

# **NutriRAG: Unleashing the Power of Large Language Models for Food Identification and Classification through Retrieval Methods**

Huixue Zhou, MBBS<sup>1</sup>, Lisa S.Chow MD, MS<sup>3</sup>, Lisa Harnack, DrPH, RD, MPH, MD, MS, Satchidananda Panda<sup>4</sup>, Emily N.C. Manoogian<sup>4</sup>, Minchen Li, MS<sup>5</sup>, Yongkang Xiao, MS<sup>1</sup>, Rui Zhang, PhD<sup>5\*</sup>

<sup>1</sup>Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA,

<sup>2</sup>Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA,

<sup>3</sup>Division of Diabetes, Endocrinology and Metabolism, Department of Medicine University of Minnesota, Minneapolis, Minnesota, USA,

<sup>4</sup>Salk Institute for Biological Studies, San Diego, California, USA,

<sup>5</sup>Division of Computational Health Sciences, Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA

## **Corresponding author:**

\*Dr. Rui Zhang, PhD

Division of Computational Health Sciences, Department of Surgery

University of Minnesota

11-132 Phillips-Wangensteen Building, 516 Delaware St SE, Minneapolis, MN 55455

Email: [ruizhang@umn.edu](mailto:ruizhang@umn.edu)

Office Phone: 612-626-420

## **Keywords:**

Large Language Model, Natural Language Processing, Retrieval-Augmented Generation, Food Classification, Dietary Analysis

Word count: 3286

## ABSTRACT

**Objective:** This study explores the use of advanced Natural Language Processing (NLP) techniques to enhance food classification and dietary analysis using raw text input from a diet tracking app.

**Materials and Methods:** The study was conducted in three stages: data collection, framework development, and application. Data were collected via the myCircadianClock app, where participants logged their meals in free-text format. Only de-identified food-related entries were used. We developed the NutriRAG framework, an NLP framework utilizing a Retrieval-Augmented Generation (RAG) approach to retrieve examples and incorporating large language models such as GPT-4 and Llama-2-70b. NutriRAG was designed to identify and classify user-recorded food items into predefined categories and analyzed dietary patterns from free-text entries in a 12-week randomized clinical trial (RCT: NCT04259632). The RCT compared three groups of obese participants: those following time-restricted eating (TRE, 8-hour eating window), caloric restriction (CR, 15% reduction), and unrestricted eating (UR).

**Results:** NutriRAG significantly enhanced classification accuracy and effectively identified nutritional content and analyzed dietary patterns, as noted by the retrieval-augmented GPT-4 model achieving a Micro F1 score of 82.24. Both interventions showed dietary alterations: CR participants ate fewer snacks and sugary foods, while TRE participants reduced nighttime eating.

**Conclusion:** By using AI, NutriRAG marks a substantial advancement in food classification and dietary analysis of nutritional assessments. The findings highlight NLP's potential to personalize nutrition and manage diet-related health issues, suggesting further research to expand these models for wider use.

## INTRODUCTION

Nutrition plays a critical role in maintaining health and well-being. It plays an essential role for preventing and managing a multitude of chronic diseases, such as metabolic syndrome, obesity, cancer, hypertension, depression, and cardiovascular disease. [1-6]. Tracking food intake in the free-living setting, however, is a challenging task. The “gold-standard” for tracking food intake is the 24-hour dietary recall, however, these recalls represent “points in time” and do not necessary capture aggregate dietary patterns over time. A potential solution for capturing dietary intake is leveraging technology to capture food images and descriptions [7,8] and use various identification and classification techniques to analyze dietary habits.

The complexity of unstructured textual food descriptions poses significant challenges to classifying food intake. Recent advancements in Natural Language Processing (NLP), especially in semantic understanding and contextual analysis, offer promising solutions for automated food classification and nutritional assessment. NLP has been used to categorize food items and analyze packaged foods [16,17] and social media content [18]. For example, Hu et al. used the XGBoost algorithm to classify packaged foods into 24 distinct categories based on Health Canada’s Table of Reference Amounts [19]. Ma et al. developed a large dataset from the USDA Branded Food Products Database to classify various nutritional components like calories and proteins using a Multi-Layer Perceptron (MLP)-TF-SE approach [17].

There remains a knowledge gap in applying NLP to analyze personalized dietary patterns due to several limitations in existing databases and traditional classification methods. Databases such as the USDA Branded Food Products Database [20] and the University of Toronto Food Label Information and Price Database [21] predominantly focus on packaged items and offer limited coverage of fresh foods and recipes that are frequently part of dietary intakes.

Additionally, traditional NLP based classification, such as XGBoost and MLP are constrained due to their inability to capture complex patterns and dependencies in large-scale and high-dimensional data typical of nutritional studies, indicating a need for more sophisticated and comprehensive classification systems [22]. The emergence of generative large language models (LLMs), such as ChatGPT, which demonstrate improved performance on a variety of NLP tasks [23-30], presents new opportunities. These models have the potential to effectively classify textual descriptions of diverse food items.

In response to these challenges, our study introduces three significant contributions: 1) **Advanced Food Classification System Development:** We created a cutting-edge classification system that defines 51 unique food categories, capturing the complexity of contemporary dietary patterns and aligning with the latest trends in nutritional science. 2) **Development and Application of the NutriRAG (Retrieval Augmented Generation-LLM for nutrition) Framework:** We developed the NutriRAG framework, a few-shot, retrieval-based named entity recognition (NER) LLM framework tailored for food item identification and food class classification. This framework leverages technology-based mobile app records (from the myCircadianClock (mCC) smartphone app) to analyze personalized dietary patterns from text-based food entries. 3) **Application and Testing of the NutriRAG Framework:** We hypothesize that LLMs can effectively identify and categorize text-based food entries into predefined categories from a nutritional database. To validate this, we applied the NutriRAG framework to data from a randomized controlled trial with the intent to detect changes in dietary patterns over time.

## METHODS

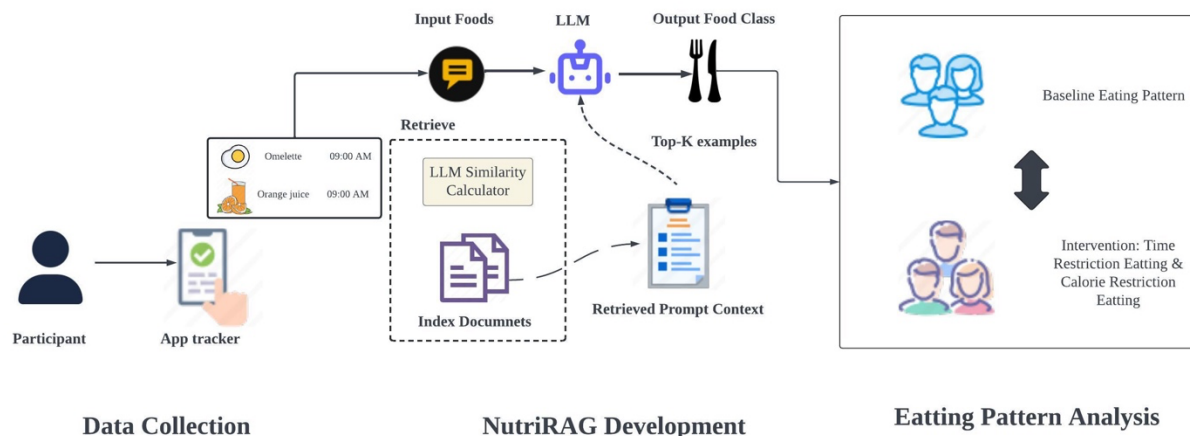


Figure 1. Overview of study.

As depicted in Figure 1, the study methodology is divided into three sections 1) Data Collection in the our previous study[31], where participants log their food intake using the mCC app [32], capturing the food textual descriptions ; 2) Development of NutriRAG, which involves utilizing a Retrieval-Augmented Generation (RAG) strategy within a Large Language Model (LLM) to accurately process and identify food items and their classifications; and 3) Application of NutriRAG to analyze eating patterns (EO), which was defined as any food or drink consumption, including non-caloric beverages, separated by at least 15 minutes from other intake events[33]. This analysis helps to examine dietary habits and trends with the context of either a time-restricted eating or caloric restriction intervention.

### Design of Parent Study and Data Collection

This study builds upon a prior investigation [31] which is a randomized controlled trial (RCT), that compared three eating patterns in patients with obesity but without diabetes over 12

weeks: 8-hour Time-Restricted Eating (TRE), Caloric Restriction (CR, 15% daily reduction), and Unrestricted eating (UR). Participants from the Twin Cities metro area used the mCC mobile application to log their food and beverage intake, excluding water and medications [32]. Those with initial eating windows exceeding 12 hours qualified for the study and continued using the mCC app throughout the intervention period. Of 88 randomized participants, 81 completed the trial, with 77 providing analyzable data for this study: 27 in TRE, 25 in CR, and 25 in UR groups. The parent study received approval from the University of Minnesota's Institutional Review Board (STUDY00008545) and the Salk Institutional Review Board (15-0003) for the mCC app. The trial was registered at ClinicalTrials.gov (NCT04259632), and all participants provided written informed consent.

For this specific study, we analyzed the data collected from the parent RCT, which largely came from the mcc app [32]. Participants logged their food consumption through images and free-text entries, receiving daily prompts to ensure consistent data input. All data, including timestamps, were encrypted, anonymized, and securely stored on a cloud server. For this analysis, we focused on the 32,825 free-text food entries recorded by 77 users.

The primary goal of this stage was to identify user recorded free text food items and classify each item into one of the predefined categories from the Nutrition Coordinating Center (NCC) Food Database, which contains 51 distinct food classes (Supplement Table 1). This NCC database provides a comprehensive classification system, meticulously detailing a wide array of food categories to encompass the diverse range of food items reported by users. To ensure our analysis was based on a robust and representative sample, we randomly selected 1,000 entries, which were manually categorized by NDSR certified staff against NCC database to serve as a “gold standard” benchmark. For instance, the input text “Carrots Short Ribs Potatoes” would be

annotated to correctly identify each food item — “Carrots,” “Short Ribs,” “Potatoes”—with their respective NCC food classes as “Vegetables,” “Meat,” “Vegetables.” The dataset was then divided into subsets for training, validation, and testing, comprising 676, 168, and 182 entries, respectively.

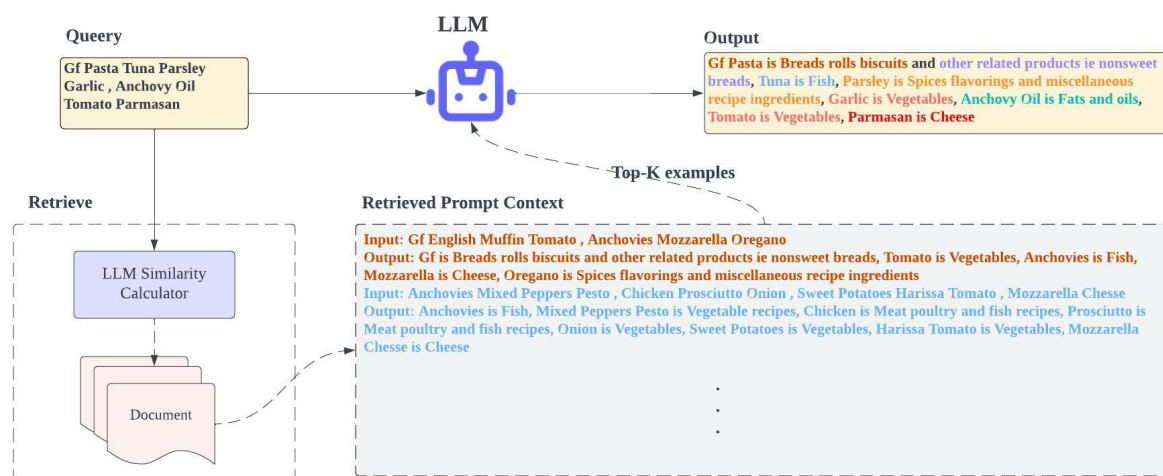


Figure 2. Framework of NutriRAG.

## Development of NutriRAG

To improve the detection and sorting of nutritional data from the mCC app, we developed a new NLP framework NutriRAG, as depicted in Figure 2, which incorporates generative LLMs within a RAG strategy. This strategy leverages relevant documents during the generation process to enrich the model’s output contextually without requiring further tuning of model parameters. Our methodical approach is structured into four primary phases: 1) Query Formulation, where nutritional inquiries are constructed; 2) Retrieval and Prompt Context, where relevant documents are sourced; 3) LLM Processing, where the model processes the queries within the retrieved context; and 4) Organization of Nutritional Data Outputs, where we systematically arrange the LLM output data for analysis of dietary patterns.

## ***Query Formulation***

As shown in Figure 2, The process begins with a “Query” phase, where a nutritional inquiry is composed, consisting of various food items and ingredients. For instance, a typical participant-entered text description might include “Gf Pasta Tuna Parsley Garlic Anchovy Oil Tomato Parmesan.”

## ***Retrieval and Prompt Context***

In the Retrieval and Prompt Context phase, we utilize an LLM Similarity Calculator to determine the score for example ( $e_i$ ) from valid dataset ( $D$ ) for each query. This involves computing the cosine similarity score between the LLM embedding of the query text ( $q_i$ ) and  $e_i$ , defined by the formula:

$$\text{sim}(q_i, e_i) = \frac{E(q_i) \cdot E(e_i)}{|E(q_i)| |E(e_i)|}$$

Here,  $\text{sim}(q_i, e_i)$  is the similarity score, and  $E(q_i), E(e_i)$  are the embeddings of the  $q_i$  and  $e_i$ , respectively. Next, we ranked the examples by their similarity scores and select the top k examples. This step ensures that the subsequent LLM processing is informed by contextually pertinent information, enhancing the accuracy of entity recognition. Once the examples are retrieved, the LLM is presented with a “Retrieved Prompt Context.” This context consists of a set of top-K examples illustrating input-output mappings of food items to their respective nutritional categories. These mappings are designed to act as a guiding framework for the LLM, providing clarity on the expected format and content of the output.

## ***LLM Processing and Output Classification***

The prompt of the LLM consists of instruction, example context and the query entry text. The instruction delineates the precise task to be executed by the model or specifies the desired response type. Meanwhile, the example context is sourced through *Retrieval and Prompt*



*Context.* The LLM then processes prompt, mapping each component of the query entry to its appropriate nutritional category. This stage is crucial as it represents the core functionality of the NER system, wherein the LLM generates outputs such as “Tuna is Fish, Parsley is Spices flavorings and miscellaneous recipe ingredients.” The final LLM output was subsequently transformed into a structured format through string matching for each food class, thereby facilitating the integration of this information into databases or for further analytical processes.

### ***Evaluation***

For our evaluation, we employed advanced generative LLMs including GPT-3.5 and GPT-4, as well as open-source models such as Mixtral 8\*7B [34] and Llama-2-70b [35], utilizing the RAG method. We tested retrieval from 1-20 examples and assessed three orders of retrieved examples: from highest to lowest similarity score, and in a random order. We compared these with base models selecting examples randomly and with fine-tuned pretrained language models such as BERT [36], PubmedBert [37], and BLUEBERT [38]. Performance was evaluated using micro precision, micro recall, and micro F1 score. The micro precision and micro recall are the measurement that considering the total number of true positives across all classes divided by the sum of true positives and false positives, and the total number of true positives across all classes divided by the sum of true positives and false negatives. The micro F1 is calculated by:

$$Micro\ F1 = 2 \times \frac{Micro\ Precision \times Micro\ Recall}{Micro\ Precision + Micro\ Recall}$$

### **Application of NutriRAG and Examination of Eating Patterns**

Our study utilized the NutriRAG in the part I, to examine the eating patterns among the study participants and evaluated the intervention effect on the number of EO, which was defined as a distinct event when food or drink was documented  $\geq 15$  minutes apart from another EO [33]. Defined by our expert, food intake was classified into four daily segments: early morning (5 am - 10 am), midday (10:01 am - 3:00 pm), evening (3:01 pm - 9 pm), and late night (9:01 pm - 4:59 am), to evaluate the timing of dietary intake. The study analyzed the variance in dietary intake, especially with regards to late-night eating between baselines. During the 12-week intervention periods (UE, TRE, CR) of the RCT from the parent study, some EOs were tagged beverage, savory snacks, and high-added-sugar events using the LLM classifier from Step 1. Following the parent study, the study timeline was divided into three phases: baseline, post-randomization (from randomization to study completion), and at post-intervention (the last 2 weeks prior to study completion). The analysis focused on comparing the baseline with the post-randomization period and the baseline with the post-intervention period. Multiple linear regression analysis was used to evaluate the associations of and EOs with BMI and HOMA, after adjustments for age, sex. Results from the linear regression models are presented as standardized  $\beta$ -coefficients. In all analyses, the level of statistical significance was set at  $P < 0.05$ . Data were analyzed using R studio 4.1.2.

## RESULTS

### RCT results

As previously published [31], the TRE group maintained a 10.0-hour eating window (9.2 to 10.9) during the 12-week intervention. In the final two weeks of the intervention (Week 10-12), the TRE group's eating window decreased to 9.1 hours (8.4 to 9.7), representing a 5.1-hour reduction (4.5 to 5.7) from baseline. Results from the 24-hour dietary recall data obtained at

baseline and end-intervention showed average caloric reductions of 16.5% (-24.8 to -8.2) in the TRE group (N=28), 8.7% (-21.8 to 4.5) in the CR group (N=19), and 1.2% (-12.7 to 10.2) in the UE group (N=26). These reductions were not significantly different between groups.

This study analyzed available data from 77 participants (57% female) with a mean age of  $43.3 \pm 10.6$  years. At baseline, participants averaged  $105.6 \pm 1.97$  kg in weight,  $173.0 \pm 9.39$  cm in height, and had a BMI of  $36.4 \pm 5.1$  kg/m<sup>2</sup>. Mean HOMA was  $4.4 \pm 3.88$ . Baseline daily eating occasions showed no significant differences between groups ( $P > 0.05$ ).

### NutriRAG

The results are described in Table 2. In the evaluation of various NER models, the baseline performances of Bert, BLUEBert, and PubmedBert were moderately effective, with Micro F1 scores of 59.38, 53.29, and 58.22, respectively. Notably, the innovative application of RAG significantly enhanced model performance across all configurations. For instance, while the standard GPT4 model achieved a Micro F1 of 73.84, its RAG augmented counterpart produced a substantially higher score of 82.24. For example, the standard GPT-4 model attained a Micro F1 score of 73.84, whereas its RAG-augmented version achieved a significantly higher score of 82.24. Likewise, retrieval-augmented versions of the Mixtral 8\*7b models demonstrated notable improvements, reaching a Micro F1 score of 77.93—the highest among open-source models, surpassing the RAG-augmented GPT-3.5. Based on the model performance of Part I, we utilized the NutriRAG with Mixtral 8\*7B, which achieved the best performance among the open-source models, to process the text-based food information for each participant.

**Table 1. Performance of Models in Food Classification.**

Model	Micro precision	Micro Recall	Micro F1
Bert	56.36	62.74	59.38

BLUEBert	49.84	57.26	53.29
PubmedBert	56.82	59.71	58.22
Random examples Llama-2-70b	37.70	71.72	49.42
Random examples Mixtral 8*7b	60.87	75.58	67.43
Random examples GPT3.5	64.54	66.43	65.47
Random examples GPT4	75.97	71.53	73.84
(UMN) Retrieval-Augmented Llama-2-70b	68.81	58.75	63.38
(UMN) Retrieval-Augmented Mixtral 8*7b	76.07	79.89	<u>77.93*</u>
(UMN) Retrieval-Augmented GPT3.5	75.00	80.29	77.55
(UMN) Retrieval-Augmented GPT4	79.10	85.64	<b>82.24*</b>

\* Note: Bold text indicates the highest values, while underscored text denotes the second.

highest.

### Examination of Eating Patterns

Post-randomization and post-intervention assessments revealed statistically significant differences in daily EOs for breakfast, lunch, late-night, and high-added-sugar events across the three groups (figure 3 and 4).

For the UE group, the average daily EOs were  $4.32 \pm 1.29$  at baseline,  $4.10 \pm 1.44$  during the post-randomization, and  $4.21 \pm 1.57$  post-intervention. There were no significant differences in daily EOs when comparing baseline to post-randomization ( $P=0.598$ ) and baseline to post-intervention ( $P=0.887$ ). Similarly, no significant changes were observed in the average number of daily EOs for savory snacks, beverages, late-night eating, and high-added-sugar foods between baseline and post-randomization (as shown in Figure 5) or between baseline and post-intervention (as shown in Figure 6).

Both the TRE and CR groups altered their dietary habits during the study. Initially, the average daily eating occasions (EOs) for TRE were  $4.31 \pm 0.93$ , which decreased to  $3.47 \pm 0.88$  during the post-randomization and  $3.39 \pm 1.03$  at end-intervention. For the CRE group, the average daily EOs were  $4.45 \pm 1.37$  at baseline, reducing to  $3.22 \pm 0.82$  during the post-randomization and slightly rising to  $3.26 \pm 0.97$  at end-intervention. Significant reductions in daily EOs were observed in the TRE group from between baseline and during the post-randomization ( $P < 0.001$ ) and between baseline and end-intervention ( $P = 0.001$ ). Similar significant changes were noted in the CR group over these periods ( $P < 0.001$  for both comparisons).

The CR group decreased their consumption of savory snacks ( $P = 0.003$ ), and high-added-sugar foods ( $P < 0.001$ ). The TRE group similarly reduced their beverage intake ( $P = 0.032$ ) when during the post-randomization. In terms of meal timing, the CR group cut down on lunch ( $P = 0.002$ ) and dinner ( $P < 0.001$ ), while the TRE group lessened breakfast ( $P < 0.001$ ), lunch ( $P = 0.011$ ), and late-night eating ( $P < 0.001$ ).

These changes persisted into the end-intervention period. When comparing baseline to post-intervention, the CR group continued to consume fewer savory snacks ( $P = 0.013$ ), and high-added-sugar foods ( $P = 0.003$ ). The TRE group maintained reduced consumption of beverages ( $P = 0.047$ ), as illustrated in Figure 6. Regarding timing, the CR group still ate less at lunch ( $P = 0.002$ ), and the TR group continued to reduce breakfast and late-night eating (both  $P < 0.001$ ).

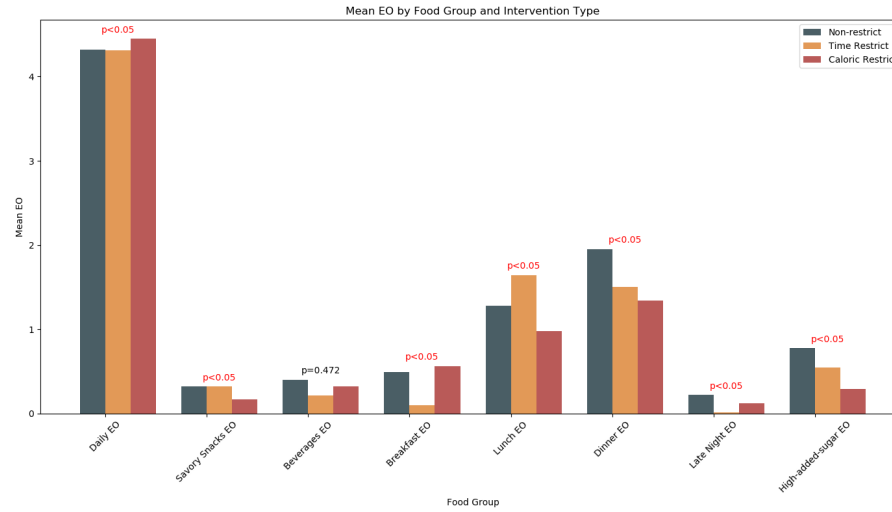
In exploring the relationship between baseline metabolic parameters and eating occasions, we observed a link between baseline insulin resistance and dietary patterns. Notably, insulin sensitivity, characterized by higher HOMA values, inversely correlated with breakfast EO ( $P = 0.019$ ) and dinner EO ( $P = 0.013$ ). Specifically, for each one-unit increase in an EO at breakfast level, there was a decrease of 1.71 units in the HOMA value. This suggests that

increased EO during the breakfast time period may be associated with lower insulin sensitivity.

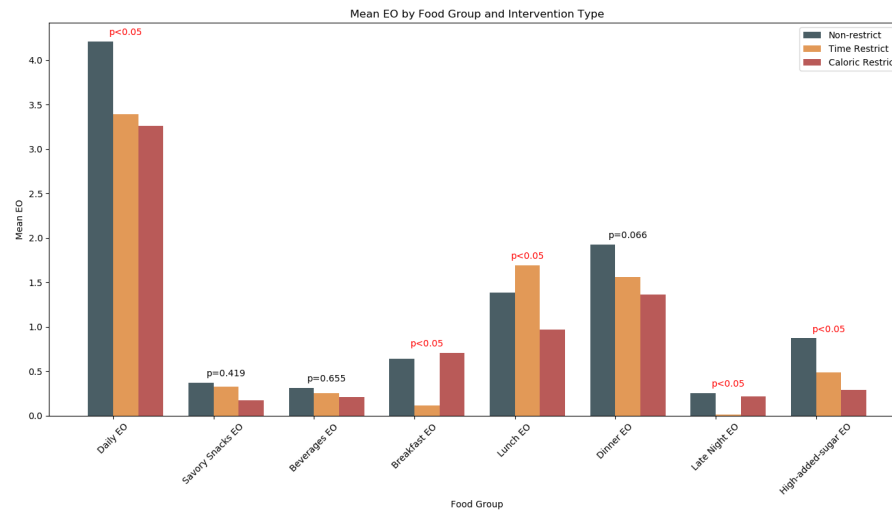
Conversely, no significant relationships were observed between eating occasions and BMI or lipid profiles.

**Table 2. Baseline data.**

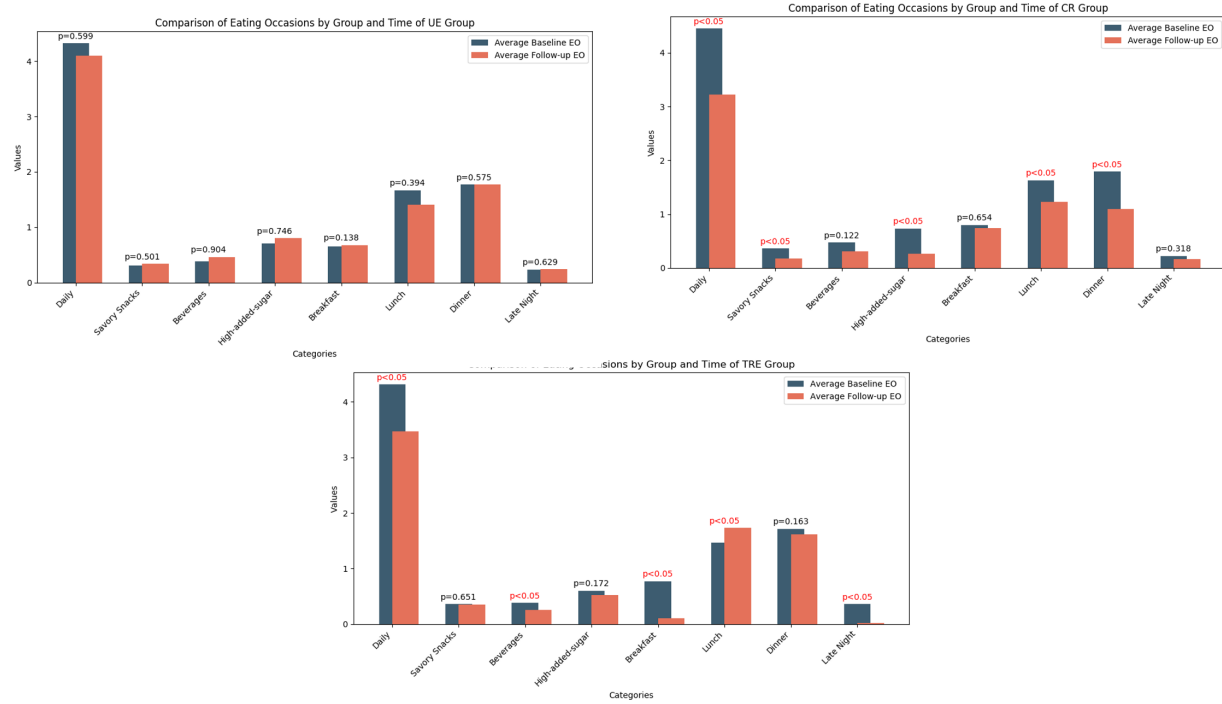
	<b>Food</b>	<b>Mean</b>	<b>Std</b>	<b>P-Value</b>
<b>Time Restrict</b>	Daily EO	4.31	0.93	0.238
<b>Caloric Restrict</b>		4.45	1.37	
<b>Unrestricted</b>		4.32	1.29	
<b>Time Restrict</b>	Savory Snacks EO	0.36	0.26	0.838
<b>Caloric Restrict</b>		0.36	0.24	
<b>Unrestricted</b>		0.31	0.29	
<b>Time Restrict</b>	Beverages EO	0.38	0.52	0.805
<b>Caloric Restrict</b>		0.47	0.48	
<b>Unrestricted</b>		0.39	0.54	
<b>Time Restrict</b>	Breakfast EO	0.77	0.37	0.551
<b>Caloric Restrict</b>		0.80	0.35	
<b>Unrestricted</b>		0.65	0.45	
<b>Time Restrict</b>	Lunch EO	1.46	0.39	0.216
<b>Caloric Restrict</b>		1.63	0.45	
<b>Unrestricted</b>		1.67	0.65	
<b>Time Restrict</b>	Dinner EO	1.71	0.45	0.795
<b>Caloric Restrict</b>		1.79	0.81	
<b>Unrestricted</b>		1.77	0.75	
<b>Time Restrict</b>	Late Night EO	0.36	0.34	0.226
<b>Caloric Restrict</b>		0.22	0.18	
<b>Unrestricted</b>		0.23	0.22	
<b>Time Restrict</b>	High-added-sugar EO	0.60	0.46	0.614
<b>Caloric Restrict</b>		0.73	0.61	
<b>Unrestricted</b>		0.71	0.47	



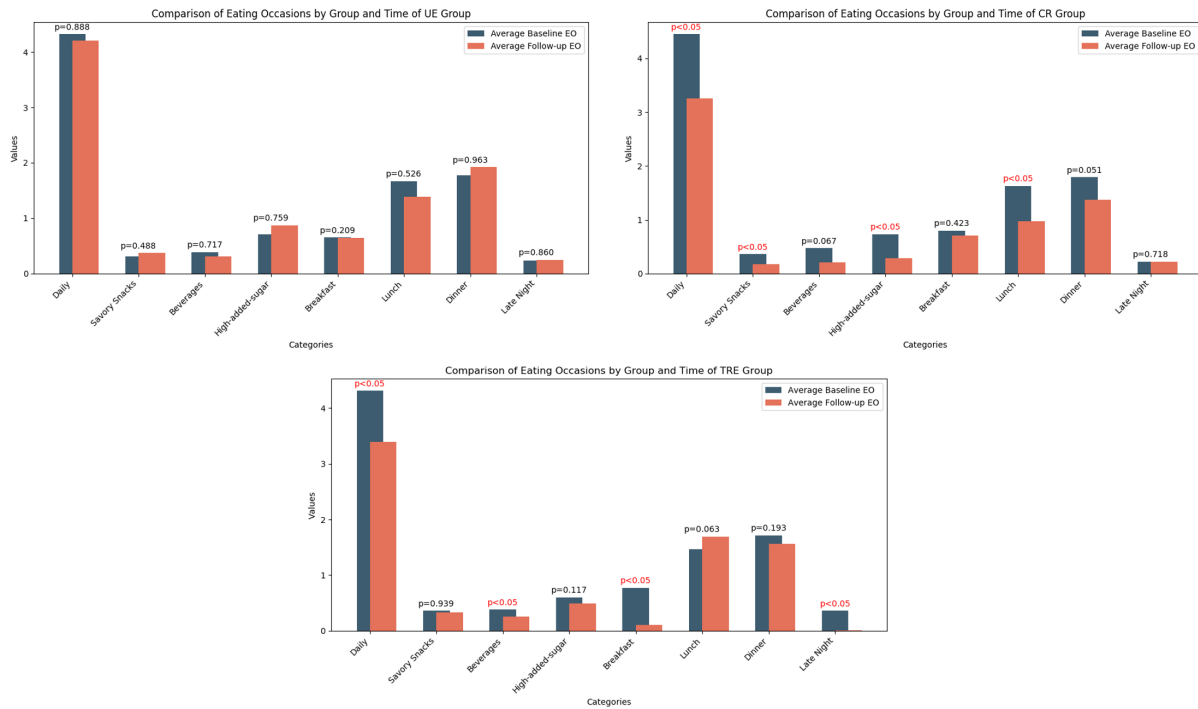
**Figure 3. ANOVA of Different Group Post-randomization.**



**Figure 4. ANOVA of Different Group of Post-intervention.**



**Figure 5. Paired T-Test of Baseline EOs and Post-randomization EOs.**



**Figure 6. Paired T-Test of Baseline EOs and Post-intervention EOs.**



**Table 3. Linear Regression Analysis Results for HOMA.**

Variable	Estimate	Std. Error	P-value
Age	-0.002	0.025	0.938
Gender Male	0.492	0.548	0.372
BMI	0.193	0.051	<b>&lt;0.001</b>
Breakfast EO	-1.708	0.712	<b>0.019</b>
Lunch EO	-0.466	0.636	0.466
Dinner EO	-1.365	0.532	<b>0.013</b>
Late Night EO	0.769	1.063	0.472
High-added-sugar EO	0.666	0.639	0.301

## DISCUSSION

This study introduces NutriRAG, a novel approach integrating retrieval augmented generative LLM for the classification of food items from textual descriptions. Our findings demonstrate that NutriRAG significantly enhances the performance of standard LLMs in the domain of nutritional science.

The improved performance of retrieval-augmented models, as shown in our results, underscores the value of contextual enrichment in model training. For example, the Retrieval-Augmented GPT-4 model outperformed its standard counterparts by achieving a Micro F1 score of 82.24, compared to 73.84 for the non-augmented GPT-4. This enhancement can be attributed to the RAG method's ability to dynamically incorporate relevant external data during the generation process, thus enabling the model to make more informed predictions based on a broader knowledge base.

The application of the mCC APP for tracking eating windows [31,39], complemented by the annotation of pre-processed entity texts by human nutrition experts [40], has underscored the integration of technological tools in dietary research. NutriRAG's deployment in real-world scenarios further highlights its utility, as it categorizes participant-entered free-text food items logged in the mCC app relative to an established nutritional database to calculate detailed nutritional profiles. Beyond its immediate application in organizing the free-text entries for the mCC app, NutriRAG's capabilities extend to a broad array of technologies that handle free-text dietary input. This makes it an invaluable tool across various platforms where users log food intake, allowing for the transformation of raw textual entries into structured nutritional data. By categorizing free-text food descriptions into detailed nutritional profiles, NutriRAG can support diverse research and health monitoring applications, enhancing dietary tracking and nutritional insights in multiple technological environments.

To evaluate the clinical relevance of the NutriRAG's capabilities to classify food groups, we analyzed eating habits in response to different types of dietary restriction. Our research noted distinct dietary pattern shifts in both TRE and CRE groups. The TRE group experienced substantial reductions in meal consumption, including breakfast, lunch, and late-night eating. In contrast, the CRE group showed marked decreases in savory snacks and high-added-sugar foods. Our results underscore the efficacy of TRE and CRE approaches in lowering energy intake [41]. Moreover, we observed that higher number of breakfast EOs are associated with lower insulin sensitivity as evidenced by lower HOMA scores. These findings highlight the importance of the timing and frequency of food consumption as key factors influencing metabolic health, particularly insulin resistance.

By leveraging daily food intake data from the app tracker, our analysis with NutriRAG between baseline and follow-up data demonstrated that both TRE and CRE groups decreased their daily meal occasions. This study provides further evidence supporting the efficacy of TRE in modifying dietary behaviors, emphasizing the potential of targeted eating schedules in enhancing metabolic health outcomes.

This study introduces a novel retrieval-based LLM framework designed for identifying food items and classifying food categories, specifically aimed at analyzing personal eating patterns. The study demonstrates the potential of retrieval-augmented LLMs in nutritional analysis. However, it also recognizes a limitation: the dependency on high-quality data for effective retrieval, which can be challenging in situations where relevant external data are limited or biased.

## **CONCLUSION**

NutriRAG represents a significant step forward in the application of AI in nutrition and health sciences. By effectively combining the advanced capabilities of LLMs with the contextually rich data retrieval, NutriRAG enhances the precision of identifying food items and classifying their categories while presenting an LLM-driven method to analyze complex dietary patterns in diverse populations. Future studies should focus on refine these models further, ensuring they are accessible and effective across different demographic and clinical settings.

## **FUNDING STATEMENT**

This study was supported by the UMN CTSA Award UM1TR004405-01A1 from the National Center for Advancing Translational Sciences. Additional support came from National Institutes of Health grants R01DK124484 to LSC, and MRI support grants P41EB027061 and S10OD017974. Funding was also provided by the National Institutes of Health's National Center

for Complementary and Integrative Health grant number R01AT009457 and U01AT012871, National Institute on Aging grant number R01AG078154, National Cancer Institute grant number R01CA287413, and National Institute of Diabetes and Digestive and Kidney Diseases R01DK115629. The project was also supported by the UMN Institute for Diabetes, Obesity and Metabolism Pilot and Feasibility grant program. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

## CONTRIBUTORSHIP STATEMENT

HZ, LC, LH, RZ conceptualized and designed the study. HZ, LC, and LH curated the data and executed the experiments. HZ and LC drafted the initial manuscript, and all authors reviewed and finalized the manuscript.

## COMPETING INTERESTS STATEMENT

The authors state that they have no competing interests to declare.

## REFERENCE

1. Jeyavani, J., Sibiya, A., Sivakamavalli, J., Divya, M., Preetham, E., Vaseeharan, B., & Faggio, C. (2022). Phytotherapy and combined nanoformulations as a promising disease management in aquaculture: A review. *Aquaculture International*, 30(2), 1071–1086. <https://doi.org/10.1007/s10499-022-00848-0>
2. Delgado-Lista, J., Alcala-Diaz, J. F., Torres-Peña, J. D., Quintana-Navarro, G. M., Fuentes, F., Garcia-Rios, A., Ortiz-Morales, A. M., Gonzalez-Requero, A. I., Perez-Caballero, A. I., Yubero-Serrano, E. M., Rangel-Zuñiga, O. A., Camargo, A., Rodriguez-Cantalejo, F., Lopez-Segura, F., Badimon, L., Ordovas, J. M., Perez-Jimenez, F., Perez-Martinez, P., Lopez-Miranda, J., ... Yubero-Serrano, E. M. (2022). Long-term secondary prevention of cardiovascular disease with a Mediterranean diet and a low-fat diet (CORDIOPREV): A randomised controlled trial. *The Lancet*, 399(10338), 1876–1885. [https://doi.org/10.1016/S0140-6736\(22\)00122-2](https://doi.org/10.1016/S0140-6736(22)00122-2)
3. Khodami, B., Hatami, B., Yari, Z., Alavian, S. M., Sadeghi, A., Varkaneh, H. K., Santos, H. O., & Hekmatdoost, A. (2022). Effects of a low free sugar diet on the management of nonalcoholic fatty liver disease: A randomized clinical trial. *European Journal of Clinical Nutrition*, 76(7), 987–994. <https://doi.org/10.1038/s41430-022-01081-x>

4. Patnode, C. D., Redmond, N., Iacocca, M. O., & Henninger, M. (2022). Behavioral Counseling Interventions to Promote a Healthy Diet and Physical Activity for Cardiovascular Disease Prevention in Adults Without Known Cardiovascular Disease Risk Factors: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*, 328(4), 375. <https://doi.org/10.1001/jama.2022.7408>
5. Waddell, I. S., & Orfila, C. (2023). Dietary fiber in the prevention of obesity and obesity-related chronic diseases: From epidemiological evidence to potential molecular mechanisms. *Critical Reviews in Food Science and Nutrition*, 63(27), 8752–8767. <https://doi.org/10.1080/10408398.2022.2061909>
6. Browne, J. D., Boland, D. M., Baum, J. T., Ikemiya, K., Harris, Q., Phillips, M., Neufeld, E. V., Gomez, D., Goldman, P., & Dolezal, B. A. (2021). Lifestyle Modification Using a Wearable Biometric Ring and Guided Feedback Improve Sleep and Exercise Behaviors: A 12-Month Randomized, Placebo-Controlled Study. *Frontiers in Physiology*, 12, 777874. <https://doi.org/10.3389/fphys.2021.777874>
7. Eldridge, A., Piernas, C., Illner, A.-K., Gibney, M., Gurinović, M., De Vries, J., & Cade, J. (2018). Evaluation of New Technology-Based Tools for Dietary Intake Assessment—An ILSI Europe Dietary Intake and Exposure Task Force Evaluation. *Nutrients*, 11(1), 55. <https://doi.org/10.3390/nu11010055>
8. Murai, U., Tajima, R., Matsumoto, M., Sato, Y., Horie, S., Fujiwara, A., Koshida, E., Okada, E., Sumikura, T., Yokoyama, T., Ishikawa, M., Kurotani, K., & Takimoto, H. (2023). Validation of Dietary Intake Estimated by Web-Based Dietary Assessment Methods and Usability Using Dietary Records or 24-h Dietary Recalls: A Scoping Review. *Nutrients*, 15(8), 1816. <https://doi.org/10.3390/nu15081816>
9. Kaur, R., Kumar, R., & Gupta, M. (2023). Deep neural network for food image classification and nutrient identification: A systematic review. *Reviews in Endocrine and Metabolic Disorders*, 24(4), 633–653. <https://doi.org/10.1007/s11154-023-09795-4>
10. Jiang, L., Qiu, B., Liu, X., Huang, C., & Lin, K. (2020). DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. *IEEE Access*, 8, 47477–47489. <https://doi.org/10.1109/ACCESS.2020.2973625>
11. Liu, Y.-C., Onthoni, D. D., Mohapatra, S., Irianti, D., & Sahoo, P. K. (2022). Deep-Learning-Assisted Multi-Dish Food Recognition Application for Dietary Intake Reporting. *Electronics*, 11(10), 1626. <https://doi.org/10.3390/electronics11101626>
12. Madhumitha, S., Magimaa, M., Maniratnam, M., & Neelima, N. (2022). Dietary Assessment and Nutritional Analysis Using Deep Learning. In V. Bindhu, J. M. R. S. Tavares, & K.-L. Du (Eds.), *Proceedings of Third International Conference on Communication, Computing and Electronics Systems* (Vol. 844, pp. 11–21). Springer Singapore. [https://doi.org/10.1007/978-981-16-8862-1\\_2](https://doi.org/10.1007/978-981-16-8862-1_2)
13. Min, W., Wang, Z., Liu, Y., Luo, M., Kang, L., Wei, X., Wei, X., & Jiang, S. (2023). Large Scale Visual Food Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 9932–9949. <https://doi.org/10.1109/TPAMI.2023.3237871>
14. Mohanty, S. P., Singhal, G., Scuccimarra, E. A., Kebaili, D., Héritier, H., Boulanger, V., & Salathé, M. (2022). The Food Recognition Benchmark: Using Deep Learning to Recognize Food in Images. *Frontiers in Nutrition*, 9, 875143. <https://doi.org/10.3389/fnut.2022.875143>

15. Mukhiddinov, M., Muminov, A., & Cho, J. (2022). Improved Classification Approach for Fruits and Vegetables Freshness Based on Deep Learning. *Sensors*, 22(21), 8192. <https://doi.org/10.3390/s22218192>
16. Hu, G., Ahmed, M., & L'Abbé, M. R. (2023). Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods. *The American Journal of Clinical Nutrition*, 117(3), 553–563. <https://doi.org/10.1016/j.ajcnut.2022.11.022>
17. Ma, P., Zhang, Z., Li, Y., Yu, N., Sheng, J., Küçük McGinty, H., Wang, Q., & Ahuja, J. K. C. (2022). Deep learning accurately predicts food categories and nutrients based on ingredient statements. *Food Chemistry*, 391, 133243. <https://doi.org/10.1016/j.foodchem.2022.133243>
18. Shankar, K., Chandrasekaran, R., Jeripity Venkata, P., & Miketinas, D. (2023). Investigating the Role of Nutrition in Enhancing Immunity During the COVID-19 Pandemic: Twitter Text-Mining Analysis. *Journal of Medical Internet Research*, 25, e47328. <https://doi.org/10.2196/47328>
19. Hu, G., Ahmed, M., & L'Abbé, M. R. (2023). Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods. *The American Journal of Clinical Nutrition*, 117(3), 553–563. <https://doi.org/10.1016/j.ajcnut.2022.11.022>
20. Ahmed, M., Schermel, A., Lee, J., Weippert, M., Franco-Arellano, B., & L'Abbé, M. (2022). Development of the Food Label Information Program: A Comprehensive Canadian Branded Food Composition Database. *Frontiers in Nutrition*, 8, 825050. <https://doi.org/10.3389/fnut.2021.825050>
21. Kretser, A., Murphy, D., & Starke-Reed, P. (2017). A partnership for public health: USDA branded food products database. *Journal of Food Composition and Analysis*, 64, 10–12. <https://doi.org/10.1016/j.jfca.2017.07.019>
22. Van Erp, M., Reynolds, C., Maynard, D., Starke, A., Ibáñez Martín, R., Andres, F., Leite, M. C. A., Alvarez De Toledo, D., Schmidt Rivera, X., Trattner, C., Brewer, S., Adriano Martins, C., Kluczkowski, A., Frankowska, A., Bridle, S., Levy, R. B., Rauber, F., Tereza Da Silva, J., & Bosma, U. (2021). Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food. *Frontiers in Artificial Intelligence*, 3, 621577. <https://doi.org/10.3389/frai.2020.621577>
23. Zhou, H., Austin, R., Lu, S.-C., Silverman, G. M., Zhou, Y., Kilicoglu, H., Xu, H., & Zhang, R. (2023). Complementary and Integrative Health Information in the literature: Its lexicon and named entity recognition. *Journal of the American Medical Informatics Association*, ocad216. <https://doi.org/10.1093/jamia/ocad216>
24. Clusmann, J., Kolbinger, F. R., Muti, H. S., Carrero, Z. I., Eckardt, J.-N., Laleh, N. G., Löffler, C. M. L., Schwarzkopf, S.-C., Unger, M., Veldhuizen, G. P., Wagner, S. J., & Kather, J. N. (2023). The future landscape of large language models in medicine. *Communications Medicine*, 3(1), 141. <https://doi.org/10.1038/s43856-023-00370-1>
25. Li, M., Chen, M., Zhou, H., & Zhang, R. (2023). PeTailor: Improving Large Language Model by Tailored Chunk Scorer in Biomedical Triple Extraction. <https://doi.org/10.48550/ARXIV.2310.18463>
26. Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., & Dagan, A. (2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digital Health*, 2(2), e0000205. <https://doi.org/10.1371/journal.pdig.0000205>



27. Tang, L., Sun, Z., Idnay, B., Nestor, J. G., Soroush, A., Elias, P. A., Xu, Z., Ding, Y., Durrett, G., Rousseau, J. F., Weng, C., & Peng, Y. (2023). Evaluating large language models on medical evidence summarization. *Npj Digital Medicine*, 6(1), 158. <https://doi.org/10.1038/s41746-023-00896-7>
28. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930-1940. <https://doi.org/10.1038/s41591-023-02448-8>
29. Zhou, H., Li, M., Xiao, Y., Yang, H., & Zhang, R. (2023). LLM Instruction-Example Adaptive Prompting (LEAP) Framework for Clinical Relation Extraction [Preprint]. *Health Informatics*. <https://doi.org/10.1101/2023.12.15.23300059>
30. Yang, H., Li, M., Zhou, H., Xiao, Y., Fang, Q., & Zhang, R. (2023). One LLM is not Enough: Harnessing the Power of Ensemble Learning for Medical Question Answering [Preprint]. *Health Informatics*. <https://doi.org/10.1101/2023.12.21.23300380>
31. Oldenburg, N., Mashek, D. G., Harnack, L., Wang, Q., Manoogian, E. N. C., Evanoff, N., Dengel, D. R., Taddese, A., Yentzer, B. P., Lysne, L., Wong, A., Hanson, M., Anderson, J. D., Alvear, A., LaPage, N., Ryder, J., Varady, K., Gao, Z., Ryu, S., ... Chow, L. S. (2025). Time-restricted eating, caloric reduction, and unrestricted eating effects on weight and metabolism: A randomized trial. *Obesity*, 33(2), 242-252. <https://doi.org/10.1002/oby.24252>
32. Gill, S., & Panda, S. (2015). A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits. *Cell Metabolism*, 22(5), 789-798. <https://doi.org/10.1016/j.cmet.2015.09.005>
33. Leech, R. M., Worsley, A., Timperio, A., & McNaughton, S. A. (2015). Characterizing eating patterns: A comparison of eating occasion definitions. *The American Journal of Clinical Nutrition*, 102(5), 1229-1237. <https://doi.org/10.3945/ajcn.115.114660>
34. D. S., Casas, D. de las, Hanna, E. B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Lavaud, L. R., Saulnier, L., Lachaux, M.-A., Stock, P., Subramanian, S., Yang, S., ... Sayed, W. E. (2024). Mixtral of Experts (Version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2401.04088>
35. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. <https://doi.org/10.48550/ARXIV.2307.09288>
36. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Version 2). *arXiv*. <https://doi.org/10.48550/ARXIV.1810.04805>
37. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240. <https://doi.org/10.1093/bioinformatics/btz682>
38. Peng, Y., Yan, S., & Lu, Z. (2019). Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets (Version 2). *arXiv*. <https://doi.org/10.48550/ARXIV.1906.05474>
39. Chow, L. S., Manoogian, E. N. C., Alvear, A., Fleischer, J. G., Thor, H., Dietsche, K., Wang, Q., Hodges, J. S., Esch, N., Malaeb, S., Harindhanavudhi, T., Nair, K. S., Panda,

- S., & Mashek, D. G. (2020). Time-Restricted Eating Effects on Body Composition and Metabolic Measures in Humans who are Overweight: A Feasibility Study. *Obesity*, 28(5), 860–869. <https://doi.org/10.1002/oby.22756>
40. Phillips, N., Mareschal, J., Schwab, N., Manoogian, E., Borloz, S., Ostinelli, G., Gauthier-Jaques, A., Umwali, S., Gonzalez Rodriguez, E., Aeberli, D., Hans, D., Panda, S., Rodondi, N., Naef, F., & Collet, T.-H. (2021). The Effects of Time-Restricted Eating versus Standard Dietary Advice on Weight, Metabolic Health and the Consumption of Processed Food: A Pragmatic Randomised Controlled Trial in Community-Based Adults. *Nutrients*, 13(3), 1042. <https://doi.org/10.3390/nu13031042>
41. Kang, J., Ratamess, N. A., Faigenbaum, A. D., Bush, J. A., Beller, N., Vargas, A., Fardman, B., & Andriopoulos, T. (2022). Effect of Time-Restricted Feeding on Anthropometric, Metabolic, and Fitness Parameters: A Systematic Review. *Journal of the American Nutrition Association*, 41(8), 810–825. <https://doi.org/10.1080/07315724.2021.1958719>
42. Apergi, K., Karatzi, K., Reppas, K., Karaglanis, E., Usheva, N., Giménez-Legarre, N., ... & Makrilakis, K. (2022). Association of breakfast consumption frequency with fasting glucose and insulin sensitivity/b cells function (HOMA-IR) in adults from high-risk families for type 2 diabetes in Europe: the Feel4Diabetes Study. *European Journal of Clinical Nutrition*, 76(11), 1600-1610.
43. Chowdhury, E. A., Richardson, J. D., Holman, G. D., Tsintzas, K., Thompson, D., & Betts, J. A. (2016). The causal role of breakfast in energy balance and health: a randomized controlled trial in obese adults. *The American journal of clinical nutrition*, 103(3), 747-756.
44. Karatzi, K., Moschonis, G., Barouti, A. A., Lionis, C., Chrousos, G. P., & Manios, Y. (2014). Dietary patterns and breakfast consumption in relation to insulin resistance in children. The Healthy Growth Study. *Public health nutrition*, 17(12), 2790-2797.