Bright Bites: a study on the impact of healthiness explanations in a recipe recommendation application

Fundamenten van Mens-machine interactie [G0Q55a]

Wout Bergmans
John Gao
Quentin St-Pierre
Jules Verbessem
r0852964
r0850136
r0855527
r0957436

ABSTRACT

This paper introduces the newly created application **Bright Bites**, which is a mobile recipe recommendation application that will provide recipes to a user according to their tastes. To encourage users to choose healthier options, additional health explanations are added with each recipe. The system first takes user preferences as input: cuisine preferences, diets and ingredients. These then serve as input for an AI model, which will generate recipes accordingly, along with healthiness explanations. Several types of explanations were used: graphical with icons, textual with motivations and detailed health benefits/risks. The aim of this study is to find how these explanations would affect users in choosing healthier recipes. To do this, a user study (N = 24) was conducted where users were asked to use the app to choose a recipe, once without explanations and once with, followed by a survey. The study showed that participants did not significantly choose healthier options when provided with the healthiness explanations.

1 INTRODUCTION

As an adult having a busy schedule trying to balance school or work with personal life, it can be a hassle to spend time trying to find a suitable recipe to cook after a hectic day. How much time is there to cook, what are you in the mood for, what ingredients are available at hand, how healthy does the meal need to be, etc. These are all questions that take time to answer. To try and solve this, the newly created application **Bright Bites** tries to simplify this process and guides the user to their desired recipe. First, the app will collect users' preferences: specific diets, favourite cuisines and ingredients. The app then presents the user several recipes to choose from, according to these preferences. Along with each recipe are also health explanations, both graphical and textual. Using the application, a study is done into **how providing explanations about the healthiness of recommended meals influences users' meal choices**. Does including explanations actually influence the user to pick a healthier recipe?

2 RELATED WORK

Several commercial recipe recommendation applications already exist, such as **Tasty, Cookpad** or **Mealime**. These apps work with a similar system, where the user first provides some of their preferences and then receives recommended recipes. These recipes will mostly not contain healthiness explanations, unless the recipe author explicitly provides some themselves. Research was conducted on what type of explanations/justifications are more effective when choosing a recipe. In a study by Musto et al. [9], users were presented with two recipes: a healthy and a popular one, along with justifications for each recipe. These justifications were about healthiness, but also user skills and popularity. Their results showed that when comparing the two recipes according to health risks and nutritional info explanations, the user was more likely to choose the healthy option over the popular one, than when no explanations were provided (across all meal types: first course, second course and dessert). Similarly, providing health benefits explanations also increased the likeliness of choosing the healthier option, although this only applied to desserts. These results show that providing healthiness explanations does in fact influence users' choices of recipes, giving an indication to what the results of this research could be.

3 IMPLEMENTATION

3.1 Data

The data used for the implementation of the recipe recommender consists of two different datasets that are joined together using an outer join on the *recipe-id* column. This column is present in both of the datasets and since they were extracted from the same website, these id's correspond to the same recipes in each dataset. Both datasets can be found on *Kaggle* and their content was originally scraped from the food.com website. [2] [7]
This joined dataset gives the following information for all the recipes that the system can recommend:

- Name
- Image (if available)
- Estimated cooking time
- Ingredients (together with their quantities)
- Aggregated review rating
- Preparation steps
- Nutrition information

The dataset consists of even more information, which is not used further. That data goes through a preprocess function that makes it

more uniform; think of a lowercase function that will ensure that the recommender is not case-sensitive.

3.2 AI system

The AI included in this project consists of two different parts. Firstly, there is the information retrieval part, which makes use of a TF-IDF matrix to compute the most similar recipes to a given user query. This retrieval system decides what recipes should be returned to the users. This method of returning relevant recipes was chosen because it is a frequently used technique in the fields of information retrieval and natural language processing. It is fully implemented by one of the team members and thus allowed more control over the retrieval system.
During training, the data is used as input to train the TF-IDF matrix, which later on will be used to match query vectors up against. These query vectors are based on what a user of the app inputs and are converted using a vectorizer, the same one that was used to create the TF-IDF matrix. Then the most similar recipes are returned; this similarity is determined using a cosine distance in the vector space, where the matrix and the query vector are defined.

Secondly, there is the healthiness explanations generation part using a large language model (further abbreviated as LLM). The LLM generates healthiness explanations in string format given the recipes generated based on the user's query. This part makes use of what seems to be almighty LLMs lately. The system provides the LLM, in this case Gemini 1.5-flash, with details of a recipe, and expects an explanation about the healthiness of that specific recipe in string format. The reason Gemini is used, is that it provides a free API-key for limited use, and it is able to handle enough data for the current implementation. [4]

3.3 XAI aspects

As explained previously, Gemini is included in the project and is used to explain why some recipes are healthy or not. Once the recipes are acquired by the algorithm, they are fed into an LLM, accompanied by a prompt asking for explanations on the healthiness of the given recipes.
The LLM takes a lot of information into account, including the ingredients and nutritional details of each recipe. After looking at all the information related to a certain recipe, it evaluates the healthiness of a recipe based on the content of the recipe, thereby explaining its reasoning. This follows the rationalisation method [5]. An example of the textual healthiness explanations produced by Gemini is shown in Figure 1.
This approach provides users with a clear understanding of the LLM's reasoning by thoroughly explaining what contributes to the healthiness or unhealthiness of a recipe. This transparency helps to build user trust in the explanation.

3.4 Interface

Results of a short unofficial study to assess platform preference, showed unanimous support for a mobile application over a desktop version. Upon opening the application on a mobile phone, the user is presented with a home screen, shown in appendix B.1. Afterwards, the user completes a form that is used to generate the recommended

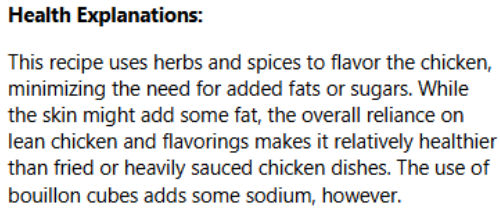

Figure 1: Healthiness explanation generated by Gemini about the Diet Herb Chicken Bake recipe

recipes according to their wishes. The form contains preferences for diets, cuisines and ingredients, of which the interfaces are added in appendices B.2, B.3 and B.4 respectively. For reasons explained in section 4.1, the interface can be separated into two parts, which have approximately the same layout.

One part of the application displays the recipes without healthiness explanations showing only basic recipe details, see interfaces 1.A and 2.A in figure 2. The second part of the application displays the recipes with healthiness explanations, showing nutritional details besides the basic recipe details, see interfaces 1.B and 2.B in figure 2. The provided textual explanations give insight into the recipe's nutritional value, highlighting both its beneficial and potentially harmful health aspects. The nutritional details include the amounts of calories, sugar, fat and sodium. A colour coded icon system (green, orange, red) is used to visually represent the relative levels of these nutrients. These icons are shown on the generated recipe cards, but also on the recipe details, this time with the numeric values of the nutrients.
In addition, the second part includes a brief motivational statement that aims to perceive the nutritional content of the meal in a more relatable way. The statement warns the user about the duration of an activity to offset the caloric intake of the recipe. An example of such motivational statement is "If you climb x flights of stairs, this meal is balanced out!".

When the recipes are generated, the application presents them as cards that the user can scroll through, see interfaces 1.A and 1.B in figure 2. Upon recipe selection, an overview of the recipe is presented, see interfaces 2.A and 2.B in figure 2.
A low-fidelity prototype sketched on paper (see Appendix C) served as the initial step for interface implementation. This was followed by a think-aloud study, with six potential users, to evaluate the initial design. The study concluded that the application appeared user-friendly, except for a few remarks, which were reviewed for the final design.
The application featured a button on the interface with the generated recipe cards that, upon activation, allowed users to review the preferences they had specified to generate the recipes. This button has been removed due to confusion that arose because the application showed the preferences without the option of modifying them. Additionally, some test users of the study did not understand what the colours of the icons were referring to. To accommodate this in the final design, a word was added underneath each nutrient icon

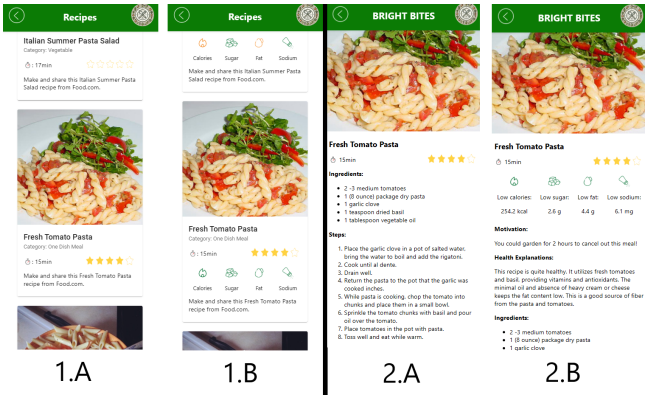

Figure 2: Interfaces for recipe selection

to provide additional clarity about the meaning of the colours (low, medium, high). With the feedback from the study, the final design was made in the collaborative web app for interface design, Figma [1].

The front-end is implemented using the JavaScript framework React [8]. It uses styling components and icons from the Material UI library [6]. The front-end communicates with the back-end via a REST API build in Python with the Flask library [3]. The application is hosted locally on the device of the developer. Mobile access to the application requires the user to open the mobile web browser and specify the following in the URL-bar: "http://developer-device-ip:3000".

4 EXPERIMENT

4.1 Experimental design

The experiment uses a within-subjects design, as each user will see both interfaces, first without healthiness explanations and then with these explanations. This allows for direct comparison of user behaviour between the two versions.
This is a lab study, as the experiment takes place in a controlled environment where participants navigate the system and perform tasks while being monitored by the researchers.

The flow of the study follows a sequence of steps:

- Start with an initial question to provide context for the test subjects.
- Make participants navigate through both versions and choose their preferred recipe:
  - Version 1: excludes the healthiness explanations, showing only basic recipe details.
  - Version 2: includes the healthiness explanations, i.e. showing details of the four nutrients (calories, sugar, fat and sodium), a motivational sentence and textual healthiness explanations.
- Answer the other questions in the survey that consist of two different formats.
  - Likert scale questions, to check for user satisfaction and user trust.
  - Ranking questions to indicate user preferences.
- A final Likert scale question about considering nutrition information in future food choices.

The reason the study has this design is to ensure that users experience both conditions, enabling the study to directly measure the impact of healthiness information on the user's recipe choices with quantitative data.

4.2 Measurements

There are a total of 11 different questions asked in the survey, which can be found in Appendix A.
The logged interactions exist of the navigation through the application, i.e. what prompt was inputted into the recipe recommender system, as well as the choice of the recipe made by the user.

The goal of logging navigation and choices is to determine whether the version with healthiness explanations encourages healthier recipe choices compared to the version without explanations. The result of this is directly related to the research question.
The Likert scale questions evaluate how clear, useful and trustworthy a user experiences the system. This gives an indication of user acceptance and understanding of the system.
The ranking questions are useful for understanding why certain users might have changed their answers depending on what they found important when choosing a recipe.
The final question is to measure long-term behaviour changes now that users have been exposed to nutritional awareness while using the system. This gives an idea of the likelihood that users would use this system or any other nutrition-based system in the long run.

4.3 Participants

24 participants were part of this study. They were recruited during social gatherings and asked for voluntary participation.
The study was open to anyone 18 years or older who frequently uses the Internet for daily tasks, such as searching for recipes. No specific characteristics were required, as the goal is to provide a broad overview representing a general user.

5 RESULTS

Although the healthiness of a meal is dependent on many person-specific factors such as gender, age, size, weight and multiple food-specific nutrients, the healthiness of a meal will be represented by four values for this research: calories, fat, sugar, and sodium. These are the same four nutrients that are included in the healthiness icons, and thus may have contributed to the choice of the users. This is an important assumption that needs to be taken into account when assessing the results of this analysis.
To compute relative values of these four nutrients, reference values with average advised consumption amounts are needed. The following nutrient reference amounts for one meal were used from the "Food and Nutrition Information Center" (FNIC) [10].

- Calories: 666.67 kcal
- Sugar: 16.67 g
- Fat: 22.22 g
- Sodium: 666.67 mg

In order to compare the degree of healthiness of the user's choice with and without explanations, the numeric, discrete data needs to be processed. For every person, the healthiness of the chosen recipe will be compared to the other nine generated recipes by the application. This comparison will be done twice per person, once without and once with the explanations.
In order to be able to compare the healthiness of different recipes with one another, nutrient values of each recipe relative to the reference values given above are combined into one number. This number represents the overall healthiness of a recipe and will from now on be referred to as the "overall healthiness ranking" of a recipe. It is important to note that this number does not accurately reflect the overall healthiness of a meal, as it considers only four nutrients. However, it is sufficient for comparing the ten meals with one another to compute a query specific ranking, enabling a quantitative measure of the user's choice. As this data is ordinal, the Wilcoxon Signed Rank Test is used to analyse the overall tendency of the users. This nonparametric test does not assume normally distributed data, which is important because this research has not enough participants to make this assumption.

The null hypothesis of the Wilcoxon Signed Rank Test states that: "there is no difference (in terms of central tendency) between the two groups in the population." [11] The test results in a p-value of 0.71 which means the null hypothesis rejection fails: there is no significant difference in the user's choice by adding healthiness explanations to the recipe recommendation application.

Besides the main research question, the answers to the survey questions can reveal additional interesting information. The questions of the survey consists of Likert scales and rankings, so those are all ordinal. The Likert scales are numerical and discrete, while the ranking questions are categorical and discrete. A first interesting aspect that can be analysed from this survey is the self assessment or consciousness of the test users about their tendency to choose healthy recipes. The only question asked before the use of the application, is to what extent the user strives to eat healthy. The last question, asked after the use of the application, investigates if the user is more minded towards healthy explanations for future food choices. Figure 3 shows a scatter plot with the results of those two questions. Notice how the trend line has a small but positive slope, which follows the intuition that people who take more care to food healthiness, are likely to take healthiness explanations more into account when making recipe choices.

That same first question, asking the user to what extent they strive to eat healthy, can be linked to the results of the experience with the application. Figure 4 pictures this relationship, where the y-axis corresponds to the difference of the two "overall healthiness ranking" numbers of a user's choices (with and without explanations). The trend line indicates a linear increase, suggesting that individuals who perceive themselves as mindful of healthy eating are more likely to select healthier recipes when healthiness explanations are provided. Some outliers are present (e.g. the point at $(4, -7.5)$) and are holding back the trend line to be steeper. Since the research does not contain a lot of participants, outliers weigh more on the trend line.

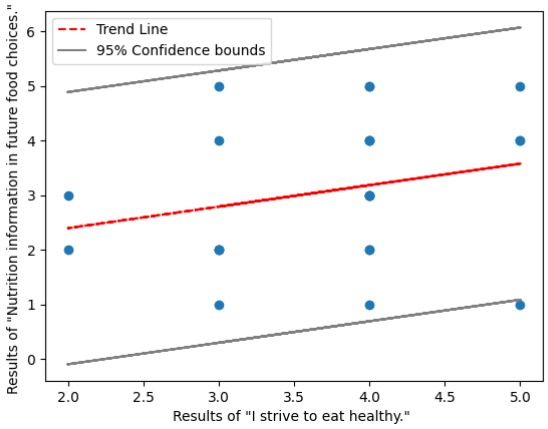

Figure 3: Relationship of healthiness consciousness and healthiness explanations in future food choices

Based on the experiment conducted, people are not choosing healthier recipes when healthiness explanations are added. However, this result gives insight in the way people choose recipes, and the survey questions further analyse this.

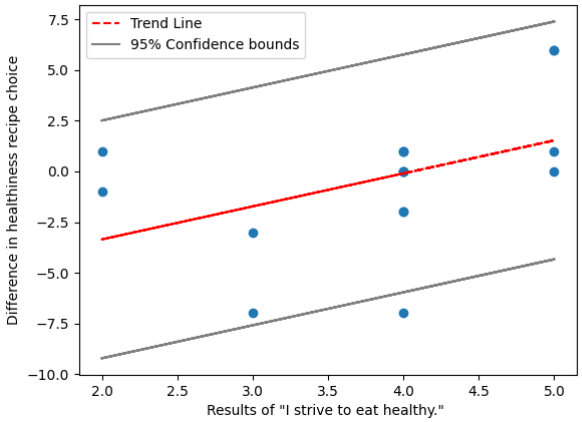

Figure 4: Relationship of healthiness consciousness and the experience results

In the survey, there was a question asking the user what specific element of a recipe influences their choice the most. In general, the image and title of a recipe were more impactful than any kind of healthiness explanation (textual, icons, motivation). This is an important observation that explains why people are not choosing healthier options with the healthy explanations. Some users may not have paid attention, while others might have ignored the healthiness explanations consciously when choosing a recipe.

The average ratings of other questions of the survey help to reflect on this. The following questions about the healthiness explanations scored relatively well (ratings were given on a scale of 5):

- “The explanations of the health benefits of the recipes were clear and informative.”: 3.92
- "The explanations clarified why a recipe is considered healthy or unhealthy.”: 4.00
- "The healthiness explanations were easy to understand.": 3.96

While the following question scored less:

- “The healthiness explanations were helpful for choosing a recipe.”: 3.67

Based on these results, it seems that the healthiness explanations were clear and well constructed, but that it may not contribute that much to the final decision of the user. In addition, notice the result of this question:

- "The explanations were detailed enough to trust the health assessment.”: 3.38

Therefore it seems the trustiness of the explanations could be improved. Providing evidence on the content of the explanations or professional advice to support the healthiness explanations could enhance users’ trust, potentially amplifying the impact of these explanations on their choices.

6 DISCUSSION

This research includes the analysis of the effect of healthiness explanations on recipe choices in a recipe recommendation application. The results of the analysis shows that users are not significantly choosing healthier recipes when healthiness explanations are included. Furthermore, the results of the survey revealed different behaviours as for healthiness diet concerns. The designed healthiness explanations seemed to be more efficacious on user profiles that are more concerned with a healthy food lifestyle. To that extent, motivational phrases were designed especially to catch the attention of user profiles that have more difficulties perceiving specific nutrient values. The survey revealed that, those motivational statements were contributing the least to the user’s decision.

Overall, this experiment highlighted the numerous complexities involved in incorporating healthiness explanations into a recipe recommendation application. Given that human health is a multifaceted issue influenced by numerous factors, integrating healthiness information while meeting the needs and preferences of the widest possible audience proves to be a challenging task. Hopefully, this research will open some doors for future work, where different kinds of healthiness explanations for recipe recommendations could be tested on a larger population. Conducting similar research on a bigger scale, i.e. a larger population, could not only provide stronger results but would also reveal many more behaviours of users when using a recipe recommendation application.

7 CONCLUSION

As part of this study, the application **Bright Bites** was created, which takes user’s eating preferences and outputs a list of recipes along with healthiness explanations, in an effort to encourage users to choose healthier recipes. After performing a user study, the results have answered the original research question: **how does providing explanations about the healthiness of recommended meals influence users’ meal choices?** The results show that users did not significantly choose healthier options when provided with explanations, this could be due to several factors, such as the types of explanations used or the limited population tested. The survey results indicate that, even though the current explanations are well constructed, users still prefer to choose based on recipe images and titles. Future work could thus focus on this aspect, trying other kinds of explanations which could influence the user more than recipe images and titles, and further stimulating users to choose healthier recipes.

REFERENCES

[1] Adobe. [n.d.]. Figma: Design, prototype, and collaborate all in one place. https://www.figma.com/

[2] Alvin. 2020. Food.com - Recipes and Reviews. https://www.kaggle.com/datasets/irkaal/foodcom-recipes-and-reviews/data

[3] Ronacher Armin. [n.d.]. A lightweight WSGI web application framework. https://flask.palletsprojects.com/en/stable/

[4] Google. 2024. Pricing models Priced to help you bring your app to the world. https://ai.google.dev/pricing?authuser=2#1_5flash

[5] Sai Gurrapu, Ajay Kulkarni, Lifu Huang, Ismini Lourentzou, and Feras A. Batarseh. 2023. Rationalization for explainable NLP: a survey. *Frontiers in Artificial Intelligence* 6 (2023). https://doi.org/10.3389/frai.2023.1225093

[6] Nguyen Hai. [n.d.]. Move faster with intuitive React UI tools. https://mui.com/

[7] Shuyang Li. 2021. Food.com Recipes with Search Terms and Tags. https://www.kaggle.com/datasets/shuyangli94/foodcom-recipes-with-search-terms-and-tags

[8] Meta. [n.d.]. React: The library for web and native user interfaces. https://react.dev/

[9] Cataldo Musto, Alain D. Starke, Christoph Trattner, Amon Rapp, and Giovanni Semeraro. 2021. Exploring the Effects of Natural Language Justifications in Food Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’21)*. Association for Computing Machinery, New York, NY, USA, 147–157. https://doi.org/10.1145/3450613.3456827

[10] U.S. Department of Agriculture. [n.d.]. Food and Nutrition Information Center (FNIC). https://www.nal.usda.gov/programs/fnic

[11] DATAtab Team. 2025. Wilcoxon signed-rank test. https://datatab.net/tutorial/wilcoxon-test

A APPENDIX A: SURVEY QUESTIONS

- To what extent do you agree with the following statement: “I strive to eat healthy.” answer 1-5
- Go to http://10.0.0.11:3000/ (link where bright bites can be found) and check the "Without healthiness explanations" box. Follow the instructions on the application to indicate your preferences, and generate recipes. Afterwards, choose one recipe you would cook/eat and enter the recipe title below.
- Now go back to home page and uncheck the "Without healthiness explanations" box. Generate again some recipes, choose one, and enter its title here.
- To what extent do you agree with the following statement: “The explanations of the health benefits of the recipes were clear and informative.” Answer 1-5
- To what extent do you agree with the following statement: "The explanations clarified why a recipe is considered healthy or unhealthy.” Answer 1-5
- To what extent do you agree with the following statement: "The healthiness explanations were easy to understand." Answer 1-5
- To what extent do you agree with the following statement: “The healthiness explanations were helpful for choosing a recipe.” Answer 1-5
- To what extent do you agree with the following statement: "The explanations were detailed enough to trust the health assessment.” Answer 1-5
- Which elements of the recipe recommendations had the most influence on your decision? Rank the following options from the most impactful (top) to the less impactful (bottom). Choose between recipe image, recipe title, healthiness icons, textual healthiness explanations, motivations.
- Which features of the healthiness explanations had the most influence on your decision? Rank the following options from the most impactful (top) to the less impactful (bottom). Choose between calories, sugar, fat, sodium.
- To what extent do you agree with the following statement: “I will probably take nutrition information more into account in future food choices after using this system.” Answer 1-5

B APPENDIX B: SCREENSHOTS INTERFACE

B.1 Home screen

Figure 5, the home screen of the Bright Bites application on a smartphone.

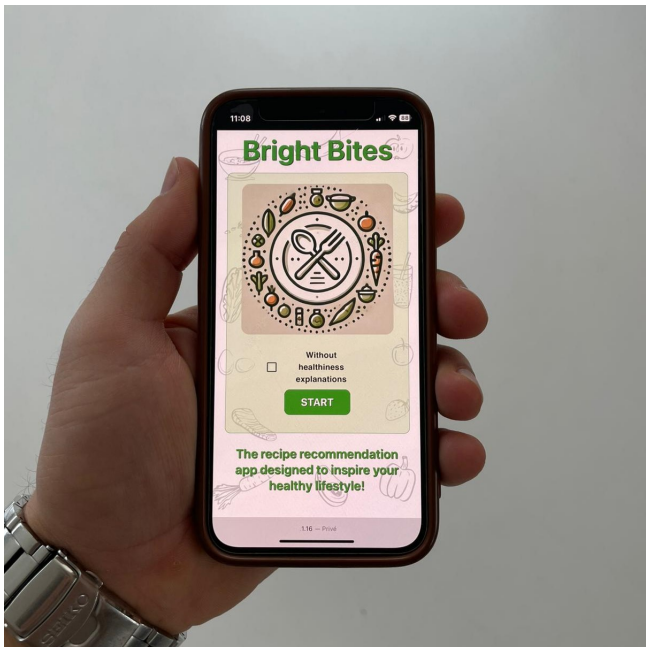

Figure 5: Home screen on smartphone

B.2 Dietary preference form

Figure 6, a screenshot of the dietary preference screen. Notice how the "SKIP" button switches to a "NEXT" button when checking at least one box. This induces that as a user, you can skip the question without indicating any preference.

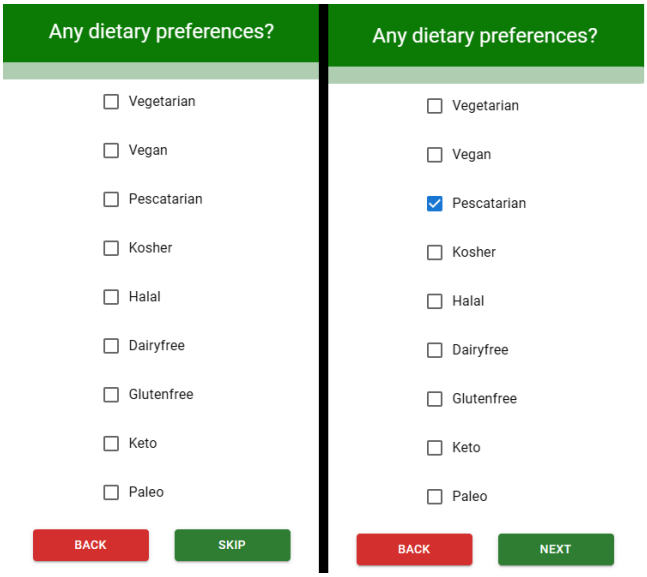

Figure 6: Dietary preference form

B.3 Cuisine preference form

Figure 7, a screenshot of the cuisine preference screen.

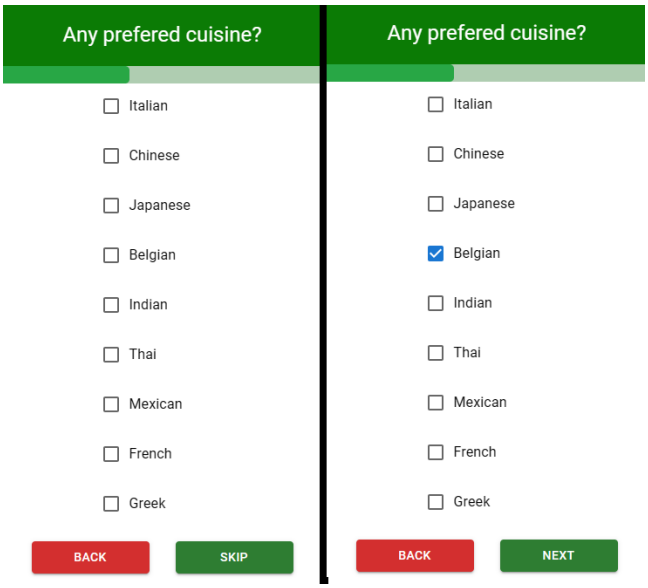

Figure 7: Cuisine preference form

B.4 Ingredients preference form

Figure 8, a screenshot of the ingredients preference screen. The user can type any ingredient they would like to include in their recipes. Ingredients need to be added one by one, by clicking on the "ADD" button each time. The "chosen ingredients" box lists all the previously added ingredients, with the possibility to remove them by clicking on the little cross next to the ingredient name. After the user added all desired ingredients, the "FIND" button will proceed to the recipe generation.

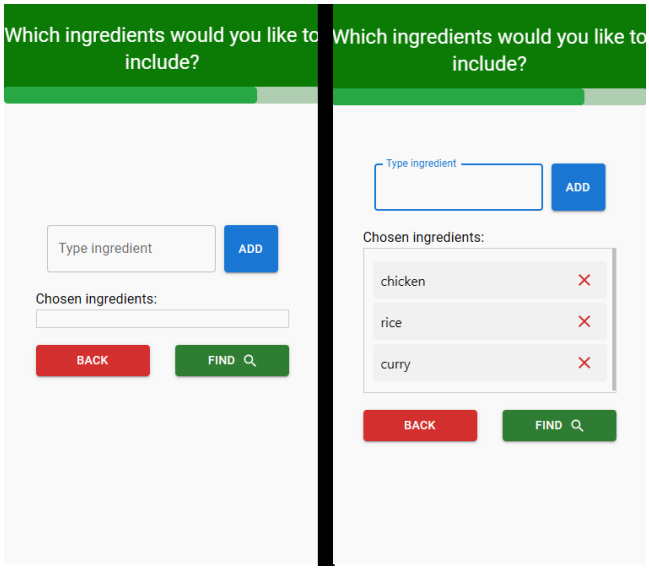

Figure 8: Ingredients preference form

C APPENDIX C: LOW-FIDELITY PROTOTYPE

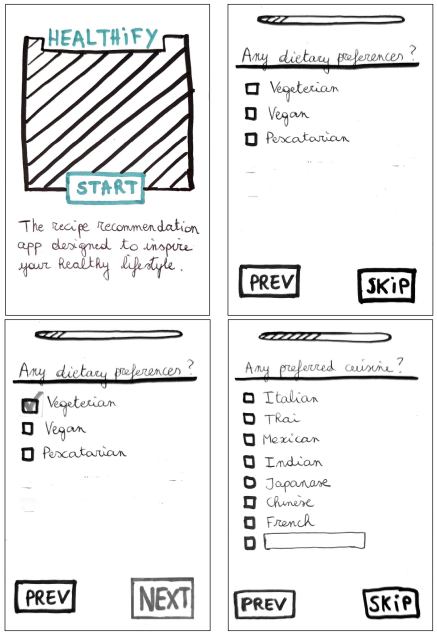

Figure 9: Low-fidelity prototype part 1

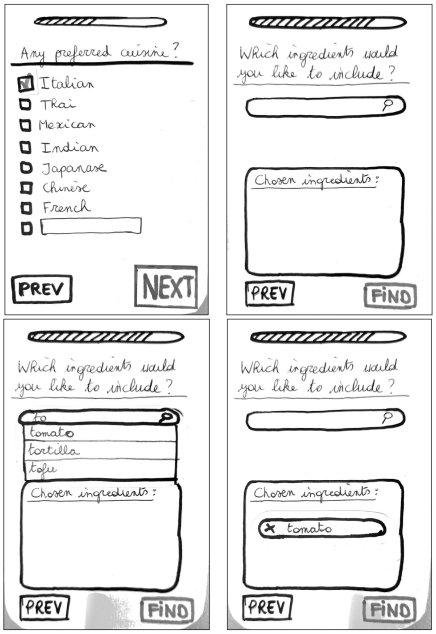

Figure 10: Low-fidelity prototype part 2

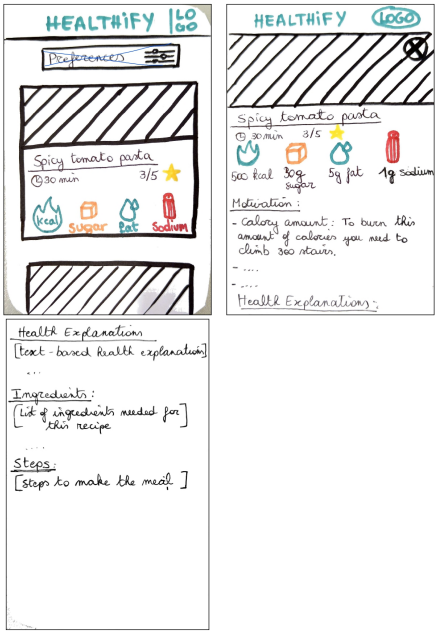

Figure 11: Low-fidelity prototype part 3