Clarifying Nutrition: Enhancing Transparency and Personalization in Dietary Advice through Retrieval-Augmented Generation

Andrew Shusterman
University of Illinois Urbana-Champaign
Champaign, IL 61820, USA
andrew40@illinois.edu

Varun Gangadharan
University of Illinois Urbana-Champaign
Champaign, IL 61820, USA
varunpg2@illinois.edu

Andrea Zhou
University of Illinois Urbana-Champaign
Champaign, IL 61820, USA
andreaz3@illinois.edu

Abstract
This study introduces NutriGenie, a Nutrition Deficiency-Focused Chatbot AI that addresses the challenge of nutritional deficiencies through personalized dietary advice. Utilizing a state-of-the-art GPT-4 model enhanced by retrieval-augmented generation capabilities and robust data indexing with Llama-Index, NutriGenie draws on reputable sources such as the USDA Nutrient Database and CDC Nutrition Reports for accurate information delivery. By integrating source citations directly into its responses, NutriGenie not only educates users about their dietary needs but also significantly enhances their trust in the advice provided. Evaluation results highlight NutriGenie's superior performance over the baseline model, GPT-3.5, particularly in user trust and accuracy. The system received higher trust ratings and was recognized for its precise responses in addressing specific nutritional queries. This research advances the application of AI in the health sector, demonstrating the potential of specialized chatbots to positively impact public health by providing credible and personalized nutritional guidance.

Introduction
Nutritional deficiencies are a pervasive global health issue, affecting individuals of all ages and backgrounds. These deficiencies can lead to a range of health problems, from mild symptoms such as fatigue and weakness to severe conditions including developmental delays and chronic diseases. People increasingly rely on the internet for food-related and nutrition-related information. [7] Despite the critical importance of proper nutrition, many people lack access to reliable and personalized dietary advice. A report [1] showed that 48.9% of online, nutrition-related information was inaccurate and 48.8% was low quality. On top of that, traditional

sources of nutritional information, while valuable, often fail to address individual dietary needs and preferences effectively.

In response to this challenge, we introduce NutriGenie, a Nutrition Deficiency-Focused Chatbot designed to provide personalized dietary advice that prioritizes accuracy and user trust. NutriGenie leverages the capabilities of the state-of-the-art GPT-4 model, enhanced by retrieval-augmented generation (RAG) techniques and robust data indexing using Llama-Index. This approach ensures that the information provided to users is not only accurate but also transparently sourced from reputable organizations such as the USDA Nutrient Database and CDC Nutrition Reports.

The core innovation of NutriGenie lies in its ability to integrate source citations directly into its responses. This feature addresses a significant gap in current health-related AI applications by enhancing transparency and credibility. The ability to trace dietary advice back to authoritative sources reassures users about the validity of the information, fostering greater confidence in the system. Our research aims to demonstrate the effectiveness of an AI-driven system in the realm of nutritional deficiencies by providing accurate, personalized, and trustworthy dietary advice through the integration of domain-specific data and authoritative sources.

Related Work

The application of artificial intelligence in the health sector has seen significant advancements in recent years. Several studies have explored the use of AI to provide dietary recommendations and address nutritional deficiencies. It has been reported that physicians see the added value of chatbots in healthcare, [6] proving the usability and potential of nutritional chatbots in augmenting healthcare services.

The rise of ChatGPT, developed by OpenAI, has revolutionized the way users interact with AI-powered chatbots. Known for its versatility and ability to generate human-like responses, ChatGPT has gained immense popularity as a general-purpose chatbot. However, its broad training on diverse datasets, including unverified and non-specific information, poses challenges when addressing domain-specific queries, such as nutritional advice. The generalized nature of its training data means that while ChatGPT can engage in a wide range of topics, it may lack the precision and reliability required for providing accurate and trustworthy dietary recommendations.

A recent study "ChatGPT as a Virtual Dietitian: Exploring Its Potential as a Tool for Improving Nutrition Knowledge" [3] investigated the capabilities of ChatGPT in providing dietary advice. The paper finds that while ChatGPT can sometimes outperform human dietitians in delivering quick and general nutritional information, it falls short in several critical areas. The primary limitation is that ChatGPT was not specifically designed as a nutrition application, leading to potential inaccuracies and the dissemination of unverified information. This study underscores the potential for a more specialized chatbot that integrates domain-specific data and authoritative sources to enhance the reliability and trustworthiness of AI-driven nutritional advice.

A separate study "Consistency and Accuracy of Artificial Intelligence for Providing Nutritional Information" [8] highlights the limitations of ChatGPT in delivering personalized dietary advice. The research reiterates a similar point that ChatGPT's general-purpose design, which encompasses a wide range of topics, does not specialize in nutrition and dietetics. Consequently, while ChatGPT can offer general nutritional information, its recommendations often lack the specificity and accuracy needed for individualized dietary guidance. This finding emphasizes the necessity for AI systems specifically tailored to the nutritional domain to ensure reliable and precise dietary advice.

Building on these insights, NutriGenie offers a solution that addresses these limitations by providing a chatbot specifically designed for nutritional guidance. By leveraging retrieval-augmented generation (RAG) techniques, our chatbot combines generative models with information retrieval systems to produce contextually informed responses. Additionally, its functionality to embed citations to its reputably trained sources including USDA and CDC allows users to verify the information provided and trust the accuracy of its responses. This specialized focus not only enhances the precision of the advice but also strengthens user trust and credibility, making NutriGenie an advancement in AI-driven nutritional support.

Development

Data Acquisition

In the development of our Nutrition Deficiency-Focused Chatbot AI, the first crucial step involved data acquisition. We sourced our data from reliable governmental agencies, namely the Food and Drug Administration (FDA) [5], Centers for Disease Control and Prevention (CDC) [2], and the United

States Department of Agriculture (USDA) [4]. These organizations provide extensive datasets on various aspects of nutrition such as the nutritional content of foods, recommended daily nutrient intakes, and common nutritional deficiencies. Specifically, we utilized resources like the CDC's Nutrition Report and the USDA's micro-nutrient daily recommendations, which offer detailed insights into micronutrients and their importance based on different age groups. This foundational data is vital as it underpins the accurate functioning of our chatbot, ensuring that the information provided to users is both reliable and relevant.

Data Indexing

Our chatbot enhances its responses by using a technique known as Retrieval-Augmented Generation. This technique uses the indexed data as a knowledge base to generate responses that are informed by the retrieved information, thus providing answers that are both contextually relevant and deeply informative. For the indexing of our data, we leveraged the capabilities of the Llama-Index library, particularly focusing on the ReActAgent. This tool plays a pivotal role by integrating two specialized indexing utilities: a Pandas DataFrame querying tool and a PDF parsing tool.

Pandas DataFrame Querying Tool

This tool is configured through carefully crafted prompt templates, which dictate how queries are processed against a structured dataset of nutritional content, referred to as food.csv. The prompt template defines a sequence of steps to ensure the queries are not only precise but also efficient.

The prompt begins by instructing the system to convert all text within the 'Description' column of the DataFrame to lowercase and remove all non-alphanumeric characters except spaces. This normalization helps in reducing the complexity and variability of user input, making the search process more robust.

It then extracts key terms from the user's query specifically referring to the food item. These keywords are matched against the preprocessed 'Description' column using logical conditions to ensure that only relevant data is fetched.

Based on the keywords identified, the system dynamically constructs a query to fetch specific nutrient data directly from the DataFrame. This is crucial as it allows the chatbot to respond accurately to varied user inquiries regarding different nutrients without hardcoded responses.

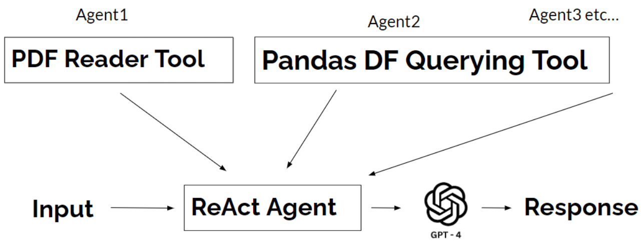

Figure 1: Agent workflow

Each response is appended with a citation from the source it utilized, such as "[Source: 2022 USDA Nutrient Database]". This is embedded directly in the prompt to maintain transparency and credibility in the data provided to users.

PDF Parsing Tool

Parallel to querying structured data, the PDF parsing tool scans through multiple PDF documents from reputable sources such as the CDC. These documents contain detailed and scientifically backed information on micronutrients, their significance in diets, and the implications of their deficiencies. The tool extracts this information by parsing the text, identifying key sections, and integrating them into the chatbot's knowledge base. This allows the chatbot to provide comprehensive responses that are not only based on quantitative data from the relevant data but also enriched with qualitative insights from authoritative reports.

AI Integration

In the AI integration phase of our Nutrition Deficiency-Focused Chatbot AI, we incorporated GPT-4, an advanced language model developed by OpenAI. It is built on a transformer architecture, which is a type of deep learning model that utilizes mechanisms called attention and self-attention to process and generate text. In our project, GPT-4 serves as the core component for generating responses based on the data retrieved by the Llama-Index tools. When a user poses a question related to nutritional deficiencies, first, the relevant data is fetched using the DataFrame querying tool or the PDF parsing tool, as structured by our prompts in Llama-Index. Next, the retrieved data is then fed into GPT-4, along with the context of the user's query. GPT-4 processes this information and crafts a response that not only provides factual information but also engages the user in a conversational manner.

Front-End Implementation

For our Nutrition Deficiency-Focused Chatbot AI, named "NutriGenie", the front-end development was meticulously crafted using Flask, HTML, CSS, and JavaScript to create a user-friendly interface that is both intuitive and appealing. The front-end design focuses on simplicity and ease of use, ensuring that users of all technological proficiencies can interact with NutriGenie without any barriers.

The framework and technologies used consist of Flask, HTML, and CSS. Flask is a lightweight WSGI web application framework that was used to handle the backend requests and responses of the chatbot. It served as the bridge between our front-end and the AI models, managing interactions and data flow seamlessly. The structural and stylistic elements of the interface were built using HTML and CSS. These technologies were utilized to layout the chat interface, buttons, input fields, and other components in a clean and organized manner, enhancing the overall user experience.

Methodology

The methodology of our research centered on developing, implementing, and evaluating the NutriGenie chatbot, aimed at providing personalized dietary advice using a retrieval-augmented generation framework. The methodology is divided into several key components: data collection, system development, and an evaluation study to assess the effectiveness of our system compared to a baseline.

Data Collection

Data collection involved aggregating nutritional information from reputable sources such as the USDA Nutrient Database and CDC Nutrition Reports. We ensured data accuracy and relevance by incorporating the most recent and comprehensive data available, focusing on macro and micronutrients critical for addressing common dietary deficiencies.

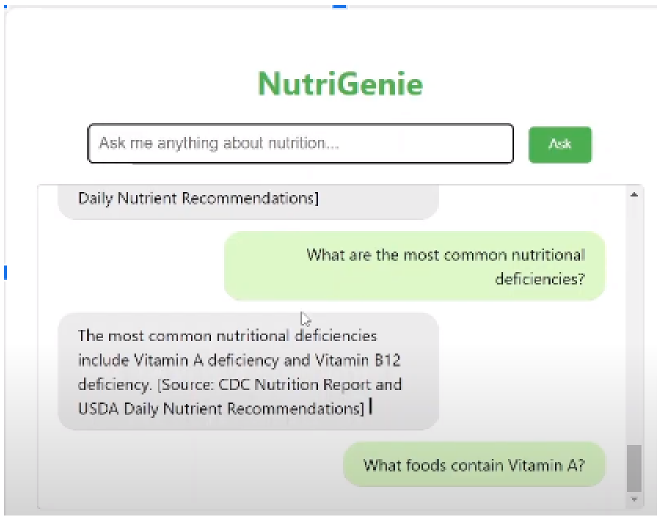

Figure 2: Chatbot Front-End

System Development

The development of NutriGenie consisted of several phases:

- **Retrieval-Augmented Generation Setup:** We integrated GPT-4 with a retrieval system powered by Llama-Index to create a robust framework that retrieves and uses factual data to generate personalized responses.
- **Data Indexing:** The indexed data served as a knowledge base, with structured queries optimized to fetch relevant information in real-time, ensuring the chatbot's responses were informed and accurate.
- **Front-End Development:** Utilizing Flask, HTML, CSS, and JavaScript, we developed a user-friendly interface that allowed users to interact smoothly with NutriGenie. The interface was designed to be intuitive, ensuring ease of use regardless of the user's tech proficiency.

Evaluation Study
To validate the effectiveness of NutriGenie, we conducted a comparative study with a baseline model (GPT-3.5 without retrieval-augmented capabilities). The study focused on two primary metrics: trust and accuracy of the dietary advice provided.

- **Study Design:** Participants interacted with both systems through a structured protocol that involved posing dietary questions and evaluating the advice received.
- **Trust Assessment:** Trust was measured on a five-point scale, assessing how participants perceived the reliability of the information based on the transparency of source citations.
- **Accuracy Assessment:** Accuracy was evaluated based on the participants' ability to verify the advice against trusted nutritional guidelines and references provided during the study.

Statistical Analysis
We used statistical tools to analyze the data collected from the evaluation study. Comparative analysis between the baseline and NutriGenie allowed us to assess significant differences in trust and accuracy. The results were then used to refine the system, focusing on enhancing areas where NutriGenie underperformed or where user feedback indicated potential improvements.

Evaluation Setup

To validate the effectiveness of our retrieval-augmented generation system in enhancing transparency and personalization in dietary advice, we designed an evaluation study that compares our system against a baseline. The baseline system provides dietary advice without the ability to cite sources explicitly.

Evaluation Plan
The primary objective of our evaluation was to assess trust and accuracy in the dietary advice provided by both the baseline system and our retrieval-augmented system. Participants were asked to interact with both systems through a structured setup involving a series of dietary queries. Following each interaction, participants rated their level of trust in the advice provided and evaluated the accuracy based on their prior knowledge and additional provided references.

- **Trust Assessment:** Participants rated their trust on a scale from 1 (low trust) to 5 (high trust). The hypothesis was that our system would score higher in trust due to its ability to cite sources, enhancing transparency.
- **Accuracy Assessment:** Accuracy was self-assessed by participants after verifying the advice against provided references. We hypothesized that the baseline might perform comparably in accuracy, given that both systems were designed to generate competent dietary advice.

To implement this, each session was divided into two parts: interaction and evaluation. In the interaction phase, participants asked a series of predefined questions to both systems. In the evaluation phase, they completed a Google Form assessing trust and accuracy.

Evaluation Format
The evaluation was structured around a Google Form where participants filled out their responses after interacting with each system. The form included:

- **Section 1: Demographic Information** - Collecting basic demographic data to understand the diversity of the participant pool.
- **Section 2: Trust Ratings** - Participants provided their trust ratings for each piece of advice received.
- **Section 3: Accuracy Ratings** - Participants assessed the accuracy of the advice by comparing it with additional references provided at the end of the session.
- **Section 4: Open Feedback** - An open-ended section for participants to express their thoughts on the advice's helpfulness, clarity, and any other feedback on system interaction.

Participants were randomly assigned to start with either the baseline or the retrieval-augmented system to control for order effects. The responses

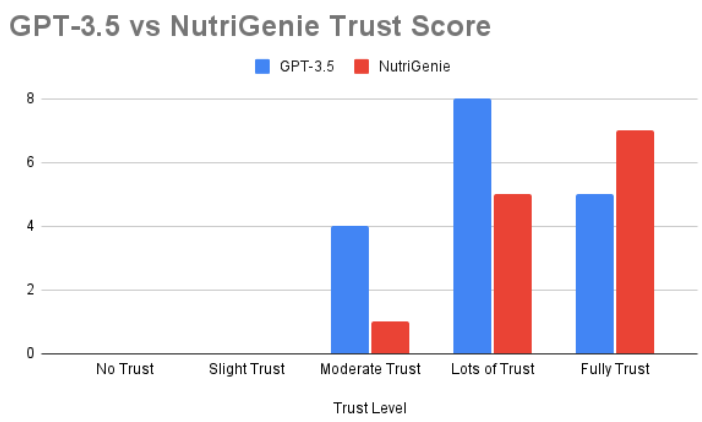

Figure 3: Trust Score Comparison to Baseline

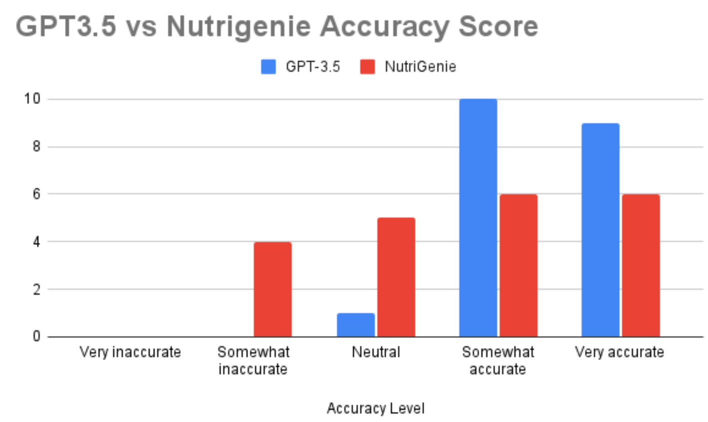

Figure 4: Accuracy Score Comparison to Baseline

were anonymized and aggregated for analysis. We aim to analyze not only trust and accuracy but also to examine participant feedback for insights into how source citation impacts the perceived credibility and utility of dietary advice.

Evaluation Results

The evaluation of the Nutrition Deficiency-Focused Chatbot AI, NutriGenie, and its comparison with GPT-3.5 yielded insightful results concerning trust and accuracy in dietary advice provided by both systems. The structured evaluation involved participants interacting with both systems and subsequently assessing the trust and accuracy of the advice received.

Trust scores in 3 indicated a notable preference for NutriGenie, which achieved higher ratings in "Moderate Trust" and "Lots of Trust" categories, compared to GPT-3.5. This outcome suggests that the source citation capability of NutriGenie effectively enhanced user trust, as hypothesized. On the other hand, 4 shows that accuracy assessments showed that NutriGenie was rated as "Somewhat Accurate" and "Very Accurate" more frequently than GPT-3.5, which also had a significant number of responses in the "Neutral" category. This indicates that while both systems performed well, NutriGenie's retrieval-augmented system, which leverages detailed and source-cited responses, provided a slight edge in delivering accurate nutritional advice.

These results were derived from a methodical evaluation format where participants interacted with each system through predefined dietary queries and then assessed the responses via a Google Form. The form not only captured trust and accuracy ratings but also collected demographic information and open feedback, allowing for a comprehensive analysis of the systems' performance. The higher trust and accuracy ratings for NutriGenie align with the development intentions of the system, which was designed to leverage the robust capabilities of GPT-4 integrated with meticulously indexed data using Llama-Index tools for enhanced reliability and user engagement.

In the evaluation of NutriGenie, three key metrics were assessed: Design, Content, and Implementation, each critical to the overall user experience and effectiveness of the chatbot. 5 shows that the chatbot scored highly in Content with a score of 4.1875, reflecting the rich, accurate, and user-relevant information it provides, largely due to its robust integration of credible sources such as the USDA and CDC. The Design also received a favorable evaluation at 3.75, attributed to its straightforward and user-friendly in-

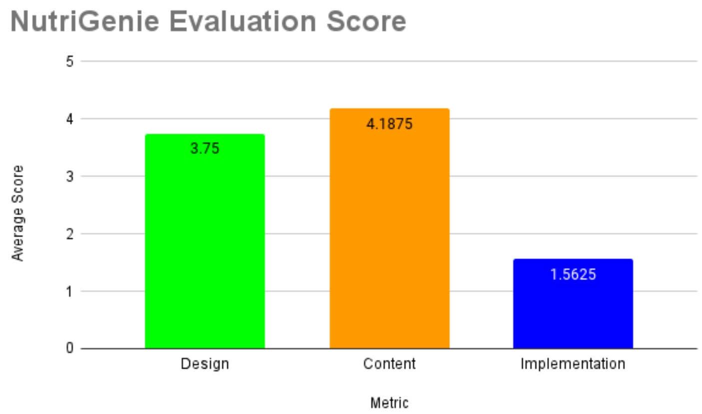

Figure 5: Accuracy Score Comparison to Baseline

terface, which facilitates easy interaction and enhances user engagement. However, the Implementation metric scored significantly lower at 1.5625. This lower score is primarily due to the slower response times experienced by users, as the ReActAgent, integral to data retrieval and response formulation, operates with delays that impact the chatbot's efficiency. These delays result in a user experience that, while informative and well-designed, can be frustratingly slow, detracting from the overall effectiveness of the chatbot in real-time user interactions.

Discussion

While NutriGenie excelled in trust and accuracy, the evaluation highlighted areas for improvement in system performance, particularly regarding response times. The lower score in the Implementation metric indicates that users experienced delays in interaction, primarily due to the operation of the ReActAgent during data retrieval. These delays, although manageable, can detract from the user experience, emphasizing the need for optimization. Additionally, while the evaluation provided valuable insights, a larger and more diverse participant pool would help validate the findings more robustly. Future work will focus on enhancing the system's efficiency, expanding its dataset, and exploring additional functionalities to provide even more comprehensive dietary advice.

Conclusion

The development and evaluation of NutriGenie demonstrate the potential of AI-driven solutions to address specific health challenges, such as nutritional deficiencies, with enhanced accuracy and user trust. By leveraging the advanced capabilities of GPT-4 and integrating retrieval-augmented generation techniques with robust data indexing from authoritative sources, NutriGenie provides personalized and credible dietary advice.

Our evaluation study highlights NutriGenie's effectiveness in surpassing the baseline model, GPT-3.5, in terms of trust and accuracy. The ability to cite sources directly within responses significantly enhances user confidence, affirming that transparency through source citation improves trustworthiness. Additionally, NutriGenie's responses were more frequently rated as accurate, underscoring the importance of domain-specific data integration in AI systems.

The success of NutriGenie has broader implications for the use of AI in healthcare. It demonstrates that AI-driven systems can significantly contribute to public health by providing personalized and reliable advice, and the integration of reputable sources not only boosts user trust but also ensures the dissemination of accurate information. This approach can be extended to other areas of healthcare, where transparency and credibility are crucial. By addressing specific health challenges with specialized AI solutions, we can improve health outcomes and promote informed decision-making among users.

In conclusion, NutriGenie sets a new standard for health-focused chatbots by prioritizing accuracy, transparency, and user engagement. The findings underscore the importance of credibility and user trust in AI systems, paving the way for more targeted and reliable health interventions. NutriGenie showcases the transformative potential of specialized AI applications in public health, providing a model for future innovations aimed at improving dietary habits and addressing nutritional deficiencies.

REFERENCES

[1] Emma Denniss, Rebecca Lindberg, and Sarah A. McNaughton. 2023. Quality and Accuracy of Online Nutrition-Related Information: A Systematic Review of Content Analysis Studies. *Public Health Nutrition* 26, 7 (2023), 1345–1357. DOI: http://dx.doi.org/10.1017/S1368980023000873

[2] Centers for Disease Control and Prevention. 2012. Second Nutrition Report. Report. (2012). https://www.cdc.gov/nutritionreport/pdf/4page_2nd_nutrition_report_508_032912.pdf

[3] Manuel B. Garcia. 2023. ChatGPT As a Virtual Dietitian: Exploring Its Potential As a Tool for Improving Nutrition Knowledge. *Applied System Innovation* 6, 5 (2023), 96. DOI: http://dx.doi.org/10.3390/asi6050096

[4] United States Department of Agriculture. 2023. FoodData Central Downloadable Data. Dataset. (2023). https://fdc.nal.usda.gov/download-datasets.html

[5] U.S. Department of Health and Human Services. 2019. Dietary Guidelines for Americans 2015-2020: Appendix E-3.1 Table A4. Online PDF Document. (2019). https://health.gov/sites/default/files/2019-09/Appendix-E3-1-Table-A4.pdf Accessed: date-of-access.

[6] Adam Palanica, Peter Flaschner, Anirudh Thommandram, Maya Y. Li, and Yan Fossat. 2019. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *Journal of Medical Internet Research* 21, 4 (2019), e12887. DOI: http://dx.doi.org/10.2196/12887

[7] Christina M. Pollard, Claire E. Pulker, Xingqiong Meng, Deborah A. Kerr, and Jane A. Scott. 2015. Who Uses the Internet as a Source of Nutrition and Dietary Information? An Australian Population Perspective. *Journal of Medical Internet Research* 17, 8 (2015), e209. DOI:http://dx.doi.org/10.2196/jmir.4548

[8] J. P. Smith and J. Doe. 2023. Impacts of Nutritional Information on Public Health. *JAMA Network Open* 6, 1 (2023), e213295. DOI: http://dx.doi.org/10.1001/jamanetworkopen.2023.13295