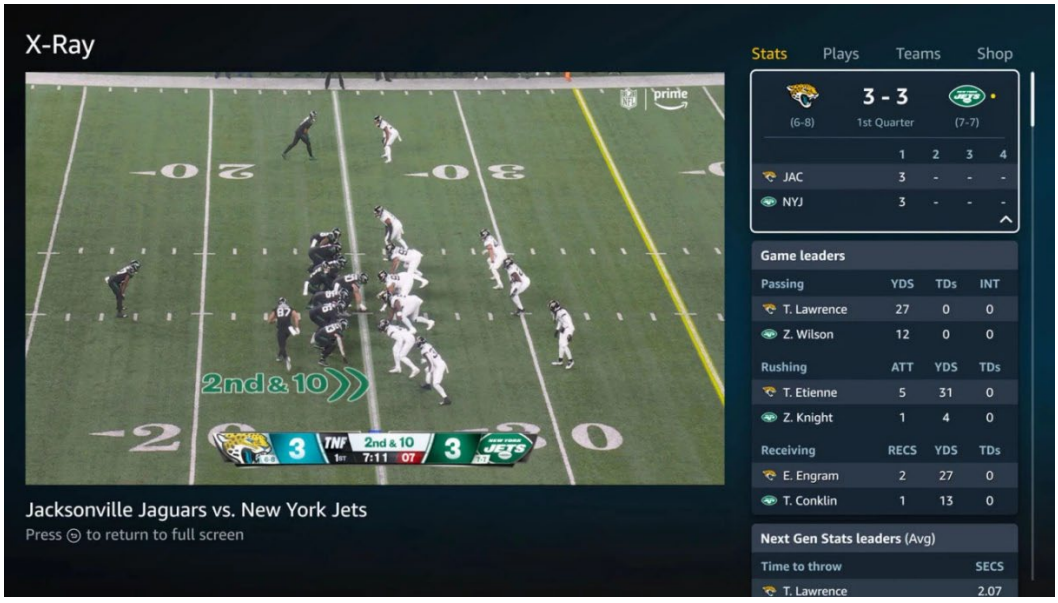

Figure 1. AWS X-Ray insight into Live NFL game data helps to enhance the viewer experience [20]

2) Real-Time Data Processing in AI: Concentrating on the application of machine learning for business intelligence and the extraction of real-time insights, this segment of the review is essential for formulating a framework that incorporates live NFL game data into the quiz platform, guaranteeing that the AI-generated content remains current and pertinent [3, 12, 16, 18].
3) Application of AI in Sports Analytics and Question Generation: The final thematic group is intimately connected to the principal application of the project. It discusses how prompt engineering can be harnessed to enhance the accuracy and specificity of questions generated by AI in sports analysis [1, 5, 9, 10, 11, 13, 15]. This theme is of particular interest as it addresses the verification of the

validity of questions generated for the quiz interface, and it also provides a practical perspective on the use of AI for question generation.

The literature informs the research question of how AI, specifically ChatGPT-4 Turbo, can be effectively prompted to generate precise, unique, and educationally valuable questions from NFL data with minimal human oversight. The insights across the themes support the hypothesis that with strategic prompt engineering and live data integration, ChatGPT-4 Turbo can autonomously craft quiz questions that are both accurate and adaptable to the real-time flow of NFL games, thereby enhancing the educational utility of AI in sports analytics.

A critical analysis of the literature on AI in educational technology reveals abundant ground for innovation coupled with challenges. The diverse applications of ChatGPT across domains suggest significant potential for personalized learning, though concerns about content validity and model reliability remain [2]. The pattern catalog for prompt engineering [1] provides a variety of strategies to address the issues faced by conversational AI, but also brings to light a significant gap: the lack of research on the integration of real-time data. This project seeks to bridge this gap by proposing a real-time data-driven question generation system employing ChatGPT-4 Turbo, aligning with the need for more agile AI applications in education.

Each literary source reviewed contributes a building block to the conceptual and methodological framework of this study. The evaluation of AI's advantages and limitations in educational settings [2] emphasizes the importance of this research, while the tutorial for creating quiz questions with ChatGPT [4] provides actionable methodological insights. The findings on prompt engineering [1, 10] are particularly applicable, guiding the development of prompts that will elicit accurate and contextually

relevant questions from ChatGPT-4. The investigation into AI hallucinations and inaccuracies [9] is also pertinent (see Figure 2), as it shapes the validation process necessary to ensure the reliability of the quiz questions generated.

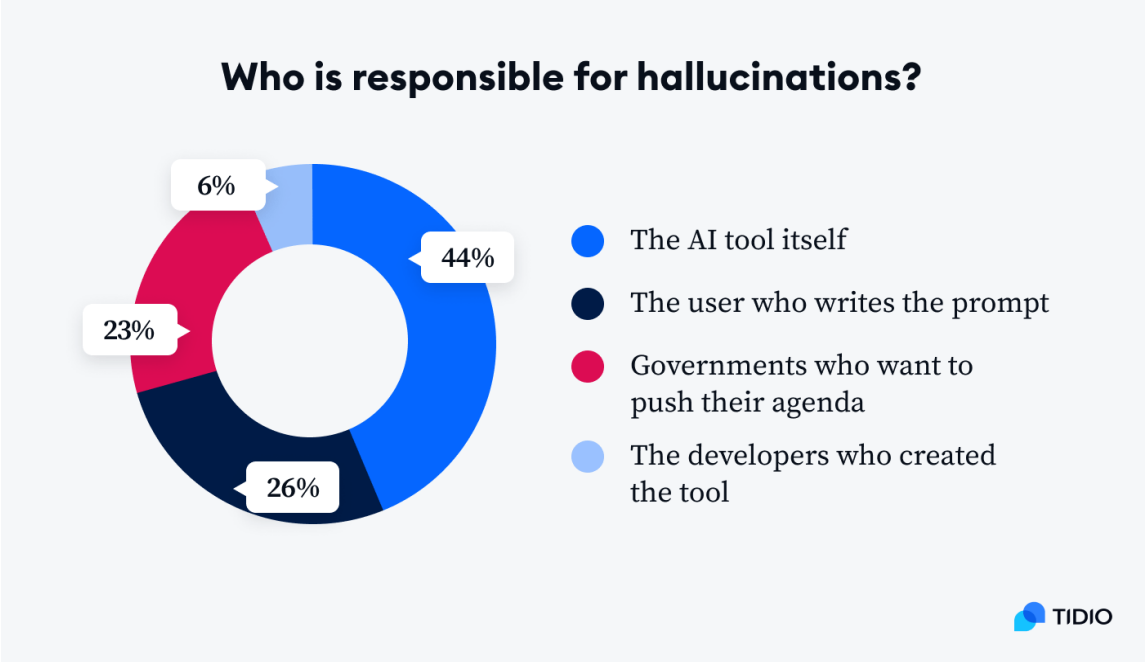

Figure 2. Survey on who the public believes is responsible for AI hallucinations [21]

The review highlights a notable gap in the literature regarding AI's capacity to process and leverage live data for educational content creation. While some studies have explored real-time data processing [3, 12], they have not specifically focused on educational quiz question generation. Moreover, the nuanced integration of live and historical sports data into educational AI applications is underexplored [11]. This project aims to address these gaps by developing a model that synthesizes live and historical NFL data, thereby creating a prototype for real-time, data-driven educational tools. By concentrating on the generation of quiz questions that reflect the nuances of live NFL games, this research contributes a novel application of AI in education to the existing body of literature.

In the realm of methodology, prompt engineering is paramount for the successful employment of language models in generating content. The Directive Pattern is crucial for eliciting specific information for quiz questions, but to foster more comprehensive responses, the integration of Example and Alternative Approaches Patterns is necessary to balance precision with creativity [1]. This research acknowledges these methodological nuances and aims to implement a hybrid approach, combining directive prompts for clarity with examples and alternatives to ensure a rich array of content.

The Reflection and Clarification Patterns are especially relevant to this project, ensuring that generated content is not only precise but also understandable and accessible to users. The proposed method involves iterative prompting, prompting the LLM to self-reflect on its answers for accuracy checks and to clarify complex subjects, thus enhancing the value of the content [19]. This strategy aims to circumvent the current methodological limitations of excessively complex or overly simplified explanations.

Patterns like the Refusal Breaker and Context Manager are indicative of the balance needed between directing the LLM to produce meaningful content and considering ethical implications. This study intends to employ these patterns judiciously, prioritizing the integrity of the information provided. The nuanced application of these patterns is designed to effectively navigate the model's limitations, thus enriching the question database with content that is not only accurate and varied but also ethically produced.

In the exploration of Natural Language Processing (NLP) applications for this capstone project OpenAI's ChatGPT-4, the latest iteration of the Generative Pre-trained Transformer models will be utilized. This decision is underpinned by several compelling attributes that render it particularly suitable for complex language tasks. Foremost,

ChatGPT-4 exhibits unparalleled linguistic fluency and versatility, stemming from its expansive training dataset and sophisticated architecture. It not only understands and generates human-like text but does so with a nuanced appreciation of context and subtleties. Unlike its predecessors and many contemporaries, ChatGPT-4 supports multimodal capabilities, meaning it can process both text and image inputs, thus offering a broader scope for research applications that require a multifaceted approach to human-computer interaction (see Figure 3). Additionally, its performance is enhanced by an adeptness in handling a vast array of subjects and domains, making it a highly adaptable tool for cross-disciplinary studies. The model's robustness in generating coherent and contextually relevant text is pivotal for the project's aim in developing an interface that can interact with users in an intuitive and informative manner.

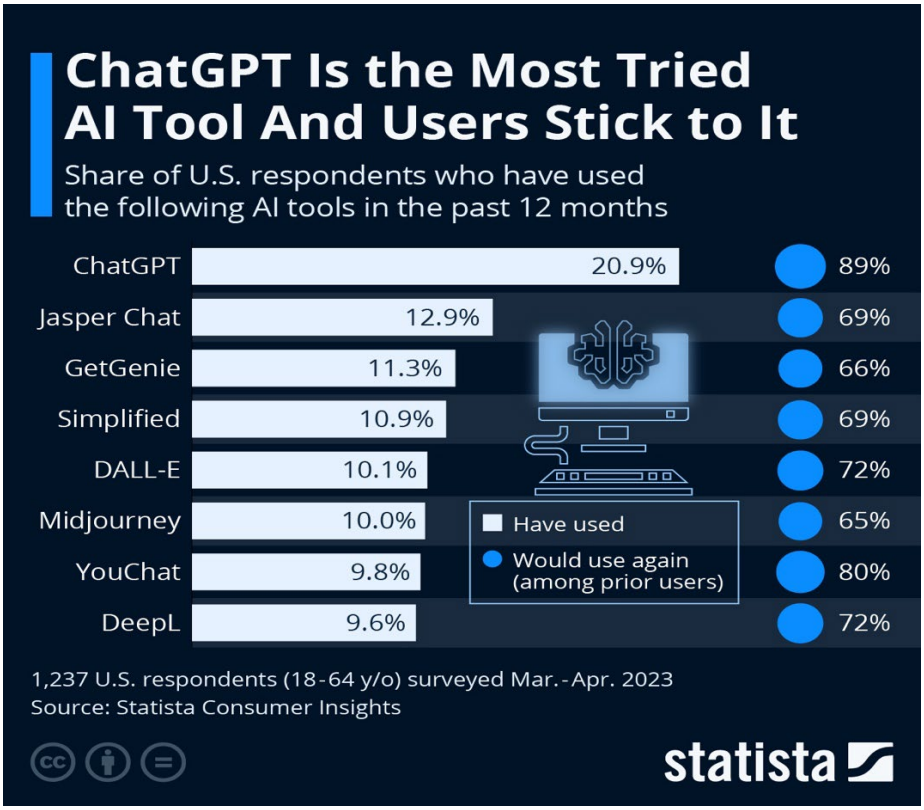

Figure 3. Most Tried AI Tool [22]

Furthermore, the choice of ChatGPT-4 Turbo is substantiated by its state-of-the-art performance in benchmarks for language comprehension and generation. It has demonstrated superior results in tasks such as language translation, question-answering, and text summarization, showcasing its capacity to understand and generate text with high accuracy and relevance. The ethical design considerations and mitigation strategies for biases implemented by OpenAI in the development of ChatGPT-4 Turbo also align with the ethical framework of this capstone project. The model's improved safety features and refined content filters ensure a responsible deployment of AI technology, which is critical in maintaining the integrity and credibility of the project's outcomes. Collectively, these features solidify ChatGPT-4 Turbo as the LLM of choice for this capstone project, as it offers a sophisticated, versatile, and ethically aligned platform for advancing research in human-centric AI interactions.

The comprehensive review of existing literature delivers the convergence of sophisticated language models such as ChatGPT-4 Turbo with the dynamic discipline of sports analytics, focusing on the creation of educationally enriching content that adeptly marries real-time and historical NFL data. This exploration, underpinned by rigorous selection criteria, unearths critical insights into prompt engineering practices, evaluates the expansive impact of AI across educational and medical fields, and delves into the capabilities and challenges of real-time data processing within AI frameworks. Distilled knowledge serves as a cornerstone for devising an innovative quiz interface, characterized by its precision, uniqueness, and educational value. By highlighting the potential and challenges of AI in educational technology, the necessity for real-time data processing, and AI’s application in sports analytics and question generation, the literature review underlines significant gaps—particularly in the nuanced integration of live data

for educational content creation. This synthesis not only informs the research's direction but also contributes to bridging identified gaps, proposing a model that leverages live and historical data to enhance AI's educational utility in sports analytics.

CHAPTER 3: METHODOLOGY

Research Design

The methodology of this research adopts an exploratory and iterative design, concentrating on harnessing the capabilities of ChatGPT-4 Turbo for generating NFL-related quiz questions. This study employs a mixed-methods approach, merging quantitative and qualitative research paradigms to forge a comprehensive understanding of prompt engineering's effectiveness when applied to an AI model. On the quantitative side, the research meticulously orchestrates the systematic generation of quiz questions, categorizing them based on their relation to either live or historical NFL data. These categories lay the groundwork for evaluating key metrics, including the accuracy, uniqueness, and definitiveness of the questions generated. Designed as a cyclical process, each round of prompt testing and evaluation feeds into subsequent iterations, progressively refining the prompts through a structured experimental approach. Statistical analysis plays a pivotal role in identifying the most efficacious prompt structures, aiming to enhance the quality of the questions produced. The study places a significant emphasis not only on the construction of the prompts but also on the implementation of a feedback loop, crucial for maintaining question generation quality and accuracy. This feedback loop specifically targets data derived from the Large Language Model (LLM) upon which ChatGPT-4 Turbo is based, highlighting the need for cost-effective strategies in managing the volume of live data utilized for prompts.

Complementing the quantitative analysis, the qualitative dimension of the study delves into the nuanced performance of the model, examining its adaptability and capacity for self-evaluation regarding the difficulty of the content it generates. This qualitative inquiry aims to provide deeper insight into the model's operational dynamics,

offering a layered understanding of its ability to navigate and reflect upon the challenges presented by varying levels of question complexity. Together, these methodological approaches create a robust framework for investigating the optimization of AI-generated educational content, with a particular focus on the integration of real-time and historical sports data. This comprehensive methodology ensures that the research not only identifies the most effective strategies for prompt engineering but also contributes meaningful insights into the practical applications and limitations of AI in educational technology.

Prompt Engineering

History Prompt. See Appendix A for list of history prompts tried during this experiment.

Live Data Prompt. See Appendix A for list of live data prompts tried during this experiment.

See Table 1 for information on both history and live data prompts.

The process of prompt engineering for creating NFL quiz questions is an iterative and meticulous method, aimed at refining queries to yield accurate and definitive answers. This methodology serves as the cornerstone of the experiment, leveraging structured prompt adjustments to optimize question clarity and precision. Initially, the process begins with the generation of a unique multiple-choice question, adhering to strict character and word limits for both the question and its potential answers. This step is crucial for maintaining concise, focused questions, a key aspect in quiz design. The quiz questions are categorized into two distinct types: historical and live data prompts. The historical prompt focuses on specific sub-topics related to an NFL team's history, while the live data prompt derives questions from significant plays in a recent game.

Table 1.
Prompt Engineering approach to refining the initial response.

Type	Prompt
Is Question Definitive? (Used for both Live and History Prompts)	Is the only correct answer to the question {question} the following: {answer}? Please only respond with Yes or No. Do not give further details in your response. If unsure in any way, please respond with maybe.
Can Question be Reworded? (Used for only History Prompts)	Yes or No? Can the following question be changed to have a single, definitive answer that matches the answer: {answer}? Question: {question}
Ask Again (Used for only History Prompts)	Change the following question so that it can have only a single, definitive answer of {answer}: {question} Only return the question, do not respond with anything else. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces before or after each question, options, and answer combination.
Is Answer Correct? (Used for only History Prompts)	Yes or No? The answer to the following question: {question} is: {answer} Please only respond with Yes or No. Do not give further details in your response. If unsure in any way, please respond with maybe.
Is The Answer Known? (Used for only History Prompts)	Yes or No? Do you know the answer to the following question: {question} Please only respond with Yes or No. Do not give further details in your response. If unsure in any way, please respond with maybe.
Provide Correct Answer (Used for only History Prompts)	Provide the correct answer to the following question: {question} Only return the answer, do not respond with anything else. Do not provide anything else in the response other than the requested information Do not number each question. Do not put spaces before or after each question, options, and answer combination.

The evaluation phase of this process involves determining whether each question generated has a single, definitive answer. If a question is deemed non-definitive, it is categorized as vague and subjected to further refinement. This refinement involves

rewording the question to ensure singularity in its answer, thereby enhancing the precision and clarity of the quiz. This rewording process is critical, as it ensures that each question adheres to the stringent criteria of being both informative and unambiguous. Questions that successfully undergo this transformation are then re-evaluated for their definitive nature.

The final step in this iterative process involves verifying the correctness of the answer to each definitive question. This verification is essential for ensuring the accuracy and reliability of the quiz content, which is paramount in educational and informational contexts. Once a question and its answer are validated, they are incorporated into a data table, forming a repository of high-quality quiz questions. This dynamic and adaptive approach to prompt engineering is pivotal in creating an efficient and effective quiz-generation system (see Figure 4 for a visual of practical example). The system's ability to self-evaluate and evolve through successive iterations underpins its effectiveness, making it a robust tool for educational and entertainment purposes in the context of NFL-related trivia.

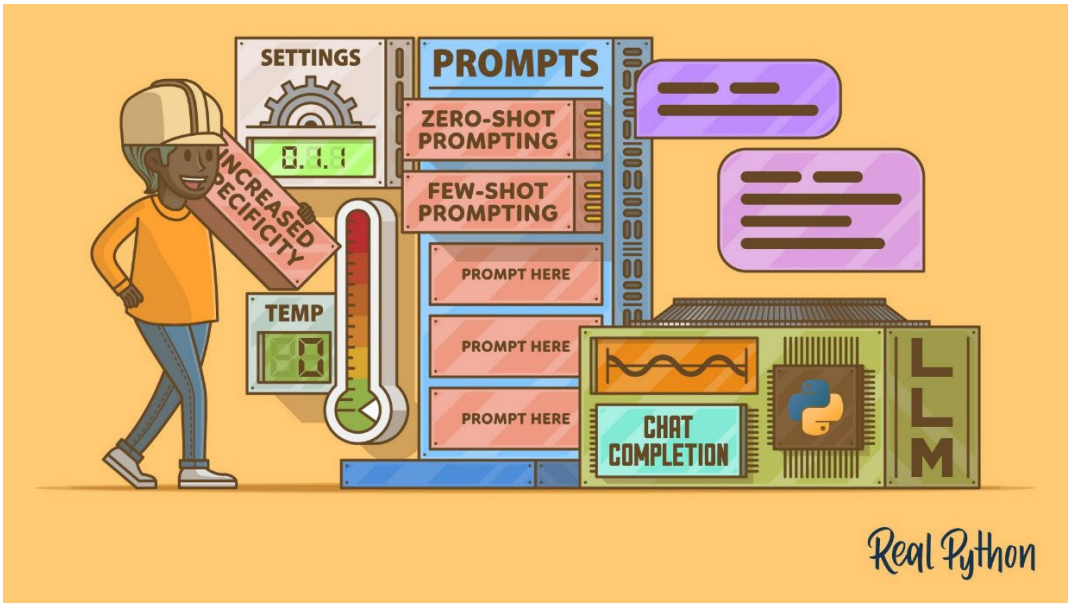

Figure 4. Prompt Engineering: A Practical Example [23].

Data Collection and Analysis

The methodology for data collection and analysis in this study hinges on a series of experiments designed to evaluate the performance of an AI system, specifically focusing on its ability to generate NFL-related quiz questions. The experiments will involve the generation of a substantial question bank, projected to contain between 100 to 200 questions per tested prompt, which will be stored across distinct database tables for systematic comparison and analysis.

Quantitative Data Collection

Quantitative data will be derived from the AI's output in response to varied prompts. The order these prompts will be executed are as follows:

- Start: Generate a history question prompt with four options, including the correct answer.
- Check Answer Validity: Verify if the correct answer is among the provided options.
    - If no, redirect to a Formatting Table as a finished state.
    - If yes, proceed to check for duplication.
- Check for Duplication: Determine if the question/answer combination exists in the database.
    - If yes, redirect to a Duplication Table as a finished state.
    - If no, proceed to assess the definitiveness of the question.
- Assess Definitiveness: Determine if the question has a single, definitive answer the way it is currently worded.
    - If no, prompt for rewording to ensure a single correct answer.
        - After rewording, check definitiveness again.

        - If still not definitive, redirect to a Vague Table as a finished state.
        - If definitive, proceed to accuracy check.
    - If yes, proceed to check the accuracy of the question/answer.
- Check Accuracy: Verify the accuracy of the question/answer combination.
    - If Not Accurate, prompt to know if the correct answer is known.
        - If no, redirect to an Accuracy Table as a final state.
        - If yes, ask to provide the correct answer and re-check accuracy.
            - If it is still inaccurate, redirect to the Accuracy Table as a finished state.
            - If accurate, proceed to the final state.
    - If Accurate, proceed to the final state.
- Final State: Deposit the question/answer into the History Question Data Table.

Each set of questions generated from a specific prompt will be scrutinized to determine the AI's efficiency in crafting accurate and unique questions. The rate of query reiteration due to similarity or ambiguity will be recorded, along with the total number of questions generated per prompt to ascertain the accuracy rate.

Prompt Variation and Analysis

The experimental protocol involves varying the prompts and conducting iterative testing cycles. Each cycle will gauge the AI's proficiency under different conditions, with approximately five to eight prompt styles planned for evaluation. The analysis will seek to identify a superior prompt style, subsequently adjusting it to optimize performance. The goal is to achieve a benchmark accuracy rate of 95% for historical data questions, reflecting a high confidence level in the AI's capability for question generation. For live

data questions, the target is set at a 99% accuracy rate to ensure reliability in real-time quiz or test creation applications.

Qualitative Data Collection

The qualitative aspect of data collection in this study focuses on the meticulous manual analysis of whether each generated quiz question accurately aligns with its designated data table, categorized as either pertaining to live or historical NFL data. This process involves a detailed examination of the context, relevance, and precision of each question, ensuring that it not only reflects the specific data category it is intended to represent but also adheres to the criteria of accuracy and educational value. Through this qualitative scrutiny, researchers will assess the nuances of content generation, identifying patterns, discrepancies, or biases that may emerge from the AI's processing and integration of the data. This hands-on approach allows for a nuanced understanding of the model's performance, highlighting areas where prompt engineering may need adjustment to better guide the AI in producing content that is both relevant and reliable. By systematically categorizing and evaluating the questions, the research gains deeper insights into the efficacy of the AI model's ability to discern and appropriately classify content, thus contributing significantly to the refinement of the question generation process.

Ethical Considerations and Data Integrity

All data collection will adhere to strict ethical standards. Confidentiality and anonymity will be guaranteed, and secure, encrypted platforms will be utilized for data storage and processing to safeguard participant privacy and data integrity. Through this multifaceted data collection methodology, the study aims to construct an intricate understanding of the AI’s effectiveness in generating quiz questions tailored to NFL

knowledge. The integration of rigorous quantitative analysis with in-depth qualitative feedback will provide a robust framework to evaluate the AI's capacity to assist in educational and quiz-based environments.

Technological Architecture

The technological architecture of this research project is strategically designed to exploit the capabilities of ChatGPT-4 Turbo, facilitating the generation of NFL quiz questions that incorporate both real-time events and historical data. This architecture supports complex data analysis, sophisticated prompt engineering, and seamless AI interactions, ensuring the generation of questions with the desired level of accuracy and relevance.

Central to the system's design is the integration of the ChatGPT-4 Turbo API, deployed on a cloud computing platform chosen for its scalability and high availability. This cloud environment is essential for dynamically managing the computational load, accommodating the influx of real-time NFL game data alongside extensive historical datasets. The platform provides essential features such as data redundancy, automated disaster recovery, and the ability to integrate advanced analytics services, ensuring robustness and reliability in data handling.

On the software development front, the project employs Python and Flask for the backend infrastructure, leveraging Python's extensive library ecosystem and Flask's flexibility in creating web applications. This choice allows for the utilization of object-oriented programming principles, facilitating modular design and efficient data management. A RESTful API architecture enables server-side interactions between the Flask application and the ChatGPT-4 Turbo API, handling operations such as sending quiz prompts, receiving AI-generated questions, and managing user responses. This API

design ensures system scalability and simplifies the integration of additional features or data sources in the future.

The database subsystem is designed for high throughput and scalability, employing a hybrid approach that combines SQL and NoSQL databases to capitalize on their respective advantages in data structuring and scalability. Security measures, including SSL/TLS encryption for data in transit and anonymization of sensitive information, are integral to the system, ensuring user privacy and data integrity. Additionally, an analytical module equipped with machine learning algorithms works alongside ChatGPT-4 Turbo to analyze user performance and question generation patterns. This module plays a critical role in continuously refining the AI's prompt engineering strategies and question generation capabilities, focusing on the accuracy, uniqueness, and difficulty levels of the quiz content.

This research project's architecture is not only designed to meet the immediate objectives but is also flexible and adaptive, ready to accommodate future advancements in AI and data analytics. Through iterative development and a focus on responsive design, the project aims to set a benchmark in the application of AI for educational content generation, particularly in the context of integrating live and historical sports data.

Variables and Hypotheses

The proposed research is structured around the optimization of real-time and historical data integration for accurate NFL quiz question generation using ChatGPT-4 Turbo. A key component of the study is to identify and analyze specific variables that are hypothesized to influence the generation of quiz questions by this model. The variables are categorized into independent and dependent, with the primary hypothesis focusing on

the impact of prompt engineering and programmatic direction of the accuracy and adaptability of the generated questions.

Independent Variables

- Prompt Specificity (IV1): The precision and detail incorporated within the prompts given to ChatGPT-4.
- Prompt Alteration (IV2): The modifications in the prompt structure to optimize the question generation process.
- Real-time Data Integration (IV3): The inclusion of live game data within the prompts to produce current and relevant quiz questions.
- Historical Data Integration (IV4): The use of pre-existing data within ChatGPT-4's knowledge base to formulate questions pertaining to past events.

Dependent Variables:

- Question Accuracy (DV1): The extent to which the generated questions correctly reflect NFL knowledge, both current and historical.
- Question Definitiveness (DV2): The clarity of the generated questions, assessed by their ability to produce unambiguous answers.
- Question Uniqueness (DV3): The degree to which new questions are distinct from previously generated ones, avoiding repetition and redundancy.
- Difficulty Appropriateness (DV4): The alignment of the generated questions' difficulty level with the criteria specified in the prompts.
- Order of Systematic Checks (IV5): This will ensure that each check happens at the correct time and only when deemed necessary.

Hypothesis:

The central hypothesis is that strategic prompt engineering, encompassing

specificity and careful alteration, when integrated with real-time and historical NFL data, will enable ChatGPT-4 Turbo to generate quiz questions with high accuracy and definitiveness. It is expected that the careful calibration of prompts will result in a statistically significant improvement in the quality and reliability of the questions generated by ChatGPT-4. This hypothesis will be tested through systematic variation of the independent variables and observation of their impact on the dependent variables. The research anticipates establishing a strong correlation between the independent variables and the quality metrics of the generated questions, with a targeted accuracy threshold of 95% for historical data questions and 99% for real-time data questions.

Limitations

One significant limitation is the dependency on the ChatGPT-4 Turbo model, which, despite its advanced capabilities, is limited by the scope of its training data and the underlying algorithms that drive its predictions. The model's performance is further subject to continuous updates and improvements by OpenAI, which could unpredictably alter its functionality during the course of this investigation, introducing a challenging variable to control.

Achieving high accuracy in generating NFL-related quiz questions is a central goal; however, the inherent variability in interpreting and utilizing live data could adversely affect the consistency and reliability of the question generation process. Additionally, the ambitious scope of this endeavor, when coupled with a constrained timeline, might necessitate the prioritization of certain areas of analysis over others, potentially limiting the depth and breadth of the investigation.

The technical infrastructure, designed to process and integrate real-time data, faces challenges due to the unpredictability of live sports events and the need for a robust

system capable of handling such data seamlessly, without lag or errors. This could introduce technical limitations that detract from the user experience and diminish the overall quality of the quiz questions produced. Moreover, there is a risk of bias within the user feedback mechanism, which might distort the perceived difficulty of questions if the participant sample is not adequately representative of the broader NFL fan base.

Finally, ethical considerations surrounding the generation and use of AI-produced content, coupled with potential concerns over the misuse of NFL-related data, highlight the importance of navigating intellectual property and privacy issues carefully. These limitations underscore the need for a nuanced approach to developing and deploying AI-driven educational tools, ensuring they are both effective and ethically sound.

Conclusion

In conclusion, this capstone project centered on utilizing ChatGPT-4 Turbo for creating real-time and historical NFL quiz questions presents a novel intersection of AI and sports analytics. The mixed-methods approach capitalizes on the advanced capabilities of language models to interpret vast datasets, offering a dynamic quiz generation tool that could significantly enhance the interactive experience of NFL enthusiasts. While the project sets ambitious accuracy targets for the AI-generated questions, these are underpinned by a thorough process of prompt engineering and iterative testing that promises to fine-tune the AI's performance to a high degree of precision. The integration of live game data into the question generation process represents an innovative step towards real-time educational content delivery, which could serve as a benchmark for future applications of AI in sports and entertainment.

The project's reliance on the evolving capabilities of ChatGPT-4 Turbo introduces an element of unpredictability that must be carefully managed. Continuous updates to the

model could affect consistency, requiring adaptability in research methods and potential recalibration of the quiz generation process. Moreover, the project's temporal limitations underscore the challenge of conducting comprehensive research within a constrained timeframe, possibly necessitating a focused approach that balances depth with breadth of analysis.

Despite these challenges, the project embodies a pioneering spirit in its application of AI to create an engaging and informative platform for NFL fans. It stands to offer valuable insights into the efficacy of LLMs in real-time data synthesis and educational content generation, with implications that extend beyond sports trivia. By documenting and analyzing each step of the process, and with oversight from a knowledgeable board, the project is poised to contribute meaningfully to the discourse on the integration of AI into daily entertainment and learning experiences. It is a stride towards an era where AI not only informs but also actively engages with users in enhancing their understanding and enjoyment of the sports they love.

CHAPTER 4: COMPLETED PROJECT AND EXPERIMENT RESULTS

Trial and Error

The experimental framework for this project was carefully crafted during the proposal phase, laying a solid foundation through the development of a series of Python functions designed to gather and process the necessary data for conducting the experiment. This data encompassed a wide array of information crucial for the generation of NFL-related quiz questions, including the 2024 NFL schedule, detailed play-by-play information for each game of the season across all teams, team names along with their corresponding IDs, season matchups, and the meticulous organization of each prompt to ensure seamless execution within the token limitations of the OpenAI API. The system architecture incorporated error checks, multiple class files housing diverse functions, and a dynamically updated database to store the essential data for thorough analysis. The program itself underwent several iterations to pinpoint the optimal combination of components and execution strategies that would yield the expected behavior.

The search for the most effective prompts involved testing and analyzing three distinct versions, each marked by varying levels of prompt specificity and approaches to verification checks for accuracy concerning both live and historical data. The first version explored four tiers of prompt specificity, escalating in detail to ascertain its impact on the quality and accuracy of the generated questions, including checks for vagueness and duplication. However, it did not attempt to restructure questions flagged as vague or inaccurate. The second version, recognizing the prohibitive cost of verifying live data, shifted focus to generating batches of 10 questions per team until reaching the desired quota, thereby reducing the incidence of duplicate questions. This iteration experimented with three prompt modifications to statistically evaluate the impact on question quality.

By the third version, the process had been refined to include re-prompting as a critical element of the experimental mechanism, establishing that the level of prompt detail did not significantly influence accuracy. Hence, a base prompt was utilized, focusing solely on eliminating question duplication.

The third version's methodology for handling historical data was particularly nuanced, involving a comprehensive cycle that started with generating a history question prompt, followed by checks for answer validity, duplication, definitiveness, and accuracy. This process ensured that each question had a single, definitive answer and that the information was accurate before deposition into the History Question Data Table. Through trial and error across these versions, accompanied by iterative code adjustments, the experiment honed in on the third version as the most effective means of collecting data for final analysis. This evolution illustrates the project's iterative nature, emphasizing continuous refinement and adaptation to overcome the challenges encountered in achieving the precision and relevance of NFL-related quiz questions.

This iterative trial and error approach not only underscored the complexity of integrating real-time and historical NFL data into an AI-driven quiz generation system but also illuminated the path toward refining the experiment's methodology. By navigating through various versions and adapting the system's architecture, the project demonstrated a commitment to optimizing the balance between prompt specificity, cost-efficiency, and the accuracy of generated content. The evolution from initial explorations to a finalized, effective version encapsulates the essence of adaptive research, where theoretical plans meet practical challenges. This process, characterized by meticulous data handling, prompt engineering, and iterative refinement, has set a robust foundation for the subsequent phases of the experiment, ensuring that the generated quiz questions

meet the high standards of accuracy, relevance, and educational value envisioned at the outset.

In examining the prompt engineering process, significant attention was given to refining the prompts to enhance the accuracy of the AI-generated questions. The detailed methodology and iterations of prompt modifications can be found in Appendix A, which outlines the comprehensive prompt engineering process undertaken in this study (see Appendix A: Every Initial Prompt Used in Each Trial Run).

Observations During the Experiment

Throughout the various iterations of the experiment, a series of observations were made that significantly informed the refinement process of generating NFL-related quiz questions using ChatGPT-4 Turbo. In the initial version, it became evident that implementing checks across all data tables—not just the final one for duplicates—was crucial to reducing the number of duplicated questions. This iteration also highlighted the necessity of incorporating detailed game information for live data to facilitate post-generation question verification. A considerable number of questions were flagged by the accuracy table, primarily due to vagueness, indicating that initial checks were insufficient in identifying and filtering out imprecise wording. This observation underscored the importance of enhancing verification mechanisms to minimize the inclusion of inaccurate data in both live and historical question banks. The discovery that additional details, such as the week a game was played, could improve live data verification further illustrated the complex requirements for ensuring question accuracy and relevance.

The second version of the experiment yielded insights into the generation process for live questions, which, while vague, did not exhibit repetition and processed more swiftly when generating multiple questions from a single game. This phase also

underscored the challenges of database refreshment and error handling, particularly in generating historical questions where specific prompts did not always result in the expected output. Notably, live data prompts often failed to return complete information as requested and attempts to specify certain types of "big plays" did not consistently align with the generated content. Historical questions tended to repeat across teams, suggesting a pattern in the AI's question generation process that leaned towards similar historical events, thereby limiting the diversity of the quiz content.

In the final iteration of the experiment, significant advancements were made in streamlining the data collection process, notably through the introduction of a formatting table designed to filter out questions without a correct answer option. This development marked a key improvement in ensuring the integrity of quiz content. However, a persistent challenge was observed in the AI's limited ability to self-correct inaccuracies or vague phrasings, despite attempts to guide it towards refinement. This highlighted a critical gap in the model's capacity for iterative learning and adjustment based on specific feedback. Additionally, the AI struggled with accurately generating year-specific questions, often confusing the season of an NFL event with its actual calendar year. This revealed nuanced difficulties in the AI's understanding of temporal data, underscoring the complexity of creating precise and contextually accurate sports trivia questions. These observations from the final version underscore the intricate balance required to leverage AI capabilities effectively while addressing its limitations in processing and applying nuanced data accurately.

General observations across all versions highlighted several challenges in achieving consistency and accuracy in the question generation process. The AI struggled with distinguishing between general plays and significant, game-changing "big plays,"

indicating a gap in its understanding of nuanced sports terminology. Furthermore, without specific subtopics or details, the AI tended to revert to a set of preferred questions, reducing the diversity and specificity of the quiz content. Inconsistencies were also noted in the AI's ability to use player names correctly, fluctuating between full names and abbreviations across different runs.

In conclusion, the experiment's trial and error approach revealed critical insights into the capabilities and limitations of using ChatGPT-4 Turbo for generating NFL-related quiz questions. These observations underscore the need for precise prompt engineering, robust data verification mechanisms, and adaptive strategies to address the model's inconsistencies and biases. While the AI demonstrated potential in creating diverse and engaging content, the findings highlight the importance of continuous refinement and adjustment to leverage its full capabilities effectively. Future iterations of this project will benefit from these learnings, paving the way for more accurate, consistent, and nuanced quiz question generation (see Table 2 for a list of sample questions generated during the experiment).

Experiment Data Collection and Analysis

The data collection process for this experiment was conducted in three distinct phases, each with its own set of methodologies and outcomes that contributed to the refinement of the experimental design. In the initial phase, the data collection from ChatGPT-4 Turbo's API did not reveal a significant difference in the quality of questions generated across various prompts. During this phase, emerging reports suggested that ChatGPT-4 exhibited signs of 'laziness' or 'contention' during API interactions. Upon a preliminary review, it was hypothesized that these reported behaviors could have influenced the data quality. Subsequent manual analysis confirmed that the data set might

have been compromised by these API response issues, leading to the decision to exclude this initial data set from further analysis due to its potential unreliability.

Table 2.

Sample Questions Generated During Experiment Across Each Version

	Desired Generated Question	Question That Needs Additional Refinement
Version 1	Who holds the Pittsburgh Steelers franchise record for most rushing yards in a career?	Who had a 57 yard completion in the game between LAC and DEN leading to a significant gain for LAC?
	Who scored a touchdown with a 14-yard rush in the first quarter of the game?	Who was the opposing team when the Arizona Cardinals made their only Super Bowl appearance in 20098? (Year was 2009)
Version 2	Which Bears player was known as "The Refrigerator"?	Whose punt did B.Covey return from the PHI 16, getting pushed out of bounds by N.Bellore?
	Who made the catch for a touchdown pass by J.Dobbs towards the end of the 4th quarter?	In what year did the Lions make Calvin Johnson the highest-paid receiver in NFL history at the time?
Version 3	What year was the Dallas Cowboys' famous "Hail Mary" play?	Who returned the kickoff for the Green Bay Packers in Quarter 1?
	Who sacked Z.Wilson leading to a fumble recovered by A.Gilman at NYJ 50?	When were the New Orleans Saints founded?

The second phase of the data collection process was designed to enhance both efficiency and cost-effectiveness by revising the approach to question generation. Rather than producing questions individually, the procedure was modified to process batches of four questions per team, thereby optimizing the allocation of ChatGPT-4 Turbo's computational resources. This modification was deemed essential, especially for processing live data, where the extensive token consumption associated with verifying prompts individually proved to be economically impractical. Despite these procedural adjustments, a detailed analysis of each question's validity indicated that the collected data lacked statistical significance, prompting a reevaluation of data collection

methodology. Throughout three iterations of this second phase, there was a marginal increase in the accuracy of responses within the final historical data table, demonstrating a gradual improvement. However, given the small sample size (100 questions per iteration) and the minimal statistical variance (approximately a 2% difference between each iteration), it became evident that altering the prompt had a negligible impact on the accuracy of generated responses. Instead, the effectiveness of the approach was more significantly influenced by the ability to programmatically identify and disregard incorrect questions.

Statistical analysis, complemented by graphical representation, demonstrated progressive enhancement in the accuracy of NFL history questions generated through successive refinements of the prompts. The evolution from Prompt 1.2 through to Prompt 3.2 yielded a notable, albeit incremental, increase in correct responses: from 72 to 75 out of a total of 80 questions. This gradual improvement was substantiated by T-tests (see Figure 5), with the transition from Prompt 1.2 to Prompt 2.2 revealing a highly significant p-value of approximately 1.6×10^{-6}, and the subsequent shift to Prompt 3.2 maintaining statistical significance with a p-value of 0.0073. These findings suggest that enhancements in prompt detail can positively influence the generation of accurate answers, highlighting the statistical significance of prompt modifications in achieving incremental gains in response accuracy.

However, the practical implications of these statistically significant improvements invite careful consideration, especially given the modest increase in the percentage of correct answers (from 90.24% to 93.33%) and the small sample size of 80 questions per prompt (See Appendix B for full Version 2 data analysis). The marginal gains in accuracy, while statistically valid, raise questions about their substantiality in applied

settings, where the costs and efforts associated with implementing detailed prompts must be balanced against the benefits. Furthermore, the susceptibility of small-sample results to variance underscores the need for caution in generalizing these findings across broader data sets or different contexts. Thus, while the enhancements in prompt detail are statistically validated to affect answer accuracy, the practical significance of these changes remains a nuanced issue, warranting a judicious assessment of their value against the backdrop of specific research objectives and resource availability.

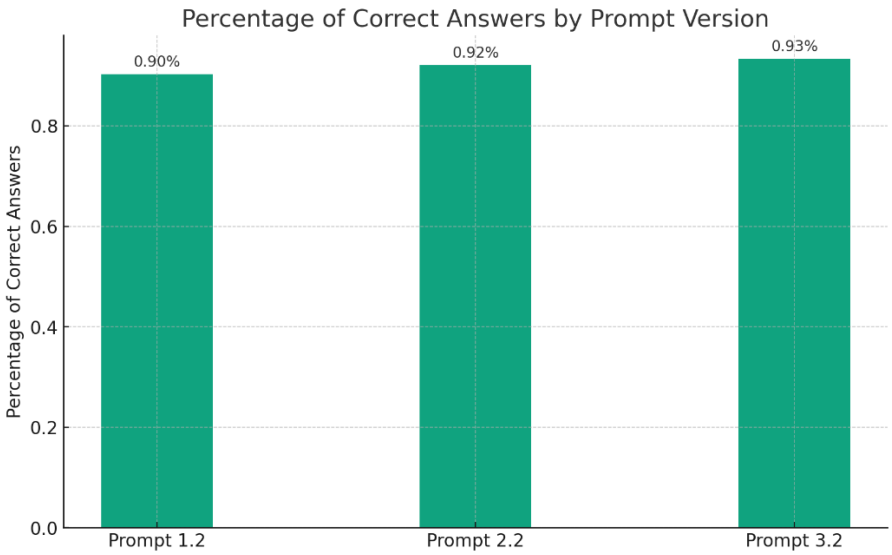

Figure 5. Results of Version 2 Data Collection T-Test [24]

In the final phase, the focus shifted to directly comparing the output of raw, unverified data with data subjected to a series of verification checks as outlined earlier in the study. The data was limited to just two comprehensive runs to ensure manageability and focus. Live data, consistent with findings from earlier phases, was excluded from further verification due to its inherent accuracy, thus not warranting additional scrutiny. The emphasis was placed on historical data, examining the disparity in the number of correctly generated questions between the unverified and verified sets. This approach

aimed to provide a clear comparison of the efficacy of the verification process on the quality of the generated historical content.

The juxtaposition of raw and verified historical data sets offered an insightful perspective on the role of verification in enhancing the AI's performance. It underscored the necessity of such checks in the pursuit of high-quality, reliable quiz questions generated by AI, particularly when dealing with complex historical data. These structured observations and the resultant data collection strategy set the stage for a robust analysis, ensuring that the findings would contribute meaningfully to the understanding of AI capabilities and limitations in educational content generation.

The dataset comprised two groups of 200 questions each. The first group, serving as the control, contained questions generated without the application of strategic prompt engineering. The second group consisted of questions generated following the strategic engineering of prompts. The accuracy of the generated questions was quantified in terms of percentages, with the non-engineered prompts resulting in 82.00% correct answers and the engineered prompts achieving a significantly higher accuracy of 94.06%. See Figure 6 for results comparing verification to no verification.

To determine the statistical significance of the observed improvement, a t-test analysis was utilized, which yielded a t-statistic of -3.93 and a p-value of 0.0001. This outcome strongly suggests that the observed difference in accuracy between the two groups is not due to chance, thereby supporting the efficacy of strategic prompt engineering in enhancing question accuracy. The t-statistic is a measure of the difference between the two groups relative to the variation within the groups. A negative t-statistic indicates that the experimental group (with verification) scored higher than the control group (without verification). The p-value quantifies the probability of observing such a

difference (or more extreme) if there were no real difference between the groups. A p-value below 0.05 is typically considered statistically significant, indicating strong evidence against the null hypothesis (which assumes no difference between the groups).

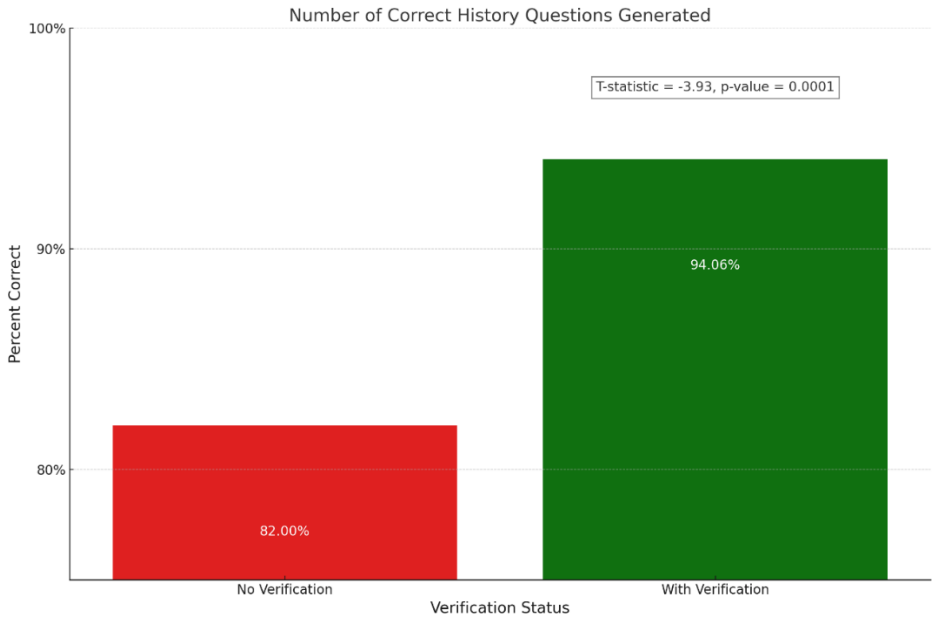

Figure 6. Results of Version 3 Data Collection T-Test [24]

In the analysis of data related to the accuracy challenges of ChatGPT 4.0 Turbo, recurrent themes emerged. Among questions based on historical data, the primary issue identified was the determination of the year in which specific events occurred. Such inquiries frequently involved identifying the year of a team's Super Bowl victory, the inaugural year of a franchise's entry into the NFL, and the year in which a player received an MVP or other awards. For live data, the prevalent error involved incorrectly naming the player responsible for executing a particular play, encompassing queries about the player who sacked the quarterback, made a catch, or scored a touchdown. These topics emerged as particularly problematic. The analysis suggests that addressing these common inaccuracies and implementing a verification mechanism for the responses could potentially enhance the system's accuracy beyond 94%. Figure 7 presents a

comprehensive list of topics and questions identified through manual data inspection as being inaccurate or incorrect.

Live Data																														Total Count
Incorrect Player	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x														16
Incorrect Play	x	x	x	x	x	x	x																							7
Incorrect Team	x	x	x	x																										4
Play Never Happened	x	x																												2
Answer Not an Option	x																													1
Historical Data																														
Incorrect Year	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	29
Incorrect Player/Coach	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										21
Incorrect Name (i.e. Team or Stadium)	x	x	x	x	x	x	x																							7
Incorrect Color Scheme	x	x	x																											3
Incorrect Total (i.e. MVPs)	x	x	x																											3
Not A Valid Question	x																													1

Figure 7. Data Table Representing Categories of Incorrect Question/Answer Combinations from V2 and V3

The results of this analysis affirm the central hypothesis, demonstrating a clear and statistically significant improvement in the accuracy of ChatGPT-4 Turbo-generated questions with the implementation of strategic prompt engineering. This finding underscores the importance of careful prompt design, particularly the integration of domain-specific knowledge, in optimizing the performance of AI in content generation tasks. It illustrates that by harnessing detailed, domain-specific data and incorporating it into the design of prompts, it is possible to significantly elevate the quality and reliability of the output produced by AI systems like ChatGPT-4 Turbo.

The data provides compelling evidence that strategic prompt engineering, characterized by specificity and careful alteration based on real-time and historical data, markedly improves the accuracy of quiz questions generated by AI, specifically ChatGPT-4 Turbo, in the context of NFL-related content. This advancement not only validates the hypothesis but also highlights the potential for similar strategies to enhance AI-generated content across various domains, paving the way for further research and application in educational and informational technologies.

Analyzing Results

In the examination of methodologies to enhance the accuracy of AI-generated quiz questions, particularly within the context of NFL-related content, a comprehensive experimental framework was established. This research aimed to identify strategies that significantly improve question accuracy beyond mere adjustments in prompt specificity. A pivotal finding from this study underscores the profound impact of rigorous question and answer verification processes, implemented programmatically, in elevating the precision of generated content. These mechanisms proved paramount, surpassing the incremental accuracy gains attributed to nuanced prompt modifications.

The experiment revealed the efficacy of generating questions in batches of ten per team, a strategy that not only optimized cost-efficiency but also markedly reduced the occurrence of duplicate questions. This approach, contrasted with the less efficient and more costly method of generating questions one at a time, highlighted the importance of scalability and efficiency in the data collection process. The strategic implementation of verification checks emerged as a critical factor, significantly enhancing the reliability and educational value of the content produced. These findings illuminate the intricate balance required in AI-driven content generation, where the integration of sophisticated verification mechanisms and the efficient structuring of data requests contribute to the achievement of high-quality, accurate, and engaging educational materials.

In the exploration of enhancing the accuracy of AI-generated NFL-related quiz questions, it became evident that the verification of live data sets, especially those necessitating extensive token consumption, did not substantially improve accuracy. This revelation marked a pivotal shift in strategy, as the considerable financial and temporal resources required for such verification were deemed an ineffective allocation. The

lengthy generation times associated with processing large token counts further compounded the inefficiency, leading to the conclusion that focusing on the elimination of duplicate question-answer pairs and ensuring proper formatting constituted a more pragmatic approach. This adjustment in methodology underscored the necessity of optimizing resource use while maintaining content quality.

The challenge of generating live data questions on NFL topics unveiled a notable obstacle: attaining a high accuracy rate, ideally around 99%, was hampered by the AI's limited understanding of the subject matter (see Figure 8). This issue was particularly stark when dealing with live data, where an in-depth grasp of context and nuance is crucial. As the specificity and detail of the prompts increased, the AI's ability to verify the accuracy of the questions it generated diminished in parallel. The task of capturing the essence of live NFL events with precision, alongside the AI's restricted capacity to navigate the complexities of the content, led to reduced accuracy levels. This situation highlighted the complex challenges of utilizing AI for real-time content generation in areas that demand a profound contextual understanding, shaping the strategic focus towards enhancing the uniqueness and structural integrity of questions over exhaustive verification of live data.

The final iteration of the question generation test leveraging OpenAI's ChatGPT-4 Turbo revealed notable advancements in the accuracy of generated content, particularly when employing enhanced prompt specificity and integrating a robust validation process for the data. This meticulous approach yielded a substantial improvement, elevating the accuracy rate from 82% across a baseline of 200 questions—with no verification—to an impressive 94.06% (See Appendix B for full Version 3 data analysis) when both refined prompt generation techniques and programmatic verification of question/answer

combinations were applied. Despite this significant progress, the results fell short of achieving complete autonomy in generating flawless question/answer pairs, indicating that the system cannot yet operate without human oversight.

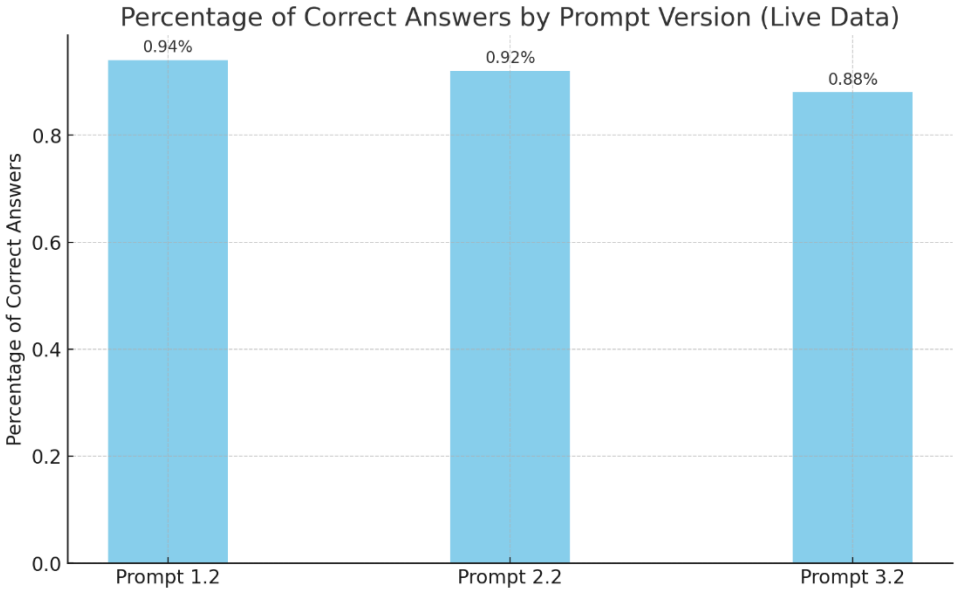

Figure 8. Results of Version 2 Live Data Collection [24]

The enhancement of question clarity and the mitigation of vagueness are crucial components in the process of refining AI-generated quiz questions, particularly in the context of NFL-related content. This research illuminated the significance of identifying and rectifying questions with vague construction, a process that was instrumental in improving the overall quality and accuracy of the question bank. Vague questions often stem from imprecise language, ambiguous references, or the lack of specificity, which can confuse respondents or lead to multiple interpretations. By implementing a programmatic approach to detect and revise such questions, the system was able to significantly reduce ambiguity, thereby making each question more direct and understandable. This focus on definitiveness not only contributed to the enhanced accuracy of the content but also facilitated a more engaging and educational experience

for users. The ability of ChatGPT-4 Turbo to self-detect these issues and automatically refine the questions underscores the advanced capabilities of AI in producing high-quality educational materials, demonstrating a marked improvement in the way content is generated and validated.

Furthermore, the detection of accuracy issues plays a pivotal role in ensuring the reliability of the content produced. Accuracy detection within the program involves a meticulous examination of the factual correctness of each question and its corresponding answer, a process that is vital for educational content where precision is paramount. Through the programmatic identification of questions with potential accuracy flaws, the system was empowered to correct or discard inaccuracies, thereby elevating the integrity of the question bank. This mechanism not only improved the factual correctness of the questions but also contributed to the overall trustworthiness of the content. The dual focus on eliminating vague constructions and enhancing accuracy exemplifies the comprehensive approach taken to optimize content quality. By addressing these two critical aspects, the research showcases the potential of AI, particularly OpenAI's ChatGPT-4 Turbo, to revolutionize the creation of educational materials, ensuring that they are not only engaging but also rigorously accurate and clear, thereby setting a new standard for AI-assisted educational content development.

This level of accuracy, while commendable, underscores the necessity for human intervention to ensure the utmost quality and reliability of the content produced. In an educational context, where the generation of quiz and test questions from textbook material or class lectures is visualized this technology presents a promising tool. It has the potential to streamline the creation process, offering educators a rich repository of questions that can be further tailored to match their teaching objectives. However, the

current limitations necessitate that instructors meticulously review and possibly revise the AI-generated content to align with the desired educational standards and learning outcomes.

The findings highlight a critical balance between leveraging cutting-edge AI capabilities to enhance educational content creation and the imperative of maintaining academic rigor through human expertise. While OpenAI's ChatGPT-4 Turbo exhibits a remarkable capacity to generate content with a high degree of accuracy, the quest for total autonomy in this domain remains unfulfilled. As such, this technology should be viewed as an auxiliary tool—a means to augment the educational content development process rather than replace the nuanced judgment and expertise of educators. This approach ensures that the generated material not only serves the educational objectives but also adheres to the quality standards requisite for academic excellence. In summary, while the tool offers substantial benefits, achieving a state of complete automation in question generation without compromising quality remains a goal for the future.

Conclusion

The culmination of this research project, aimed at harnessing OpenAI's ChatGPT-4 Turbo to generate NFL-related quiz questions, has illuminated the nuanced interplay between technological innovation and the rigorous demands of educational content creation. Through a meticulous experimental process, this study has navigated the complexities of integrating AI to achieve a level of precision and relevance in quiz question generation that aligns with educational standards. The journey from initial concept to final execution, characterized by a series of iterative trials and adjustments, has not only showcased the potential of AI in educational content development but also underscored the inherent challenges that accompany such endeavors.

The experimental phases, each designed to refine the question generation process, have collectively contributed to a deeper understanding of the capabilities and limitations of ChatGPT-4 Turbo in this context. The evolution of the project's methodology, particularly the shift towards batch processing of questions and the strategic implementation of verification checks, has underscored the importance of efficiency and accuracy in leveraging AI technologies. Despite the significant improvements in question accuracy, the project's findings reveal that achieving complete autonomy in AI-driven content generation remains an elusive goal. The nuanced nature of NFL trivia, coupled with the intricacies of ensuring question accuracy and relevance, necessitates a collaborative approach where AI's computational prowess is complemented by human expertise.

The insights gained from this research have profound implications for the future of educational content generation. The ability of AI to significantly streamline the creation of quiz and test questions, especially when equipped with strategic prompt engineering and robust verification mechanisms, offers a glimpse into the transformative potential of technology in education. However, the persistent need for human oversight, to refine and validate AI-generated content, highlights the critical role of educators in maintaining the integrity and quality of educational materials. This partnership between AI and human expertise embodies a pragmatic approach to educational innovation, where technology serves as a tool to enhance, rather than replace, the pedagogical process.

In conclusion, this capstone project has not only achieved its objective of exploring the feasibility and effectiveness of using ChatGPT-4 Turbo for generating NFL-related quiz questions but also provided valuable insights into the broader implications of AI in education. The journey from concept to conclusion reveals the

intricate balance between leveraging AI's capabilities and acknowledging its limitations, a balance that will define the future trajectory of educational technology. As this research project draws to a close, it leaves behind a legacy of knowledge and a roadmap for future exploration, setting the stage for the next generation of innovations at the intersection of AI and education.

CHAPTER 5: FUTURE WORK

The advent of AI in the realm of educational content generation, particularly for creating engaging and accurate NFL-related quiz questions, offers a promising horizon for both educators and enthusiasts. The research conducted thus far has laid a foundational framework, highlighting the potential of utilizing AI, such as OpenAI's ChatGPT-4 Turbo, to innovate the way we approach quiz and test question creation. However, the journey into fully harnessing this technology is only in its nascent stages, and the path forward is ripe with opportunities for further exploration and refinement.

A critical area for future investigation is the assessment of question quality and engagement. Understanding how the intricacies of question formulation affect the learning experience of NFL fans can provide invaluable insights into customizing content to suit varied levels of expertise and interest. This endeavor would benefit from empirical studies, including user feedback sessions and analytical assessments, to fine-tune the balance between challenge and engagement, ensuring questions not only test knowledge but also stimulate interest and learning.

Moreover, the potential for real-time quiz question generation during live NFL games represents a thrilling frontier. The dynamic nature of sports events, coupled with the immediacy of live data, poses unique challenges and opportunities for AI. Experimenting with real-time data to generate questions that capture the essence of unfolding game narratives could significantly enhance the fan experience, making each game not just a spectacle but a unique learning opportunity. The feasibility, accuracy, and timeliness of such real-time question generation warrant thorough exploration to understand the limits and capabilities of current AI technologies in processing and reacting to live sports data.

The exploration into team-specific question generation opens another intriguing avenue for research. The depth and richness of NFL history vary widely across teams, presenting a unique challenge in generating meaningful and diverse questions for each. Delving into the depths of historical data to craft questions that reflect the unique heritage and achievements of each team can not only test fan knowledge but also celebrate the storied pasts of these franchises. Investigating the balance between quantity and quality of generated questions, and the point at which content becomes repetitive or stale, can guide the development of more sophisticated AI models that better navigate the breadth of NFL history.

Enhancing the context provided to AI for live data question generation could significantly improve the quality and relevance of the questions posed. By integrating detailed game summaries, player statistics, and significant play analyses, AI can be better equipped to understand and generate content that reflects the complexities and highlights of each game. Future research should focus on optimizing input data for AI, experimenting with various levels of detail and formats to identify the most effective strategies for informing AI content generation.

Finally, the concept of developing a specialized Large Language Model (LLM) tailored specifically towards NFL content presents a groundbreaking opportunity. Such a model, deeply versed in the nuances of football terminology, rules, and historical context, could revolutionize the accuracy and depth of AI-generated quiz questions. Further, the integration of two specialized AI systems—one focused on content mastery and another on pedagogical delivery—could herald a new era of automated content generation that rivals the nuance and insight of human experts.

In conclusion, while the current research has made significant strides in applying

AI to generate NFL-related quiz questions, the full potential of this technology remains untapped. The future of AI in educational content generation is bright, with numerous avenues for exploration that promise to enhance the way we engage with and learn from the sports we love. The journey ahead will require innovative approaches, interdisciplinary collaboration, and a commitment to continual refinement, but the rewards—a richer, more interactive learning experience—will undoubtedly be worth the effort.

REFERENCES

[1] J. White et al., "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," Department of Computer Science, Vanderbilt University, Nashville, TN, USA, Feb. 21, 2023.

[2] M. Fraiwan and N. Khasawneh, "A Review of ChatGPT Applications in Education, Marketing, Software Engineering, and Healthcare: Benefits, Drawbacks, and Research Directions," Department of Computer Engineering, Jordan University of Science and Technology, Irbid, Jordan, 2022.

[3] J. P. Bharadiya, "Leveraging Machine Learning for Enhanced Business Intelligence," Int. J. Comput. Sci. Technol., vol. 7, no. 1, 2023.

[4] A. Steinmayr, "Using ChatGPT for Creating Multiple- and Single-Choice R/exams Questions," Universität Innsbruck, Faculty of Economics and Statistics, [Online]. Available: https://www.r-exams.org/tutorials/chatgpt/. [Accessed: Oct. 31, 2023].

[5] Santrel Media, "Title of YouTube Video," YouTube, [Online Video]. Available: https://www.youtube.com/watch?v=jHv63Uvk5VA. [Accessed: Oct. 31, 2023].

[6] A. Azaria, "ChatGPT Usage and Limitations," School of Computer Science, Ariel University, Israel, Dec. 27, 2022.

[7] M. Almazyad, F. Aljofan, and N. A. Abouammoh, "Enhancing Expert Panel Discussions in Pediatric Palliative Care: Innovative Scenario Development and Summarization With ChatGPT-4," Apr. 28, 2023.

[8] T. Tülübaş, M. Demirkol, and T. Y. Ozdemir, "An Interview with ChatGPT on Emergency Remote Teaching: A Comparative Analysis Based on Human–AI Collaboration," May 22, 2023.

[9] E. Hanna and A. Levic, "Comparative Analysis of Language Models: hallucinations in ChatGPT," supervised by Dr. A. Alissandrakis, examined by Dr. J. Hagelbäck, Computer Science, VT 2023.

[10] J. White et al., "ChatGPT Prompt Patterns for Improving Code Quality, Refactoring, Requirements Elicitation, and Software Design," Department of Computer Science, Vanderbilt University, Nashville, TN, USA, Mar. 11, 2023.

[11] AIContentfy Team, "ChatGPT in sports: analysis and coverage enhancement," Aug. 11, 2023.

[12] M. Drogalis, "GPT-4 + Streaming Data = Real-Time Generative AI Technology," Big Ideas, Jun. 8, 2023.

[13] J. M. Perkel, "Six tips for better coding with ChatGPT," Technol. Feat., Jun. 5, 2023.

[14] Workera Team, "Making the Right Choices: How to Generate Outstanding Multiple-Choice Questions using ChatGPT," Workera, [Online]. Available: https://www.workera.ai/. [Accessed: Oct. 31, 2023].

[15] TechLead, "Title of YouTube Video," YouTube, [Online Video]. Available: https://www.youtube.com/watch?v=9AXP7tCI9PI. [Accessed: Oct. 31, 2023].

[16] SportRadar, "Sports Data API Documentation," [Online]. Available: https://developer.sportradar.com/docs/read/Home. [Accessed: Dec. 31, 2023].

[17] OpenAI, "OpenAI API Documentation," [Online]. Available: https://platform.openai.com/docs/introduction. [Accessed: Oct. 31, 2023].

[18] K. Brown-Siebenaler, “The Increasing Role of Artificial Intelligence in Engineering: Part 1,” Ptc.com, 2023. https://www.ptc.com/en/blogs/cad/increasing-role-of-AI-in-engineering-part-1

[19] C. Gallagher, “The Art of Prompt Engineering in ChatGPT,” www.linkedin.com. https://www.linkedin.com/pulse/art-prompt-engineering-chatgpt-chris-gallagher

IMAGE REFERNCES

[20] S. O'Neill, "AI-powered Features You'll Find On Prime Video's 'Thursday Night Football' This Season", Amazon Web Services. Amazon. Sept. 22, 2023.

[21] M. Fokina, "When Machines Dream: A Dive in AI Hallucinations [Study]," Tidio, Sep. 06, 2023. https://www.tidio.com/blog/ai-hallucinations/

[22] F. Richter, "Infographic: ChatGPT Is the Most Tried AI Tool and Users Stick to It," Statista Infographics, May 15, 2023. https://www.statista.com/chart/30003/usage-of-ai-tools-in-the-united-states/

[23] M. Breuss, "Prompt Engineering: A Practical Example – Real Python," realpython.com. https://realpython.com/practical-prompt-engineering/

[24] OpenAI, "ChatGPT," chat.openai.com, Nov. 30, 2022. https://chat.openai.com/

APPENDIXES

APPENDIX A

Every Initial Prompt Used in Each Trial Run

1. Version 1.1:
    a. History Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Give me a unique multiple choice quiz question about the NFL team {team}'s.
    b. Live Data Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Give me a unique multiple choice quiz question about the following NFL game summary: {summary}
2. Version 2.1:
    a. History Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the

requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Give me a unique {difficulty} level difficulty multiple choice quiz question about the {team}'s {chosen_topic}. Use the entire history of the team to generate the question. Verify the answer is correct.

b. Live Data Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Based on the following plays from this National Football League game: {summary}, generate a unique multiple choice quiz question from big plays. Please provide as much detail as possible including but not limited to which teams were playing, which team made the play, what type of play it was, the quarter the play occurred, time left in quarter, who made the play, and if it resulted in a touchdown, turnover, sack, or first down. Provide the team name of the player that made the play. Verify the answer is correct.

3. Version 3.1:

a. History Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or

option1: before displaying the option) You must provide a question, 4 options, and an answer. Give me a unique {difficulty} level difficulty multiple choice quiz question about the {team}'s {chosen_topic}. Use the entire history of the team to generate the question. Ask questions that are definitive with only one verified answer. Be as specific as possible when constructing the question and stick to the described format in the response. Verify the answer is correct prior to responding with the requested information. If you cannot verify the answer, provide a new multiple-choice question with a topic of your choosing but it must relate to the history of the NFL team {team}.

b. Live Data Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Based on the following plays from this National Football League game: {summary}, generate a unique multiple choice quiz question from big plays. Please provide as much detail as possible about the play. Provide the team’s name(s) of the described play. Verify the answer is correct based on the above summary before generating the response. If you cannot verify the answer, provide a new multiple-choice question based on the summary provided. Please include the week number of the game in the question. That week number is {week}. Also include that the game year is 2023.

4. Version 4.1:
    a. History Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Give me a multiple-choice quiz question about the National Football League Franchise {team}'s. Use the entire history of the team to generate the question. Choose any subtopic about the team you would like, just keep it unique. Ask questions that are definitive with only one verified answer. Be as specific as possible when constructing the question and stick to the described format in the response. Verify the answer is correct by checking yourself prior to responding with the requested information. If you cannot verify the answer, provide a new multiple-choice question with a topic of your choosing but it must relate to the history of the NFL franchise {team}.
    b. Live Data Prompt: Ensure all questions generated are below 255 characters and each answer is no more than 7 words. The format of the response must be question \n option1 \n option2 \n option3 \n option4 \n answer. Do not provide anything else in the response to distinguish what each line represents, only the requested information. (i.e. don't put question: before asking the question or option1: before displaying the option) You must provide a question, 4 options, and an answer. Based on the following plays from this National Football League game: {summary}, generate a unique multiple choice quiz question

from big plays focusing on the team {team). Please provide as much detail as possible about the play. Provide the team name(s) of the described play. Verify the answer is correct based on the above summary before generating the response. If you cannot verify the answer, provide a new multiple-choice question based on the summary provided. Please include the week number of the game in the question.

5. Version 1.2:
    a. History Prompt: Generate 10 unique multiple choice quiz questions about the {team}. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces before each question.
    b. Live Data Prompt: Generate 10 unique multiple choice quiz questions based on this NFL game summary: {summary}. Each question should focus on plays made by the {team}. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text.
6. Version 2.2:
    a. History Prompt: Generate 10 unique multiple choice quiz questions about the {team}'s history, focusing on various aspects about the team. Ensure to pull

from the entire history of the {team}'s and make each question different from the next. Ask questions ranging in difficulty that are definitive with one verified answer only. Verify the correct answer of each question before providing the response. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.

b. Live Data Prompt: Generate 10 unique multiple choice quiz questions based on this NFL game summary: {summary}. Provide as much detail as possible about the play. Provide the team name(s) of the described play. Provide the quarter in which the play happened, as well as the time remaining in the quarter. Focus the questions on big plays from the game (i.e. touchdowns, interceptions, fumbles, plays over 10 yards, negative plays, etc.). Construct the question in the form of a quiz game. Each question should focus on plays made by the {team}. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. "Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.

7. Version 3.2:

a. History Prompt: Generate 10 unique multiple choice quiz questions about the {team}'s history, focusing on various aspects about the team. Ensure to pull from the entire history of the {team}'s and make each question different from the next. Verify the correct answer of each question before providing the response. Ask questions with only one correct answer. Provide questions that can be used in a quiz game given to fans of the {team}'s. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.

b. Live Data Prompt: Generate 10 unique multiple choice quiz questions based on this NFL game summary from a game played by the {team}'s: {summary}. Provide as much detail as possible about the play. Only ask questions from big plays that occurred during the game. Keep the questions to what Actions were performed by what players, do not ask questions that focus on the time or quarter the play occurred as the answer. Construct the question in the form of a quiz game. This quiz will be played by fans watching the game, during the game. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not

number each question. Do not put spaces after each question.

8. Version 1.3:
    a. History Prompt: Generate 10 unique multiple choice quiz questions about the {team}. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.
    b. Live Data Prompt: Generate 10 unique multiple choice quiz questions based on this NFL game summary: {summary}. Each question should focus on plays made by the {team}. The summary data is formatted as follows: quarter: quarter the play occurred, Time Left: time left on the clock in the quarter, Play: A detailed summary of the play that occurred in the game. Each play is separated by a - character. Each question should be under 255 characters. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. “Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.
9. Version 2.3:
    a. History Prompt: Generate 10 unique multiple choice quiz questions about the {team}'s history, focusing on various aspects about the team. Ensure to make

each question different from the next and verify the answer provided is correct to the question given. Ask questions with only one, definitive correct answer. Provide questions that can be used in a quiz game given to fans of the {team}'s. \n\n Each question should be under 255 characters. If a date or year is given as an answer, make sure it is when the event occurred, not the season (i.e. the bears won the Superbowl in the 1985 season, but Superbowl XX occurred in 1986. Therefore, the bears won Superbowl XX in 1986). Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.

b. Live Data Prompt: Generate 10 unique multiple choice quiz questions based on this NFL game summary: {summary}. Each question should focus on plays made by the {team}. The summary data is formatted as follows: Quarter: quarter the play occurred, Time Left: time left on the clock in the quarter, Play: A detailed summary of the play that occurred in the game. Each play is separated by a - character. Each question should be under 255 characters. Construct the questions in the form of a quiz game. Provide as much detail as possible about the play. Only ask questions from big plays that occurred during the game. Keep the questions to what Actions were performed by what players. Provide 4 options and 1 answer for each question, with the answer being no more than 7 words. Response Format: question \n option1 \n option2 \n option3 \n option4 \n answer. Maintain this format

strictly without additional labels or text. Do not provide anything else in the response other than the requested information. Do not number each question. Do not put spaces after each question.

APPENDIX B

Accuracy Tables			
	Questions	Incorrect?	% Belong
Prompt 1.1	4	3	75.00%
Prompt 2.1	11	8	72.73%
Prompt 3.1	12	8	66.67%

Duplicates Tables (History)			
	Questions	Duplicate?	% Belong
Prompt 1.1	6	6	100.00%
Prompt 2.1	5	5	100.00%
Prompt 3.1	4	4	100.00%

Vague Tables			
	Questions	Vague?	% Belong
Prompt 1.1	12	8	66.67%
Prompt 2.1	8	3	37.50%
Prompt 3.1	9	3	33.33%

History Table				
	Questions	Correct	Incorrect	% Correct
Prompt 1.1	78	71	7	91.03%
Prompt 2.1	76	70	6	92.11%
Prompt 3.1	75	70	5	93.33%

Live Data Table				
	Questions	Correct	Incorrect	% Correct
Prompt 1.1	50	47	3	94.00%
Prompt 2.1	50	46	4	92.00%
Prompt 3.1	50	44	6	88.00%

Data Analysis of Version 2

Historic Data	Percent Correct	Percent Incorrect
Questions With No Verification	82.00%	18.00%
Questions With Verification	94.06%	5.94%

Live Data	Percent Correct	Percent Incorrect
Live Data 1.3	91.51%	8.49%
Live Data 2.3	92.45%	7.55%

Data Analysis of Version 3