Usage of data science techniques to personalize and optimize nutrition recommendations and information via a micronutrient focused application

Master Thesis in the 4th semester
at the program M.Sc. Business Data Science
AAU Business School
Spring Semester 2024
written by Nils Bayer

AALBORG UNIVERSITY
STUDENT REPORT

AAU Business School
Aalborg University
http://www.aau.dk

Title:
Usage of data science techniques to personalize and optimize nutrition recommendations and information via a micronutrient focused application

Theme:
Differentiating a novel nutrition application product with artificial intelligence technologies in a viable manner

Project Period:
Spring Semester 2024

Participant(s):
Nils Bayer

Supervisor(s):
Roman Jurowetzki

Program:
M.Sc. Business Data Science

Number of strokes:
122,793

Page Numbers: 64

Date of Completion:
June 2, 2024

Abstract

This paper explores the usage of data science techniques, like retrieval augmented generation chatbots, graphRAG, object detection machine learning models, and generally LLMs as well as speech-to-text to create a new nutrition application. This new product shall be able to offer a new value proposition and thus, differentiate itself from preexisting solutions. After regarding prior research in the field, publishing and analyzing a survey of potential users, and talking to an actual nutritionist, a difference can be drawn between an aesthetical fitness and a health focus in applications. While most products in the market concentrate on the first one, this paper's application represents a health- and micronutrient-focused solution. The central element is a food recommendation system based on the concept of vector similarity. This way, the application is not only a food tracker, but actually offers more functionality, namely the recommendation of foods based on their nutrient profile. Interpreted from the survey, the logging of foods was furthermore perceived as a pain point in usage of other applications. Hence, computer vision and speech-to-text was successfully used to offer an alternative to the slower, manual process of typing in food names. Throughout, LLMs are a central technology to quickly implement new artificial intelligence-based functionalities, whether solely text-based or multimodal.

Contents

List of Figures

1 Introduction

Nutrition applications allowing users to track their diet and helping them to lose weight or gain muscle are data intensive. Every user has an individual profile of food preferences, lifestyle choices, and health statuses. Beyond the user, there are thousands of food items to choose from every single day, each with a unique nutrient profile and production history. Finding the ideal diet is therefore a very individual process dependent on a large variety of data points.

The most known nutrition apps, like MyFitnessPal, are however breaking down this complexity by focusing on macronutrients and calories - fats, carbohydrates, proteins and the energy we consume each day. While this simplification supports users that have goals to primarily track their calorie intake to lose or gain weight, users interested in their health and whether they cover all their micronutrient needs are left out or need to find this information in deeper, almost hidden interfaces.

Therefore, this paper explores the opportunity to build a more health-centered nutrition application. It focuses on the tracking of micronutrients and formulates food recommendations based on what users have already consumed in a day and what they should when looking at micronutrient intake recommendations by the World Health Organization (WHO) or the European Food Safety Authority (EFSA). These recommendations are based on vector similarity search where each vector is representing foods and their nutrients or the user's nutrient demand.

Beyond being a tool for analysis, the application shall also offer the possibility for the users to learn more about nutrition. Therefore, a chatbot is implemented which has access to academic nutrition papers and thus, can formulate science-based explanations and advice.

Moreover, the application shall be developed and run in close collaboration with nutritionists and dietitians, increasing its quality and value to the user. These nutrition professionals are also expected to create more content in the application whose display to the user can be personalized by the user's interactions with the application.

Research approach and the concept of viability

Beyond the practical implementation of a new nutrition application, this research paper deals with the usage of data science techniques and especially, artificial intelligence (AI) models, and the connected concept of viability. Common pain points in nutrition applications, like the time intensiveness of food logging, are explored and data science techniques, like computer vision or speech-to-text leveraged to decrease these.

The central question is how more net value can be viably generated to potential users by using these different data science technologies. The benefit of using a certain technology as well as necessary resources for its implementation are being investigated and a conclusion drawn whether it is worth implementing, especially in the context of a new entrepreneurial venture. Starting a new business and finding a product-market-fit requires fast iteration and hence, the research question is stated as the following:

"How can modern data science practices and the usage of artificial intelligence augment the value of nutrition apps and what is a quickly-validating and viable approach to building one?". This is especially interesting against the background that the emerging Large Language Models (LLM) have capabilities which required specific, and highly fine-tuned machine learning models whose creation was time and capital intensive. These possibilities are nowadays only an API call away and very affordable, especially with the increasing variety of providers and models.

Digital affordances, like the internet, made new entrepreneurs independent of economic clusters, and intermediaries as well as granted them more freedom and control over directly accessing and interacting with customers. This changed the amount and variety of entrepreneurial opportunities (Autio et al., 2018).

In this paper, the importance and great variety of use cases of LLMs is being displayed. Beyond serving as text-only machine learning models, multimodal LLMs or LLMs in connection to e.g. speech-to-text technology enable the processing of other data types as well, like e.g. images and audio, and hence, allow to build complex and smart applications with fewer knowledge and time investment. Specific and complex models, and their training processes are broken down to prompts consisting of human everyday's language. Just like other technologies described by Autio et al. (2018), the emergence of LLMs helps to break complexity and constraints and enables different types of entrepreneurship and the pursuit of greater entrepreneurial opportunities. Applications, like this paper's nutrition application, would have required bigger teams of engineers with knowledge in the different machine learning model types and distinct modalities to quickly create a well-working product. With the use of versatile LLMs, trained, hosted and run by third party services, this operation can be limited to fewer engineers, even only one, enhancing the chances for smaller teams and solopreneurs.

2 Literature review

The idea of a nutrition app or food tracker is not completely novel and neither is the usage of artificial intelligence and data science in nutrition. Especially in the recent past, use cases to leverage data science to enable new applications and enhance value delivered to patients was covered in academic research.

In the paper by Yang et al. (2024), ChatDiet, a chatbot to learn and inform yourself about your own optimal diet, is presented. Important aspects to the researchers to enable through their work on ChatDiet are a high degree of personalization of the recommendations, their explainability as well as good interactivity of the entire system which primarily means good responses to follow-up questions. Beyond using a LLM, more specifically OpenAI's gpt-3.5-turbo, two models are used to classify which foods will be recommended. One of these models focuses on population data while the other one is based on personal data. The results of these will be fed to the LLM which consequently, will create the response in human-readable natural text. The data needed for this infrastructure includes food items and their nutritional characteristics as well as data from a wearable device, namely an Oura ring, tracking user's health data. Interestingly, also a causal knowledge graph based on tracked user data is used to understand logical connections between items, however the main importance lies in a Retrieval-Augmented Generation (RAG) system. To evaluate their system, Yang et al. (2024) are manually defining queries and expected answers to these. Consequently, they check how far the generated results by their system align with the expected one.

The paper by Farrokhi et al. (2021) deals with the topic of smart fitness. It presents different Internet of Things (IoT) devices as well as artificial intelligence and how value can be created through their effective use. While the researchers primarily focus on fitness tracking and recommendation for physical exercise, the paper also includes macronutrient-based food tracking applications as well as the topic of diet prediction and timing. Its authors lay out the different paradigms of supervised, unsupervised and reinforcement learning and show different use cases of model types of all of these categories. While the paper includes a great part of what could be referred to as additional context for this paper's use case, it is impressive to see the complexity and potential growth plans for a nutrition based app and how closely they can be linked to each other. It supports the mentioned need and nowadays' viability of personalization in nutrition expressed by Yang et al. (2024) and shows how it can also work in real-time, especially with IoT devices in use.

The complex nature of nutrition and overall health recommendations and their personalization are also covered in the paper by Verma et al. (2018). The researchers focus on the challenges of implementing a data-driven approach leveraging AI, the required technical infrastructure and its load when computing personalized results with personalized models, the need of data standardization and training for staff working with the data, as well as data sparsity and usefully imputing missing data. They describe the use of AI technologies as promising, but also warn about models' biases and privacy concerns. These challenges are great aspects to consider when planning to build a new application and the concern shall be governed effectively to ensure trust in potential users.

AI is also used in a more clinical setting to assess and manage nutrition of patients as laid out by Sudersanadas (2021). It also deals with the personalized approach to food recommendation, especially based on the patients' nutrient intake. The author highlights that commercial applications do exist, however, they are not based on nutrients and their data sources are not validated. Also complexity is mentioned, especially in the context of hospitalized patients with more complicated health conditions.

A systematic review of the use of AI in nutrition science is created by Theodore Armand et al. (2024). The researchers showcase the great breadth of application use cases "from dietary assessment to food recognition and tracking, predictive modeling for disease, disease diagnosis and monitoring, and personalized nutrition" (Theodore Armand et al., 2024). Through the review, the authors found that the number of relevant studies increased over the recent years. Beyond effectively dealing with the complex data from a person's diet, health conditions and lifestyle (like expanded on also in the other papers), AI is also used to simplify the food logging process. To achieve this, computer vision models are implemented to automatically spot different food items.
Besides showcasing the great potential and the already published works on AI in nutrition science, the authors do highlight that most existing applications are prototypes or very early in their development and hence, these more modern systems do not have a long track record of real world user experiences and product improving iterations.

Bond et al. (2023) focus on the application of AI in clinical nutrition. As other papers mentioned above, the authors state the great ability of AI models to analyze large and complex datasets. This capability enables them to find new links between nutrition and diseases to prevent diseases or help in managing them well. Real-time and interactive applications could also help with motivating users to choose healthier food options. Beyond the promising side of AI in nutrition, the researchers also call for attentiveness and careful usage since concerns like data privacy issues, biases in models, lacking user trust and little research density need to be addressed.

Miyazawa et al. (2022), similarly to papers above, also mention the use of computer vision technologies in nutrition. They state that self-reported food and nutrient intake can be prone to errors and therefore, a smart intake solution is highly beneficial. In that context, they specifically talk about the tracking of alcohol consumption, an angle not discussed in papers above. An interesting and powerful model mentioned in the paper is called goFOOD™ which uses computer vision and 3D calculations to estimate nutrient and calorie amounts in photographed foods. The researchers state that this model even outperformed trained professional nutritionists in estimating nutrient amounts. Beyond food logging, they also discuss the use of AI in food safety and production and the increasing importance of these technologies in the new commencing field of multidisciplinary research.

Knights et al. (2023) focus on obesity and the prediction of health outcomes. In their paper, they mention the grand rise in data collection in nutrition and thus, develop a machine learning model to forecast weight loss progress based on 200 study participants' anthropometric and biochemical data as well as age, and gender. By using a Deep Neural Network as well as a RandomForest model, the researchers successfully forecast weight loss results and hence, discuss the usage of such a model in everyday's nutrition practice. Beyond helping individuals with their diet plans, the authors mention the possibility to also optimize populations' dietary guidelines. While the field is rapidly evolving, the researchers advise to use this technology in combination with professional guidance, like doctors and nutritionists.

Balloccu et al. (2024) talk about the data scarcity in the field of nutrition, specifically for nutrition counseling. Since most data lies within private companies which hold their data confidential, the motivation of the paper is to create a dataset of nutrition counseling dialogues. For its creation crowd workers, nutritionists and OpenAI's LLM is leveraged. The initial input of patients' concerns comes from crowd workers whose contributions are checked by experienced nutritionists. This curated data is then used to impute similar data with the help of the LLM. While data was successfully created in a great amount, professional nutritionists critiqued the behavior of the LLM for giving counseling advice due to reinforcing harmful stereotypes as well as lack of true patient understanding.

Zheng et al. (2023) deals with building a healthy nutrition supporting application including sophisticated computer vision technology for food detection and resulting food logging. The application is focused on the population of Singapore. Interestingly, the food detection model does not only provide food items' names, but also nutrient amounts directly based on the uploaded image by the user. In the paper, the authors state that food images can vary vastly in different locations and that there are two major issues. Intra-class variation is the problem of dishes of the same class looking differently. The researchers show examples of sushi, which can vary strongly in looks, but not much in contained food items and hence nutrients. The second mentioned issue is inter-class resemblance which means foods that do look alike, but are actually very distinct in contained food items and nutrients. A good example might be sushi with salmon and grilled salmon.
Beyond this tracking and analysis functionality, the application also offers a chat functionality where the user can talk to professional nutritionists as well as to other users that have similar health conditions or interests.

Han et al. (2023) create, as well as other researchers above, a computer vision model to accurately estimate the amount of nutrients present in a photographed dish. Similarly to Miyazawa et al. (2022) they calculate a 3D representation of the food items, based on a depth map made available by using a machine learning model. They state that their model's performance is comparable with the one of best-practice, state-of-the-art models when evaluating the model on the so-called Nutrition5K image dataset. However, food detection, especially if willing to estimate nutrients immediately, common and difficult issues are food stacking, basically food items being hidden by other overlapping items, as well as mostly invisible ingredients, such as oil and sugar. These food items do contain great amounts of calories, however can often not be spotted by pure vision.

Summarized, and most important aspects relevant for this paper's project
Research in applying data science technologies in the field of nutrition includes a great range of different use cases as well as various types of machine learning models. However, many papers deal with similar aspects, issues and name same motivations and value propositions.
The probably most mentioned detail is personalization of nutritional advice. The usage of AI enables applications to not only apply general best practice, but also find specific advice tailored to an individual's anthropometric and biochemical data as well as cultural background and taste. Since experiences with nutrition and health can vary strongly from person to person due to e.g. different genes, an individualized approach seems more effective and thus, more valuable. Therefore, for this

paper's application, an important learning is to enable and showcase a solution to the potential customer that is tailored to the individual and therefore, can apply high quality advice.

Beyond personalized recommendations, real-time dietary advice is highly applicable. Instead of merely suggesting generally healthy food choices, it is valuable to recommend specific foods at particular times. Nutritional needs vary throughout the day based on what has already been consumed. For instance, if an individual has already consumed sufficient amounts of a specific nutrient, the application should recommend foods containing other necessary nutrients that have not been adequately consumed. This approach ensures that nutritional advice is timely and tailored to the individual's current dietary needs.

The aspect of nutrients is another interesting point of interest. Han et al. (2023) explicitly express that most known applications in the food recommendation and tracking field are focused on calories and macronutrients. Calories and macronutrients are important to these apps' customers which primarily face issues like being overweight or the struggle to grow muscle. While these are legitimate customer problems to be tackled, these apps do not put a great spotlight on a general, healthy nutrition covering all micronutrients. They can be seen as fitness apps, but not really as health-focused apps. This rather novel differentiation shall be achieved in this paper's application.

One very dominant factor in creating the application is data that all the machine learning models as well as the general functioning of the application is based upon. Multiple papers mention the lack of dense and high qualitative food log data sources to train useful models applicable in real-world scenarios. Due to the distributed and non-exhaustive data existence, official sources should be considered in the best case. Therefore, for this paper's application and nutrient estimations, nutrient data of foods from the United States department of agriculture (USDA) is being used.

Regarding data science techniques, which are applicable as viable-sounding possibilities to this paper's case, two options are mostly covered. Large language models enable valuable and interactive chatbot applications that with the use of RAG, like applied in the paper by Yang et al. (2024), can provide nutritional advice 24 hours everyday. The simple usage of the out-of-the-box LLM however is partially questionable, as experienced by Balloccu et al. (2024). From their work, also the value-add of a collaboration with professional nutritionists becomes clear. They can assess the application's behavior and like in the case of Zheng et al. (2023) be an important, supporting element of the application itself. For the necessary food logging process, computer vision models seem to generate a lot of value through automating a very manual process. A highly sophisticated food detection model based on computer vision, however, does seem complex, difficult and not very viable for a first, quick iteration of a low-resource venture. Yet computer vision shall also be used in this paper's application due to its high value proposition. An interesting experiment in this context to be executed is also the use of a multimodal LLM which because of its enormous training data might be able to solve issues of invisible ingredients in photographed dishes which cannot be spotted by models only relying on visual input (OpenAI, 2024a).

The last important aspect and take-away from previous literature is the importance of transparency and explainability. Especially in the paper by Yang et al. (2024), these qualities were highlighted. For the users to fully trust the application and its recommendations, it has to be understood where the data comes from, how entered user data is treated and stored as well as how recommendations are calculated or advice reasoned. Of course, due to the majority of end-users not having a deep technical background, especially in regards to machine learning, these workflows and IT infrastructures have to be broken down to everyday's language. Ways to find information should be easily accessible.

3 Commercialization and product differentiation

To be able to discuss the value proposition of a product, it is relevant to establish and evaluate a potential business model. It describes the system how the application, the product, can be run economically and thus, how value can be sustainably generated for the users. Important, initial questions are to whom the product is being catered, how it is differentiated from competitors, how it can be monetized and what investment is needed to start and run the application.

3.1 Lean canvas (business model canvas)

The lean canvas is an alteration of the business model canvas introduced by Osterwalder and Pigneur (2010) in their book called Business Model Generation. The lean canvas retains the same structure as the business model canvas. However, it does exchange certain aspects to fit better to very early-stage business endeavors. It primarily focuses on the problem solving aspect and the competitive advantage. These canvases are created with the goal to effectively communicate a new endeavor's business model and value proposition. They help with creating an overview of the different aspects in a business, like e.g. the customer, the costs, the income channels, as well as with finding a strategy suitable for the planned endeavor and its ecosystem. Recognizing that long, more traditional business plans are not being fully read and not enjoyed when written either, a quicker concept of analysis and presentation was required. Additionally, since most elaborations in a business plan are based on initial assumptions which will inevitably be overwritten by real market feedback, the lean canvas portrays a simpler, and more flexible method. It can be quickly adapted and hence, used to run iterations of continuous innovation in an uncertain environment. Entrepreneurs tend to be solution-centered and obsessive over their product. The lean canvas moves the attention to a higher conceptual level to understand and solve customer problems effectively (Maurya, 2011).
Due to the entrepreneurial nature of the practical implementation of this paper, the lean canvas serves as a great tool to easily create an overview and a common understanding for how the new software can solve a relevant problem as well as to whom. It shows the entirety of the business and sets the stage to fully grasp the value generation which is valid in a specific environment of conditions being true.

3.1.1 The problem

Health-conscious individuals are inspecting and optimizing their behavior to live a healthier life and strive for longevity. Nutrition is a great part of this equation and thus, different diets are being tested upon new insights in research. When changing a

diet and restricting food options, individuals might encounter worries regarding the lack of nutrients and wonder whether they need to supplement vitamins, minerals or other micronutrients.
Especially at the beginning of a new diet, the individual needs an adaptation time to find new recipes supporting an overall balanced diet and thus, a tracking and recommendation system is of great need.
This problem is enlarged if considering cultural backgrounds and personal liking regarding the recommendation of different foods. An easy, one-fits-all approach is less valuable in this setting.

3.1.2 The solution

The solution to the above stated problem is an app including a tracking and recommendation system as well as a self-learning platform. To make sure all nutrients are covered by the daily intake of foods and beverages, the individual has to track the items. This shall be a smooth and not time-intensive process. Based on the data which foods and thus nutrients were already consumed, the app could give recipe recommendations to fit nutrient needs as well as personal taste and incorporate recipes from someone's cultural background.
This personal system is especially needed if, moreover, the individual has a certain health condition potentially altering recommendations for nutrient intake.
For these self-optimizing individuals who are interested in learning more about nutrition, a tutor or companion in the form of a chatbot would also be of great value, especially if it knows about nutrition papers and links between foods, micronutrients, organs, and health statuses.
Beyond this interactive information functionality, also small curated articles can be featured in the app. These can be written by cooperating nutritionists and thus, offer high quality insights and best practices.

3.1.3 Unique value proposition

While most well-known nutrition tracking apps or more commonly called calorie trackers, are focused on tracking calories and macronutrients (protein, carbohydrates and fats) to help users with weight loss or muscle gain, there is less supply for an approach focused on micronutrients and nutrient deficiency prevention. In this manner, also most recipe recommendations are shown primarily based on popularity or likings or generally just as they are considered healthy. In a micronutrient based approach recipes can be recommended on the current needs of the user's body alongside taste and popularity.
Another component which can be included in the app can be the ecological footprint

of the foods to be consumed. Besides nutrient coverage, popularity and taste, factors for recommendations can also be local seasonality and average water consumption per food.

3.1.4 Customer segment

The last aspect above works well with the target customer group in mind. Initially, vegan and vegetarian individuals will be in focus. They do actively manage their diet and choose to exclude certain foods, potentially, if done wrong, leading to a nutrient deficiency. Especially new vegans and vegetarians could be interesting prospects since they are not used to the new diet and might not have tried a lot of new foods and collected knowledge in the topic. This can increase the chances for having worries and being willing to use an app to help with the transition.

As indicated above, vegans and vegetarians are moreover also not only not eating meat or animal products in general, but also, depending on their reasoning, are interested in having a lower ecological footprint and choosing more sustainable ways to live. Thus, an app that connects the aspects of nutrition and its ecological impacts might have an even bigger attractiveness.

3.1.5 Unfair advantage

Since the concept of a nutrition app or a nutrient tracker is not novel, unfair advantages are of high value. I am a data science student with web development experience and valuable contacts in nutrition studies which are essential for executing this idea. The concept of a nutrition app is data-driven and requires a lot of domain as well as technical expertise. While there are common, open access data sources, supplementary data can be leveraged to gain an edge as well as expertise of contacts or business partners. To better understand the target group and the associated worry of lacking micronutrients, a deeper understanding and empathy for plant-based people is beneficial. Marketing messages as well as features for the application can be better formulated and prioritized.

3.1.6 Channels

There are different methods to reach the above mentioned audience. Organic, not paid ways include leveraging platforms like Reddit, blogs or community websites based around the topic of plant-based nutrition. Valuable contributions on these platforms can generate trust and thus, enable a promotion of the own product. Since many individuals are looking for answers to a correctly executed plant-based diet online, search engine optimization (SEO) as well as social media posts can generate a lot of traffic to the nutrition app.

Besides organic marketing methods, also paid opportunities can be used. This can include paid ads on social media and search engines, like Facebook and Google. People actively searching for nutrient supplements (e.g. for magnesium or vitamin B-12) might experience the worry or the actual problem of lacking micronutrients and hence, might be interested in an app guiding them along. Moreover, there are many creators on social media promoting plant-based diets. These figures have authority in the space and a loyal following which through promotions (which are usually paid) can be leveraged.

3.1.7 Key metrics

Beyond generating a revenue stream from the app, usage is probably one of the most effective measures to check whether there is demand for the app and whether it can actually generate value and solve a problem. Usage is normally tracked in how often users interact with the software (this includes logins, button clicks etc.). By tracking this data, a business can tell how much percent of the users are monthly or even daily active users. The higher this percentage, the better the app performs.

3.1.8 Cost structure

To build an app with a chatbot and a data-savvy backend there are multiple cost centers to be named. For the technical realization, a server is needed to run the application programming interface (API) and a web-application version which will be accessible through the web browser. The expenses for this can vary extremely depending on the supplier. Cloud computing of big enterprises like Amazon Web Services (AWS) are exhaustive, but expensive. Smaller web hosting companies in contrast offer a smaller product range, however are a lot cheaper. A linux server of such a supplier can only cost 5-20 Euro per month. To realize a sophisticated chatbot nowadays, the use of a LLM is obligatory. You can host such a model yourself, however due to its expensive requirements in GPU power, this will not end up cheap. On the other hand, you can use providers like OpenAI, Google or together.ai. These companies price based on usage, are usually easy to use and quick to implement, making them attractive options for validating a business idea without great overhead structure.
For a nutrition app, data on nutrient values of a lot of foods is required. While there are open-access and free databases (e.g. data from the USDA), data quality or ease of use can be common issues. Leveraging paid data providers later in the process can thus make a lot of sense.
Beyond these costs for technicalities, there are further business expenses to be expected. These include the creation of contracts, potentially needed freelance con-

tractors or employees as well as marketing budgets for the above mentioned paid marketing techniques.

3.1.9 Revenue streams

A nutrition app can be monetized in different manners since it can be interesting to B2C as well as B2B parties. B2C customers are the ones actually using the app to track their nutrition and they could be monetized by using a subscription model. To increase the general number of sign-ups a freemium model can be considered which means offering free components of the app and only charging for certain features.
B2B parties can be nutritionists and dietitians that would like to add value to their offerings by including a subscription of such an app. Furthermore, the app could be white labeled for these customers, meaning colors and the design is adapted to the B2B customer's corporate identity, creating the illusion that the B2B customer created the app and thus, seems to be more exclusive.
Alternatively, both approaches could be merged. Simple B2C customers who found the app by themselves pay a subscription fee while customers using the app through their nutritionist are not paying directly to the app maker, but through the third party, namely the nutritionist.
Nutritionists can also be listed in the app for B2C customers if they decide to consult a professional. This could generate leads to B2B customers and the business could generate revenue through a commission or a pay per session booked. Quality of the nutritionists' insights and experience in nutrition counseling can also be assessed by the users who can interact with nutritionist curated recipes and articles.

3.2 Activity in the space, show competitors

Since this is not the first nutrition app helping users to follow a better diet and reach their nutrition goals, it is important to name competitors and analyze their value proposition in the marketplace. In many conversations with survey respondents, nutritionists and users of nutrition apps, the application called MyFitnessPal was mentioned. Even for calorie tracking tasks at a university's dietitian program, the application was recommended. However, while it seems to be the primarily known nutrition application, there are also others, for instance FatSecret, My Diet Coach, Cronometer and MacroFactor. A great red line that can be seen in most of these applications, is the focus on fitness in an aesthetical sense. The main value propositions communicated by these apps center around getting in the desired physical body shape, whether that might be losing fat, building muscle or both. Moreover, these applications are usually generalized in regards to diets. The applications are offered for people following a keto, low-carb, and low-fat and many more diets.

While a generalized application can cater a bigger, overall market, a potential user following a specific diet could identify better with a product focused on their specific diet and resulting conditions. This is especially true with vegans as well as vegetarians because their motivations usually are not centered around aesthetics. A plant-based diet (including vegans, vegetarians and flexitarians) is usually chosen for cardiovascular and other health benefits, ecological sustainability, and ethical concerns regarding the killing of animals and animal products. These are aspects which will be interesting to be addressed in a plant-based focused nutrition application, however, might rather be seen as random or irrelevant to the great majority of users of the fitness centered nutrition applications. To illustrate this positioning, figure 1 is mapping out the above-stated competitors as well as this paper's application on a coordinate system. Its x-axis describes the focus of the application on aesthetical fitness versus on more holistic health. Its y-axis describes the differentiation of the applications in terms of their focus on specific diets. While this map is not exhaustive, the desired differentiation of this paper's application is visible.

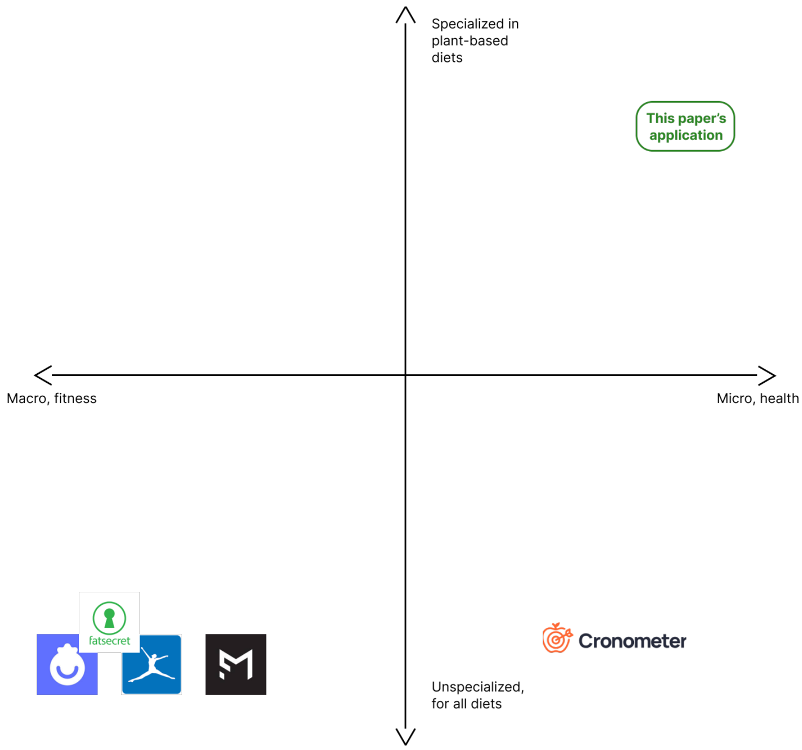


Figure 1: Competitor map for nutrition applications

3.3 Survey results

In the context of this paper, a survey was created to get quantitative feedback from potential future users. Beyond understanding how a new nutrition application with new AI functionalities could be run as a valid business and analyzing this and its

differentiated value proposition in a business model representation, data from potential, real world users which fit the described customer segment from above, is of high importance to underline statements with actual insights. With the concept of viability as a main research focus, which is the interplay of technical attainability and worth to the user, survey data can be used to identify potential user problems and liking and hence, deliver supporting evidence over what can be considered more generated value.
This merger of a quantitative method with more qualitative assessments of a practical implementation can be seen as an approach of mixed, inter-supporting methods. The survey questions can be found in the Appendix B and the raw data in the supplementary zip folder. Data was collected by reaching out to personal contacts as well as publishing the survey link in a group for vegans and people eating a plant-based diet on the website Reddit. Since this paper's application shall be differentiated and especially catered to a plant-based target market, it is great to see that the survey was welcomed and in fact, 87.5% of the 72 respondents follow a primary plant-based diet, including vegans, vegetarians as well as flexitarians. Also a sufficient part of the respondents have suffered from nutrition-related issues which solution shall be assisted by the use of this paper's application.

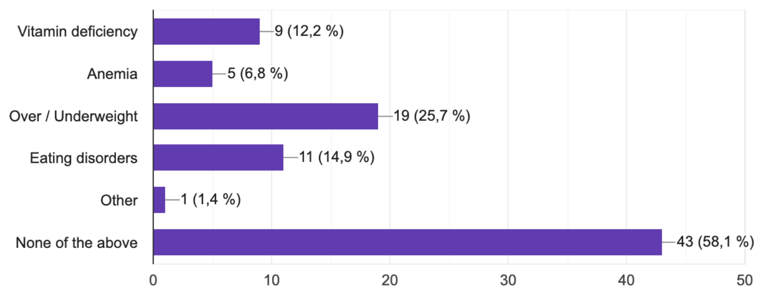


Figure 2: Distribution of nutrition-related issues

Vitamin deficiency, anemia and eating disorders are especially relevant problems for this paper's use case.
Looking at general demographic factors, the sample was approximately split into 50% female sex and resulting 50% in male sex. Most of the respondents belong to young age groups. More than 50% are less than 30 years old.
Regarding whether respondents have visited a nutritionist before, 27.4% actually have done so in the past, however only 3.1% do so currently. A visit to a nutritionist is usually associated with a current dietary problem, so it makes sense that the current rate is low. Important however, is to see that there is a skewness to the right when letting a nutritionist access data in the nutrition app. From a B2B2C business opportunity perspective that is a great result to see since it means that people are

open for collaboration.
Interestingly, the respondents self-report that they are well-educated in the field of nutrition. This statement is of course vague and respondents could interpret knowledge states with the five steps in the provided Likert scale differently. However, it is interesting that while most respondents see themselves educated, they still have the interest to learn more about it. This shows a curious target market which may benefit from a knowledge hub inside of this paper's app.
As further aspects beyond the food logging and the food recommendation process, the respondents evaluated the ideas to see the relationship between the food choices and moods and sensations as well as including sustainability aspects into the food recommendations as useful. That is especially true for mood and sensations where 63% scored the highest option on the Likert scale.

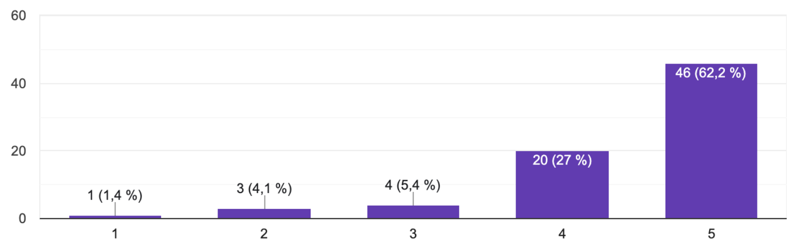


Figure 3: Willingness to know the relationship of food choices and sensations

Arguably, the most relevant retrieved data from the survey is the data on previous nutrition app usage to see why people used it and how their experience was with the available options in the market. Results here can be leveraged for feature development and prioritization as well as for good positioning and messaging in the application's marketing. Luckily, 61.1% of all respondents have used a nutrition application in the past and 34.1% are currently still using one.
Regarding the motivation to use a nutrition application, there is general health and longevity, aesthetics, and specifically weight loss as options. Of course these goals do overlap, however it is interesting to see that fitness in an aesthetical sense did not seem to be a major reason to most respondents. Only 22.2% chose 5, as the strongest agreement, in the context of aesthetics as a driver to use an application. Higher support, if also not completely unambiguously, was given to the more holistic motivation of reaching a good health condition. And, regarding weight loss the results look, even if measured on a five point Likert scale, binary. This makes sense since in comparison to health or aesthetics, weight loss is a more specific, less vaguely interpretable goal. Closely related to the motivation is also the importance put on macro- and micronutrients. As mentioned above, micronutrients are more important in a health-focused usage while macronutrients are relevant in any case.

The data supports this argument because the importance to track macronutrients is highly skewed to the right, meaning great agreement.
Additionally, an interesting insight beyond nutrients is the relationship users can have with the calorie count. Friends in previous conversations, and also respondents from qualitative feedback on the survey mentioned that the tracking of calories (the energy equivalent) is associated with a mindset of restriction and thus, can lead to the development of an eating disorder. It is an important learning which can be used in this paper's nutrition application. Since the application shall be health-focused, which includes the mental health of the users as well, and promote healthy eating patterns, the calorie count can be hidden by default and only be shown if a user actively turns its display on in the settings. This would also be a tangible way to show new potential users that this application truly tries to be different since the most common applications are not only showing calorie count, they are centered around it.
A very strong and rather unsurprising result is the disliking of the food logging process.

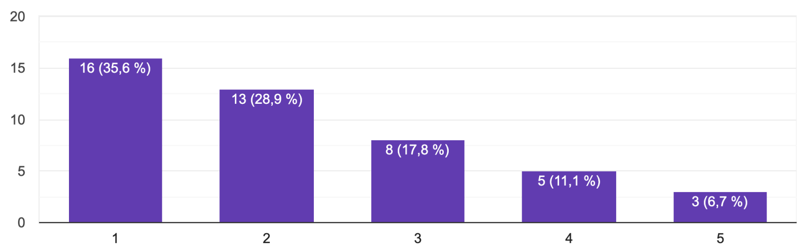


Figure 4: Enjoyment of food logging process

As you can see in figure 3 from the great skewness to the left, indicating less enjoyment, the food logging process in previous nutrition application usage was a great burden for most respondents. Also in the qualitative question asking why respondents stopped using the application, a lot of the reasons have to do with the annoying process of food logging, especially entering foods manually. This is a great sign for this paper's application since this paper is about exploring the viable use of artificial intelligence to create more value for users. If the application can successfully minimize the effort needed for food logging by using image based object detection, or speech-to-text models, it does effectively generate a better solution. More surprisingly in this aspect is to see that quite a large portion of respondents used the application for a long time. While the distribution of continuous usage times is fragmented, meaning there is not one primary usage time frame, 44.4% of respondents with prior application usage have reported to have used an application for more than 3 months.

For more details on food logging and to evaluate better food logging processes, the survey collected data on the previously used and desired methods for food tracking. The two main previously used methods are manually typing in food items and scanning barcodes of products. Only 4 respondents indicated prior usage of taking images to recognize foods and nobody has ever used speech-to-text to log dishes. When inspecting which methods are desired, we can see that the availability of all possible methods is well appreciated. The main desired methods are barcode scanning and detecting foods by images. However, also the speech-to-text has been indicated as useful by 35.6% of the respondents. Since it seems to be a new method, respondents might not yet be familiar with the exact way that this could work.
An important question in the context of this paper deals with the willingness to use a health-centered nutrition application. The indications speak in favor of a potential usage since the distribution of the Likert scale results is skewed to the right, with 56.1% choosing a willingness of 4-5 out of 5.
A less favorable insight from the survey results is a stronger aversion towards the usage of an AI chatbot to talk about one's nutrition and progress.

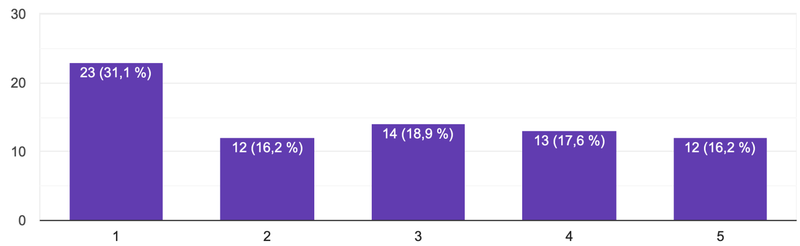


Figure 5: Willingness to talk to an AI chatbot for nutrition advice

Also only 22.2% of respondents have ever used a chatbot, like ChatGPT or BARD, to discuss nutrition. When inspecting specific concerns, misinformation and unreliability are often named worries. However, data privacy and a lacking understanding of AI have also been selected as such. In total, only 9.7% have marked that they do not have any concerns.
While the low willingness to use a chatbot does not seem great e.g. for the implementation of a chatbot in the nutrition application, the usefulness is yet more complex. Primary concern by survey respondents was the misinformation. This concern will be tackled by the use and implementation of a RAG chatbot which additionally has access to academic nutrition papers and thus, is expected to not hallucinate. Since the users of the nutrition app will most likely not have the technical understanding it is therefore important to communicate effectively and transparently how RAG enables science-based chat responses. Points of trust should be placed throughout the application's design. For instance, every chat response by the AI can be accompanied by the information on how far the statement is based on scientific literature.

Besides the chatbot, linear content (e.g. curated articles) is available in this paper's application, too.
For further feature development two more questions were included in the survey, namely the willingness to find workout as well as period tracking in the application. These features, similarly to the tracking of sensations and moods, could generate more data to predict certain symptoms and just overall improve the recommendations in the application. Both features were majorly supported by the respondents, indicating that users would find their implementation as useful.

3.4 Expert interview with a nutritionist

By surveying potential users of this paper's application to understand common pain points in nutrition applications and their expectations of a good version, decisions on the design of the application as well as the development and prioritization of functionalities can be taken based on an initial, real market feedback. Since this paper's application is supposed to cater value to users, but also to nutritionists and dietitians, the conversation with them is equally of importance. Hence, in Appendix C you can find a translated conversation with a Spanish nutritionist who focuses on clients following a plant-based diet. She was prompted with questions regarding common clients' issues, their usage of nutrition applications as well as the experience of nutritionists' lead generation. Through this qualitative feedback, insights can be generated and procedures and designs of the application can be validated.
Regarding common clients' issues which she is facing, she mentioned lack of knowledge in the process of turning one's own diet to a plant-based one. Clients often report confusion and unawareness when stopping to consume certain products and finding fitting alternatives. Her customers partially also struggle with cooking typical plant-based foods, like e.g. tofu, and their tasteful preparation. She also mentioned the worry of missing important nutrients. While vegan clients usually are more aware of potential lacks, her vegetarian customers are oftentimes less educated since their diet change does not include as much as a cut off in food options as the vegan one. This less severe diet change leaves them oftentimes more ignorant when it comes to the change's health impact. Generally, the interviewed nutritionist states, however, that vegans pay high attention to their micronutrient intake and encounter worries. Also the usage of nutrition applications is common among her clients. She specifically mentioned MyFitnessPal as an often used market player. However, she also mentions that sharing nutrition information including photos of eaten dishes as well as screenshots of nutrition apps does come with pain points since consolidating the information from these media often is not smooth and streamlined. She would enjoy a solution where her client can track their foods and she can seamlessly see

the entries and aggregated data based on the client's inputs.
With respect to the use of artificial intelligence in nutrition, she, unfortunately, has not yet used newest chatbot technologies, like e.g. ChatGPT. As expected, also to her, the potential misinformation is a concern and she highlights that she thinks nutritionists can offer a more individualized and holistic approach to nutrition counseling.
Regarding the business opportunity, the nutritionist classified finding new clients as hard and named word-of-mouth as the strongest marketing channel. Her colleagues and herself are also present on social media and have websites where she publishes an article once a month. While she enjoys writing articles, she also mentions that it requires a significant investment of time. Recipes and videos are also commonly created media types, however, not her personal favorite. In the interview, she mentioned that lead generation through a platform, like this paper's application, is a valuable method to pay money for, if the platform is effectively providing her with new potential clients. Colleagues of hers are already using paid medical registries, like Doctoralia or Doctorlib. For her, however, some type of guaranteed value is necessary to consider joining such a software.
As a limitation, the nutritionist mentioned that she does not work with a great number of clients and experiences can vary with every nutritionist.
Yet, the insights gained through the interview are valuable in the creation of this paper's application. Especially in terms of validating the interest of nutritionists to be present on the platform, the interview reaffirms the lead generating opportunity for them. A success-based pricing model is preferable over a fixed monthly price since there is no downside if nutritionists only pay for actual users reaching out. Furthermore, making the content and idea generation as simple as possible on the software side could increase the amount of content created. Generally nutritionists are used to creating content for lead generation already, hence, doing it for the platform would not be a new circumstance for them.

4 Technical implementation

Having laid out a potential business model and investigated the customer segment as well as validated assumptions with the help of an expert in nutrition, it can be well evaluated what generates value for users of this paper's application. On this basis, problems and possibilities for product differentiation have been identified and technologies pointed out. Subsequently, to draw conclusions on how viable the implementation of AI technologies is in a new nutrition application, these technologies have to be presented and their specific use case explained. Furthermore, their implementation, and its resulting degree of effort and difficulty in relation with their value proposition in this paper's use case has to be discussed to be evaluated in viability for a new venture.

Technically, a proof of concept of the envisioned nutrition application is created. This means a fully functioning web application is programmed. In the following, it is further explained and visualized in figure 6. Besides the creation of the application, small experiments are run in Python notebooks to explore functionality and performance of these technologies. The code of this exploration phase as well as the nutrition application itself can be found in the supplementary zip folder.

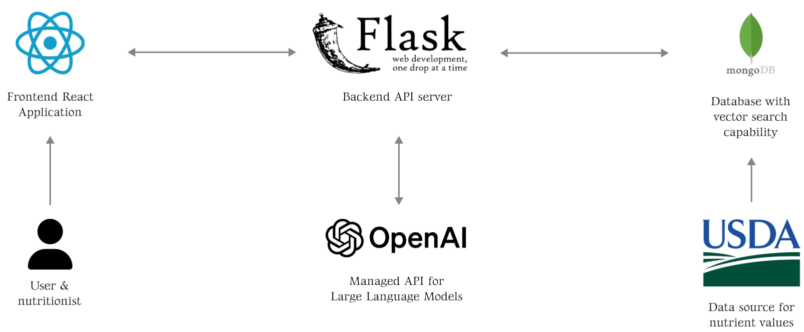


Figure 6: IT infrastructure of this paper's application

4.1 Design & IT infrastructure of the application

The web as well as mobile application consists of multiple parts, namely a frontend (visual user interface), a backend, e.g. computing recommendations, a database, storing users' and food's data as well as a web server, coordinating and managing web traffic to ensure availability.

4.2 React.js

React.js (or simply React) is a popular JavaScript library for building a dynamic and performant frontend. It is known for its efficiency and flexibility in creating interactive web applications. One of its primary advantages is its component-based architecture which promotes reusability and modular development (Meta, 2024).

4.3 Flask Backend and API

For the backend of the web application, a Flask server is used. Flask is a Python framework for building an API. Its advantage is to be simple, lean and flexible which is great to realize a proof of concept as fast as possible. Furthermore, it has a great developer community which creates useful libraries to handle all of nowadays' complex requirements (Ronacher, 2024).

4.4 MongoDB

To store all information about food data, like nutrient values, as well as user data, MongoDB is used as database technology. MongoDB is a NoSQL database and data is stored in documents. These are similar to JavaScript Object Notation (JSON), a widely used way to structure data in web applications. Its structure is flexible making it a great choice for a newly developed application and changing data structures that have to be stored (MongoDB, 2024). Beyond simply storing data, the team of MongoDB also introduced vector search capabilities in 2023. This is highly relevant to do quick vector similarity searches which usage will be further explained in section 4.6.

4.5 Nginx web server

Nginx is an efficient, fast and reliable web server technology. It serves as a reverse proxy and load balancer to ensure that traffic can be handled well and the web application is always available to the users (Nginx, 2024). While Flask can host its own development server, it is not recommended practice in a production environment and a proper web server, like Nginx, is to be used.

4.6 Vector database of foods and nutrient similarity search

A central promise to the user of this new nutrition app is the functionality of receiving food item recommendations to satisfy current nutrient demands to sustain a balanced, and healthy diet. The system enabling this value-add is based on vector similarity. The concept of vector similarity works as follows: vectors can be compared in various ways, most commonly and especially for high-dimensional vectors by cosine similarity which portrays the angle between vectors. The smaller the angle between two vectors, the closer related are the contents which are represented by the vector, and of course, vice-versa. In the case of this paper's application, vectors

are made up of different nutrients' values and, thus represent the nutrient profile of a given food.
For the creation of these nutrient profile vectors, the USDA's free database on food items and their respective nutrient values has been queried to create a custom dataset on foods and nutrient values. These values along with the metadata to find the original source on the website of the USDA are being stored in the above-mentioned MongoDB database.
Subsequently, 17 specific nutrients, including the most important ones and which were available in most data entries, are collected in a Python list, or as referred to above, the resulting nutrient vector. The values in the list are all transformed to the microgram unit. An important observation is the strong difference in scale of different nutrients. While only a couple of micrograms of one nutrient are existent in a specific food, multiple million micrograms of another one are given in the same food item. This difference in scale of the values in the nutrient vector can mean that the angle between two vectors is primarily impacted by a few nutrients only and not by the entirety of all nutrients which is an undesired behavior. To counteract this issue, all stored nutrient vectors are standardized upon insertion into the database as well as standardized again upon every time that a new food item is added. This is the case since an additional nutrient vector changes the elements of mean and standard deviation per nutrient stored in the vectors which are parts of the here applied standardization formula.
Beyond nutrient vectors of foods, another essential nutrient vector has to be created, namely the user's nutrient vector. This vector, in contrast to the foods' ones, is not representing the nutrients and their amounts the user contains or consumed, but instead, which nutrients the user needs based on the already intaken foods throughout the day. That means that every day the vector starts off with a fixed, base vector containing the 17 nutrients and the recommended amounts published by e.g. the EFSA. There are different authorities spreading distinct recommended amounts, including national institutions or the WHO. This base vector is then altered throughout every day by deducting the nutrient values of foods that the user ate. It has to be this way around since vector similarity is used to find the food options which supply the most similar seeming nutrient profile compared to the one requested by the user's body, based on health organizations' recommendations. On a "perfect" day, the user would have a nutrient vector storing only zeros at the end of the day, meaning all nutrient recommendations were met.
Technically, with a large number of food items stored and necessary to be ready for a cosine similarity search, this cannot be handled efficiently in the RAM of the server. To ensure a quick and clean process a vector database should be used which can nowadays be handled by MongoDB as explained above. This way, we can rely on

one database technology and do not need to manage and monitor multiple database types.

4.7 Machine learning models

Beyond the use of vector similarity search as a data science technique, machine learning models are being leveraged to add value to the customer. Since tracking is an unenjoyable process, as seen from the results from the survey, cutting the time to track all foods can be achieved by using object detection. This allows the user to simply take a quick photo of the food or plate, and the model recognizes and adds the different food items that the user is about to consume. While some of the survey respondents have already experienced logging foods by making pictures, a new method to log eaten foods is leveraging speech-to-text technology. Simply talking into your device can be simple and time-efficient. In fact, the availability of this type of logging technology would be well appreciated by survey respondents.
Besides tracking and recommending foods, the app shall also be able to offer information on nutrition science. Therefore, beyond the possibility to start a conversation with a nutritionist, a chatbot is available to the user to answer questions and explain interrelationships between foods, nutrients, organs and pathologies.

4.7.1 Object detection model

To allow users of this paper's application to scan their food by uploading a photo which consequently will be analyzed and food items extracted, the implementation of an object detection model makes a lot of sense. Since a dish usually consists of multiple foods in varying amounts, a simple image classification can quickly experience issues of complexity. An object detection model however is specialized in finding specific items in a picture and also counting the number of them. Moreover, the found items can be identified on screen because the model defines a box around the found object. With the x and y positions of the four box corners, items can not only be recognized and numbered but also located.
A well-known and proven model for this object detection task is the You only look once (YOLO) model, firstly presented by Redmon et al. (2015). Since 2015, the model went through multiple iterations and different development teams have published various versions. A quick version of this model which is known to this paper's author is the so-called yolov5, created by ultralytics, a US-based company with a second location in Madrid, Spain. The yolov5 model is pre-trained on Microsoft's Common Objects in Context (COCO) dataset (Lin et al., 2014). This dataset includes 80 object classes among which are for example people, cups, laptops, phones. While the pre-trained model can theoretically detect sandwiches and hamburgers, it

is not ready to be used as a food detector right from the start. To enable the model to also pick up on all imaginable food items, transfer learning has to be applied. Transfer learning is a technique in machine learning where a model trained on one task is adapted to work on a second related task. This way, the model does not need to start the learning cycle at zero, but instead has a starting point with prior knowledge which is transferable to the new, but related task. In this paper's case, the yolov5 model has the capability to already detect objects. However, it will be fine-tuned to recognize further, different objects as well.

To make the pre-trained yolov5 to detect food items, a lot of examples of different, labeled food images have to be shown to it. These pictures have to contain all the foods which the fine-tuned model shall be able to recognize in the end. To label these images, the Python application labelImg is used (HumanSignal and Tzutalin, 2024). This tool allows users to label images for object detection models. As a user, you open images in the tool, draw a rectangular box around the item to be detected, and finally provide the class name that the identified object belongs to. This way, the model can see where certain object classes are located in the images, and how these items look, to finally learn to recognize the food items in the fine-tuning process.

An important aspect in creating the food image dataset is its variety. The same foods can look very different in distinct situations. A banana, for example, has a different visual appearance with the peel than without it. Furthermore, it can be cut into smaller pieces which again, without the understanding of a banana, would look like a different object. Also the setting of the image is highly relevant. In the final application of the object detection model, users are not going to upload pictures of bananas in front of a white wall. The entire banana can be found alone with or without peel on a plate, however, it could also be mixed in a granola or placed on top of a pancake. In the secondly named settings, the banana might be sliced and its pieces might be partially covered with grains or maple syrup. The dataset needs to include a variety of images where the food items appear in their different shapes and are sometimes covered, and sometimes not. Of course, as mentioned in the section on the previous literature, there are also situations where food items are not only covered, but invisible to a spectator's eye. For instance, foods in a soup or a smoothie are mostly not recognizable anymore. As announced above, the usage of a multimodal LLM can be tested to augment the visual input with further context knowledge and predict possible food items beyond the image's content.

LLMs can do that since they have seen recipes and similar online content during their training process (OpenAI, 2024a). While that literally means guessing ingredients, it is of great value to this paper's application user. As seen from the survey results, the manual entry of foods is a strongly-perceived customer pain. For the user it is a quicker action to just deselect a falsely recognized food item, precisely one click, than typing out its name and selecting it, which includes multiple clicks. This way, the friction to log foods is lowered which means more consistent data is available which can be effectively displayed or further processed for analysis. This enlarges the perceived value since the application can successfully create an overview of the user's nutrient intake, give better recommendations, and fulfill its promise.

Above, counting objects is mentioned as a benefit of the object detection model. This benefit can however only be leveraged to customer value if certain details are met. With a well performing model, foods can be detected successfully in any shape that they may come in. Beyond recognizing foods and saving the user from manually typing in food names and selecting the correct item from a list, the other vital piece of information to be logged is the amount of the consumed food. In prior literature, machine learning models able to estimate depth in a picture were leveraged. Based on the depth information and the detections of which food is given, a 3D representation can be calculated and amounts better estimated. This is, however, a very technically complex approach and not a viable, and quickly implemented method. An alternative to achieve better weight estimates of food items with less complexity would be to further split food classes. Instead of only having one banana class meaning detecting that a certain object is a banana, there would exist multiple banana classes, e.g. full banana, half a banana and banana slice. Then each of these classes would be associated with a weight. Taking the count of each class found in the image, a final estimate for the weight could be created. Of course, this approach hardly predicts the exact amount, however it could be a more viable method to estimate the amount at all, based on the image input.

Generally, regarding estimating amounts and usage of them to calculate nutrient intake, a disclaimer to the user is required since nutrient amounts in foods do not only depend on the consumed amount, but also on the degree of maturity and how long a food has been cooked.

4.7.2 Speech-to-text model

Another way to easily log foods and one that is novel as well as desired by survey respondents is the input via voice message. From personal experience and interpreted from the survey results, this option is usually not available. It does however, offer a low friction method since the user can simply talk into the application without having to pay attention to specific expressions and because multiple or even all foods consumed in a day can be logged by one interaction.

Technically, the speech-to-text model from OpenAI, called Whisper, is used in the application (OpenAI, 2024b). Whisper is available as an easy-to-use API. This paper's nutrition application simply needs to make a request to the Whisper API which includes the audio file that was recorded by the user. In the next step, the text is extracted from the voice files and returned to the application as a string. This text can then again be sent to OpenAI's LLM with an engineered prompt to extract all foods and their estimated amounts in a JSON structure.

4.7.3 Large Language Model

Large language models have, due to their large training dataset and their ability to generate human-like texts, a great potential to be leveraged in a nutrition app. Beyond tracking one's diet, checking for nutrient coverage, and finding the best food items to sustain a balanced and healthy diet, users want to expand their knowledge regarding nutrition and move from applying advice to understanding it. One way to allow them to do so could be an e-learning platform within the app. This manner of providing information is, however, static. It is not adapted to one's specific prior knowledge and understanding. Chances are high that the materials are too difficult to understand or, the other way around, too easy and boring. A dynamic approach would be a consultation, like e.g. a consulting session with a nutritionist. The health professional can ask questions and adjust their explanations and jargon to the user's level of understanding. While this offers more value to the user due to the individuality of the approach and learning experience, it is hardly scalable nor a simple all-time solution. While a chat interface can be included into the app, the availability of a nutritionist around the clock is difficult and in a beginning phase close to impossible. Such a chat might be interesting to nutritionists if this can be a legitimate lead generation process. However, if user numbers are low, checking the interface regularly is financially less attractive. This is where LLMs can be leveraged. They possess the ability to have a human-like conversation with the user and adjust their vocabulary to the user's level of understanding, just like a human nutritionist. In contrast to the human counterpart however, a LLM can be available all time around and does not need to be aligned financially.

While that sounds like a great way to achieve the individualistic approach of great value with no operational downside, LLMs encounter the issue of hallucination. This describes their behavior of generating untrue or incomplete responses, also in the medical field (Wang et al., 2024). Therefore, the implementation of a LLM in its pure form in the nutrition app would be a potential risk. While informing the user about the possible downsides of the chatbot is one way to mitigate the risk, another one would be the implementation of a system that ensures (more) truthful responses.

4.7.4 Retrieval Augmented Generation (RAG)

One method to ensure more truthful responses is called "Retrieval Augmented Generation" (RAG). The idea behind RAG is to add information from a reliable source to the user query that is being sent to the LLM. In this paper's case, this information includes scientific papers on nutrition and medicine. This way the model has access to relevant information and instead of hallucinating, can simply rely on the reliable source. A major variable in this procedure is the context window of the LLM which is a synonym for the maximum number of tokens that can be supplied to the query. If there would be a huge or even unlimited context window, it would be possible to simply add the entire additional text data available to the query. Due to models being limited to a certain number of tokens however, the most relevant passages of the available text data in regards to the user's query have to be chosen before running the complete query to the LLM.

To achieve this behavior, all supplementary documents, like the nutrition papers, are being split in passages of similar size (110 words in this paper's case). Neighboring text passages have overlaps in the beginning and the end to mitigate/avoid losing context by randomly cutting a text into pieces. In the following step, these text passages are being encoded into vector representations with the use of a machine learning model. This allows the text passages to be compared contextually by the computer system. The texts along with their vector encodings are being stored in a vector database. This type of database allows storing data with their respective vector embeddings and metadata in one place. It is optimized for querying data based on an input vector. In our use case it would look like the following: a user submits a query (e.g. what shall I eat to support my health?) which consequently is being encoded by the same machine learning model as the supplementary text passages. The resulting vector representation of the user's query is then used as an input vector to query the vector database. This lookup in the database is using cosine similarity which is a commonly used method to compare the similarity of high-dimensional vectors. This query returns, in the following, the text passages which are most contextually similar to the user's query. In this example, the text passages probably deal with food recommendations or the highlighted importance of certain nutrients to achieve a healthy diet. A certain number of these best-fitting text passages will consequently be added to the user's query and finally the LLM is consulted. With the relevant extra information supplied, the LLM is now enabled, even though limited by context window size and a potentially huge database of supplementary documents, to provide a truthful and in actual science founded response to the user.

In this paper's case and regarding the question how to use this in a viable manner

as a new venture, there is one limitation to be mentioned. The effectiveness of RAG is based on the quality and quantity of the supplementary documents. In the field of nutrition and medicine, however, many journals do not publish their papers with open access. This means either paying license fees or finding another viable approach to make a LLM more truthful. Since a large amount of data should be accessible, for a good, and broad knowledge foundation, multiple journals would have to be paid and easy data mining access (e.g. through an API) would need to be given or customly coded, increasing the cost of the dataset creation.
In many cases, if data is available and programmatically accessible, the text is restricted only to the abstracts. While main findings can partially be retrieved from abstracts, the data would probably lack granularity and details.

4.7.5 graphRAG

Another approach to make a LLM more truthful and solve the above stated issue is the use of graphRAG. Introduced by a team of Microsoft in February, 2024, graphRAG follows the same idea of leveraging knowledge outside of the LLM's knowledge, however implemented in a distinct manner (Larson and Truitt, 2024). Instead of taking documents, cutting them into smaller text passages and encoding them into vector representations to be queried and compared in similarity, graphRAG involves portraying knowledge from experts or documents in the view of knowledge graphs. Graphs are a more natural way to model different elements and their relationships to each other (Ji et al., 2022). They exist as entities, also called nodes, and relationships which are also called edges. In the example of this paper's case, nutrition science, entities could be e.g. micronutrients, macronutrients, organic compounds, organs, muscles, diseases and foods. In this sample graph, examples for relationships could then be as follows: CONTAINS_A_LOT_OF between foods and micro- and macronutrients, ENHANCES_FUNCTIONING_OF between nutrients and organs. Please refer to fig. 7 for an example graph.

If a great amount of data exists, a LLM can be leveraged to extract the different entities and relationships between these that have been stated in the given documents. In the example, a LLM can find the statements of researchers acknowledging how different items in the human body impact others. These entities and relationships can be returned by the LLM in a JSON structure which consequently can be turned into the statements in the query language of the graph database. For this paper, a Neo4j database is used and hence, the query language is called Cypher. Once Cypher statements were created, the graph can be enriched with the knowledge of the documents. With a sophisticated graph established, graphRAG can then be used by users. While in normal, baseline RAG the user's query is en-

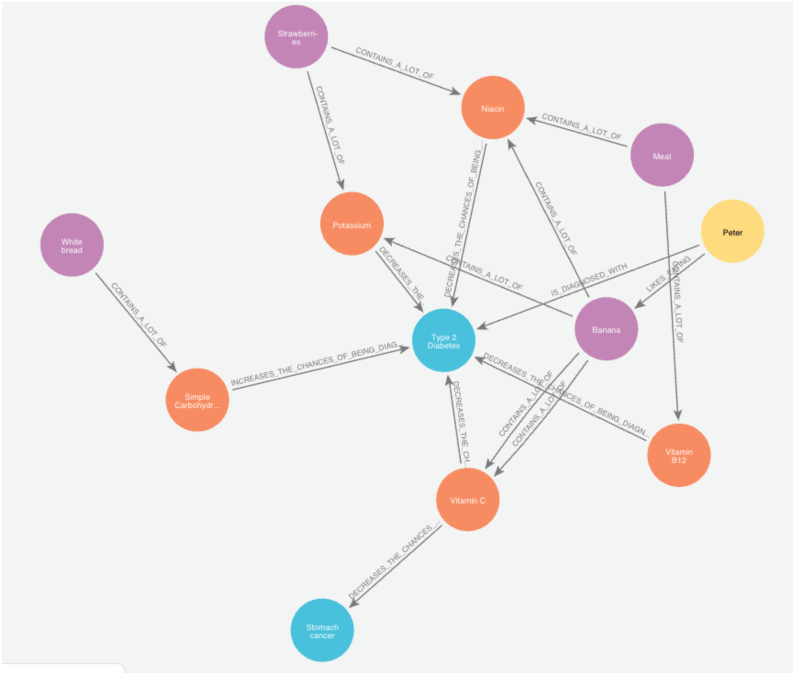


Figure 7: Example Neo4j knowledge graph

coded and compared to stored vectors, the user's query in graphRAG is used to create a Cypher statement to retrieve relevant information in form of entities and relationships from the database. Based on these results from the graph database, a response text is generated in human language.

The research by Microsoft found better performance by graphRAG compared to normal, baseline RAG, especially regarding the aspects of combining different pieces of information to find synthesized insights as well as holistically understanding documents' data (Larson and Truitt, 2024). This improved capability can be of great value in nutrition science since the combination of different studies which might include groups of different ages and sex, thus can generate a more comprehensive, and holistic view throughout the entire data.

Beyond mitigating or even solving issues of baseline RAG, graphRAG can also solve the issue of paper texts being hard or expensive to retrieve. Since graphRAG is not based on the actual text of documents, but rather on the knowledge within these, a graphRAG supported chatbot can still work without underlying papers. Of course, however, the graph database has to be enriched with data anyway. To do so, instead of extracting entities and relationships from papers programmatically, a manual process executed by domain experts can be chosen. In this paper's case, nutritionists can be taught how to enter entities and establish relationships in a Neo4j database. The fully managed database product of Neo4j called Neo4j AuraDB possesses a no-code interface to execute the two above-mentioned actions. With a graph database filled with information, the normal graphRAG querying process can be utilized to answer users' questions on an individual level. Moreover, along the journey of a new venture leveraging this technology, it is also quite simple to manipulate the chatbot's behavior towards a specific goal. Since there is a better overview of connections and knowledge within a graph instead of long, full text, a change can be implemented quicklier, especially if the contents have not been revisited for a longer time period.

4.7.6 Use of chat history embeddings for content recommendation

A great use case of AI is the personalization of the experience a user has with an application. This is especially true when a lot of contents are available and need to be filtered for the user. With the best fitting and most interesting contents dominantly positioned, the user finds relevant content quickly and uses the application for a longer time as well as connects a positive experience with it (Konstan and Riedl, 2012). The basis for achieving this value-add is data that can be used for this content-based filtering approach. In this paper's application, there are multiple events where users leave data points. This includes logging foods and thereby revealing information about their dietary patterns and liking as well as nutrient coverage. Furthermore, for good food recommendations, the users are prompted to enter their sex and age in the sign up process. Beyond these data points, also chatbot queries are interesting since these reveal actual interests and worries. These recent chatbot queries can be collected together to one document which consequently can be encoded to a vector embedding using an embedding machine learning model. Besides queries, also other information, like the age and sex, can be added to the document to be encoded.

Beyond creating a vector representation of the user data, the other relevant part is the encoding of the available content in the application. The primary planned media types are articles, and recipes. Since these are text-based, they can be immediately encoded to vector representations. If short explanatory videos shall also be offered as well, these need to be transcribed with a speech-to-text technology and then the transcribed text transformed to a vector embedding.

In the following step with user data and content encoded, vector similarity search will be used again to find the closest related content to the user's interest. Leveraging the in-built vector search capabilities of MongoDB, this can be done on every render of the content and hence, be always up-to-date. Needless to say, this also assumes a rerun of the user data encoding process every time the user triggers a new event, like logging a food or querying the chatbot.

5 Discussion

In this section, the implementation of the above explained technologies is discussed. The technologies' use case and hence, added value is opposed to the difficulty of creating a well-running proof of concept. From this evaluation, conclusions can be drawn subsequently whether the implementation of a technology is recommended or not in this paper's new venture case.

5.1 Additional value and implementation of nutrient vector recommendation system

The nutrient vector system enables the application's functionality to recommend food choices to the user to cover all nutrient demands. This feature is not available in other, inspected applications and hence, can be a point of product differentiation. When running test queries with distinct input vectors, food recommendations changed as desired. In the real application use case, it is not relevant which element is listed on the top position. It is more interesting to see a couple of food recommendations since certain foods might be out of the user's physical reach. A professional evaluation of these food recommendations can however only be done by a nutritionist.
When reflecting on the vector system, two aspects have been considered, namely the nutrients used for the vector representation, its standardization and the order of nutrients in the vector. Standardization of data is a common, default procedure in data science and hence, was implemented as laid out above. For the selection of nutrients for the vector representation all micronutrients which were available in all collected USDA food items were used. This was more of a technical than nutrition science based decision. Hence, the creation of a nutritionist reviewed vector makes sense and it has to be reassured whether more or less micronutrients should be considered as well as if macronutrients should be included in the vector. Calories, as a strongly present metric in other nutrition applications, is less useful in this vector context since the macronutrients together form the calorie count. The order of nutrients in the vector can be arbitrary since the pure order does not have an impact on cosine similarity if all vectors, nutrients from food items and the user's nutrient demand vector, are structured the same way.

5.2 Feedback on object detection model and Whisper + LLM to collect eaten foods data

Since image-based as well as voice-based food detection contain a lot of value in time savings to the potential user, as seen from survey results, it is highly interesting to see the associated efforts needed to enable the functionalities.

Unfortunately, the creation of the food detection model was more complex and time intensive than expected. The manual labeling process takes a long time if the food image dataset shall include images of a food in various conditions. This multiplied by the amount of different foods is not very viable in a one person operation. Also the training process of the yolov5 model takes long with a greatly sized dataset. From the first iterations, it can be observed that foods are better recognized, the more images of it are available in the image dataset. In a second run, there were vastly more images of potatoes labeled than of bananas and this led to a lot of cases of confusion. The object detection model was prone to detect a banana slice as a potato slice. Due to the long training times for a great reduction of the classification loss metric, quick iterations of the dataset and the model were not possible.
Considering this experience, the above-mentioned technique to split classes further down in smaller classes to estimate weights upon that can be classified as having low viability in this context. While technically, the complexity of a weight estimating model is broken down, the manual effort is large.
While the food detection model based on a yolov5 object detection model is time-intensive, but possible, it is a better technology to be developed later in the product roadmap. With more actual user feedback and the knowledge which foods are being usually logged, the dataset can be created efficiently for the new venture and optimally suited for the users.
However, to enable the users to log foods based on image inputs in the meantime, the multimodal LLM represents a better alternative, at least in regards to setup and implementation time. Both the methods, namely using an image for the multimodal LLM and transcribing an audio file and entering its detected text into the LLM, only require API calls to get results. The only adjustable aspects that have to be configured are the prompt accompanying the input data and if necessary, the LLM model, its temperature and the token output. In this paper's case, the later settings did not need to be finetuned since OpenAI's gpt-4-turbo performed well on the extraction task. When running the first prompt, foods were correctly listed in the supplied, desired JSON format. The only issue was encountered when complete dishes, like e.g. a hamburger, were not listed as single food items. Instead, the LLM responded with a one item JSON list of a food item called "hamburger". When altering the prompt and instructing the LLM to split dishes into whole foods, this behavior can be bypassed. Beyond naming dishes instead of whole foods, users are given more freedom and can also use this technology to log foods in another unit than grams, like otherwise required in the application, but in common objects, like plates or bowls. While nutrient amounts cannot be calculated based on plates or bowls of a certain food, the LLM is additionally prompted to estimate the amount of grams of e.g. a bowl of white rice. This works well and the estimations are usually within a

logical range. Furthermore, users are given the freedom to name any food they come up with and the LLM can generate any naming for a given food. In the database however, foods are only logged under one name. While this also incorporates details like e.g. raw or mature, LLM generated food items might not be found in the existing food items stored in the database by direct match. Therefore, all foods in the database have an additional field which displays the food's name encoded with the sentence BERT model (Reimers and Gurevych, 2019). When looking up LLM generated food names in the database, not the names themselves but the embeddings are queried. This way, there is always a match. This works great if for instance, the user would say "bell pepper", but the database only stores a food called "paprika". Due to MongoDB's native vector search capability, this search is also executed fast. The JSON output is robust and slight errors can easily be caught by checking for correct formatting and existing, expected keys within the JSON objects.
When manually evaluating the technology in action, the use feels natural and especially the speech-to-text is handy to e.g. log all foods from a day in the evening with one button click in the application.
Another success is the effective detection of potential, invisible ingredients. If the LLM is only prompted to detect foods on an image, this behavior is not obtained. However, just by including an additional sentence prompting the LLM to estimate probable, invisible ingredients and providing the examples of olive oil and sugar, the LLM does indeed include invisible items. In a trial, the LLM did estimate oil as well as pepper and salt, which all were present in the trial's dish. The most impressive aspect in this case is not that the LLM has the capability to guess invisible items, but rather the speed of implementation. Adding one short sentence, which literally takes seconds, enabled this paper's nutrition application to detect more food items based on image inputs, hence improving the user experience, since items like oil do not need to be entered manually. In a time before the existence of nowadays multimodal LLMs, for each new food item which shall be detectable, images of the items would have had to be added to the dataset as well as labeled, and the object detection model would have had to be retrained. The same value would have been achieved by a substantial time difference and would have required much more knowledge in machine learning. Another benefit of this LLM based detection implementation is its support of multiple languages. The LLMs as well as Whisper handle a great amount of languages and hence, an international target audience or a later geographical expansion is well assisted technologically. The only downside of the LLM based support is the associated payment if managed services are used. Since these features to log foods were indicated as useful by the survey respondents, a great usage can be expected. This will lead to higher operating costs per user which might make a freemium model less economically attractive and in case of low

capital even inviable.

5.3 Data mining academic nutrition and health papers for RAG and graphRAG

To enable RAG and graphRAG, open-access, academic papers covering nutrition and health science had to be accessed and their contents retrieved and then for the case of RAG, split and encoded into vector embeddings, and for the case of graphRAG, sent through LLM queries which extract entities and their relationships.
Data was accessible through OpenAlex which allows a user to programmatically see a lot of data on papers of different research fields, also including nutrition, health and medicine. Open-access online PDF versions are listed in the received data and hence, the entire paper's content can be viewed and with PDF reading Python libraries be made usable for further programmatic procedures. Partially, there are issues in using this technique. While it is straight-forward and very viable to generate a lot of data in a short time frame, many URLs to online PDFs are not leading directly to the PDF file, but rather to the online version of the paper on a certain paper publishing website, like e.g. ScienceDirect. Sometimes, an intermediate step is necessary to finally receive the URL to the PDF file which can look slightly different for each paper publisher and thus, causes extra work. This consumes more time to generate a satisfying amount of data, however, while it is an obstacle, it does not reduce the economic viability of implementing RAG or graphRAG.
There is something else that is much rather impacting the economic viability of this procedure - this is the necessity of previous knowledge in the domain for implementing graphRAG. Since nutrition and the human body are very complex concepts and contain a large depth of processes and different items influencing the entire system, it can be hard to answer the question of what elements shall be extracted from the papers. Due to the observation from my programming trials that it is advantageous to predefine elements to be extracted and provide examples to the LLM, it is necessary to know previously which categories of elements shall be extracted (e.g. micronutrients, organic compounds, enzymes, etc.). Without a great expertise in the field however, you might easily miss important elements and hence, create an oversimplified representation of reality. This is problematic because we implement and use RAG and graphRAG especially for this exact reason - to prevent hallucination and showcase an answer based on real findings in real research projects. An oversimplified knowledge graph could therefore lead to less accurate and half-baked results which does not fulfill its purpose.
Furthermore, it is crucial to pay attention to the search query made, since a search for nutrition and health also resulted in results regarding animal's nutrition and

health which is not applicable in this paper's use case.

5.4 Prompt engineering the entity and relationship extraction for graphRAG

When running the entity and relationship extraction on the academic papers, multiple complications have been noted. To be able to evaluate how well the extraction performed, the initial idea was the primary manual extraction of entities and relationships. Contrary to the first assessment, this was more complex since a graph existent of only two types of entities is not useful in a nutrition context. To recognize all entities manually, prior domain knowledge in medicine or nutrition science is necessary. Also labeling distinct papers is desired since the extraction prompt should not be overfitted to one paper. Each academic work covers different topics and thus, includes types of body elements and foods, such as enzymes, organic compounds, or micronutrients. All this leads to a great complexity which would require high time investments and the inclusion of healthcare professionals. Since this exceeded available resources, the extraction process was started without a labeled evaluation dataset.

In the first step, three different prompts were defined and run on the academic papers and the results saved in JSON files for later inspection. In the first prompt all entities to be extracted were mentioned and then an example formulated for one of the entities. This was done because the formatting looks the same for all entities. Also for the relationships a few explaining sentences were created and an example attached. In the second prompt, a much shorter version was explored. The prompt only included explanations of the concepts of entities and relationships, but did not show examples. It is an interesting prompt since in case of successful extraction, a lot of tokens and consequently, money could be saved when creating a knowledge graph. The third prompt was the strongest contrary to the second. Like in the first prompt explanations for entities and relationships were given, but in contrast to the first prompt, examples for all entities as well as more relationship examples were formulated.

After running these three prompts, strong differences between the results of each could be seen. The more examples were made in the prompt, the more entities were usually extracted. This means prompt 2 performed the worst and prompt 3 slightly better than prompt 1 when manually inspecting extraction outcomes. Generally however, a non-optimal overall performance was visible. The extraction was run with OpenAI's gpt-3.5-turbo as well as with its gpt-4. For example, vitamins, which are classified as micronutrients, were partially extracted as macronutrients. This confusion in entity classes could lead to inaccurate representation in the knowledge graph and finally results in misinformation in the chatbot application. Additionally, no relationships between the extracted entities were created.

To fix this issue, the text of the longest prompt was reviewed and more examples of potential relationships were written. This, however, also in a second run did not lead to better results. In another attempt, the temperature of the LLM was altered from 0 to 0.5 and even 1. Increasing the temperature of the LLM allows it to create more random outputs and, colloquially said, be more creative. While this seemed to have created more entities, it also led to a few duplicates in entities and yet no relationships.

What did however work well was the robust formatting of the output. The LLM was prompted to create a JSON structured output which was always achieved. Only partially, single instead of double quotation marks were generated which when loading in the Python JSON library led to an error. This error was, however, simply caught by replacing all single for double quotes before encoding the LLM's output string to JSON.

5.5 Performance of baseline RAG

To evaluate the performance of this paper's RAG application, the Python RAGAS library is used (ExplodingGradients, 2023). RAGAS stands for RAG assessment. It provides different metrics to judge the capability of the retrieving part of RAG as well as the generative part. The data needed for an evaluation includes a question, which could be asked by a user of the RAG application, a list of texts retrieved from the vector database, the generated answer by the application and finally, the expected answer also referred to as ground truth. Based on these inputs, RAGAS leverages a LLM to analyze the data and provide the following metrics. The two retrieval relevant metrics, which are applied for this paper's case, are the context recall and the context precision. Context recall is a method to check how many of the retrieved texts are contextually close to the provided ground truth answer and thus, supported successfully in creating the desired answer to the user.

$$\text{context recall} = \frac{|\text{GT sentences that can be attributed to context}|}{|\text{Number of sentences in GT}|} \tag{1}$$

RAGAS breaks down the statements made in the ground truth and consequently checks whether these can be refound in the retrieved texts. If all statements' origin can be found in retrieved texts, then context recall returns 1.0. In contrast, if no statements can be traced back to retrieved texts, the context recall is zero. An

important side note regarding context recall is that by increasing the number of retrieved texts, usually the recall metric can be positively impacted. Therefore, in this paper's case we limit the number of retrieved texts to five.
In the first iteration of the RAG implementation, the context recall is of the value 1.0 if querying contents available in papers from the vector database. This is a great result, especially considering that initially, the sentence BERT encoder was used which arguably is not state-of-the-art anymore and only has embeddings of 384 dimensions. For queries regarding topics not covered by papers in the vector database, the context recall results in zero. This is probably due to the low amount of data provided.
In the later iteration with more papers added, the context recall is increased in the cases where before it was at zero. However, with the question that resulted in a 1.0 before, the context recall is decreased. This might be due to the open nature of the question (e.g. asking for connections between nutrition and health) and the higher number of potential documents to be considered. It is also important to mention that the ground truth is manually crafted and not exhaustive. It usually is much shorter in length than the generated answer which often seems like the more all-encompassing response. Also for questions outside of the database knowledge, the metric is very low and even zero. This was especially true when asking for differences in meat options while most analyzed and stored papers deal with plant-based nutrition.

Context precision is the metric to investigate, beyond whether retrieved texts are relevant, how high these texts rank. The vector search in the vector database results in a list of texts whose first element should be the most contextually similar to the user's query and hence, be the one that, in the best case scenario, can deliver all necessary information to answer the query. Context precision basically checks whether this fact is true by seeing "if [the retrieved text] is relevant or not relevant to arrive at the ground truth for the given question" (ExplodingGradients, 2023) and subsequently, dividing the number zero in case of not relevant data and the number one in case of relevance by the position number in the retrieval list. These resulting numbers for each retrieved text are then summed up and divided by one to calculate the final context precision score. Therefore, similarly to context recall, it returns a value between zero and one, with one being the best score possible.

$$\text{Context Precision@K} = \frac{\sum_{k=1}^{K} (\text{Precision@k} \times v_k)}{\text{Total number of relevant items in the top } K \text{ results}} \tag{2}$$

$$\text{Precision@k} = \frac{\text{true positives@k}}{(\text{true positives@k} + \text{false positives@k})} \tag{3}$$

In the early iteration of this paper's RAG application, context precision is, similarly to the context recall, at the maximum value of 1.0 if the query topic is based on papers saved in the vector database. This means that the embedding model and the similarity search metric of cosine similarity do work fine enough to rank the text chunks effectively. As above, if the topic of the query is outside of the saved vector database content, the metric is worse, usually around 0.33.
In the later iterations with more paper content added, the context precision sees an increase in value meaning a better performance. For a question where before context precision was about 0.33, it is now at 1.0 or close to it. Only for questions truly beyond database knowledge, like the above-mentioned question regarding meat options, the context precision is still zero.

To evaluate the generation part of this paper's RAG application, the two metrics of faithfulness and answer relevancy are used.
Faithfulness defines the share of claims made in the generated answer which can also be found in the provided contexts. If all of these claims can be found, the faithfulness score is one, while if none can be found, it returns zero. Consequently, if a share of the claims are found, it returns this share.

$$\text{Faithfulness score} = \frac{|\text{Number of claims in the generated answer inferred from given context}|}{|\text{Total number of claims in the generated answer}|} \tag{4}$$

In this paper's case, faithfulness can be considered the primary metric. The main reason to apply RAG in the nutrition app instead of simply only using the LLM as it is, is the generation of answers based on true facts stated in academic research and thus, reducing hallucination and misinformation. The metric faithfulness is exactly checking for this. In the first iterations as well as in the later ones, the faithfulness scores were very high indicating a successful implementation for this use case.

Answer relevancy looks at the generated answer and the potential question written by the user and defines how relevant, specific, or closely-related the answer is to the query. If the LLM in the RAG application is given too much temperature and responses are very broad, and hence, unuseful, the answer relevancy score would be low. It is calculated by reverse-engineering the process, namely making a LLM gen-

erate questions based on the generated answer. Between each of these newly created questions and the actual user question, cosine similarity scores are computed and the mean of them returned as the final result.
In this paper's case, answer relevancy scores were throughout high, namely between 86% and 98%. For the usage in the nutrition application, this means that the RAG application can be implemented without much worry whether the chatbot would provide irrelevant or non-specific responses to the user. In contrast, the chatbot provides valuable answers.
Beyond using quantitative and hard metrics, the final manner of evaluation is the manual usage and assessment of the RAG application. This is especially necessary to uncover unexpected behaviors or factual incorrectness that might exist due to old, outdated academic papers.

5.6 How to improve RAG performance

There are different elements impacting the RAG application's performance, namely the embedding creating model, the similarity measurement, the amount of documents and their relevance for the user queries, the range of potential user queries, the text chunking method, and whether there is metadata embedded with the text chunk.
Already from the first step on, we can impact the RAG behavior. Cutting the provided texts from the nutrition PDFs into smaller chunks can be handled in different manners. Besides deciding on a different length of each text chunk, the text can also be split based on different conditions. In this paper's case, the texts are simply split based on word count, identified by spaces. Similarly the texts could also be separated by token count, which is similar, however uses the LLM's tokenizer as reference and not spaces between words. Both these techniques are simple, however as seen above, do deliver good enough results in this use case. To improve this part of the RAG pipeline, semantic chunking could be used. In this method, the texts are split and then each part encoded into embeddings which consequently, are compared to neighboring parts with cosine similarity. If two neighboring parts show high similarity, they will be merged in the same text chunk in the final splitting process. This way, texts are not cut by the position of the words or sentences in a document, but rather by context.
Once chunked, each text part is being embedded. For the creation of a vector representation of a text chunk, different models are available. The initial model used in this paper is the sentence BERT transformer model, established by Reimers and Gurevych (2019). Compared to embedding models provided by LLM creating companies, the sentence BERT is not a state-of-the-art model and only encodes text

into a vector of 384 dimensions. For instance, the text-embedding-3-large model by OpenAI provides a vector of 3072 dimensions, hence able to contain more information. However when looking at the trade-off, the sentence BERT model performs well enough and the venture could, due to sentence BERT being free to use, benefit from small monetary savings if a huge corpus of nutrition papers should be embedded.

Another aspect when looking at the embedding process, is the creation of text chunks that themselves lost context of the entire nutrition paper. While these might not include the most relevant information, it would still be beneficial to add the overall context to the specific text chunk. This can be achieved by integrating a descriptive piece of metadata, like e.g. the title of the academic paper, into the text chunk before embedding it.

Regarding the similarity measurement, cosine similarity is usually used due to the high-dimensionality of the embeddings. In theory, it could be changed to e.g. euclidean distance, however best practice is the cosine similarity.

The two lasting influences on RAG performance are the high-conceptual variables of the topics of typical user queries and the content of the stored documents. While many user queries can be anticipated, especially in an application that is focused on a specific target user and a specific problem to be resolved, some queries might deviate from the expected topics. As seen above, these queries, therefore, do not have a well-fitting context in the vector database and consequently, lead to lower scores in context retrieval. These metrics can be improved by adding more academic papers and simply enlarging the context available.

To do this in the most effective manner and with a better methodology, it is important that queries from users of the application are being saved. By inspecting the list of queries and, in case of high amounts of queries, an overview leveraging topic modeling, inspiration can be won to find the most relevant, new nutrition papers to answer user queries with better performance. Furthermore, the saved queries can be used for ongoing evaluation of the model. While there will be no ground truth given for these questions until they are reviewed, faithfulness and answer relevancy can still be calculated.

5.7 Baseline RAG vs graphRAG in a new venture's nutrition application

In comparison to baseline RAG, graphRAG's creation is more exploratory and less of a robust, fixed process to follow. Instead of only choosing the tools, like the embedding model and the LLM, the creation of the fundamental knowledge graph includes a great prompt engineering process. This process is time-intensive and

needs prior expertise in the field. While the exact contents of the academic papers do not need to be known by the programmer for baseline RAG, it is important to know which relevant elements are included in the list of papers to successfully and effectively extract them to realize a graphRAG application.
As described above, the first iterations of the prompt engineering of the graphRAG extraction script did extract entities, however usually missed all relationships between them. The need for medically educated people to label a dataset for evaluation and the long prompt engineering process are factors making graphRAG greatly unviable when compared to the baseline RAG implementation. However, fast-forwarding to a state of a running system, graphRAG would be the system which allows quicker and easier fine-tuning by adding or deleting connections between entities or eliminating entities entirely. This way, new advancements in research and their strongly, academically supported views can be unambiguously represented and enforced in the chatbot. This might be a desired functionality if the understanding of a biochemical process is well researched and hence, can be modeled in a knowledge graph. In cases of ambiguity in research, however, where researchers can make suggestions based on correlations from studies, but not based on biochemical processes in the body, graphRAG might not have a great advantage over baseline RAG.
To fine-tune a baseline RAG application, new papers need to be added or existing ones deleted. While that is technically not a difficult process, primarily creating an overview of what is being stated in which papers and then selecting papers or text chunks to be deleted or altered is more time-consuming. That is especially true if this alteration procedure is done by a non technical person. Neo4J AuraDB allows graph mutation easily by using a no-code tool while querying a document-based database and then editing fields encounters more technical steps.
To sum it up for a tactical procedure, there is probably not one RAG technology that has to be left out and the other one implemented since both hold individual benefits. The decision lays much more in when to implement which technology. After considering this paper's implementation experience, it can be recommended to use baseline RAG in the first iterations of the product. It is the quicklier and easilier built technology and does perform well, especially when looking at faithfulness. Consequently, in later iterations graphRAG can be given more attention and its successful creations and implementation be prioritized to profit from its easy tunability. Beyond adjustability, another aspect regarding this technical decision should be user feedback. Collecting user feedback in terms of likes of chatbot answers indicating good results, amount of chatbot usage of a certain technology or generally qualitative feedback is crucial to getting an idea of which technology delivers most value to the actual end users. This should be the primary concern in this endeavor.

5.8 Ethical considerations when implementing the above technologies

Building a nutrition application means dealing with sensitive and personal data. Hence, a transparent and professional appearance is of high importance. This aspect is especially true considering the additional use of artificial intelligence in the software.

A clear basic assumption of users has to be that their data is stored safely with this paper's application. To be able to offer this, managed database services are used where the database company is expert on the technology and takes responsibility for security updates and backups. Furthermore, important data, such as passwords, are being hashed before being stored.

Beyond these infrastructural technicalities, nutrition science-specific concerns need to be addressed. Since the application's chatbot is connected to academic papers in health and nutrition, it is relevant to continuously monitor the stored papers and determine their factual correctness. News regarding new research outcomes as well as serious critiques on existing papers need to be followed closely. These events should trigger an alteration of the database content, whether that means adding new papers or deleting old and irrelevant ones. This way, the chatbot is being kept up at the latest insights in nutrition science and can answer user questions truthfully.

Concerning the chatbot technology, the use of AI should be clearly communicated. Especially, since in-application chats with real human nutritionists shall be enabled as well, it is necessary to transparently display when the user is talking to another human being and when the answer from the application comes from AI. Moreover, it should be stated where data, like the academic papers or data on micronutrients in different foods, originates. In the application this can be done by incorporating pop-ups which inform the user about data origins and can show links to other websites, e.g. the website of the USDA. In the chatbot specifically, another piece of information would add a lot of value and trust through transparency. Besides the actual response, every chatbot answer should show the faithfulness score. It can be displayed as a percentage and accompanied by additional text communicating the meaning of the score. It could, for instance, say "100% based on academic papers". This way, users are warned if the RAG application produces a response which differs from content in academic papers and thus, might be hallucinated. While it does slow down the latency time until the user receives a message back, it is a scalable approach to provide the user with transparency.

This aspect of creating an understanding and awareness is highly relevant when taking into account the data from the survey on the willingness of users to interact with a chatbot about their nutrition and diet, which as above shown, is not indicated as

high.
Beyond the monitoring of academic papers in the database, the queries to the RAG application and the resulting answers with faithfulness scores should be saved for later inspection. This way the venture's team can better understand demands of users and see whether the chatbot is continuously generating relevant, non-hallucinated responses. This quality assurance can be displayed in a dashboard, accessible to team members, which displays the topics of queries, the faithfulness scores, and allows to zoom in on single queries, while not showing which user has prompted them.
With the application collecting user data on diet, moods, sensations, interests, and worries, further machine learning models could be trained to predict mood and symptoms. or aggregated reports could be created and catered to the user. When merging all of this existing data, procedures need to be double-checked by professionals since the results of any data science application on this data could potentially impact a user's dietary patterns and psychology. Due to these implications, the new venture's team should be advised by healthcare professionals. Lawyers in healthTec need to be counseled before publishing the final product to the market.
An important aspect, touching upon the European General Data Protection Regulation (GDPR), is the opt-in meaning that users actively accept the way how their personal data is processed. Even better for the users' privacy and power about it would be the availability to configure the privacy settings. Being able to more specifically agree to certain processes, the users can choose which features generate value for them and hence, make it a fair deal to share personal data for its effective use.
Another interesting and holistic health-relevant aspect that was able to be seen from survey responses and casual conversations with contacts who have used a nutrition application in the past was the impact of tracking calories on individuals' relationship to food. It was mentioned that counting calories was a driver for an eating disorder. This is the case because, if the person's goal is weight loss, the calorie count can be seen as a restriction and not as a simple guiding number. An obsession about limiting oneself and one's food choices drastically based on the calorie count led to these individuals not seeing food as an energy supplier, but rather as a necessary evil in the way of a dream body shape. Being aware of this issue, and having the goal, and differentiation of a more holistic health-focused nutrition application, the calorie count can be hidden by default. The sum of all macronutrients also represent the calories, hence, the users can see whether they eat enough based on how far they have eaten sufficient fat, carbohydrates and proteins. This view however shows the higher complexity underneath a simple calorie count and logging a new food, especially if full of nutrients, will be shown rather as an achievement

with progess bars being filled and turning green. This way a healthy relationship to food shall be supported and the responsibility be actively accepted and used for the good.

6 Conclusion

Having discussed all the technologies used in this paper's application, conclusions can be drawn to answer the initially raised question. The additional value as well as the technical viability is highlighted to determine whether the inclusion of the specific technology is recommended and at what stage of the new venture's journey it shall happen.

6.1 Viability for nutrient vector recommendation system

The nutrient vector recommendation system allows this paper's application to add a different aspect in nutrition applications and thus, create a product differentiation and cater better to a specific target market. While the investigated market players collect food logs and provide overviews on which a strategy for losing weight or living healthier can be based on, the nutrient vector system can formulate more direct recommendations beyond establishing an overview. Also technically, the creation of the system is not complex. The biggest challenge is data acquisition. Data on many foods with as many micronutrient entries as possible is necessary to make the vector recommendation system as useful and valuable for the end user.
With the combination of great value added for the user as well as the new venture's positioning, and a rather simple implementation, the deployment of the system in the final new nutrition app is highly recommended.

6.2 Viability for object detection model

This paper's survey showed that food logging is perceived as an upsetting and annoying process. Hence, any technology which can be leveraged to decrease the associated pain of customers to use a nutrition application is cherished. Providing time saving and naturally user-friendly seeming methods, can lower the threshold to try out this paper's nutrition application when there are no user reviews yet available. Moreover, the ease of use could potentially lengthen the time of usage. This means more data collected which makes overviews exhaustive and enables the database for future models to predict further relevant elements.
In the survey, respondents scored the image detection method as desirable and hence and due to the above aspect, a value-add by implementing the technology can be expected. Technically, the implementation is highly time-consuming and due to the nature of food's appearance complex. A labeled, big, and high quality dataset is required to create a model which can satisfy the needs of the application's user base. For a new, low-resource venture the training of such an object detection model is not viable. A much greater seeming alternative for food detection, implemented in

a fraction of the sophisticated object detection model's time to deployment is the use of a multimodal LLM. While error handling needs to be established to handle different food naming between the database of foods and nutrients and the LLM, the multimodal LLM reaches similar or even better performance on food detection. As a result, the usage of a multimodal LLM for food detection is recommended, especially in the beginning of the venture when the initial product iteration is being validated with the first users. If the venture is successful, the models can still be switched and a dedicated object detection model fine-tuned and implemented. This would make sense in case the usage of the multimodal LLM becomes, due to high usage, too expensive, or highly specific foods, which often experience confusion, shall be able to be detected with greater precision.

6.3 Viability for speech-to-text technology

Based on the above explanation for the value-add of the food detection model, also the inherent value of speech-to-text technology can be reasoned. Furthermore, its implementation is simple due to today's existing APIs to transform speech to text and subsequently, extract foods from the transcription. High value-add and quick implementation makes it an obvious choice to include it in the final application.

6.4 Viability for RAG and graphRAG

In this paper's survey, respondents expressed the desire to learn more about nutrition. While static content can satisfy this demand, its personalization in the form of a chatbot could potentially add more value. An important aspect to mention is the fear of misinformation of survey respondents. Suitable for this concern, the implementation of RAG with academic nutrition papers makes sense and should, in the best case scenario, resolve worries, especially if faithfulness is transparently shared. Even though relevant open-access papers first have to be identified, which can take some time, a RAG chatbot should definitely be made accessible to the user. Setting up the embedding-creating and querying processes is a simple step by step process. Thus, RAG is recommended.

Less clear is the evaluation of the graphRAG powered chatbot. A graph is a good choice to represent the body's biological elements and their interconnected processes. Hence, a method for querying it in natural language is valuable to eager learners. Concerning the implementation, graphRAG is however highly explorative in this paper's context. This means that a higher amount of time have to be spent to create a graph which is useful in the final application. To quickly iterate and create an initial product, graphRAG is therefore not representing the tool of choice. Baseline RAG is recommended over graphRAG for the first product iteration, while graphRAG can be implemented later when a sophisticated graph was able to be created and yields a benefit in chatbot responses.

7 Limitations

This paper and its nutrition application have limitations which are important to be pointed out. When using the application, incorrect user entries, regarding sex, age, but also wrongly logged foods impact the recommendations and make them less applicable. Also data sparsity through users forgetting to log foods leads to less and less valuable data to evaluate one's nutrient intake and coverage. In a next iteration of the application, features can be added to motivate and remind users of logging foods and an exhaustive error prevention system to make sure, correct information is being entered.

Furthermore, nutrition science is a relatively new, complex, and evolving field. The vector system can potentially oversimplify circumstances and thus, lead to less valuable advice and recommendations. Future literature needs to be closely followed to adapt the application and its behavior respectively.

Moreover, it is important to mention that the application is targeting worried, but healthy users without complex pathologies. It has not the capability to replace any professional, medical advice. Another limitation is the development of the application in a western, European country. Available food options and the use cases the app has been programmed for, and nutrient intake recommendations can differ in different countries all over the world. Due to cultures and religions, diets look very distinct in different places. Therefore, machine learning models, like food detection models, and food entries in the database would need to be adjusted to fit a worldwide audience.

Regarding the analysis of the product differentiation, other frameworks can still be applied to gain deeper insights. For instance, the application of Michael Porter's five forces could be a more specific tool for analyzing competitive advantages and threats (Porter, 1989).

8 Future extensions to the product

From previous literature as well as by building out the nutrition application, certain functionalities were discovered which however, were not able to be built in the time frame of this paper. Especially since this paper focuses on more viable technologies, these extra features did not make it into the first iteration due to their lower viability.

Many works from the previous literature include the prediction of weight loss or gain as well as risk for certain diseases based on a user's biochemical as well as dietary data. An existing application which in a production environment possesses real users can collect data. Based on this data and continuous updates by the users, this data can be leveraged to train further machine learning models. Of course, this has to be done carefully and in accordance with data privacy laws.

Another extension can be realized in the context of image based food detection and amount estimation. In prior literature, computer vision models were explored which leverage depth estimation beyond detection of food items. Given the fact that multiple researchers had to work together to enable these models just displays again the low viability of these models. However, an effective model with this architecture would be a great value-add to exchange the more costly multimodal LLM in detecting as well as accurately estimating foods and their amounts. This can lower the operating costs which can be reinvested into greater customer acquisition or higher pay-out for the founders.

An issue that would however not be resolved by the use of the above depth-detection-model is the recognition of invisible ingredients. However, it could be a useful action to create a classification machine model that predicts existing, invisible ingredients based on detected food items. This should work ground on the assumption that recipes that include specific ingredients will also include other certain invisible ingredients to process these.

An important annotation is as well the support only for whole foods. In the survey the barcode scanning was noted as highly useful, however this is a technology to track processed groceries. While implementing a barcode scanner is not highly difficult and food item information via barcode is available with APIs, like e.g. OpenFOODFacts, these data sources only provide data on macronutrients. The lack of micronutrients is vital and a hybrid solution to accept processed foods and yet be able to give good food recommendations is demanded.

9 Bibliography

Autio, E., Nambisan, S., Thomas, L. D., and Wright, M. (2018). Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems. *Strategic Entrepreneurship Journal*, 12(1):72–95.

Balloccu, S., Reiter, E., Kumar, V., Recupero, D. R., and Riboni, D. (2024). Ask the experts: sourcing high-quality datasets for nutritional counselling through human-ai collaboration.

Bond, A., Mccay, K., and Lal, S. (2023). Artificial intelligence clinical nutrition: What the future might have in store. *Clinical Nutrition ESPEN*, 57:542–549.

ExplodingGradients (2023). Metrics | ragas. URL: https://docs.ragas.io/en/latest/concepts/metrics/index.html/, Last accessed: 02.06.2024.

Farrokhi, A., Farahbakhsh, R., Rezazadeh, J., and Minerva, R. (2021). Application of internet of things and artificial intelligence for smart fitness: A survey. *Computer Networks*, 189:107859.

Han, Y., Cheng, Q., Wu, W., and Huang, Z. (2023). Dpf-nutrition: Food nutrition estimation via depth prediction and fusion. *Foods*, 12(23).

HumanSignal and Tzutalin (2024). labelimg: image annotation tool created by tzutalin. URL: https://github.com/HumanSignal/labelImg/, Last accessed: 02.06.2024.

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514.

Knights, V., Kolak, M., Markovikj, G., and Gajdoš Kljusurić, J. (2023). Modeling and optimization with artificial intelligence in nutrition. *Applied Sciences*, 13(13).

Konstan, J. A. and Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction*, 22:101–123.

Larson, J. and Truitt, S. (2024). Graphrag: A new approach for discovery using complex information.

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

Maurya, A. (2011). Running lean: Iterate from plan a to a plan that works.

Meta (2024). React. URL: https://react.dev/, Last accessed: 02.06.2024.

Miyazawa, T., Hiratsuka, Y., Toda, M., Hatakeyama, N., Ozawa, H., Abe, C., Cheng, T.-Y., Matsushima, Y., Miyawaki, Y., Ashida, K., Iimura, J., Tsuda, T., Bushita, H., Tomonobu, K., Ohta, S., Chung, H., Omae, Y., Yamamoto, T., Morinaga, M., Ochi, H., Nakada, H., Otsuka, K., and Miyazawa, T. (2022). Artificial intelligence in food science and nutrition: a narrative review. *Nutrition Reviews*, 80(12):2288–2300.

MongoDB, I. (2024). Mongodb - the developer data platform. URL: https://www.mongodb.com/, Last accessed: 02.06.2024.

Nginx (2024). Nginx - http and reverse proxy server. URL: https://nginx.org/en/, Last accessed: 02.06.2024.

OpenAI (2024a). How chatgpt and our language models are developed. URL: https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed/, Last accessed: 02.06.2024.

OpenAI (2024b). Openai platform - whisper. URL: https://platform.openai.com/docs/guides/speech-to-text/, Last accessed: 02.06.2024.

Osterwalder, A. and Pigneur, Y. (2010). *Business Model Gen- eration: a handbook for visionaries, game changers, and challengers (The Strategyzer Series)*. Wiley.

Porter, M. E. (1989). *How competitive forces shape strategy*. Springer.

Redmon, J., Divvala, S. K., Girshick, R. B., and Farhadi, A. (2015). You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.

Ronacher, A. (2024). Flask documentation. URL: https://azcv.readthedocs.io/en/latest/, Last accessed: 02.06.2024.

Sudersanadas, K. (2021). Application of artificial intelligence on nutrition assessment and management. *European Journal of pharmaceutical and medical research*, 8(6):170–174.

Theodore Armand, T. P., Nfor, K. A., Kim, J.-I., and Kim, H.-C. (2024). Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review. *Nutrients*, 16(7).

Verma, M., Hontecillas, R., Tubau-Juni, N., Abedi, V., and Bassaganya-Riera, J. (2018). Challenges in personalized nutrition and health. *Frontiers in Nutrition*, 5.

Wang, J., Yang, Z., Yao, Z., and Yu, H. (2024). Jmlr: Joint medical llm and retrieval training for enhancing reasoning and professional question answering capability.

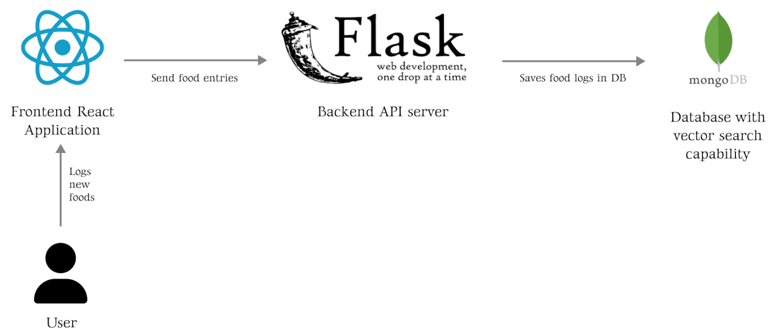
Yang, Z., Khatibi, E., Nagesh, N., Abbasian, M., Azimi, I., Jain, R., and Rahmani, A. M. (2024). Chatdiet: Empowering personalized nutrition-oriented food recommender chatbots through an llm-augmented framework. *Smart Health*, 32:100465.

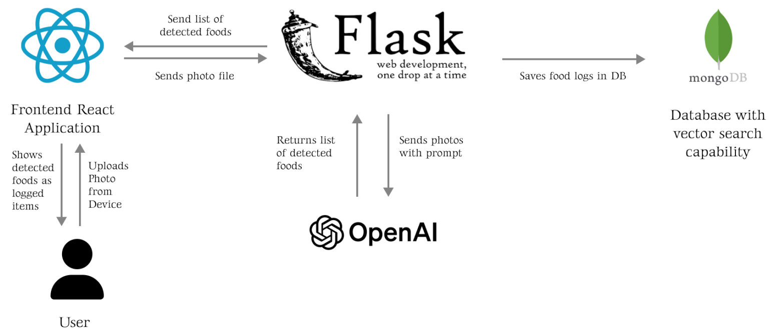
Zheng, K., Nguyen, T., Chong, J. H. S., Goh, C. E., Herschel, M., Lee, H. H., Liu, C., Ooi, B. C., Wang, W., and Yip, J. (2023). From plate to prevention: A dietary nutrient-aided platform for health promotion in singapore.

10 Appendix

10.1 Appendix A

Further visualizations of processes in IT infrastructure





10.2 Appendix B

Survey questions

- What sex are you?
- How old are you?
- Which of the following diets do you follow?
- Have you ever tried any of the following diets for longer than 3 months?
- Have you experienced problems related back to your nutrition
- Have you ever visited a dietitian/nutrionist?
- Are you currently seeing a dietitian/nutrionist?
- How high is your knowledge level on nutrition?
- How much would you like to know how your food choices impact your feelings, such as mood, energy,

- health, and sensations like bloating?
- On a scale of 1-5, how much would you like to learn about nutrition?
- In how far would you like to follow a diet / eating habit that is as environmentally sustainable as possible (e.g. eating local products or eating products with low water consumption)?
- Have you ever used an app on your phone or computer to help you with your nutrition?
- Are you currently using an application to help you with your diet (goals)?
- **How much do you identify yourself with the following statement regarding your usage of a nutrition app:**

  - I use(d) an app to reach an aesthetic fitness goal (e.g. looking strong or toned)
  - I use(d) an app to specifically reach weight loss
  - I use(d) an app to reach a health goal (feeling good in your body and reaching for longevity)
  - I use(d) an app to track my macronutrient intake (Protein, carbohydrates, fats)
  - I use(d) an app to track my micronutrient intake (like vitamins, potassium etc.)
- What is the maximum time you have continuously used the tracking app?
- How much do you enjoy the food tracking process?
- Which method for entering consumed foods have you been using?
- Which method for entering consumed foods would you like to use in an app?
- If you have stopped using the app, why did you stop?
- How likely are you to use/try a food tracking app in the future that focuses on health outcomes (not on aesthetics)?
- Have you ever used an A.I. chatbot, like ChatGPT or Google Bard, to learn about the impact of your food choices on your nutrition/health?
- How willing would you be to talk to a Chatbot about your nutrition goals and your progress so far?

- Which concerns would you have when using a food tracking app using artificial intelligence?
- On a scale of 1 to 5 how much would you like to find the ability to track work outs in your nutrition app and its impact on food recommendations?
- As a woman: On a scale of 1 to 5 how much would you like to find the ability to enter whether you have your period in your nutrition app and its impact on food recommendations?
- On a scale from 1 - 5, how much would you like to share the data from your nutrition app with your nutritionist?

10.3 Appendix C

Transcribed questions and answers of expert interview

Question:
What are common issues of your plant-based clients?

Nutritionist:
Well, I am going to tell you about it. Well, in relation to the problems, perhaps the main problem is that when they start following a plant-based diet, they do it, that is to say, they continue with their usual diet and simply eliminate the protein part, mainly meat and fish. So that can mean that there is never a tremendous risk of a lack of protein. But well, it is true that in the end, well, maybe it would be a little poor in that sense because we could be including, well, vegetable protein sources that maybe are not included. That could be one of them. Then another very common problem is the lack of B12 supplementation, especially in vegetarians because perhaps vegans, which still happens, but vegans may be a little more aware, but vegetarians often, since they consume eggs and dairy products, think it is not necessary and that is quite a serious problem. Then, I don't know, sometimes maybe there is a sense of loss in the sense that I am starting to follow the diet, but I don't know how to do it or eliminate foods that I used to eat. I don't know what other foods to include or how to cook them. Maybe tofu I'm just not familiar with. A little bit more so in terms of how to organize the dishes or how to organize the menus.

Question:
What do you think of modern chatbots, like ChatGPT, and their usefulness in nutrition?

Nutritionist:

Regarding chatbots, the truth is that I don't know too much about them either, I mean, I have heard things and so on, both that they answer questions and that they do your diet, but the truth is that I wouldn't dare to get too involved because I don't know them. I mean, I would say, obviously I think that they do not offer a service even close to what a nutritionist can give you, right? I mean, obviously. Then I don't know to what extent they can wrong or maybe even say barbarities. From what I understand, which is not much, I would say that perhaps it could be a starting point that, supervised by a dietician-nutritionist, could be something decent. But without that supervision I do think it can be a bit catastrophic because in the end I imagine that it won't be very personalized no matter how much input you put into it. Or I don't know, there will be a lot of things that it might not take into account, right? Maybe a kind of relationship with food and everything that is a little bit related to food beyond kilocalories, macronutrients, etc.

Question:

Are your clients using applications to track their diet?

Nutritionist:

Look, in relation to the applications that people can use to monitor their diet, I would say that it depends, in the sense that, for the people I have dealt with, perhaps omnivorous or a little more of, let's say, a more conventional way of eating, not so much, but it usually happens, often vegan patients go a little hand in hand with the fact that maybe, I do not know, maybe they are more concerned about their diet or, of course, they have had to investigate more and, therefore, they have it more in mind. Well, for whatever reason, I do have the feeling that in this type of patients, well, simply vegan patients, well, that there are things that they do enjoy, that they have much more in mind, for example, food portions, weight, etc, and one of the things is also the use of applications, that, well that, they sometimes record the food with MyFitnessPal. Well, I know there are several apps, but that's the one that has come to my mind primarily, I know they use it to record all the meals they eat and so on, so yes, I couldn't tell you a percentage or anything, but well, it is used.

Question:

Do you have an application to access shared data from your clients? What would you think of such an application?

Nutritionist:

In relation to accessing client results, some patients do mention it to you about backing up, send me pictures of what you eat or write it down and send it to me and so on, and they say ah, well if you want I'll send you the screenshot of the application. Well, I think that yes, it could be useful. That is, if it were easy and convenient to access, because for me, for example, a screenshot seems to me like a pain, I do not want to have a screenshot of the application on my mobile, I prefer that you complete the record that I send you. But if it is a common application that we both share and that is registered there, then yes, it can be quite convenient.

Question:
Do you create content, like articles, on nutrition and do you enjoy this process?

Nutritionist:
Let's see, in relation to the topic of articles, I mean, in fact one of the things I do most, a little bit in line with teaching, but let's see, I mean, if I want to make a good article, I mean, there are two types, right? An article that you do a little bit for the sake of doing it, because sometimes, if you have to publish every week on the blog, I try to keep it active, I can not be doing every week an article, reading a lot of biography, because it would take me a lot of time for something that I'm just doing it so that my website has movement, right? The article, well, I do it once a month. I don't know if I'm making myself clear, it's as if writing articles is cool, but you have to take into account that if you do it with a low frequency, I think you can get good articles and if that is going to give you some kind of remuneration, profit, whatever, but if it's about writing articles every two weeks, in the end, that demands a lot of time, especially since then in the end you are going to do very simple articles, so in the end they are going to be of low quality.

Question:
Do you create also other content, like videos and articles, on nutrition and do you enjoy this process?

Nutritionist:
And then, well, when it comes to videos and recipes, let's see, I don't know, it's not that I'm passionate about it, I do it because, well, it's good for traffic and I think I have to be there and so on, but well, it's not that I love making videos. I do enjoy the articles more, but the videos are like preparing it, maybe if you are good at it and it comes to you spontaneously, then great, but for me, which I would say is not my case either. And recipes, let's see, yes, but well, it's a little bit the same, it also takes work, I mean, if I have a database of great recipes, from there I can get recipes, well, but to be inventing recipes every now and then, it's a time you have to sit down

and do it, and maybe you can invest that time in something more productive, so well.

Question:
How do you and your colleagues find new clients and how difficult is it?

Nutritionist:
I would dare to say that, well, it is not easy. Customers do not come to you, no, you do not sit in the office and the clients come, but you have to move, you have to really work actively so that they come, and well, and how they find them, in the end it is word of mouth, that is, much more, so I tell you, from what other colleagues and others have told me, much more than being in networks, doing one thing or doing another, which is still important because in the end I think it is a bit your letter of introduction, right? , But above all word of mouth, that is, doing a good job with one person and that person recommends it to another person, and that person, and so on.

Question:
Would you be willing to be listed as a professional nutritionist contact on a nutrition/tracking app, so that users can contact you?
Would you be willing to pay for a user to contact you from this app? How much would you pay?

Nutritionist:
If you are looking for patients and so on, yes that could be a good option. It occurred to me that it's a little bit similar to Doctoralia, right? I know some colleague using it, I think it's paid, I'm not really sure either. But I understand that you have to pay to be there and I think it works well for them, actually. So yes, I understand that it would be something similar and well, it could be a good option. And paying for a user to contact you? Well, I don't know. I mean, of course, in case of doctoralia for example, you pay to be available on the application. I do not know, I would rather want to have a guarantee that the application works, yes, I don't think I would want to register, otherwise, you know. If it is a small amount, I'd say, well, I'll go and try it. But if I don't have minimum guarantees or if I'm not sure that it works and that it's going to pay me back, then no, I would not want to use it.