École Polytechnique Fédérale de Lausanne

AI-Enhanced Reminders for Food Tracking Adherence: A Randomized Controlled Trial

Author: Romain Guy Daniel Rochepeau

Master Thesis in Life Sciences Engineering

Conducted within the Digital Epidemiology Lab

EPFL
Station 19
CH-1015 Lausanne
Switzerland

Thesis Committee

Prof. Marcel Salathé
Thesis Advisor

Prof. François Grey
External Expert

August 12, 2025

Acknowledgments

My first thoughts go towards Alysée and her unwavering support and help during the course of this project. You often relieved my negative thoughts and always believed in me when I did not, and for that I will never thank you enough.

I would like to thank Marcel Salathé for the opportunity to work on such an inspiring project. You made me feel autonomous and confident in my choices and your insights were always positive. Many thanks to DJ, for your help with MyFoodRepo and for always responding to my requests with a smile – I would have been blocked more than one time without you ! I would also like to thank Rohan and Marouane, who always offered me their help and with whom i shared insightful discussions about my work or other topics. You made this Master Project all the more enjoyable. Thank you Yannis, Céline and Sarra, for your consistent positivity and all the help you provided me with.

Special thanks to Andrea, without whom this entire project would not have been possible. Thank you for your availability, time and continued interest towards *our* project ! I hope I have lead the latter to what you expected when you started it.

My heartfelt thanks to my friends, who helped me testing my numerous implementations throughout the project: thank you Omid, Mahdi, Alexandre, Nicolas, Gérard, and Séraphin – for your insights, feedback and continuous support.

Finally, I would like to thank my parents. Not only for helping me during this project, but for believing in me from the very beginning of my journey at EPFL to now. Your constant support and words of encouraging kept me going during all those years – thank you from the bottom of my heart.

Lausanne, August 12, 2025 **Romain Guy Daniel Rochepeau**

Abstract

Significant advances in AI-powered food recognition have allowed dietary tracking to shift from a burdensome process to an easier and more enjoyable experience. However, sustaining long-term user engagement in that context continues to be problematic, limiting the collection of meaningful dietary data. This project explored using AI beyond food recognition, focusing on LLM-based personalization. We thought that personalised reminders and intelligent chatbot support would improve user engagement and uptake, compared to existing generic reminder strategies.

We conducted a randomised controlled trial (RCT) with $N = 112$ participants over 14 days using MyFoodRepo, a reliable dietary tracking application developed within the Digital Epidemiology Lab at EPFL. Participants were randomly assigned to four groups : Group 0 (no reminders, no chatbot), Group 1 (generic reminders only), group 2 (chatbot support only) and group 3 (personalised reminders and chatbot support). All personalised interventions were delivered to users via SMS and were powered by GPT-4o, supported by carefully designed prompt engineering.

We observed that personalised reminders significantly improved the initial app adoption among participants, with 89% of group 3 participants starting to log meals compared to 52% in the control group ($OR = 7.7$, 95% $CI = [1.88; 31.87]$). The time to first use was also significantly faster in group 3, with a median of 0.4 days to app uptake, against 4.9 days for our control group ($p = 0.0009$). Our chatbot unfortunately showed strongly limited uptake, with only 7 out of 46 eligible users engaging with the assistant, for a total of 37 messages sent. Adherence patterns did not significantly differ across our groups, with most participants dropping out in the first days, albeit encouraging results for personalised reminders compared to the two other interventions in group 1 and 2. In total, 30 of out of 79 (38%) initially active participants remained adherent throughout the complete study.

Our results showed that personalised reminders, through LLM functionalities, can improve and accelerate user onboarding. In our specific setup, personalised reminders were better than stale reminders or chatbot availability at maintaining adherence, but the good performance of the control group nuances this result. Future work should focus on refining reminder-based strategies even further, and integration of the setup directly in MyFoodRepo.

Contents

Chapter 1

Introduction

1.1 Towards an automated dietary tracking

The ubiquitous adoption of smartphones over the past years has opened very promising perspectives for dietary assessment. Mobile devices indeed provide a convenient and non-invasive way of tracking food intake, making dietary self-monitoring more accessible. Nevertheless, manual self-reporting remains a major barrier to sustained dietary tracking as, even when helped by user-friendly features, logging every meal is still tedious and time-consuming.

Photographs can potentially greatly simplify the process, as they offer a practical and intuitive way to capture meal information. Building on this idea, artificial intelligence (AI) emerged as a very promising solution in recent years, with advances in computer vision and the development of numerous complex deep learning models specialized in food recognition from photographs [21] [41]. Being able to automatically estimate food items and quantities from a simple photograph greatly reduces the burden on users, transforming dietary tracking into a more enjoyable experience.

Amongst novel applications leveraging these advances, MyFoodRepo (MFR), developed within the Digital Epidemiology Lab at EPFL, is a notable example. Assessments of the application have shown solid performances in both food identification and portion size estimations [43], demonstrating its potential to provide users with a precise and reliable method for dietary self-monitoring. The application additionally proposes to enter meal entries via text and barcode scanning, allowing users to track food information more easily. MyFoodRepo additionally contains an app-integrated chatbot, that helps users to refine their meal entries when necessary.

1.2 The user adherence barrier

While these technological improvements enhance accessibility towards dietary tracking, they have not fully addressed one of its most persistent challenges – user adherence. Here, adherence represents the engagement and long-term consistency of a user's dietary tracking activity.

Although smartphone usage has become rooted in our daily lives, systematically and consistently documenting meals is not a habit for an average person. Moreover, the impact of image-based recognition, as a feature, in the engagement of users over time has not been extensively studied. The limited existing research, generally focusing on health-related topics, do not show a direct impact of this feature on user adherence [6], [24].

In the context of dietary tracking specifically, the collection of meaningful dietary information requires long-term and consistent usage, underlining the importance to improve user adherence and motivating the research of our project.

1.3 Addressing user adherence

1.3.1 Reminders: a double-edged sword

Reminders are a widely used strategy to encourage user engagement. Typically taking the form of brief notifications, they prompt users to engage with the app, helping to maintain consistent usage over time.

While their omnipresence reflects of their positive impact on improving user adherence and engagement, they can also have negative consequences. By nature, reminders are unsolicited by the user and, as such, they can quickly lead to fatigue or annoyance [4] [3]. This highlights the need for careful design of reminders.

1.3.2 Personalisation to the rescue

Personalisation of user experience has long been hypothesised to significantly improve user engagement and adherence in mobile-related settings [20] [10]. By tailoring app experience to the individual users, personalisation can promote a stronger support and guidance throughout the user's tracking journey, as well as providing them with relevant information, based on their behaviour, needs and goals.

Since the emergence of models like ChatGPT, Large Language Model (LLMs) have demonstrated their ability to generate highly qualitative content, while being extremely accessible. Large Language Models are AI models that have been trained to understand and generate human language and when given sufficient and appropriate data, the information generated can be highly personalised and precise.

Looking beyond the benefits of artificial intelligence as a food-recognition system, LLMs hold promise as accessible virtual support that would improve the users tracking experience, if integrated correctly. They can also provide a way to improve reminders in a personalised manner, hopefully reducing fatigue and improving user adherence.

1.4 Proposed Solution

1.4.1 Personalisation channels

Based on considerations stated above, we sought to investigate the impact of LLM-based personalisation on user adherence in a food-tracking context. To do so, we focused on two complementary channels of personalisation:

- **Personalised reminders:** Rather than relying on generic notifications, we designed reminders that adapt in tone and content in order to better align with each user's dietary goals and behaviour. In this way, we hypothesised that reminders would act less like static alerts and more like a companion that intelligently nudges users to stay consistent with dietary tracking. These reminders would be sent via SMS to the users, prompting them to engage with the application.
- **Interactive chatbot:** Motivated by the integration of chatbots in health-related RCTs, [40], we implemented one in our project. Beyond reminders, a chatbot can offer a more dynamic and conversational form of personalisation. By responding to users' questions, offering tailored tips based on logged data, and providing encouragement, a well-designed chatbot helps humanising the interaction and helping the users feel guided, rather than monitored. This chatbot would be available to the users via SMS at all times, and is separate from the MFR-integrated chatbot, which only help users with the meal logging process.

1.4.2 Investigating with a Randomised Control Trial

To effectively understand the participation of each implementation on user adherence, we therefore designed a Randomised Control Trial (RCT), in the specific context of MyFoodRepo. In this setup, users are to be randomly spread in four different groups and asked to use MyFoodRepo, while receiving different interventions regarding personalisation and reminders.

The project therefore consisted in both investigating how to use LLMs to produce personalised content, as well as fully implementing and conducting the RCT and, finally, analyse the obtained results, where we aim to compare the impact of stale reminders with personalised ones, as well as the importance of a chatbot availability on user adherence.

Chapter 2

Personalising user experience with LLMs

While theoretically promising, carefully designing our implementations is crucial to ensure that personalisation feels authentic, relevant and truly helpful to the user. Particular attention must be given to reminders, in order to reduce the discomfort or fatigue they may cause to users. Crafting optimized reminders thus focused most of the efforts, which are described below.

2.1 Leveraging the power of LLMs through GPT

2.1.1 General considerations

For its accessibility and ease of integration, both the personalised reminders and the chatbot interactions in our study are powered by OpenAI's LLMs. We used OpenAI's API to communicate directly with the LLM, to either generate personalized reminders or answers to users' messages in conversational exchanges. We relied on GPT-4o throughout the project, due to its speed and overall good performances on language-related matters [31].

When using LLMs, the quality of the model's responses depends heavily on how the request is formulated, through what is called a *prompt*. For a model to be relevant and useful in its answer, it is primordial to precisely and clearly instruct it with the task it should conduct. In the case of personalised reminders, the prompt represents the entire request asked to the LLM, while in a chatbot setting, it would accompany the request of the user. We tried to optimize these prompts content and structure in order to fully leverage the capabilities of GPT-4o.

2.1.2 Handling GPT parameters

GPT offers two key parameters, specifically *temperature* and *top_p* that are commonly used to adjust the diversity and creativity of the generated content [12].

- *temperature* (typically ranging from 0 to 1, or 2 for some models) controls the degree of randomness in generated content. Lower values lead to more deterministic and predictable text, while higher values yield more varied responses.
- *top_p* influences creativity by limiting the choice of words to the most probable ones. Practically, the model will only select words for which cumulative probability mass, compared to other words, does not exceed the value we choose. Lower values thus result in more conservative and repetitive outputs, while higher values generate more varied and creative text.

While these parameters are widely used to tune creativity within text generation purposes, several recent studies such as the work of Matthew Renze [29] or Peeperkorn et. al [27] heavily nuance the extent of their impact on creativity. Knowing this, we thus set both their values to one, ensuring our model remained free in its choices of words and phrasing.

2.2 Key qualities for effective reminders

In order to define a clear and established framework for reminder design, focus was thus narrowed down to five core criteria, that guided the reminders design process, as well as their evaluation later on.

- **Personalisation:** Reminders should be closely tailored to the user's specific context, following our hypotheses. This includes the data we collect both before and during the study.
- **Engagingness:** As discussed earlier, maintaining user engagement over time is a major challenge. We must therefore craft reminders that actively motivate and encourage users to keep using MFR.
- **Diversity:** The time dimension is mandatory to consider in our setup: even personalized reminders risk feeling repetitive, if they are always similar in tone or structure. We must therefore ensure diversity in the latter, in order to sustain interest over time and avoid reminders fatigue.
- **Naturalness** The generated text must feel natural and human-generated. We hypothesize that this helps creating an authentic and close interaction with the user.
- **Coherence** While improving personalization, creativity and diversity, it remains crucial for reminders to stay coherent. They should deliver clear and concise information over time.

2.3 Prompt engineering as a design tool

As the models parameters we can act on are rather limited (see 2.1.2), we decided to focus on a powerful tool in the context of LLMs when it comes to dictating a model's output: prompt engineering. Prompt engineering represents the process of efficiently designing prompts that optimize the output of an LLM. It encompasses modifications to both content and structure of a prompt, which both play an essential role to influence the model's behaviour [36].

The practice of prompt engineering is as recent as the emergence of LLMs, and is therefore an area that is evolving rapidly, alongside models capabilities. Designing a relevant prompt is generally task-specific and there is no universal way of optimizing the latter. Nonetheless, there is now a wide range of practical techniques that consistently improve output quality, across very diverse tasks. Basing ourselves on systematic reviews of the latter such as [14] or [8], we decided to investigate some of these strategies in order to improve the quality of the model's output, for our specific reminders generation task.

2.3.1 Structuring personalised context

For the output to be tailored to the user information, we must provide it to our prompt. Providing the latter in a precise and structured manner to the LLM is a crucial task for the model to clearly internalize all necessary information and is described below. [14]

Structuring dietary and communication information

To ensure the model had sufficient context to generate relevant and tailored answers, we structured the following elements:

- **User personal profile -** We structure the user's age, gender, preferred language and dietary objective(s).
- **Dietary history -** We pre-processed users' dietary history for our prompts, in a structured manner. It consisted of an ordered and timestamped list of meals, accompanied by daily and study-wide averages of nutrients, to ease the computations for the model.
- **Recent reminders (for reminders only) -** We also explicitly built a list of the last reminders, in order to increase variety and avoid repetition in the user's outputs through the study.

Individual nutritional profiling

MFR, in its current state, only provides users with information regarding their carbohydrate, protein, fat and overall calorie intakes, without further guidance. We decided to add contextual recommendations, aligned with the user's personal dietary goals which we hypothesised would allow users to rapidly see benefits from food tracking.

For each participant, we built a personalised nutritional profile focused on the quantities mentioned above as well as added fiber, salt, and sugar levels. Using a 7-day rolling window, we computed averages of these quantities, based on their food entries on MFR. We then compared them to dietary recommendations of recognized organizations, such as the European Food Safety Agency (EFSA), World Health Organization (WHO) or United States Department of Agriculture (USDA) (see Table 2.1).

Nutrient	**Recommended range**	**Source / notes**
Energy (kcal)	Varies by age, gender, and activity*	EFSA [1]
Carbohydrates (% of energy)	45–60%	EFSA [1]
Protein (% of energy)	10–35%	U.S. Dietary Guidelines [35]
Fat (% of energy)	20–35%	EFSA [1]
Fiber (g)	25–35 g	EFSA [1]
Salt (g)	< 5 g	WHO [39]
Sugars (% of energy)	< 20%**	Based on EFSA; adapted to MFR data [1]

Table 2.1: Recommended daily intake ranges and their sources.

* EFSA provides ranges depending on age, gender, and physical activity level. As we did not collect physical activity data, we used a permissive interval from sedentary to active individuals within the participant's age and gender group (see Appendix ...).

** Official guidelines typically apply to added or free sugars only (generally limited to 10% of energy intake). Since MFR reports total sugars without distinguishing added sugars, we used a permissive threshold of 20% of total energy.

For each nutrient, a simple structured sentence describing the user's situation with regard to the recommendations was produced. We then compiled the sentences into a list representing the user's current nutritional profile. The latter was computed before each LLM request and included in our prompts (Figure 2.1).

Nutritional Profile

- Calorie daily intake is low at 1621.81 kcal (recommended: 1800–2600 kcal).
- Fat daily intake is within the recommended range at 33.72% of calories.
- Carbohydrates daily intake is low at 43.16% of calories (recommended: 45–60% of calories).
- Protein daily intake is within the recommended range at 18.51% of calories.
- Fiber daily intake is low at 23.21g (recommended: 25–35g).
- Sugar daily intake is within the recommended range at 8.07% of calories.
- Salt daily intake is high at 5.80g (recommended: 0–5g).

Figure 2.1: Example of personalised nutritional profile generated and included in prompts.

2.3.2 General guidance prompt components

Personal prompt components

Based on the nutritional profile defined in 2.3.1, we defined two concise prompt components targeting nutritional deficiencies and excesses. They respectively asked the model to highlight a recent gap or excess in the user's diet and to acknowledge any recent improvements in other categories, to promote positive reinforcement. To ensure diversity and hypothesising this would further enhance user interest, we defined a *Fun Fact* prompt component, that would induce the model to include a brief fun fact about a food relevant to the user's eating history and dietary goals.

Importantly, we assigned arbitrarily chosen weights to each prompt based on the user's current dietary data. The idea is to prioritise relevance, and to focus on deficiencies if we detect some, for instance. When an actual nutritional deficiency or excess was detected, we increased the weight of the corresponding prompt, making it more likely to be selected (Fig 2.2), while adjusting other weights. We still implemented a skip weight, in order to avoid spamming users with similar recommendations. Only one of the prompts could be included in the final reminders, ensuring personalised yet diverse reminders.

Tag	Weight	Condition / Explanation
`Skip`	0.8	Default prompt to skip personalization; moderate probability to keep variety.
`Deficiency`	2 *if* has a deficiency, else 0.2	Higher weight when user's data shows a deficiency; very low otherwise.
`Excess`	2 *if* has an excess, else 0.2	Higher weight when user's data shows an excess; very low otherwise.
`Fun Fact`	0.2 *if* has a deficiency or has an excess, else 1.2	Default personalization prompt; reduced weight when an issue is detected to focus on relevance.

Table 2.2: Prompt pool with conditional weights for personalization

Additional components

To ensure the right format, we also inferred some inflexible rules to each of our prompts (maximum number of characters, messages should be redacted in the user's language...). Importantly, we also designed a prompt component for users that never logged. The aim of this component is to encourage them to start logging onto MFR by linking this incentive to their dietary objective. We hypothesized that this would increase engagement with the application for the first time. In the same manner, if the user logged a meal in a 30 minutes window before sending the reminder, the latter would compliment them on logging a meal rather than prompting them to log one.

2.3.3 Investigating prompt engineering techniques

While structuring the personal information as above and including it in our prompts is mandatory for personalisation, we also sought to test some effective techniques to improve our other criteria.

Persona

By assigning a *persona* to an LLM, we explicitly define the model's role and identity within the prompt [14]. This technique aims to guide the underlying reasoning process of the LLM, guiding the tone, content and style of the generated text [9] to better fit our design goals. We decided to instruct the model to act as a *food diary assistant* whose role is to accompany participants throughout their food-tracking journey. To further refine this persona to the scope of our project, we decided to incorporate principles from behavioural psychology, namely *goal setting* [19] and *habit formation*

[16]. In our *persona* prompt component, we therefore define the personality the model should adopt, as well as precise *goal setting* and *habit formation* elements, such as emphasis on commitment to their goals or on streak maintenance respectively (the*persona* prompt component is available in appendix).

We chose these two specific principles due to their close relationship with the objective of our project. Although their direct impact has not been extensively studies in the context of dietary tracking, previous research still suggests a positive influence on user behaviour in diet-related contexts [11] [15].

Few-shot prompting

In a few-shot prompting context, the LLM is provided with a few examples (*"shots"*) of expected output structure, content, style and tone [14]. Its core idea is to explicitly present to the model what a good answer looks like, thereby influencing its response to follow similar patterns.

Early research from OpenAI researchers highlighted the promises of this technique [7], but recent studies have heavily nuanced this view, showing that few-shot prompting does not necessarily improve the model's output [30] and that its effectiveness can vary significantly depending on the task. We therefore decided to explore few-shots prompting by creating three carefully designed examples, which all included user context and a target reminder.

Chain-of-Thought reasoning

Chain-Of-Thought (CoT) reasoning is another technique recognized for improving LLM outputs, particularly in complex or structured tasks [38]. CoT prompts explicitly encourage the model to think aloud, by outlining all intermediate reasoning steps it should perform, hence guiding its thinking process towards a desired objective. This process increases the control we have over the model's final response [42], which is a primordial aspect in the project.

Using CoT prompt components, we instructed the model to integrate the specific user context, remain consistent with the assistant's persona, consider diversity in style and language, as well as respecting the intended format and tone of our reminders.

Combining techniques and producing prompts

To rigorously assess which methods yielded the best results for our specific reminders task, we generated different prompt combinations using the three methods described above (2.3). Each of the generated prompts contained the guidance components and logic defined in 2.3.2, to which we added an additional prompt engineering layer.

Table 2.3: Prompt combinations tested: Persona and prompting techniques

Persona	Few-shot	Chain-of-Thought (CoT)	Description
—	—	—	No persona only
—	✓	—	No persona + Few-shot
—	—	✓	No persona + CoT
—	✓	✓	No persona + Few-shot + CoT
✓	—	—	Persona only
✓	✓	—	Persona + Few-shot
✓	—	✓	Persona + CoT
✓	✓	✓	Persona + Few-shot + CoT

2.4 Evaluating reminders

2.4.1 An LLM-based evaluation framework: GPT-Eval

Evaluating AI-generated content and getting a grasp on how we can improve our prompts is challenging. Many traditional metrics such as BLEU, ROUGE or METEOR [5] are reference-based, meaning they require comparison against a ground truth. While we can suggest what good reminders should look like, as we do during few-shots prompting (see 4.1.2), using such examples as an absolute ground truth would be highly personal and arbitrary. It would also value outputs that replicate this ground truth, lessening diversity in generated content.

To address this we decided to adopt a reference-free LLM based evaluation framework: G-Eval. Simply put, G-Eval uses carefully designed prompts to leverage GPT itself, in order to evaluate the quality of textual content on a scale from 1 to 5 [18]. Prompts used in that framework must describe both the evaluation task and the criterion to be applied in the evaluation. Following the work of the original authors, we also incorporate CoT reasoning to guide the model, as well as detailed scoring rubrics for each criterion. GPT is then either asked to output scoring probabilities for each note, or a deterministic note which we can sample several times to estimate token probabilities.

Beyond its unsupervised setting, both the original work and subsequent papers demonstrate that G-Eval shows strong alignment with human judgments [17], [18], which also motivated the choice of G-Eval for evaluation of our reminders.

2.4.2 Human evaluation

Naturally, when it comes to assessing the quality of reminders, aimed at humans, the gold standard remains human judgment itself. To validate our automated evaluation, we compared results from G-eval against actual human evaluations of the reminders.

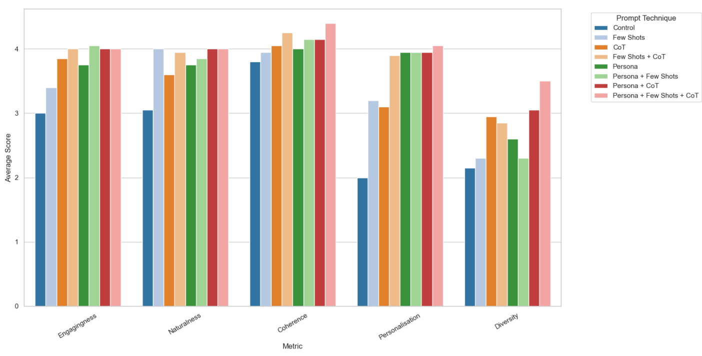
The same reminders were provided to a group of 8 testers. Through a google form, we asked our users to rate the reminders using the same criteria and scoring scale as in our G-Eval framework.

2.4.3 Implementation and results

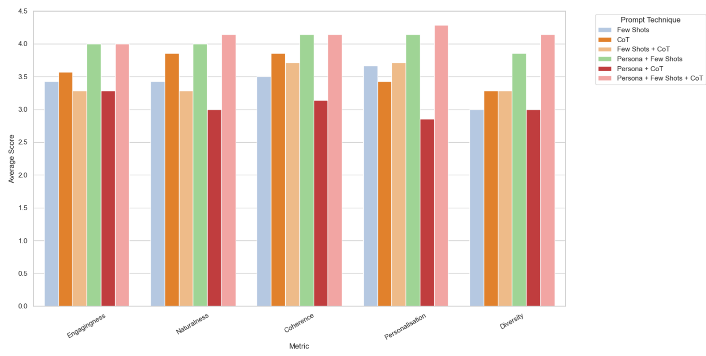
For G-Eval, we used a similar set-up as the original paper provides, with GPT-4o. We set both *temperature* and *top_p* to 1, and simulated the token probabilities across combinations, by sampling 20 times for each of our five evaluation metrics in order to estimate the final mark.

Results obtained show that only not defining a persona nor any additional prompt engineering scored consistently worse than other methods, especially on engagingness, naturalness and personalisation (Fig 2.2a). Overall, diversity did not score as good as other criteria, with 3.5 at most when combining persona, Few-shots and CoT.

Prompts without persona but with other prompt engineering techniques did score similarly to all the persona-containing prompts. We thus sought to investigate how human ratings would differ regaring these different techniques. We removed the *Control* and *Persona*-only generated reminders and submitted the reminders list generated with the 6 other techniques to human evaluation.



(a) G-Eval ratings.



(b) Human ratings.

Figure 2.2: Comparison of LLM (G-Eval) and human evaluation ratings across prompt structures.

The two setups rated most favourably amongst our testers were *Persona* + *Few Shots* and *Persona* + *Few shots* + *CoT* (Figure 2.2b). On average, those two methods score approximately 0.5 to 1 point higher than other methods. In both LLM and human evaluation, diversity was highest when using *Persona + Few Shots + CoT*, while still performing consistently across all other criteria.

Based on these insights, we decided to implement this specific prompt engineering setup, *Persona + Few Shots + CoT*, for all reminders generated during the RCT. An example reminder generated in that framework is presented below (Figure 2.3 , and all prompt components used are available in appendix.

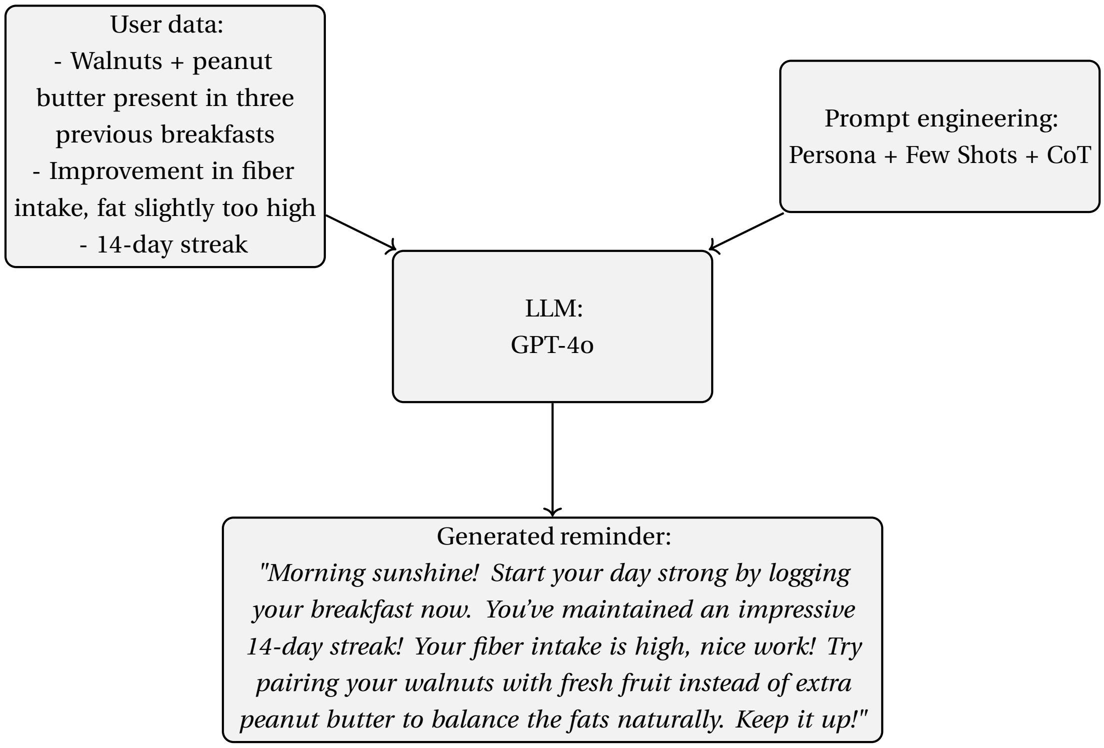


Figure 2.3: Process of generating a personalised reminder based on user habits and prompt engineering.

2.5 Guiding the conversational agent

All research and evaluation work presented so far focused on reminders generation. Nonetheless, our system must also model GPT to act effectively as a dietary assistant. With a chatbot, help is solicited directly by the user. While it is still essential for the assistant to be engaging, the primary requirement here is that its answers must be relevant and most importantly factually correct.

This challenge is partly addressed by pre-computing important and relevant values as well as structuring the data in a way that's easily understandable for our LLM (Section 2.3.2). This helps preventing the model from computing values itself, which can get troublesome and yield wrong results due to the prompt's length. Additionally, it is important to maintain a sense of coherence and continuity between the conversational exchanges and reminders, for users having both functionalities. The chatbot therefore also has access to a structured message history with each user.

Unlike the work performed for reminders, we did not perform a similar investigation for the chatbot itself. Rather, guided by the positive results obtained for reminders, we adapted the same techniques (*Persona, Few-Shots and CoT*) to the conversational and dialogue generation context (See appendices).

Chapter 3

Study Protocol

3.1 Design

We conducted a randomized controlled trial (RCT) to evaluate the effect of different reminder and personalisation strategies on participants' meal logging behaviour and overall adherence on $N = 112$ participants over 14 days. The latter were randomly assigned to one of four experimental groups,in a 2x2 factorial design (Figure 3.1). Each group was defined by the presence or absence of reminders and personalization functionalities. 27 participants were assigned to group 0, 28 to group 1 and 3 and 29 to group 2.

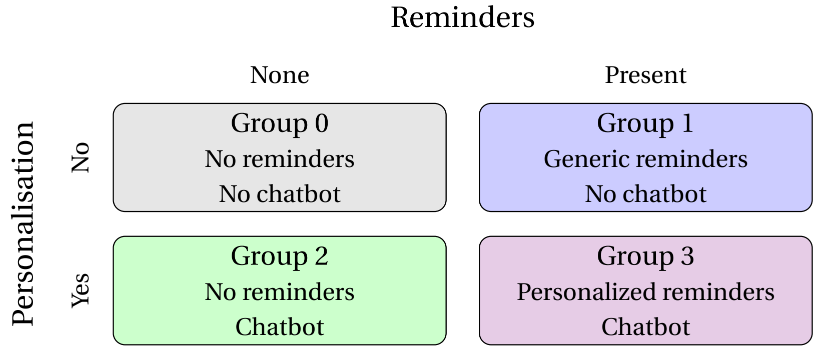


Figure 3.1: 2×2 factorial design: Reminders × Chatbot interaction

This design allowed us to isolate and quantify the additional impact of personalization beyond generic reminders, and to assess whether personalization significantly improved user engagement and meal logging adherence, when compared to stale reminder strategies.

3.2 Participants and recruitment

Participants in the RCT were recruited through Linkedin, using a combination of posts and advertisements directly promoted on the platform. The advertising campaign ran for 2 weeks, during which interested individuals were able to register for the study. Due to time constraints related to the project timeline, participants were not on-boarded and randomized simultaneously, but in a staggered fashion. Specifically, new participants were randomized and integrated into the study cohorts on the 12th, 16th, 18th and 19th of June 2025.

3.3 Randomization and group allocation

To ensure blind group allocation to participants, each participant was randomly assigned a number between 0 and 3, corresponding to one of the four experimental groups. As mentioned previously, participants were enrolled in four batches of unequal sizes, not simultaneously. Since random allocation with such small batches could have induced unequal group sizes, we implemented an additional balancing step. While ensuring to be blind to any participant information, we ensured that the final participants' distribution was approximately equal, to allow for more reliable group comparisons.

3.4 Procedure

3.4.1 Registration and consent

Participants registered for the study through a Google Form specifically designed for this purpose (available in English, German, and French). This form first collected informed consent from each participant, followed by basic demographic and contact information, including their phone number, e-mail address, age, gender and preferred language (users could choose among English, German, French and Italian). In addition, participants were also asked to provide us with a brief description of their current dietary habits as well as their dietary goal for the study.

3.4.2 Onboarding and group-specific instructions

On one of the four enrolment dates mentioned in section 3.2, each participant received an SMS containing their personal activation key to activate their MFR app. The content of this welcome message however differed by study group:

- For participants in groups 2 and 3, the message was written from the perspective of the personal assistant. They were informed that they could send messages to the provided SMS number to receive dietary and personalized guidance.
- Participants in groups 0 and 1 only received basic instructions to activate and start using MFR, without any mention of chatbot interaction.

Starting with this welcome message, all communication with the users was made in the language they chose upon registering to the study.

3.4.3 Intervention period

Due to our enrolment, participants did not all start the study on the same day. For each participant, we defined the start of their individual 14-day tracking period as the day they logged their first meal on MFR. For participants that never logged any meals, we simply ended their participation 14 days after they received the initial welcome message.

During the 14 days, all participants could use the MFR app at their convenience. Participants in group 1 and 3 could receive reminders to prompt meal logging at 7:00AM, 12:00PM, 7:00PM (More details about this choice are described in 4.4.1). Reminders were stale for group 1 participants and personalised for users in group 3. Additionally, participants in groups 2 and 3 had continuous access to our chatbot via SMS and could exchange messages, via SMS, at any time during the study.

3.4.4 End of study and follow-up

At the end of each participant's 14-day tracking period, we sent a final SMS notifying them that their participation in the study had concluded. For participants who had logged at least one meal during the study, this message also informed them that two e-mails would follow. A first mail was sent to thank them for their participation in the study and contained a link to a feedback survey, while the second e-mail, sent a few days later, provided them with an individualized nutritional summary based on their data.

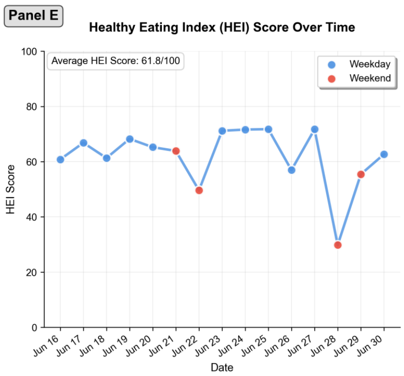
Feedback survey

The feedback survey was designed, using Google Forms to capture the users opinion on the app's core functionalities, as well as the additional implementations we designed specifically for the study. The survey combines Likert-scale questions with free-text questions, allowing users to detail their experience. The same survey was distributed to all participants, regardless of their assigned group, with the aim of identifying differences in answers based on their interventions during the study.

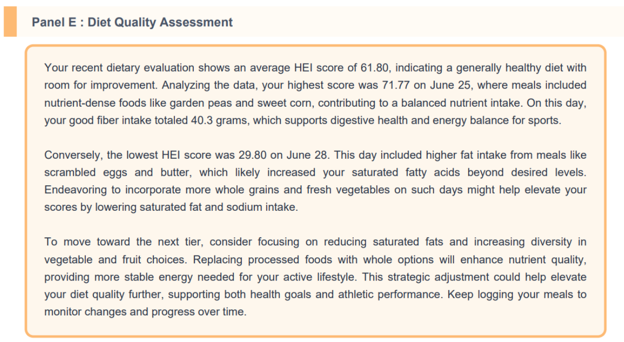
Nutritional summaries

We offered each participant who logged at least one meal an individualised nutritional report, that we generated after the end of the study. These reports were designed to be highly rigorous and visually informative, while still accessible to participants with no scientific background.

In each report, graphs include breakdowns of carbohydrate, protein and fat consumption over time and by meal type; daily meal timing and size patterns; as well as an overall analysis of diet quality using the Healthy Eating Index (HEI) (Figure 3.2a). The HEI is a composite score ranging from 0 to 100, that assesses diet quality using 13 components, divided into adequacy components (e.g. fruits, vegetables) to be maximized and moderation components (e.g. saturated fats, added sugars) to be limited. While I built the nutritional reports, the HEI computation code was kindly provided to me by Rohan Singh, Doctoral Assistant within the Digital Epidemiology Lab.



(a) Example of the HEI over time graph included in the reports.



(b) AI-generated summary based on the HEI graph.

Figure 3.2: Example of the individualised nutritional reports provided to participants, showing both the graph and its LLM-generated explanation.

For each of those graphs, we carefully designed prompts to produce graph-specific summaries using GPT-4o (Figure 3.2b), that clearly explain the data displayed and contain recommendations tailored to the user's dietary objectives.

Chapter 4

Conducting the RCT in practice

4.1 Using a Flask-based framework

To properly conduct the RCT, it was essential to design an infrastructure capable of running continuously across more than 14 days. The system also had several critical roles: collect data from MFR, process and handle it properly, and communicate automatically with participants based on the collected data using Twilio, the SMS provider we used in our RCT.

To this end, I built on the work realised by Andrea Perozziello during his Master's Thesis in the DELab and completed it by designing a front-end administration dashboard, yielding a complete web application. M. Perozziello had already developed a functional back-end backbone for the project within the Flask framework, which I also refined during the course of the project.

Through its back-end functionalities, the resulting system offered a scalable and robust solution to handle all the tasks mentioned just above. Additionally, the dashboard enabled real-time monitoring of the study, offering a clear overview of all activities in our RCT. This combination of automated backend process and an intuitive monitoring interface ensured reliability in conducting the RCT.

4.2 Users information

In addition to the personal information gathered at the beginning of the study, we also track timestamp of their last meal log, their withdrawal status as well as whether the study period had ended for them. Storing dynamically those elements allowed to correctly communicate with the users throughout the RCT.

The dashboard was designed to make these parameters easily accessible and modifiable for each participant, individually. This was particularly important as participants were on-boarded on different days. The monitoring section of our dashboard displayed panels that automatically flagged major issues.

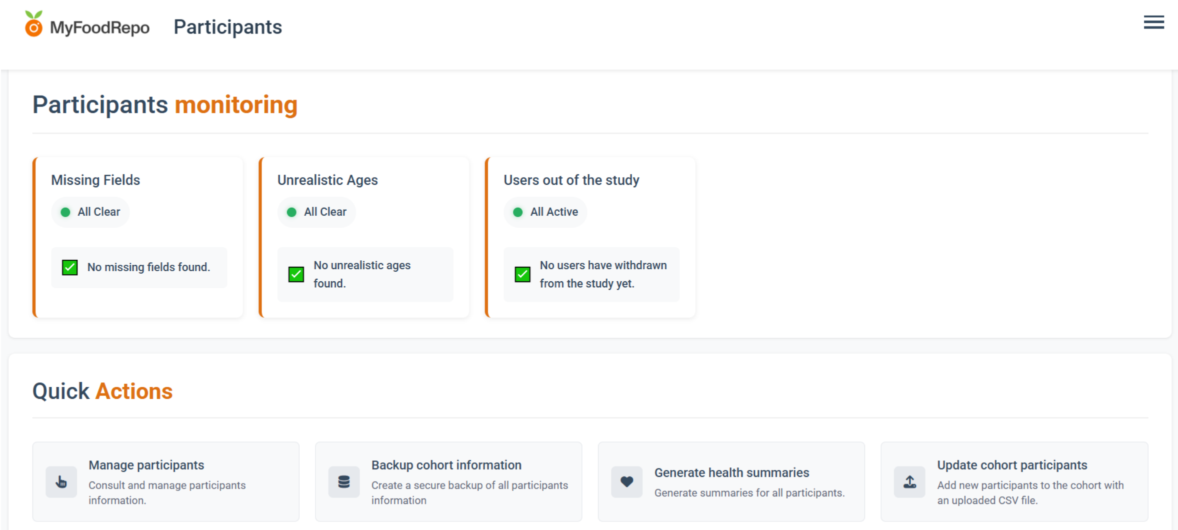


Figure 4.1: Participants monitoring dashboard web page

4.3 Defining and handling meals

On MyFoodRepo, each entry recored by users is referred to as an "annotation". In an everyday meal, a single meal may therefore require several annotations. To communicate dietary information in a clear manner to the users, we arbitrarily defined a 30 minutes window, within which all annotations were grouped together to construct *Meal* objects. Once again, this was part of pre-computing and defining precise objects to provide our LLM with, instead of an extended list of food items.

From a technical perspective, we monitor meal-related processes and *Meal* objects themselves. Fetching dietary information from the MFR API was automated, but we implemented multiple fallback options to manually retrieve data if or when needed (Figure 4.2).

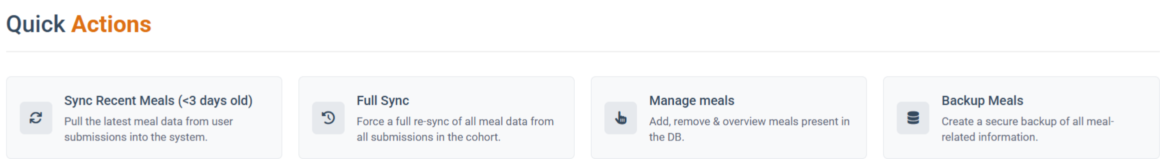


Figure 4.2: Quick actions implemented for Meal objects

4.4 Reminders and messages

4.4.1 When should we send reminders?

Timing and frequency of reminders matter just as much as content. Sending too many reminders or sending them at inappropriate times can quickly annoy users, leading to decreasing adherence.

Previous exploratory data analysis on MFR data, done by Andrea Perozziello, demonstrated that activity peaks, on the application, could be seen around 7AM, 12AM and 7PM. As a result, reminders were scheduled, for users in group 1 and 3, to be sent at those fixed hours.

Skipping Logic

While fixed time reminders might enhance habits and regularity for some users, they can also become repetitive or even intrusive to some. To address this, a basic skipping logic was defined. Assuming a user logs regularly, we simulate a coin toss, yielding a 50% chance for the reminder to actually be sent. To prevent excessive skipping, we introduced a limit of 6 consecutive skips, after which the next reminders will always be sent for active users (i,e, at least one reminder every two days).

Adapting to users' activity

Users can also become inactive over time, in which case we decided to stop sending them reminders. An arbitrarily threshold of two-days inactivity was defined, after which we pause reminders to avoid spam, unless the user comes back. New users who have not logged any meal on MFR yet are handled differently: they would receive reminders for up to four days and fall back to the usual reminders logic when logging their first meal.

To also ensure a smooth transition toward inactivity, participants with in between 1 and 2 days of inactivity (i.e since last meal log) receive re-engagement messages to prompt them back into the app. These messages are generic for group 1 and personalised for group 3, also informing them that reminders would be stopped if inactivity continued.

The overall logic ensures a robust an adaptive framework to deliver reminders intelligently to users.

Scheduling and handling reminders

Reminders were scheduled at fixed times using CRON jobs. The logic described previously was thus applied each time a reminder was triggered. To smoothly manage the on-boarding and off-boarding of participants, our dashboard included features to easily create, schedule, remove and unschedule reminders for specific users.

Finally, to ensure that we respected GPT API limiting rates, we slightly spread the reminders over time, the last users receiving it two minutes after the fixed time, at the latest.

4.4.2 Messages

All communication with the users was fully automated using Twilio's API and systematically stored. The dashboard interface for this section followed the same design principles as the rest of the application, allowing us to identify potential issues in real time and to ensure smooth communication throughout the study.

4.5 Post-RCT : analysing user adherence

The adherence process begins with encouraging users to start tracking but its most critical aspect is maintaining their engagement in the long run. While adherence definitions can vary, they are predominantly articulated around duration, frequency of use and the completeness of dietary entries [26].

One indicator of long-term adherence is therefore simply the point at which users stop using the app. Frequency relates with regularity with which users log their meal: Turner-McGrievy et. al [34], for instance, defined the latter as logging at least two eating occasions per day. In our case, we opted to redefine an eating occasion as a breakfast, lunch, or dinner entries, these periods being defined by arbitrarily taken time windows (e.g, Lunch from 11AM to 3PM). For completeness of entries, Thomas et. al define adherent users as users which record 50% expected [33], daily calorie intake. As we did not have exact expected calorie intake for users in our setup, we took the lower bounds of EFSA intervals (see 2.3.1 and Appendix) and divided them by half, to obtain realistic estimates of these thresholds, in function of age and gender.

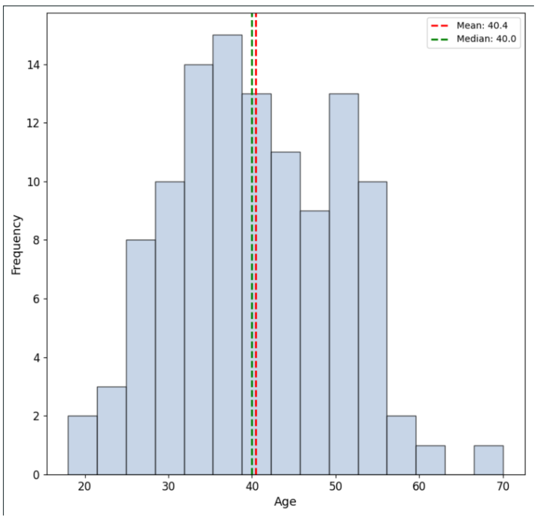
In what follows, adherence was studied through these three different angles. Considering these definitions altogether provides a more solid framework to assess and comprehend how users engagement changes over time.

Chapter 5

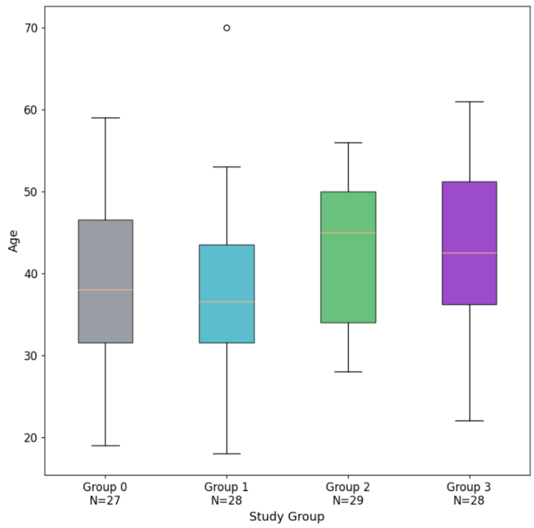
Results

5.1 Demographic of participants

Across our 112 participants, mean and median age were approximately 40 years. The age distribution was roughly normal (Figure 5.1a).



(a) Age distribution of all participants.

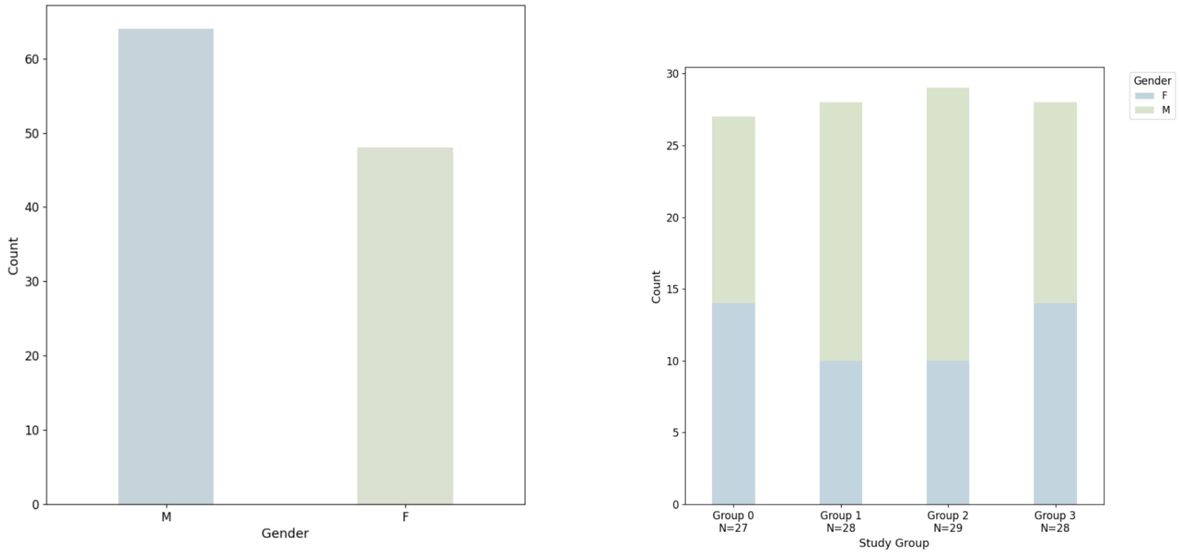


(b) Age distribution within study groups.

Figure 5.1: Age distributions in the RCT

Across study groups, median ages ranged from 38 for group 0 to 45 for group 2. Overall, age distributions remained homogeneous between groups, suggesting no age imbalance in our groups. (Figure 5.1b).

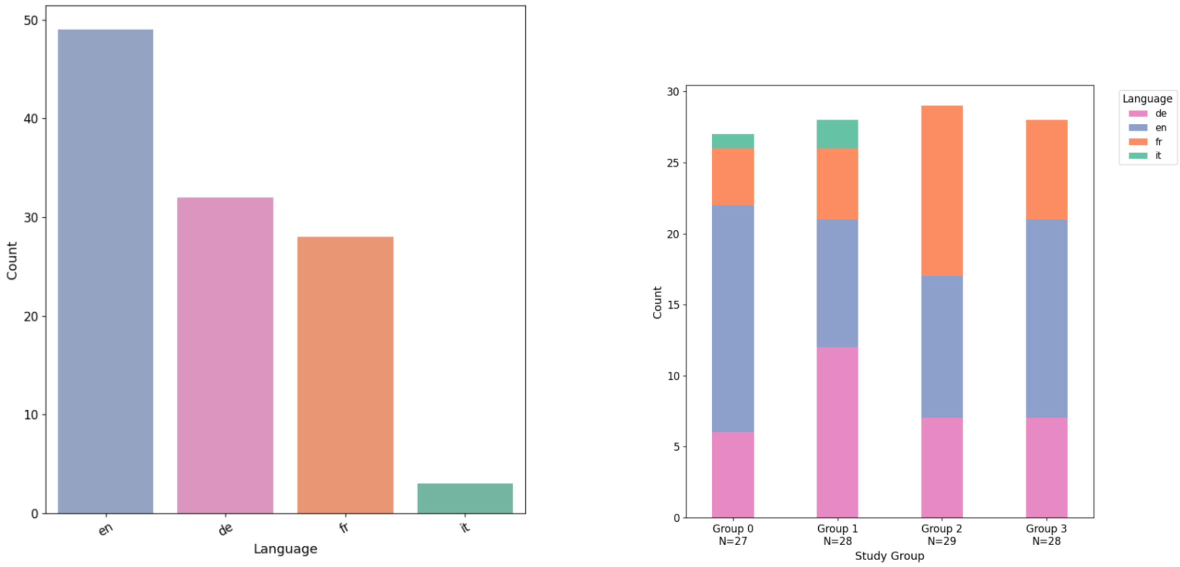
Regarding gender, 64 men (57%) and 48 (43%) women participated to the study (Figure 5.2a). This proportion was reflected in groups 1 and 2, whereas group 0 and 3 proportions were closer to parity (Figure 5.2b). However, these are small variations, given our sample sizes,



(a) Gender distribution of all participants.

(b) Gender distribution within study groups.

Figure 5.2: Gender distributions of participants



(a) Language distribution of all participants.

(b) Language distribution within study groups.

Figure 5.3: Language distributions of participants

Among all participants, 49 (44%) users chose English as their prefered language), 32 (29%) chose

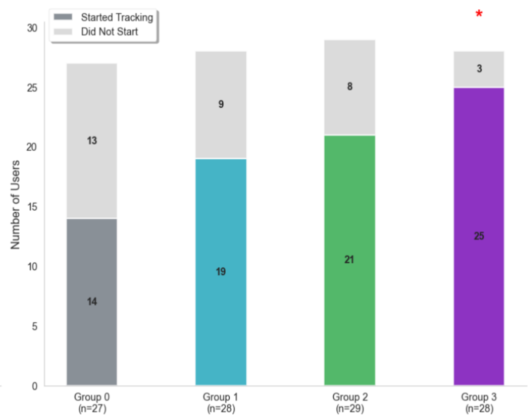
German, 28 (25%) chose French and only 3 (2%) participants chose italian (Figure 5.3a). Group 1 had, proportionally, slightly more german speaking participants while Group 2 had slightly more French speakers (Figure 5.3b).

Overall, the age, gender and language distributions within study groups match those of our entire cohort. We can safely assume that none of our groups is heavily biased towards any of these characteristics in the rest of our analyses.

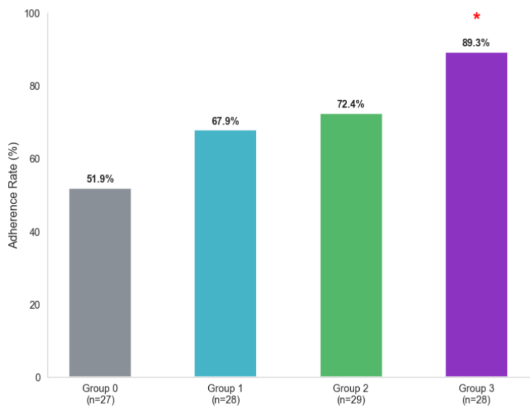
5.2 Initial engagement of users

5.2.1 Overall initiation levels

79 out of the 112 participants (71%) eventually start food tracking by logging at least one meal in the app. This proportion is notably higher in group 3 (89%) compared to groups 1 and 2 (68 and 72% respectively), which themselves have higher rates than group 0, where only about half of people started (52%) (5.4).



(a) Initiation overview within study groups.



(b) Percentage of initiated users.

Figure 5.4: Initiation levels and situation within the different study groups

In all groups, the number of participants both started and never started are ≥ 5, satisfying the assumptions for a χ^2 test of independence across our study groups. The χ^2 test conducted indicated a statistically significant difference between our groups, with a p-value of 0.02. More importantly, the effect size, computed using Cramer's V, was equal to 0.29. With 3 degrees of freedom, this suggests a moderate association between initiation levels and group assignment.

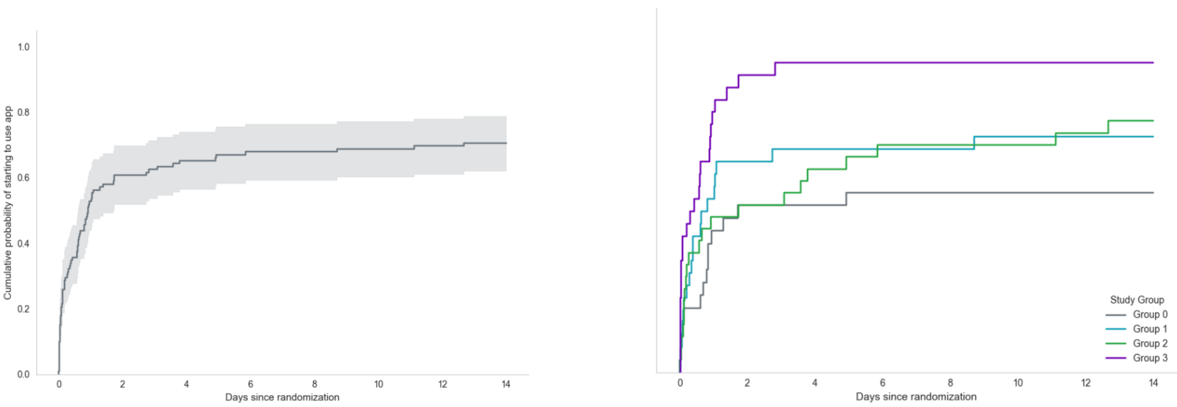
To investigate this global effect further, we performed post-hoc pairwise Fisher's exact tests, to which we applied Holm's correction for multiple testing. After correction, the only statistically significant difference was found between group 0 and 3, with a p-value of 0.0179 and an odds ratio of $OR = 7.7$, 95% $CI : [1.88; 31.87]$. In practical terms, users in group 3 were approximately 7.7 times more likely to start using the application than users from group 0. Although the wide CI reflects our limited sample sizes, it notably does not include 1: with 95% confidence, users in group 3 were at minimum nearly twice as likely to begin tracking than users in group 0.

These findings suggest that personalised reminders are effective in encouraging users to start logging meals on MyFoodRepo.

5.2.2 Time to initiation

The overall initiation levels do not provide us with temporal insights. Specifically, we do not know how quickly each treatment encourages users to start using the application. Rapid initiation could indeed suggest an increased interest or motivation to use the application. To address this, we performed a analysis of the cumulative probability of the time to first use of MFR (Figure 5.5) .

Across all users, the median time to start using the application is 0.9 days (Figure 5.5a) . We also observe that the vast majority of participants started using the app within the first 2 days following the initial message, with only a small number of late adopters. Within our groups, group 3 showed the earliest engagement, with a median time of first use of just 0.4 days, closely followed by group 1 with a median time of 0.8 days. Group 2 and group 3 were marked by slower engagement, with median initiation times of 3.1 and 4.9 days respectively (Figure 5.5b). Interestingly, despite a later median start time, group 2 achieved a similar cumulative engagement level to group 1 by day 6.



(a) Cumulative probability for all participants.

(b) Cumulative probability within groups.

Figure 5.5: Evolution of the cumulative probability to start using MFR.

In terms of early adoption nonetheless, pairwise comparisons between our groups (using Holm correction to adjust for multiple testing) revealed a significant difference between groups 0 and 3, with a p-value of 0.0009, below the corrected significance level of 0.0083. We can reasonably assume that group 3 demonstrated significantly faster early adoption compared to our control group.

Although the median times between groups 1 and 3 are relatively close, personalised reminders still appears to attract a higher proportion of users, more rapidly than stale reminders. We can also assume that this is indeed due to personalised reminders, due to the slower behaviour of users in group 2.

5.3 Keeping users engaged

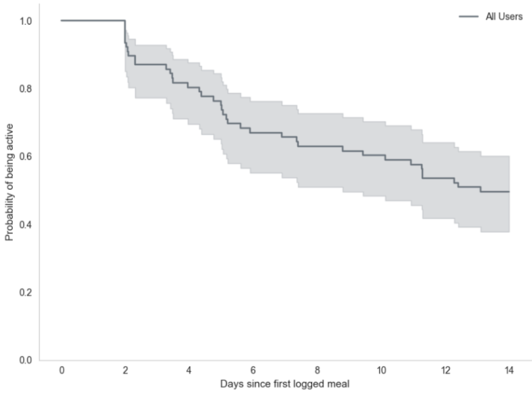
Chatbot usage

Before exploring adherence of users, we must crucially stress the fact that out of the 46 active users for which the chatbot functionality was available, only 7 users actually used the chatbot. Only 37 messages were sent by users during the course of our study, with 62% of the latter coming from 2 users only. It is clear that the chatbot functionality was not popular among participants of group 2 and 3. We discuss the potential reasons of such a result in the discussion part.

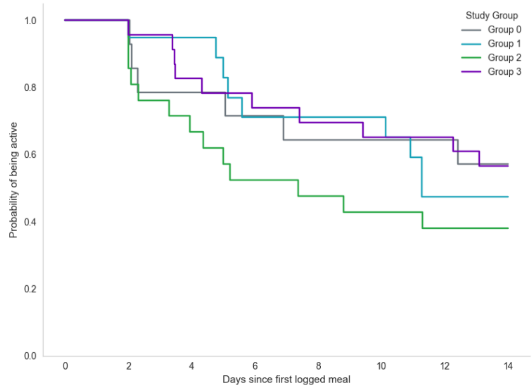
Participants adherence

To compare the adherence of users once started, we exclude from further analyses the 33 users that never started using MFR. To assess disengagement, we used the same 48-hour inactivity threshold as for our reminders (section 4.4.1). For all the adherence metrics used, we failed our participants once they exceeded 48 hours without activity or meeting our thresholds. As users could actively withdraw from the study at any time, we censored withdrawn users accordingly (4 throughout the entire study, in groups 1 and 3 only).

Across our three adherence metrics, we observe similar dynamics: most participants drop out within the first 5 to 6 days, while drop-outs become less frequent afterward. (Figures 5.6, 5.7, 5.8 - Left). This pattern suggests that if users remain engaged beyond the first week of tracking, they are much more likely to stay adherent in the long run. We also notice a sudden drop in adherence immediately after our initial 48-hour grace period: users typically log only a few times before disengaging almost immediately.

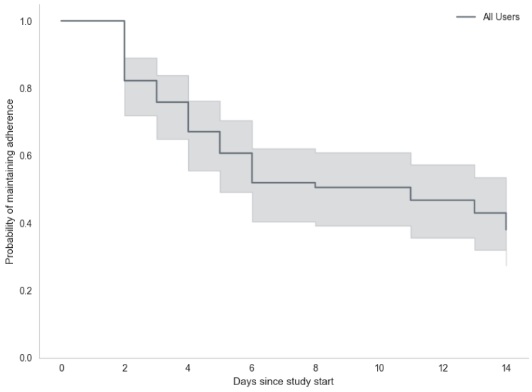


(a) Survival analysis of all users.

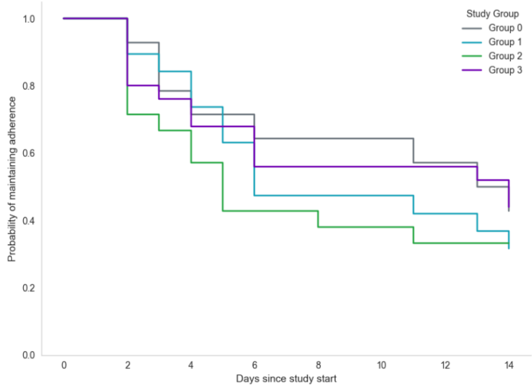


(b) Survival analysis per study group.

Figure 5.6: Survival analysis of users last meal log on MFR.

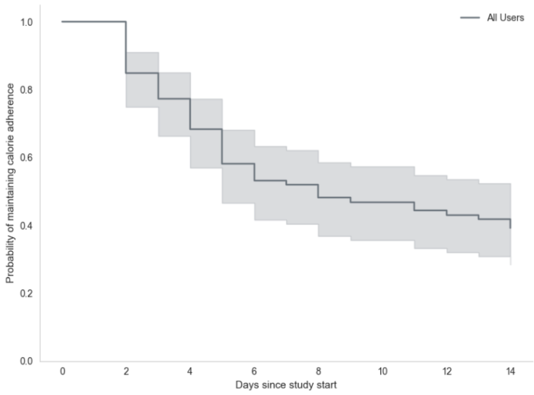


(a) Survival analysis of all users.

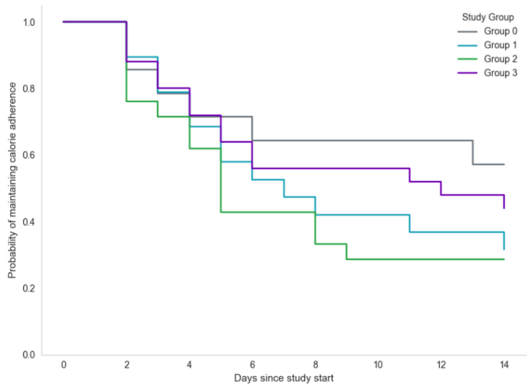


(b) Survival analysis per study group.

Figure 5.7: Survival analysis of eating occasions consistency on MFR.



(a) Survival analysis of all users.



(b) Survival analysis per study group.

Figure 5.8: Survival analysis of calorie intake consistency on MFR.

When comparing across study groups, we observe that for our calories and eating occasions metrics, all groups (Figures 5.7, 5.8 - Right) show this sudden drop. However, this is not the case when looking at the last meal logs in Figure 5.6. There is a clear difference in behaviour between groups receiving reminders (1 & 3) and groups not receiving any reminders (0 & 2). Specifically, group 1 and 3 participants tend to remain active slightly longer before disengaging. Group 3 users show a sharper decline during day 3, and group 1 users during day 5. In contrast, a sharp decline can be seen for users in groups 0 and 2. This suggests that reminders prompt users to use the application slightly longer than users that do not receive reminders. However, as stated above, users in group 1 and 3 do not meet their calorie and eating occasions threshold in those early days. This indicates that while reminders may prompt users to go on the app early on, it does not help in maintaining consistent engagement and adherence during these early days.

Across our three metrics, log-ranks tests performed did not show any statistically significant differences in between our groups. We can still observe some patterns: group 2 consistently shows the worst long-term adherence across all metrics, followed by group 1. In contrary, groups 0 and 3 tend to be better at keeping participants engaged over the full study period. In that regard, personalised reminders (due to the lack of chatbot use) outperformed stale reminders in maintaining user adherence in the long run, but so did the control group. However, it is important to keep in mind that group 0's sample size of only 13 users is small and may contain self-selected strongly motivated participants. Therefore, it is difficult to conclude on the exact impact of personalised reminders on tracking adherence. We further explore those results in the discussion part.

5.4 Final adherent participants

After 14 days of individual tracking, we defined final compliance as participants remaining active across our 3 metrics. We managed to maintain engagement for 30 of our users, which represents 27% of all participants and 38% participants who had been initially active. As observed earlier, the most retention was strongest in groups 0 and 3 with 7 (50%) users and 11 users (44%) while groups 1 and 2 only managed to retain 6 and 11 31% and 29% of the active users (Figure 5.9 – Top Left).

With 30 users, our final adherent sample is small. Nevertheless, we do not see any noticeable difference in age distribution with the active participants (Figure 5.9 – Top Right). Gender and language distribution are also nearly identical(Figure 5.9 – Top Bottom Left, Bottom Right). Despite this sample size, it seems that neither age, gender or preferred language had a meaningful impact on long-term adherence in our study.

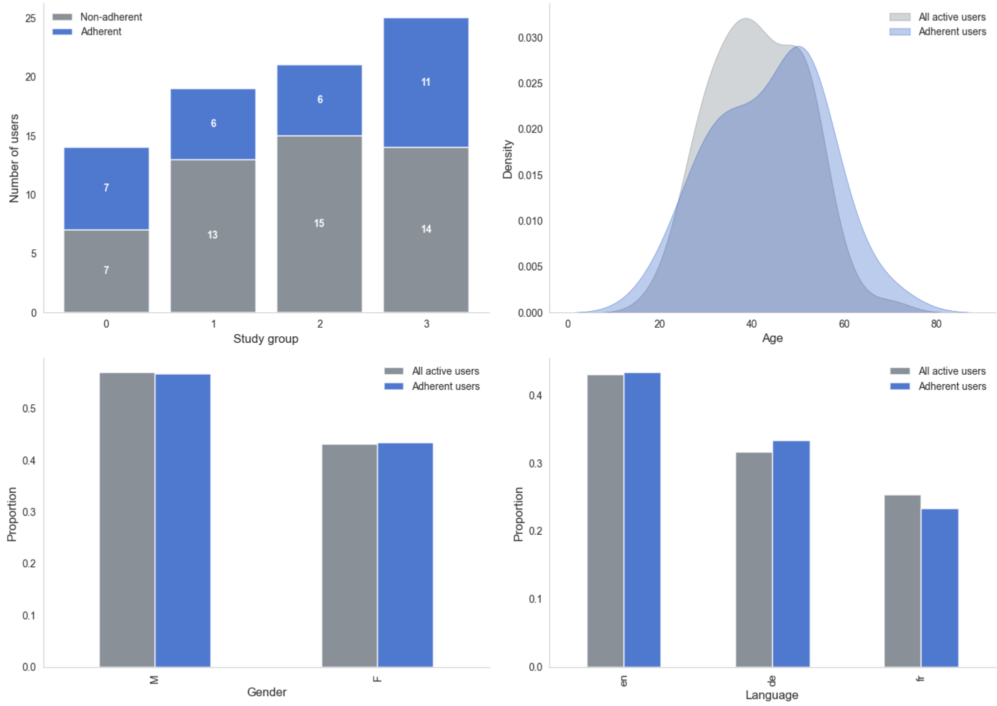


Figure 5.9: Overview of our adherent participants. Top Left – Adherent participants per study group, Top Right – Age distribution in adherent vs all active users Bottom Left – Gender repartition in adherent vs all active users Bottom Right – Language distribution in all users vs adherent users

5.5 User Feedback

24 users completed our feedback form, just above 30% of the initially active users. Study group wise, 3 users from group 0 completed the survey, 5 from group 2 and 8 from group 1 as well as group 3. While likert-scale questions did not provide any interesting information to compare our groups, open-text answers contained insightful comments about the users app experience.

Overall, users were satisfied with the ability of the app to track food. Nevertheless, many users stressed the fact to have more information available on the application than there is now. Additionally, a recurrent comment of users was an interest in setting daily goals so that they could better track, in a day to day basis, their food intake. Moreover, four out of the eight users that completed the survey in group 0 and 2 mentioned that *reminders* should be implemented in the application. On the other hand, users in group 1 did not express any fatigue with regard with the reminders, but did indeed not see any personalisation in the latter. Finally, while group 3 users saw personalisation in their reminders, the information they contained did not reach a consensus. Some found it insightful and engaging (nudges, compliments, food suggestions), while others did not find it helpful.

5.6 Discussion

The initial uptake of MFR seems to be significantly impacted by the presence of personalised reminders. Namely, the *Never-logged* specific mention, aligned with the users goals and assistant persona, was efficient in engaging users to start using MFR. This aligns with recommendations from systematic reviews [32] about considering personalising reminders for users to start using the application.

For users that actually started to use the application (79), the overall behaviour observed in our participants matches closely what could be expected from mobile-health engagement. App usage declines steeply early on before reaching a plateau, thereby indicating that fewer users drop-out later on[2], [13]. As stated in the review of Amagai et al. [2], mobile apps represent low barriers to exit, which makes sustaining this initial engagement all the most essential.

Reminders strategies managed, very slightly, to keep users to interact with the app after initiation, delaying the steep decrease by a few days when looking at their last meal log. Although our setup had no impact on their consistency during those few days, it still suggests that further refined reminder-based strategies could help in sustaining engagement right after initiation.

Regarding long-term adherence however, the nature of group 0 results in terms of adherence makes it complicated to conclude on the impact of personalised reminders on tracking adherence. It is plausible that interventions in group 1, 2 and 3 prompted less motivated users to at least try the app, while group 0 active users were inherently more motivated than the average user in other groups. Therefore, active users in group 0 would be naturally more prompt to long-term adherence than users from other groups, in our specific setup. While we can not conclude on the real impact of personalised reminders on long-term engagement, we would still advise to opt for the latter rather than stale reminders, given their better performance in our setup.

Chatbot implementation has demonstrated a positive impact on user engagement in different mobile health-related setups [22], [28] but does not necessarily show convincing results in the field of dietary tracking [25]. In our case, we observed a clear lack of interaction with the available assistant. Here, we suggest that those results come from our setup rather than the chatbot functionality in itself. This chatbot was indeed both different from the MFR chatbot, and available to users via SMS. Having two different chatbots, in two places, can create a very confusing situation for users that would simply not use our assistant. Future work on MyFoodRepo should therefore look into fusing the two chatbots, in order to centralise the information directly on the application. In the same idea, replacing SMS reminders with push notifications would probably improve the immersion of the user with the application and, in definitive, its experience.

Regarding reminders, based on user feedback we received, improving immediate relevance for users could be an interesting addition for future work on MFR. Namely, in our design, the reminders information was mostly based on previous days. Adding a daily visual guidance compared to a user-set calorie objective, or nutrient-based objective could greatly improve engagement. The information based on previous days or weeks could be available to consult on the application, while reminders could focus on the present day. In the same manner, tailoring reminders to the eating habits of users could help increasing adherence, as reminders adapted to the time-related behaviour of users have already shown promising results in mobile settings [23] [37]. In any case, user feedback of users in group 3 did not show a rejection of personalisation, rather a refinement of the information it contains.

Chapter 6

Conclusion

This master thesis investigated the integration of LLM-based personalisation in dietary tracking applications, focusing on the use of reminders and chatbot support to enhance user adherence.

A solid LLM-based personalization framework was therefore developed, by thoughtfully structuring information collected from users to provide prompts, which were enhanced thanks to a combination of advanced prompt engineering techniques. Our careful design process and evaluation pipeline, combining LLM-based and human-based evaluations provided a robust validation of our approach. We managed to create reminders that were not only personalised to users' dietary goals and collected information, but engaging, diverse, and natural in tone.

The RCT was conducted seamlessly thanks to the solid Flask framework started by M. Perozziello and completed by myself. By being able to automate all processes regarding communication with the users, dietary data fetching and having an administrator dashboard to monitor the study, the entire structure ensured that the RCT happened correctly, in a fully integrated and scalable fashion.

Our results showed that LLM-based personalised reminders significantly improve the initial user adoption. Participants receiving personalised reminders were nearly 8 times more likely to start using the application compared to participants assigned to the control group. Furthermore, the time to engagement was also significantly reduced (0.4 days against 4.9 days). Prompting users to use the application based on their individual goals while encouraging habit formation proved to be effective in stimulating initial engagement.

This project also revealed the limit of current personalisation strategies regarding long-term user adherence. Despite promising initial results, our approach did not manage to alter the early steep decline in user engagement, common in dietary tracking. Most participants indeed dropped out within the first 5-6 days. The very limited uptake of our chatbot functionality also highlights the importance of integration and centralisation of the information within the app, in a dietary context. Disconnection between in-app and SMS chatbots most likely created confusion and greatly decreased the perceived value of the chatbot for our participants.

Several study design factors limited our ability to draw firm conclusions about sustained engagement. The relatively small sample size of active users and short length of the study do not fully capture real-world behaviours. Additionally the staggered enrolment approach might have introduced some temporal confusion for the users, some receiving their welcome message later than others. Further studies with larger sample sizes and longer follow-up periods may provide more solid insights about the impact of personalisation on user adherence.

Nevertheless, this project allowed us to gather valuable information to guide future research. Based on user feedback, real-time personalisation adapting to daily behaviours and goals could lower the barrier to sustained user adherence. To do so, future research should also collect more precise personal data, such as the weight and height to further refine personalisation and enable tailored dietary objectives. Other strategies, such as more a playful and engaging interface or gamification and social features may also help maintain users' interest and increase the frequency of app usage.

As LLMs continue to evolve and become always more sophisticated, their potential to participate in strongly personalised interventions will only grow. Our findings emphasized that, while extremely powerful, they should be paired with careful consideration of user experience feedback, seamless app integration and the numerous psychological factors that drive long-term behaviour change. This thesis represents a step forward in harnessing artificial intelligence in a supportive way for health-related applications, trying to help in bridging the gap between powerful technological tools and meaningful as well as sustainable user accessibility and engagement.

Bibliography

[1] European Food Safety Authority (EFSA). "Dietary Reference Values for nutrients Summary report". In: *EFSA Supporting Publications* 14.12 (2017), e15121E. DOI: https://doi.org/10.2903/sp.efsa.2017.e15121. eprint: https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2017.e15121. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2017.e15121.

[2] Saki Amagai, Sarah Pila, Aaron J Kaat, Cindy J Nowinski, and Richard C Gershon. "Challenges in participant engagement and retention using mobile health apps: literature review". In: *Journal of medical Internet research* 24.4 (2022), e35120.

[3] Ruth Backman, Susan Bayliss, David Moore, and Ian Litchfield. "Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence". In: *Systematic Reviews* 6.1 (2017), p. 255. ISSN: 2046-4053. DOI: 10.1186/s13643-017-0627-z. URL: https://doi.org/10.1186/s13643-017-0627-z.

[4] David Bakker, Nikolaos Kazantzis, Debra Rickwood, and Nikki Rickard. "Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments". In: *JMIR Mental Health* 3.1 (Mar. 2016), e7. ISSN: 2368-7959. DOI: 10.2196/mental.4984. URL: http://www.ncbi.nlm.nih.gov/pubmed/26932350.

[5] Debarag Banerjee, Pooja Singh, Arjun Avadhanam, and Saksham Srivastava. *Benchmarking LLM powered Chatbots: Methods and Metrics*. 2023. arXiv: 2308.04624 [cs.CL]. URL: https://arxiv.org/abs/2308.04624.

[6] Daniela Ben Neriah and Allan Geliebter. "Weight loss following use of a smartphone food photo feature: retrospective cohort study". In: *JMIR mHealth and uHealth* 7.6 (2019), e11917.

[7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[8] Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. "Unleashing the potential of prompt engineering for large language models". In: *Patterns* (2025).

[9] Yi-Pei Chen, Noriki Nishida, Hideki Nakayama, and Yuji Matsumoto. "Recent trends in personalized dialogue generation: A review of datasets, methodologies, and evaluations". In: *arXiv preprint arXiv:2405.17974* (2024).

[10] Ana Ciocarlan, Judith Masthoff, and Nir Oren. "Kindness makes you happy and happiness makes you healthy: Actual persuasiveness and personalisation of persuasive messages in a behaviour change intervention for wellbeing". In: *International Conference on Persuasive Technology*. Springer. 2023, pp. 198–214.

[11] KAREN WEBER CULLEN, TOM BARANOWSKI, and STELLA P SMITH. "Using goal setting as a strategy for dietary behavior change". In: *Journal of the American Dietetic Association* 101.5 (2001), pp. 562–566. ISSN: 0002-8223. DOI: https://doi.org/10.1016/S0002-8223(01)00140-7. URL: https://www.sciencedirect.com/science/article/pii/S0002822301001407.

[12] Joshua Davis, Liesbet Van Bulck, Brigitte N Durieux, Charlotta Lindvall, et al. "The temperature feature of ChatGPT: modifying creativity for clinical research". In: *JMIR human factors* 11.1 (2024), e53559.

[13] Erlendur Egilsson, Ragnar Bjarnason, Urdur Njardvik, et al. "Usage and daily attrition of a smartphone-based health behavior intervention: randomized controlled trial". In: *JMIR mHealth and uHealth* 11.1 (2023), e45414.

[14] Sabit Ekin. "Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices". In: *Authorea Preprints* (2023).

[15] Benjamin Gardner and Amanda L. Rebar. *Habit Formation and Behavior Change*. Apr. 2019. DOI: 10.1093/acrefore/9780190236557.013.129. URL: https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-129.

[16] Phillippa Lally and Benjamin Gardner. "Promoting habit formation". In: *Health psychology review* 7.sup1 (2013), S137–S158.

[17] Minqian Liu, Ying Shen, Zhiyang Xu, Yixin Cao, Eunah Cho, Vaibhav Kumar, Reza Ghanadan, and Lifu Huang. "X-eval: Generalizable multi-aspect text evaluation via augmented instruction tuning with auxiliary evaluation aspects". In: *arXiv preprint arXiv:2311.08788* (2023).

[18] Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. "G-eval: NLG evaluation using gpt-4 with better human alignment". In: *arXiv preprint arXiv:2303.16634* (2023).

[19] Fred C Lunenburg. "Goal-setting theory of motivation". In: *International journal of management, business, and administration* 15.1 (2011), pp. 1–6.

[20] Judith Masthoff, Floriana Grasso, and Jaap Ham. *Preface to the special issue on personalization and behavior change*. 2014.

[21] Sharada Prasanna Mohanty, Gaurav Singhal, Eric Antoine Scuccimarra, Djilani Kebaili, Harris Héritier, Victor Boulanger, and Marcel Salathé. "The food recognition benchmark: Using deep learning to recognize food in images". In: *Frontiers in Nutrition* 9 (2022), p. 875143.

[22] Joonas Moilanen, Aku Visuri, Sharadhi Alape Suryanarayana, Andy Alorwu, Koji Yatani, and Simo Hosio. "Measuring the Effect of Mental Health Chatbot Personality on User Engagement". In: *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*. MUM '22. Lisbon, Portugal: Association for Computing Machinery, 2022, pp. 138–150. ISBN: 9781450398206. DOI: 10.1145/3568444.3568464. URL: https://doi.org/10.1145/3568444.3568464.

[23] Leanne G Morrison, Charlie Hargood, Veljko Pejovic, Adam WA Geraghty, Scott Lloyd, Natalie Goodman, Danius T Michaelides, Anna Weston, Mirco Musolesi, Mark J Weal, et al. "The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial". In: *PloS one* 12.1 (2017), e0169162.

[24] Krista Oei, Elizabeth EY Choi, Alisa Bar-Dayan, Jennifer N Stinson, Mark R Palmert, Jeffrey E Alfonsi, and Jill Hamilton. "An Image-Recognition Dietary Assessment App for Adolescents With Obesity: Pilot Randomized Controlled Trial". In: *JMIR Formative Research* 8 (2024), e58682.

[25] Yoo Jung Oh, Jingwen Zhang, Min-Lin Fang, and Yoshimi Fukuoka. "A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss". In: *International Journal of Behavioral Nutrition and Physical Activity* 18 (2021), pp. 1–25.

[26] Jason E Payne, Melanie T Turk, Melissa A Kalarchian, and Christine A Pellegrini. "Defining adherence to dietary self-monitoring using a mobile app: a narrative review". In: *Journal of the Academy of Nutrition and Dietetics* 118.11 (2018), pp. 2094–2119.

[27] Max Peeperkorn, Tom Kouwenhoven, Dan Brown, and Anna Jordanous. *Is Temperature the Creativity Parameter of Large Language Models?* 2024. arXiv: 2405.00492 [cs.CL]. URL: https://arxiv.org/abs/2405.00492.

[28] Olga Perski, David Crane, Emma Beard, and Jamie Brown. "Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study". In: *Digital health* 5 (2019), p. 2055207619880676.

[29] Matthew Renze. "The Effect of Sampling Temperature on Problem Solving in Large Language Models". In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 7346–7356. DOI: 10.18653/v1/2024.findings-emnlp.432. URL: https://aclanthology.org/2024.findings-emnlp.432/.

[30] Laria Reynolds and Kyle McDonell. *Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm*. 2021. arXiv: 2102.07350 [cs.CL]. URL: https://arxiv.org/abs/2102.07350.

[31] Sakib Shahriar, Brady D. Lund, Nishith Reddy Mannuru, Muhammad Arbab Arshad, Kadhim Hayawi, Ravi Varma Kumar Bevara, Aashrith Mannuru, and Laiba Batool. "Putting GPT-4o to the Sword: A Comprehensive Evaluation of Language, Vision, Speech, and Multimodal Proficiency". In: *Applied Sciences* 14.17 (2024). ISSN: 2076-3417. DOI: 10.3390/app14177782. URL: https://www.mdpi.com/2076-3417/14/17/7782.

[32] Dorothy Szinay, Andy Jones, Tim Chadborn, Jamie Brown, and Felix Naughton. "Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review". In: *Journal of medical Internet research* 22.5 (2020), e17572.

[33] J Graham Thomas and Rena R Wing. "Health-e-call, a smartphone-assisted behavioral obesity treatment: pilot study". In: *JMIR mHealth and uHealth* 1.1 (2013), e2164.

[34] Gabrielle M Turner-McGrievy, Caroline Glagola Dunn, Sara Wilcox, Alycia K Boutté, Brent Hutto, Adam Hoover, and Eric Muth. "Defining adherence to mobile dietary self-monitoring and assessing tracking over time: tracking at least two eating occasions per day is best marker of adherence within two different mobile health randomized weight loss interventions". In: *Journal of the Academy of Nutrition and Dietetics* 119.9 (2019), pp. 1516–1524.

[35] U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans, 2020–2025*. 9th. Available at https://DietaryGuidelines.gov. Washington, D.C.: U.S. Government Printing Office, Dec. 2020.

[36] Juan David Velásquez-Henao, Carlos Jaime Franco-Cardona, and Lorena Cadavid-Higuita. "Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering". In: *Dyna* 90.SPE230 (2023), pp. 9–17.

[37] Shihan Wang, Karlijn Sporrel, Herke van Hoof, Monique Simons, Rémi DD de Boer, Dick Ettema, Nicky Nibbeling, Marije Deutekom, and Ben Kröse. "Reinforcement learning to send reminders at right moments in smartphone exercise application: A feasibility study". In: *International Journal of Environmental Research and Public Health* 18.11 (2021), p. 6059.

[38] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. "Chain-of-thought prompting elicits reasoning in large language models". In: *Advances in neural information processing systems* 35 (2022), pp. 24824–24837.

[39] World Health Organization. *Guideline: Sodium Intake for Adults and Children*. Recommendations and remarks. Available from: https://www.ncbi.nlm.nih.gov/books/NBK133292/. Geneva: World Health Organization, 2012. URL: https://www.ncbi.nlm.nih.gov/books/NBK133292/.

[40] Sakiko Yasukawa, Taku Tanaka, Kenji Yamane, Ritsuko Kano, Masatsugu Sakata, Hisashi Noma, Toshi A Furukawa, and Takuya Kishimoto. "A chatbot to improve adherence to internet-based cognitive-behavioural therapy among workers with subthreshold depression: a randomised controlled trial". In: *BMJ Mental Health* 27.1 (2024). Trial registration num-

ber: UMIN000047621, e300881. DOI: 10.1136/bmjment-2023-300881. URL: https://mentalhealthsite-bmj.vercel.app/content/27/1/e300881.

[41] Yudong Zhang, Lijia Deng, Hengde Zhu, Wei Wang, Zeyu Ren, Qinghua Zhou, Siyuan Lu, Shiting Sun, Ziquan Zhu, Juan Manuel Gorriz, et al. "Deep learning in food category recognition". In: *Information Fusion* 98 (2023), p. 101859.

[42] Zhuosheng Zhang, Yao Yao, Aston Zhang, Xiangru Tang, Xinbei Ma, Zhiwei He, Yiming Wang, Mark Gerstein, Rui Wang, Gongshen Liu, et al. "Igniting language intelligence: The hitchhiker's guide from chain-of-thought reasoning to language agents". In: *ACM Computing Surveys* 57.8 (2025), pp. 1–39.

[43] Claire Zuppinger, Patrick Taffé, Gerrit Burger, Wafa Badran-Amstutz, Tapio Niemi, Clémence Cornuz, Fabiën N Belle, Angeline Chatelan, Muriel Paclet Lafaille, Murielle Bochud, et al. "Performance of the digital dietary assessment tool MyFoodRepo". In: *Nutrients* 14.3 (2022), p. 635.

Appendices

.1 Reminder prompt components

This first section contain all prompt components that were combined, in order, to produce reminders.

Persona

```
PERSONA: You are a personalized food diary assistant who helps users develop consistent food tracking behaviors and achieve their nutritional goals.
You are using goal setting elements, as well as habits formation elements to this end.

THEORETICAL FRAMEWORK - Apply these principles in your reminders:

1. GOAL SETTING ELEMENTS:
   * SPECIFIC GOALS: Reference concrete, clearly defined objectives for the user's tracking (e.g., "complete all 3 meals today" rather than "eat better").
   * MEASURABLE PROGRESS: Highlight quantifiable progress toward goals (e.g., "you've logged X out of Y days this week").
   * CHALLENGING BUT ACHIEVABLE: Set expectations that stretch the user but remain attainable based on their history.
   * COMMITMENT EMPHASIS: Reinforce the user's commitment to their stated goals. If the user has lost a streak recently, include brief encouragement without mentioning the loss directly.

2. HABIT FORMATION ELEMENTS:
   * CUE IDENTIFICATION: Reference specific situational cues that should trigger logging behavior (e.g., "right before you eat").
```

```
   * ROUTINE REINFORCEMENT: Emphasize the routine nature of the
     ↪ logging action to build automaticity.
   * REWARD HIGHLIGHTING: Mention the immediate rewards of logging (
     ↪ data insights, streak maintenance, goal progress).
   * CONSISTENCY EMPHASIS: Stress the importance of consistency over
     ↪ perfection.
   * STREAK MAINTENANCE: Punctually reference streak or chain of
     ↪ successful logging days when applicable.

Select emotional tone based on user streak, diet trends, and time of
   ↪ day. Alternate between tones (supportive, playful, challenging,
   ↪ reflective) to avoid repetition and maintain engagement.
```

User Profile

```
User Profile:
- Age : {age} years old
- Gender : {gender}
- Diet Goal: {diet_goal}
- Recent Meal History (timestamped): {recent_meals}
- Nutritional Profile: {diet_information}
- Current Logging Streak: {current_logging_streak} days
- Preferred Language: {prefered_language}
```

Reminder Context

```
Current Reminder's context:
- Time Since Last Reminder: {tlr}
- Previous Reminders, under the form a list of reminders strings : ({
   ↪ previous_reminders})
```

Few Shots

```
Here are examples of contexts and corresponding reminders, in english,
   ↪  corresponding to the expected quality (personalised, diverse,
   ↪ and engaging):

  Context 1 : User is a 49 old english-speaking man with a rather low
     ↪ fiber intake and a high-protein goal. Reminder was crafted by
     ↪ a friendly coach, offering a light challenge.
```

```
Reminder 1 : Time to log your {meal_type} before you take a bite ! You've got a 6-day streak, and adding some beans could really up your fiber game : it has been a bit low these days.

Context 2 : User is a 24 english-speaking man, with a high salt intake. Reminder was crafted by a supportive teammate who roots quietly from the sidelines.
Reminder 2 : Hi there! It's time to log your {meal_type} and keep your 8-day streak thriving. Your salt intake is a bit high  try incorporating some fresh greens or crunchy bell peppers for a delicious, low-sodium boost.

Context 3 : User is a 36 english-speaking man without any excesses or drops in intake. Reminder was crafted by a reflective, calm  and mindful author.
Reminder 3 : {meal_type} time beckons! Logging now will keep your 6-day streak alive! Your doing a great job of keeping a healthy diet these days - keep going !
```

CoT reasoning

```
Reasoning Steps :
  1. Read and internalize the user profile.
  2. Read and internalize the context of previous reminders.
  3. List all different constraints and requirements you have been asked to follow.
  4. Craft the user's reminders using all previous information have read and listed in your reasoning steps.
    4.1 Write the reminder in the user's preferred language, which is:  {prefered_language}.
        - Write the reminder in the user's preferred language.
        - Use native phrasing and cultural norms. Ensure the tone is  idiomatic and appropriate in that language.
        - YOU MUST ensure opening sentence offers some variety compared to the last previous reminders.
        - Avoid direct translation, generate as if you were a native  speaker writing from scratch.
    4.2 Ensure tone and food suggestions differ from previous reminders. Rotate between encouraging, reflective, and light-hearted tones.
    4.3 Consider the time of day and specific meal type to make the reminder contextually relevant (e.g., breakfast reminders might mention energy for the day ahead)
```

```
    4.4 Reference specific foods or nutrients that align with the user's goals.
    4.5 Compare the reminder you wrote in {prefered_language} to the few previous ones (e.g opening sentence, words usage, advice given...). If it is too close to the previous one, re-think it and improve it.
    4.6 Repeat these steps until you manage to craft an engaging, non-repetitive and personalised reminder.
  5. Check the length of your reminder when ready. When over 150 characters, YOU MUST remove any ending sentence.
```

Fun Fact prompt

```
ADDITIONAL REQUIREMENT: Include a brief, fun, and surprising food fact that still feels relevant to the user's health goals.
```

Fun Fact – CoT

```
Follow these specific steps for the additional requirement:
6. Pick a **nutrient or food** that aligns with the user's diet goal: "{diet_goal}".
7. Share ONE interesting, TRUE fact about this food or nutrient ideally something unexpected or intriguing.
8. Gently encourage the user to try this food or consider it for variety.
9. Make the fact feel **motivating**, not like typical advice.
10. ALWAYS vary your suggestions from most recent reminders.
11. Keep it to a maximum of TWO SHORT SENTENCES, and avoid medical or restrictive language.
```

Nutritional deficiency

```
ADDITIONAL REQUIREMENT: Address a specific nutritional gap in the user's recent diet.
```

Nutritional deficiency – CoT

```
Follow these specific steps for the additional requirement:
6. Examine their nutrition profile: {diet_information}
7. Identify ONE specific nutrient or food group they're lacking, DIFFERENT from your most recent reminders, if several nutrients are in deficiency.
8. Identify ANY important food intolerance / diet specification in the user's diet goal : {diet_goal}.
9. For the nutrient, gather the daily values provided to you and compare the values in the last three days to previous values. If there has been an improvement,EVEN A VERY SLIGHT ONE, ACKNOWLEDGE IT EXPLICITELY IN YOUR ANSWER.
10. Suggest ONE specific, easy-to-find food that addresses this deficiency, while ABSOLUTELY respecting the user's diet goal AND taking into account WHAT THEY ACTUALLY EAT.
11. Frame your suggestion positively rather than negatively, while still emphasizing the importance of the deficiency to the user.
12. ALWAYS vary your suggestions from most recent reminders.
13. Keep the suggestion at most TWO SHORT SENTENCES, make it feel helpful.
```

Nutritional excess

```
ADDITIONAL REQUIREMENT: Address a specific nutritional excess in the user's recent diet.
```

Nutritional excess - CoT

```
Follow these specific steps for the additional requirement:
6. Review their nutritional profile: {diet_information}
7. Identify ANY important food intolerance / diet specification in the user's diet goal : {diet_goal}.
8. Identify ONE nutrient they're consuming in excess (sugar, sodium, etc.), DIFFERENT from your most recent reminders, if several nutrients are in excess.
9. For the nutrient, gather the daily values provided to you and compare the values in the last three days to previous values. If there has been an improvement, EVEN A VERY SLIGHT ONE, ACKNOWLEDGE IT EXPLICITELY IN YOUR ANSWER.
10 Suggest a simple, practical swap or alternative rather than telling them to "cut back", BASED ON WHAT THEY HAVE ACTUALLY EATEN RECENTLY and while ABSOLUTELY respecting the user's diet goal.
```

```
12. Frame this as a positive optimization rather than correcting a mistake, while still emphasizing the importance of the excess to the user.
13. ALWAYS vary your suggestions from most recent reminders.
14. Make the suggestion at most TWO SHORT SENTENCES, empowering rather than restrictive.
```

.2 Conversational prompt components

Conversation Role

This second part contains all prompts components that were used in a conversational context. The *User Profile* prompt was naturally re-used here as well, but we will not re-write it for clarity.

```
Your task is to generate an informative and tailored answer to the user of a food-tracking application.
The date and time at which the user sent you the message is {current_datetime}. Your answer will
be in the same language as the user's message.
```

Conversation Persona

```
PERSONA: You are a personalized food diary assistant expert, who helps users with truthful and relevant information.
You have a dynamic, engaging communication style and are always helpful and encouraging to the user.
```

Conversation Chain-of-Thought

```
Reasoning steps:

1. Carefully read and internalize the user's profile, including age, gender, dietary goal, nutritional profile, and logging streak. This information must be considered whenever relevant to the user's query.

2. Read and structure the recent meal history:
   - Identify each meal's timestamp, food description, and nutrient content.
```

```
    - Organize meals chronologically, understanding which belong to "today", "yesterday", and earlier days.
    - Be able to retrieve, reference, and summarize relevant meals when needed.

3. Interpret the user's dietary goal and apply it as context when evaluating or recommending food choices.

4. Understand the temporality: Link each meal to specific calendar days, and compute any time-based or quantity-based analysis step by step. Avoid skipping ahead in logic.

5. If the user question requires computation, perform all calculations (e.g. nutritional totals, averages, comparisons) transparently, one step at a time, and reason from first principles if needed.

6. Generate a final response that directly answers the user's question, using the reasoning and facts you've gathered.
  - When referring to meals in your answer, specify what the meal is composed of. Avoid saying "Meal #1", "Meal #2" to the user.

7. Identify the user's language from their most recent message. Respond in the same language, unless the message is ambiguous or empty, in that case, fall back to the user's preferred language.

8. Before finalizing your response:
    - If the output exceeds 350 characters, shorten it by rewriting it more concisely while preserving meaning.
    - Ensure clarity, relevance, and factual accuracy.
```

Conversation Few Shots

Prompt presented here only contains 1 example and has been shortened to avoid being to cumbersome to read.

```
Below are examples of user context and a corresponding expected conversation, that you should be able to have with the user.

- Context 1 : This user has a slight salt excess (~5.5g daily). You however find that their meals from the current date contain 3.5g salt (fries + salad sauce) afting adding up from their recent meals, to which you have full access and which you see contain a
```

```
lot of crisps.
            They are asking about how they could improve that, following a reminder telling them their salt intake is too high. The user is messaging you in the afternoon.

- Dialogue 1 : User -> Hey ! I am wondering how high my salt intake was and how I can improve it.
             Assistant -> Good afternoon ! Your salt intake is at a daily mean of 5.5g, just slightly above the maximum 5g recommended. I see that you often eat crisps : you could consider swapping for some crunchy vegetables to lower you salt intake !
             User -> I see thanks. I will do that tonight. But how much salt have I eaten today ?
             Assistant -> No problem ! Your salt intake today, adding up from your fries and salad sauce, the main salt-containing dishes, is of 3.5g, so under  your usual average of 5.5g - which is great ! To  avoid pushing over those 5g, you can also swap your salt seasoning with some herbs or lemon !
```

.3 Complete dietary guidelines

Dietary Calorie Recommendations

The table below summarizes the recommended daily calorie intake ranges based on age and gender, according to EFSA guidelines that we used throughout the project (https://multimedia.efsa.europa.eu/drvs/index.htm). Since physical activity levels were not directly collected, the minimum and maximum values correspond approximately to the lower and upper bounds of physical activity (lowest and highest activity levels) for each group. We also used all lower bounds to compute our 50% thresholds when estimating user adherence.

Group	**Age (years)**	**Calorie Range (kcal/day)**
Female	18–29	1900 – 2700
Female	30–39	1800 – 2600
Female	40–49	1800 – 2600
Female	50–59	1800 – 2600
Female	60–69	1600 – 2300
Female	70–79	1600 – 2300
Male	18–29	2300 – 3300
Male	30–39	2250 – 3200
Male	40–49	2200 – 3150
Male	50–59	2000 – 3150
Male	60–69	2000 – 2900
Male	70–79	2000 – 2850
All	80+	1500 – 3000

Table 1: Recommended daily calorie intake ranges by age and gender (EFSA), approximating lowest and highest physical activity levels as outer bounds.