Chatting about data - a conversational interface for meal tracking

Lorenzo Martinico

MInf Project (Part 1) Report

Master of Informatics
School of Informatics
University of Edinburgh

2018

Abstract

Increases in smartphone usage have made recording habits easy for anyone, enabling self-tracking of vitals and preventive medicine applications. Effective results have been achieved with food logging, but the User Experience (UX) of mainstream applications is still not perfect. We propose and describe the implementation of a conversational interface for meal tracking, which combines features of modern messaging platforms to enable capturing food via text or images, and give verbal feedback about nutrition. We enumerate some structural issues that complicate a successful implementation and our attempts to address these problems. To measure the effectiveness of our solutions, we conducted a user trial, comparing the chatbot with a popular meal logging application. Despite some disappointing results in its usability, we highlight evidence that a conversational interface might eventually be successful in this area, and some proposals for future directions of reseach. However, in light of recent events, we recommend that future work first take into account the confidentiality of user data.

Acknowledgements

I have to thank my supervisor, Stuart Anderson, for being an endless source of new ideas, giving me a new angle to look at the problem from, whenever I was stuck, and for agreeing to go along with my radical plan changes.
Nikita and Sumit, for being a bouncing board for new ideas, answering my machine learning questions, acting as guinea pigs during testing phase, and above all putting up with me for a year.
Yanna, for an amazing last minute readthrough of an unreadable first draft, and Em, for her always *invaluable insights*, and validating my love of commas.
Finally, to my family, for constantly going out of their way to help me succeed, and being a call away when I needed help.

Table of Contents

Chapter 1

Introduction

The *Chatting about data* project aims to explore the potential of using chatbot technologies to empower users by increasing their understanding of their own health through data. The project explores different areas of human health: sleep, mental health, and nutrition. The latter provides an interesting variety of challenges that have been explored in academic and clinical trials, but never for the purpose of a comprehensive, general purpose consumer application. This may explain why its still not common to discuss your dietary habits with your favourite personal assistant, regardless of how ubiquitous the technology has become in the last decade. But while the trend of consumer voice assistants has been developing products as all-round helpful tools, which do not specialise in any one task, the lack of a clear winner in the space of textual digital assistants has resulted in the proliferation of programs with smaller scope.
Known as *chatbots*, they usually specialize in accomplishing a small set of specific tasks, akin to the app model of the early smartphone era, and are similarly distributed through a centralized platform. This means food logging, a facet of the growing *Quantified Self* movement that has exploded with the smartphone, might well be suitable for the scope of operations of chatbots. And just as the smartphone has not completely solved food logging, with adoption still relatively low compared to tracking of other vitals and abandonment being high, the relatively unexplored area of natural language assistants in the context of *m-health* (mobile health) still presents several unsolved problems.

A nutritional assistant must contain two essential features: the storage of foods consumed by the user, and the retrieval of such information for the purposes of informing dietary choices. There is much variation in how these two tasks are accomplished among various existing solutions, and the aim of our project is proposing an implementation that, using a Natural Language interface and the ubiquity and capabilities of modern messaging platforms, makes interactions easier and more engaging.

Automated food logging in most commercial solutions is predominantly text based, with the user typing in the name of what they had on the day, selecting the closest match from a database of foods and their nutritional value, and estimating the amount of portions they consumed. It has also become popular to add an option to scan the barcodes that are present on all commercial food packaging, linked to their database entries. Some experiments have been conducted to estimate calorie content through image recognition, but there are few commercial products available to the general public, and most experimental implementations require a marker for dimension, which is a significant detriment to usability.
We propose implementing a hybrid textual and photographic input food assistant, where portion size can be specified by typing in a quantity, or by characterizing consumption in comparison to previous meals. This eliminates the cognitive load of using a complicated interface or trying to remember precise portion sizes, and will prompt the user to start reflection on how much they have been eating recently, rather than focus on number of calories tracked. We hope that thinking about past nutritional history in relative terms and exposing overall trends could be more useful than rigorous data collection for the purposes of maintaining long term engagement.

The harder problem is information retrieval, or rather how to extract value from the data and display it in a useful manner. The traditional techniques include calorie counting and breakdown of nutritional information based on macro and micro nutrients, usually shown as a table, a graph, or a progress bar. These strategies have been disputed by the nutrition community, and do not seem to be effective on the general public, resulting in poor engagement and little understanding of underlying dietary issues.
Our solution involves handcrafting a small set of simple heuristics to detect harmful dietary patterns, and initiating a conversation by highlighting the issue and giving some suggestions to the user. While this method is unsatisfactory and “does not scale” to deal with the whole gamut of possible dietary problems, there are so far no better solutions in the literature, and we hope having a natural language dialogue will be more accessible to a non-expert than having to interpret the traditional markers in other fitness apps.

We describe our process in creating the Healthbot chatbot on Facebook Messenger, from designing its functionality to implementing its logic, and running a user evaluation comparing it to *MyFitnessPal*, one of the most popular food logging consumer technologies. We also explain the various challenges we have encountered during our implementation, and how they relate to the wider field of chatbot design, as well as the insight we gained from our participants. Chapter 2 provides some background information on the state of the fields of chatbot development and its relationship with the Quantified Self movement and nutrition tracking in particular. Chapter 3 describes the design and architecture of Healthbot, along with a review of features that we originally set to implement, and why it was impossible to complete them. Chapter 4 shows the steps we have taken to review the chatbot's functionality, both during development and after, describing our evaluation study and participants' responses. Chapter 5 elaborates on the results gathered from the survey and observation of how participants interacted with the prototype, and identifies future problems that will have to be tackled before distribution to the public.

Our contributions to the field are implementing and deploying the first integration of textual and photographic food logging in a chatbot interface, as well as a comparative experiment between the chatbot and traditional diet tracking app on subjects from a similar demographic group. Our chatbot is also the first to utilise relative units of measurment for capturing logs, a simplification that has been well-received by testers and shows some promise for future implementations.

Chapter 2

Background

2.1 The chatbot revolution

Much has been said about how the rapid reduction in cost of the semiconductor has changed the world in a significant way in the last 60 years. The rapid spread of inexpensive and energy efficient computers, networks, and storage facilities, has revolutionised the way we access information, exchange goods and services, and communicate with one another. The proliferation of the smartphone, specifically, has brought forth an explosion in the amount of information generated globally, with more than two thirds of the world population owning one [32]. Adults in the United Kingdom spend an average of 2 hours per day interacting with their phones, browsing the web, using applications, generating tracking data, and chatting, the latter being one of the most popular applications, with 42% of mobile users [29] being active on social media. The top downloaded messaging applications, as of January 2018, are *Messenger* and *WhatsApp* (1.3 billion active users each), both owned by *Facebook*, and *WeChat* (980 million), owned by China-based *Tencent* [28].

Besides keeping up with friends and professional contacts, business transactions are also conducted through chat, either arranging sale and delivery of goods, or for customer assistance, the former being much more prevalent in Asia, where most medium and large size companies, as well as some smaller ones, have a WeChat presence and conduct what is known as *conversational commerce*. Increasingly, many of these transactions are being automated through the deployment of *chatbots* (bots), an evolution of classic conversational interfaces that have become popular in the last decade for commercial and entertainment applications [54]. The most popular bot platform outside China is Facebook Messenger, which introduced the functionality to developers in April 2016 [17], and it has since taken off with more than 200,000 bots on the platform as of December 2017 [16]. The development of Voice Assistants such as Google Assistant, Siri, Cortana, Alexa, or open-sourced Mycroft [1], have also pushed the deployment and popularity of conversational interfaces, either through voice or textual input. But if the above mentioned technologies try to be a general purpose digital assistant, most chatbots are typically concerned with a smaller domain problem, such as booking flights, checking the status of a bus, or telling a simple story [77].

2.1.1 Natural Language Processing

The recent uptick in chatbot usage can not be attributed only to marketing, but also to significant advances in Natural Language Processing (NLP) and Natural Language Understanding (NLU). While early chatbot implementations relied on simple pattern matching rules based on recognition of specific words (entity recognition) or part-of-speech (POS tagging), most of today's chatbot frameworks can leverage large corpora to apply machine learning algorithms, such as Intent analysis. Conversation can follow a slot, or a flow model: the latter is a hardcoded scripted flow diagram that guides the user through a preset conversation; the former specifies "slots" that contain some data the developer is interested in, and the chatbot will use NLP techniques to fill the slots from conversations with the user. Responses are typically pre-written and matched to an intent, but advances in deep learning are opening up possibilities for generative models, which create the answer from scratch [63]. Particularly successful can be combinations of several approaches, such as Serban et al. (2017)[94]'s use of reinforcement learning to combine the approach of a generative deep learning model and a template-based retrieval model. Critical to the success of the chatbot is a good context management system, to ensure that a multi-turn conversation doesn't feel disjointed, and that previously entered information remains available to the chatbot throughout the session. All of this functionality is implemented by a growing variety of open source and commercial tools available today [68].

[1] https://mycroft.ai

2.1.2 Chatbot use cases

From a service provider's perspective, the potential of using a chatbot instead of a human to provide customer service or present content can be very appealing, offering the opportunity of automating repetitive tasks. A successful example of a chatbot deployed in a classroom setting is Georgia Tech's *Jill Watson*, which complemented a team of Teaching Assistants to answer questions on a high traffic class forum [58]. But if chatbots can decrease the load for a smaller team, and improve customer experience by decreasing response times, they can also cause increases in customer dissatisfaction. In fact, conversational skills and friendliness are important elements when interacting with a customer service representative [69]. Within these interactions, a particular emphasis on tone should be given [80], an aspect of conversation that until recently had not been developed in bots [67]. While a complete replacement is still far off, and would cause concern for the livelihood of a still large number of people, chatbots are likely to soon be adopted by more companies to deal with clients' initial queries, while having a human agent supervise the conversation and be ready to intervene when necessary. The centralization of services under a single interface might also address the phenomenon of "app fatigue": smartphone owners are no longer installing new apps, and when they do retention rates are abysmal [1]. While giving up apps for chatbots might limit some kinds of deeper integration into users' devices, it also significantly increases the company's reach to everyone who is registered on a messaging platform, rather than the few who will choose to install new software.

Chatbot users report their main motivation to be the increase in productivity in information retrieval tasks, compared to installing an app, scanning a long webpage, or placing a phone call, as well as the possibility of receiving customised replies based on their own preferences [41]. The decrease in attention spans in younger generations [103] and addictive design might explain why young users of social media crave immediate feedback [42]. Thus, as a synchronous form of communication, chat might be perceived to increase productivity over an asynchronous medium like email, and is reflected in users favouring conciseness in the chatbot's personality [76]. Other reasons for people to use chatbots, to a lesser extent, are the entertainment value, the social aspect of conversation, and the novelty value, with users actively selecting chatbots as a tool because they fulfil some form of psychological pleasure, according to the *use and gratification* theory [41].

2.1.3 A new User Interface

Given the necissity of using chatbots productively, user needs will cause significant consequences for the field of Human-Computer Interaction (HCI): new paradigms of interface design will have to be invented, and novel approaches to combine different types of output will be possible [60]. Undoubtedly, conversational interfaces provide a great advantage to the less tech-savy, who might have trouble understanding the user interfaces of many bespoke applications, but will already be familiar to the “unifying” chat window, a philosophy promoted as Zero UI [45]. The blank canvas offered by the chat interface offers an unlimited potential for User Interface Design, to the point that, if the chatbot application is aware of enough information about its users, it might be able to create personalised interaction tailored to their preferences. Doing so may alleviate the growing digital divide that some segments of the population experience, because of the implicit biases of tech companies’ employees when designing user experience without considering people who are less comfortable around computers [43].

2.2 Advances in m-health

Since the earliest days of chatbots, such as Eliza, which was modelled after a Rogerian psychotherapist [102], it is clear that conversational agents can have significant impact on health. Their applications, however, initially remained limited, in part because of the restriction of having to teletype into an inaccessible computer, but mostly because the still primitive state of the technology made freeform chatting, and, crucially, the ability to understand anything about mental health, impossible.
More recently, the development of cheap sensors and widespread connectivity through smartphones has spurned a growing sector of m-health applications. From the 2000s, mobile phone use made doctor-patient communication quicker and more frequent, as well as enabling some initial forms of monitoring and providing a hub for Body Area Network, including different medical sensors [87]. The 2010s have seen further developments for these usages, and also the advance of ”Artificial Intelligent” applications. These are made possible by the vast quantities of data that can be collected through mobile devices, as well as advances of other Computer Science and Engineering disciplines, such as image recognition, virtual reality, robotics, drones, 3D-printing, and the Internet of Things (IoT) being applied to the medical field [88].

The vast quantities of data being collected have helped to advance the state of the art on several medicinal application - but they also provide a valuable monetisable resource, both from the perspective of Big Data analysis, and for the number of for-profit advertising companies that will sell medical information to marketers and insurance companies [98]. And whenever personal data is put online, it will also attract hackers, who might be interested in patient data for its resale value on the black market and for identity theft [104]. The sensitive nature of this kind of information makes it difficult to operate without breaking patient confidentiality, since anonymised medical records can be re-identified by correlating with outside sources [97], and even large experienced medical institution and data collection companies can breach existing regulations [20].

2.2.1 The Quantified Self

Smartphone and wearable device users have also been collecting their own personal information, giving birth to the idea of the Quantified Self, or Personal Informatics. [90] Much of the Quantified Self movement is based on the idea that the automation of data gathering will lead into greater insight and improve our own health and behaviours by making us reflect and evaluate our past experiences. Data collection for Quantified Self purposes in activity and sleep tracking has boomed in the last decade, with the proliferation of inexpensive inertial measurement units and heartrate monitor sensors combined in a wristworn factor, as well as step counting applications being bundled in many smartphones [53].

Rivera-Pelayo et al. (2012)[93] propose a framework necessary for a self tracking app, based on the three activities of tracking cues, triggering and recall. Tracking can be done through software logging or hardware sensors. Triggering can be active, when the user is prompted to reflect in a suitable context, or passive, where the information is simply displayed in a location the user can observe and notice significant changes. Recalling can be aided through different techniques, which usually involve a considerable amount of post-processing, enhanced by access to large datasets. There is still much work that could be done in the area of contextualization, by associating the collected data with other sources, and data fusion, comparing your own data with independent self or peer reporting, as well as better data visualization in attractive and intuitive ways.

One example where the Quantified Self phenomenon can have a real impact is preventive medicine, promoting healthy lifestyle to alleviate future medical issues. Diet, in particular, has been shown to benefit from open ended food-logging more than other methodologies [38]. While Turner-McGrievy et al. (2013)[100] found little advantage in replacing a paper diet tracking system with a mobile application, Personal Activity trackers in their study did receive an advantage; this leads to speculation that the lack of improvement in the first group might have been caused by the diet tracker's UX. In fact, using the My Meal Mate app over a 6-month trial, Carter et al. (2013)[49] reported increased adherence, usage, convenience, social usability, and overall satisfaction compared to traditional diet tracking.

Good examples of currently active commercial Quantified Self apps for fitness and nutrition are MyFitnessPal [14] and Google Fit [12], whose designs have been shown [96] to prioritize continuance intention (the willingness to continue using the app), usability qualities (directness, informativeness, learnability, efficiency and simplicity), and user value features (satisfaction, customer need, attachment, pleasure and sociability). The usability of these interfaces, which include both a textual input and a barcode scanner for commercial products, has increased the number of DIY food loggers [36]; however, there is still some friction to a seamless logging experience [40].
The output of fitness tracking apps is also of questionable usefulness: using calories as a primary metric is an oversimplification, as different kind of foods provide more nutritional value for the same caloric amount [71]. This might mislead a user in consuming more food than they should be, just to hit their calorie goal, or to leave some essential nutrient off their table, which can have negative consequence on overall health and impact their weight goals. In fact, Hebden et al. (2014)[65] reports that user engagement with m-health food logging solutions tapered off after a month of sustained use. Within the study, patients were most engaged by the text messaging component of the food logging system: this suggests that some of the issues with current interfaces might be alleviated by the use of a chatbot.

2.2.2 Medical chatbots

Chatbots have been speculated to provide a useful tool as a behavioural intervention technology, used to complement human practitioners in reaching a larger number of users and automate personalised messages [61]. Experimental and clinical trials using simpler informational chatbots have been made in various medical fields, such as

counselling [48], mental health intervention [59] and sexual health information [44], generally providing positive results, or at least giving indication that the medium might be used to address the specific issues.
Among the more successful experiments in Behavioural Health Intervention, the *MobileCoach* open-source platform [18] was originally designed as a text messaging based system, which did some parsing on the backend but mostly relied on practitioner's interventions. It has now been redesigned as a full fledged online chat platform, which was perceived positively in clinical trials where participants treated for obesity interacted with a chatbot that exhibited a distinct personality [72].

One successful attempt at making a dietary tracking chatbot is Forksy [13], which seems to be still actively developed, but there are no statistics on its usage and success rates. From direct experience, Forksy is very aggressive with its reminders to log your meals, and effectively displays the food diary through a graphical interface, but seems to have no "smart" features, and is inconsistent with the quality of its parsing.

2.3 Making a smart chatbot application

As researchers in many other fields in Computer Science are realising in the last decade, the collection of large quantities of data can have other uses besides record keeping, by leveraging the booming fields of machine learning and data science. This could eventually lead to make a truly useful chatbot, which in the future could replace or complement professional dietitians in a larger measure than today's solutions.

2.3.1 Nutritional knowledge

Ever since Richards (1902)[92], attempts have been made to algorithmically use food composition values to maximise food value per money spent. However, as Richards herself notes, "we know too little of the effect on digestibility, of cooking, and of the combination of two or more foods in one dish, or at one meal, to permit of very close calculation".
Even a century later, nutritional science still struggles to establish criteria to categorise individual food as "good" or "bad", because of the large number of nutrients each is made up of [101]. While dietary guidelines exist based on nutrients rather than

foods, they often fail to be effective, because of the difficulty for consumers to find options that combine all nutritional recommendation, while also being economically accessible, convenient, and conforming to their taste and cultural preferences [62].

To achieve a truly smart dietary assistant, given a vast amount of information about users, their habits and goals, we should be able to recommend an effective strategy to achieve the latter by analysing their choices in the former. As Gregori (2017)[63] describes, the architecture of a chatbot requires four components: a front end, a knowledge base, a back end and a corpus. While there are many tools that can be used for chatbot front end and back end, finding an appropriate corpus and generating a knowledge base are domain specific tasks, with far less options available.

Even for a restricted dietary task such as reducing fat consumption, current AI expert systems are based off a set of handcrafted rules [89], and more generally require experts in the field ("knowledge engineers") to add in known facts [51]. While the medical community has made efforts to solidify their field into knowledge bases, there are no prevailing standards to read and interpret them, and although some attempts have been made to use knowledge graph representation to power a symptoms identifier chatbot [78], there does not seem to be a canonical dietary knowledge base. Current commercial apps use a combined approach of total calorie counts and macro/micro nutrient percentages, but this is often insufficient to initiate healthy behaviours [55].

Despite the criticism for the occasional sensationalism, the emerging field of dietary epidemiology advocates a holistic approach to nutrition studies, by taking into account genetic, lifestyle and metabolic information as much as dietary records, making the mere tracking insufficient to draw anything but the most casual inferences on the users' health [47].
But until this branch of the field develops enough to provide us with effective personalised nutrition (some recent startups would like us to believe they already provide that option [22]), it is possible to use a more restricted approach based on recognising unbalanced diets from the lack or excess of certain key nutrients, abstracting the mechanics of quantifying exact measures from the users by providing more immediate advice through food recommendation. Data analysis techniques on food composition can be used to draw networks of complementary foods (foods that together fulfil nutritional needs) [70], which could be used to give suggestions based on what users have already eaten.
There are plenty of choices for nutritional value composition datasets (the most pop-

ular compiled from the United States Department of Agriculture [21]), and free or commercial APIs [24].

2.3.2 Leveraging social connections

Success in activity tracking is influenced by demographics, with older and lower income subjects having lower rates of initial activation and retention [86]. This problem may be caused by the bespoke user interface each fitness tracker comes with, which we believe a universal chat interface will alleviate. However, there are other factors worth considering for integration.

Popular fitness tracking apps often providing social networking functionalities, which have helped participants achieve their fitness goal through a combination of competition with their peers and social accountability [50]. Gamification has also proven useful [82], and so have financial involvement, but only when profiled as a loss and not for modest gains [81].
One company who successfully integrated diet tracking with all these aspects (gamification, monetary incentives and social accountability) was Gym-Pact (later Pact app), which rewarded users for tracking their calories, eating enough fruits and vegetables, and exercising, but took money from them if they didn't, and allowed users to post their progress to social media and compare it with their peers. The app reached a sufficiently large number of participants [75] to sustain itself for several years, and a high percentage of users was frequently able to achieve their goals without cheating thanks to progress reviews from other users (but their business model was not profitable enough, and they closed in Summer 2017).

2.4 Image recognition

Most messaging apps today come with media functionality integrated; in particular, it's easy to take pictures and send them as a message from within the app itself. A diet tracking chatbot might benefit from the users' ability to take pictures and instantaneously receive feedback on their nutritional value. While, to the best of our knowledge, this has never been attempted within a chatbot interface, photographic diet diary have complemented food logging for many years, both in a traditional paper form to aid recollection [66], and, more recently, electronically. For most of the photographic food logging app, portion size estimation requires placing a fiducial marker, an object with a distinctively recognisable pattern, in the frame of the image, to enable fitting a geometrical model on the entire picture [34]. A slight twist on this has been given by Smartplate, a startup that uses a distinctively shaped plate to implement image based food tracking [25].

A different approach was used by Google research with the *Im2Calories* Android app [83]. Besides using a convolutional network based off a newly collected multilabel dataset to classify what the food in the picture is, different CNNs are also used to segment images and to estimate their 3D volumetry. This allows the app to assign calorie counts to images that contain different foods in the same plate, and to have more precise estimation of size. Unfortunately, neither the app nor the datasets have been publicly released.
More recently, small startups like *Calorie Mama* [5] and *Bitesnap* [3], as well as Samsung's digital assistant *Bixby* [99], have also implemented similar functionality, although it is still not clear how their models were trained or how effective they will be.

Among social media users, especially on the Instagram photo sharing platform, it's common to photograph images of aesthetically pleasing food. While this does by no mean provide an exhaustive nutritional history, it can be used as a further automation to save users from having to manually log their meals and extract nutritional information, as well as another potential avenue to establish social accountability to log healthy food [95]. We will still have to use computer vision algorithms on this data, because Instagram tags are unreliable in identifying the content of the picture due to the large number of slang-related false positives [91].

Chapter 3

Implementation

3.1 Design

As a dialogue agent, a chatbot needs to be designed for conversation with a human user, each turn in the exchange supplying or retrieving information from the system.

To determine what kind of interactions the users and the chatbot should have, we started our design phase using the Botmock[1] tool to draft some conversation graphs. Botmock is a simple web application that allows its users to draw textboxes on a canvas and connect them to form a flow diagram, using a library of design elements that are adapted from popular chat interfaces, to give the impression that each box is part of a chatbot conversation.

One consideration when determining the tonality of our conversation sketches was what personality the chatbot should exhibit. Given the novelty of the medium in its modern form, we still lack comprehensive guidelines to craft what conversing with a bot should "feel" like, although as more chatbots are designed and deployed, some heuristics are beginning to emerge [31]. Since chatbots users have shown to appreciate the fun aspect of designed chatbot personalities [76], we tried to imbue our conversations with some colloquial aspects, like using slang, calling the user names, and using emojis.
As final touches in the design, we added some level or randomisation to the stock responses we had written, to make sure the chatbot did not feel repetitive, especially for

[1] botmock.com

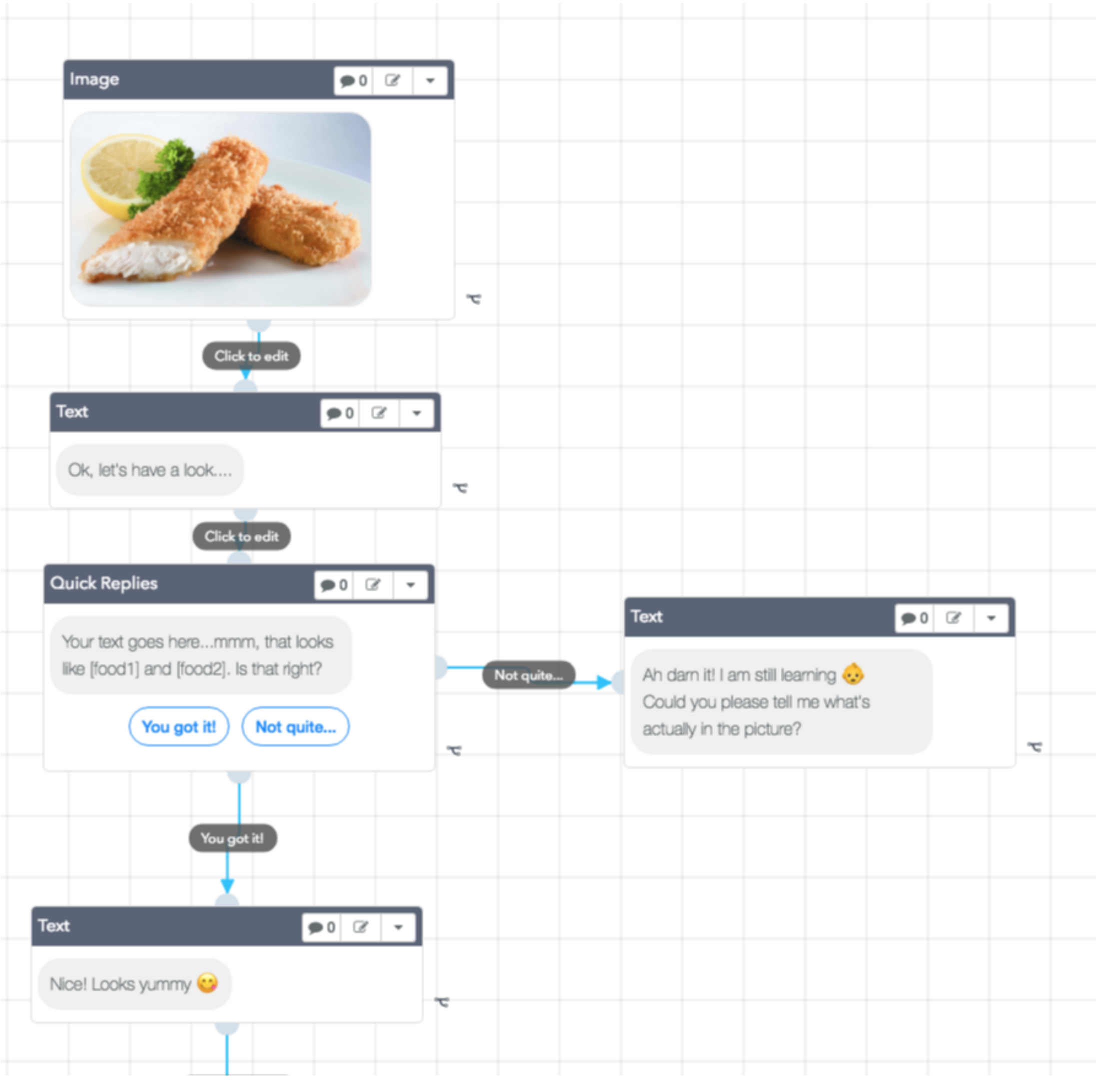


Figure 3.1: An example conversation flow drafted in Botmock

frequent expressions. We also provided our bot with a cartoonish image of a robot (available on the internet under a free license) as a profile picture, and gave it the name *Healthbot*, to give people an impression of talking with a real Facebook user.

3.1.1 User functionality

To establish a level of familiarity, rather than fetching a user's name from the client's account information, we initiate our first conversation by asking the user how they would like to be called. The rest of the introductory messages just describe the chatbot's functionalities.

Besides asking for help, users have three main available actions they can take: tell

the chatbot what they are eating, send a picture of food, and ask the chatbot to recap what they had on a specific date. If the user doesn't specify a quantity or measurement to their meal, the chatbot will ask for details. To avoid bogging down the user with measuring their portions, and since we mostly want them to think about their diet in general terms, we encourage the use of relative rather than absolute quantities (*more, less, same as usual*). We hope that using these measurements we will be able to maintain longer term user engagement, without sacrificing understanding of how much food is being consumed. Once we have this information, we send it through the Nutritionix API, to find out the nutritional content of that meal.
Sending a picture to the chatbot triggers a call to the *clarifai* vision API, to understand what the contents of the image is, using their food model [6]. *clarifai* returns a list of guesses for a picture, and their confidence value. If only one guess has confidence above the arbitrary threshold of 97%, we save the food in the picture in our database; in any other case, we present the user with a list of the top three guesses as interactive buttons, so we can store the correct value in the database.

3.1.2 Nutritional analysis

Once the user starts logging details about their food consumption, we will need to start analysing what they are eating to give them advice. Lacking comprehensive nutritional knowledge, the best we can do is crafting some elementary heuristics.

The food retrieval intent queries the chatbot's database for all foods stored on a certain date. To aid user's reflection, we provide some basic data analysis, specifically on the food's quantity:

- if there are more than 10 foods logged, with at least one having been consumed in large quantities and less than $\frac{2}{3}$ consumed in small quantities, we warn the user they might be overeating
- if more than $\frac{2}{3}$ of a users' total logged foods on the day are smaller portions than usual, or if they have eaten less than 3 foods, we warn them about undereating. If the food log requested is for the current date, however, a small number of foods could be caused not by undereating, but from the user not having logged their meals yet. In this case, we send no warning if the time is before 10 PM

One example of food category that is often praised by dietitians for its high nutritional value [39] is leafy green vegetables. To demonstrate the capabilities of chatbots as nutritional advisers, we handcrafted a list of leafy greens, which we try to match with every meal. If we do not observe the user eating any of these foods in 2 days, we prod them with a reminder.

Prodding in a chatbot context requires finding a very delicate balance: it can be too infrequent, making it less useful for users who are interested in being reminded; but if too frequent, it might soon become annoying. For example, the Forksy nutritional chatbot [13] sends a reminder to record your food for every single meal (at least 3 a day). This causes the bot to feel overbearing, and it might actually drive the user away. Forksy's solution is asking the user what time they want to receive their next reminder; but if no food is logged, it gives no option to deactivate reminders completely, and keeps asking for meals every day of inactivity.

We attempted to solve this problem by staggering the no usage messages at increasing intervals, first after one day of inactivity, then on the second day, and after that the fourth, seventh and fifteenth, to maximise our chances of getting a forgetful user to start messaging again, without being too annoying if they choose not to. Larger amount of data analysis might be used in the future to determine when, based on current dietary habits or external inputs, the notification might be more effective in nudging the user into resuming logging or eating vegetables, and eventually provide an overall more active prompting system that takes into account correct dietary practices.

3.2 Architecture

Our chatbot's architecture is composed of a natural language interface in *Google Dialogflow*, hooking up to an *Express.js* server running a *Heroku* instance, storing data on a hosted *MongoDB*, and retrieving other information from external APIs.

The agent itself is served as a bot on Facebook Messenger. While there are many possible client integrations for Dialogflow, and we considered using *Slack* as a distribution channel because of its first class support for bot users. Eventually we chose Facebook for its popularity, as it allowed us to serve the bot to any of its billion users without them having to install any additional software.

3.2.1 Dialogflow components

There are several Natural Language Understanding platform specifically designed to create and deploy chatbots easily and quickly. Our choice fell on *Google Dialogflow* (at the time *API.AI*), because of its ease of integrations with most popular messaging platform, ease of development, responsiveness and solid NLP functionality.

A Dialogflow agent is set up with a library of patterns, *intents*, that use example sentences or templates to parse inputs to the chatbot. Templates include parameters whose types are called *entities*, some of which are already defined by Dialogflow, but new ones can also be added by the programmer. One of the standout capabilities of Dialogflow is that a parameter can be marked as required by the intent, so if the initial utterance does not contain it, the chatbot will prompt the user to specify the new parameter.
To maintain the flow of conversation, an intent can be followed by another, based on what the user replies, with a context object which stores all parameters passed from the original intent down to all its followups.
Each intent can trigger an immediate response, as defined on the Dialogflow console, or it can trigger an *Action* to be fulfilled by the *Webhook*. *Actions* are functions held on the server that can access the parameters of the current intent, execute their own logic, call external APIs etc. They usually are triggered as a POST request to the webhook's address, with the request JSON object as its body. The HTTP response determines the behaviour of the chatbot, either by replying with more text, sending media or rich text, or triggering an *event*, which executes an intent just as if one of the triggering sentences had been sent by the user.
Small talk intents are also defined as part of Dialogflow to take care of common sentences unrelated to the chatbot domain. Dialogflow's machine learning engine can be interacted with through a Training console (still in Beta at the time of writing), which can be used to correct intent and parameter recognition (and indirectly, to gain insight into how the engine's model classifies new text).

Our chatbot defines the following intents:

- *First interaction* is the initial intent, triggered by the Facebook Messenger ”Get Started” button or with the ”Greetings” keyword for debugging purposes. This asks the user what their name is, and is followed up by
  - *Name save* which waits for a name to be given, and saves it to the database

  - *Name confirmed* replies with a welcome explanation message

- *Food log - text* is triggered when the user adds a meal to the log. It takes required parameters of food (as a list), quantities, and optional parameters of meal name and date

- *Send food pic* waits for a Facebook Messenger Media object (a picture in our case) to trigger an action which runs image recognition. If no unique match is found,

  - *Send food pic - no* is triggered by the webhook whenever there is no clear candidate for classification of a picture

  - *Clarify food pic* recognises user's sending a clarification for what the food was, through a Facebook Messenger button. To make sure the button's message matches this intent and not the generic food logging, a *messenger_button* token is appended to the message payload

  - *Send food pic - yes* is launched whenever a picture is classified correctly, and just triggers the Analyse food pic intent

  - *Analyse food pic* takes the food content found in a picture received from the user, analyses and stores the result into a database

- *Help* matches a request for help with a few reminders of the chatbot's functionalities

- *Date retrieval* is used to query for past food logs, taking just a date as a parameter. For logging purposes, with removal functionality to be implemented in the future,

  - *Date retrieval - false* will recognise when the user declares the food log for that day to be incorrect

- *What's my name* is mainly a debugging intent, to check if the chatbot has managed to successfully save the name for the current user in the database

- *Sinkhole* is used for training purposes to redirect all the intents that were misclassified

Our chatbot defines the following entities:

- *meal* enumerates four different meals: breakfast, lunch, dinner and snacks

- *quantity* contains a list of all the different measurement units for food
- *meal-quantities* combines quantities with quantifiers (a, some, integers etc.)
- *approximate-quantifier*
- *date-ext* extends the @*sys.date* object with today and tomorrow

3.2.2 Backend

By default, the Dialogflow interface includes a small inline editor to implement some simple webhook logic. While the web interface is limiting for creating a backend of the complexity required, it's easy to export this example code to *Google Cloud Functions* [7], Google's serverless cloud function service, and their own database, storage and analytics tool, Firebase.
The sample code consists of an Express.js [9] web server, which listens to POST requests sent from Dialogflow to route */webhook* route. The server then crafts the appropriate response as a JSON object, which triggers a reply through the chat client. The largest portion of the code is the *action-handlers* dictionary, which associates a different function to provide a response for any of the Dialogflow intents.

We started developing our webhook from this starting example in GCF, but we soon realised that a core requirement of our design, the ability to call up external APIs, was not possible under the free tier of Google Cloud. Thus, it was necessary to find a replacement. Some options that were considered were Apache OpenWhisk [2], Captain Duck Duck [3], Amazon Web Service Lambda or Elastic Beanstalk [4], Dokku [5], 1backend [6].
In the end, Heroku [7] was selected as a solution because of its many productivity advantages. To select the popular Platform as a Service (PaaS) solution over all other alternatives, we took into consideration the mature tooling, the easy to use deployment infrastructure, which consists of simply pushing the code to a version controlled repository, the automatic inclusion of a free domain name, and simple (but barebone) scheduling functionality. Heroku offers both a free and a paid tier; for our purposes of

[2] https://openwhisk.apache.org
[3] https://github.com/githubsaturn/captainduckduck
[4] https://aws.amazon/com/products
[5] https://github.com/dokku/dokku
[6] https://github.com/1backend/1backend
[7] https://heroku.com

creating a prototype, the free tier offers all required functionality; however, it would not be sufficient to power a chatbot infrastructure in a production setting, as there are limits to free users' capabilities, most notably a temporary suspension (and extensive wake up times) after an hour of inactivity.

Having moved to a full PaaS implementation, we were able to expand our program from a single file to several modules, which was necessary to avoid a large unwieldy single file, and allows us to compartmentalize between different types of functionality.

- *index.js* is the main function for the Express server, it runs in a loop waiting to receive any requests, and serves a response for any predefined URLs. On a POST request to */webhook*, it will read the request body to find the action's name and parameters, and calls a function as defined in the *actionHandlers* dictionary. Most intents match one-to-one to an inline function in this object, although some are defined in an external module
- *messenger.js* contains output functions to send a reply to the user, either going through the Dialogflow API, or directly through the user's Facebook Messenger account (necessary for sending a message on a predetermined schedule, without a user initiating the conversation)
- *picture.js* handles all intents related to pictures, from querying the *clarifai* to handling incorrect or imprecise guesses.
- *mongo.js* abstracts some of the low level database, like connecting to the database and querying the username
- *analysis.js* deals with all food related activity, like querying the Nutritionix API and analyse users' dietary records
- *strings.js* contains all messages the chatbot will send to the user, all collected in one file for ease of editing
- *worker.js* defines a set of functions to run periodically, which will verify some information about each of the chatbot users, and send an appropriate message

The latter module, unlike all the others, does not export any of its functions, but runs every day at 8 AM GMT by the Heroku scheduler add-on.

We considered training our own food recognition model by training on the Food-101 dataset using the DeepFood Caffe model [74], but after an initial evaluation it seemed ike precision would not be much better than *clarifai*, so we opted to use the free service and save ourselves from training the model and store it on our server.

3.2.2.1 Database

When considering which database to use, the variety of offerings is truly staggering, so we decided to optimise for speed of development, by using a NoSQL database, which combines speed and scalability with the flexibility of not having to predefine a schema, an essential feature when building a prototype which has to undergo through several changes before finalising its design. While some chatbots designs integrate a graph database, which allows rapid aggregation over different entities, it was unnecessary for our current usagee, and we preferred to adopt a more widely used and better supported product.

Our final choice was a MongoDB database free instance hosted on *mlab.com*. The database consist of a single collection, users, with a unique document for each user.
Besides containing identifying information such as a unique Dialogflow ID, a session ID to initiate messenger conversations and a name, we also store a list of *meals*, which we defined as objects containing a list of food items, their quantities, a date, and an optional meal name (lunch, dinner, breakfast or snacks).
Finally, we have a *counter* object, which stores values for the number of days since the user has logged any food, the number of days since the user has had any green leaf vegetables, and the total number of days. These are used to determine whether a reminder should be sent to that specific user about any of those issues (the total count reminder is used to ask for feedback about our experiment after three days, to collect some data while participants are still engaged with the chatbot).

3.3 Roadblocks

While we have developed and delivered a fully functional chatbot, there are many features that we would have liked to have but could not implement for lack of time. Other features were started, but could not be completed.
During our planning stage, for instance, we envisioned Healthbot to be able to answer some facts about the nutritional properties of any food the user would ask about, and to have the ability of adding recurring meals, so that if a user had a habit of having the

same food every day they would not need to add it again and again.
We also started scripting some "challenges" the user could ask to be set, to improve their habits through gamifying the tracking activity; we explored options for both individual and group challenges.
Eventually, all of these projects had to be suspended because we had to prioritise the core features; for the other aspects listed below, we started working on an implementation, but encountered some obstacles.

3.3.1 Instagram integration

A unique feature of our initial chatbot design was its integration with *Instagram*. If a user had an account on the social network, we would have asked them during the onboarding procedure to log into their account to generate a unique access key, which we would later have used to crawl their image history for food pictures.
We registered an Instagram developer account, created an applet on the platform, and developed the onboarding dialogue sequence, giving the user a choice on whether to sync their account. If logged in through a unique generated URL, they would be redirected to a custom address on our server which recorded their access token.

We completed our implementation of this first phase, and successfully tested the token generation on our own account. But as we were developing the crawler implementation, our application was suspended for Terms of Service violation. Speaking in person with Facebook employees, they speculated that the reason for the ban was that the picture retrieval API is only intended for building alternative clients, making our idea infeasible.
In hindsight, it was for the better we stopped working on this feature: during the evaluation period, following the fallback from journalistic investigation over Facebook's excessively permissive data sharing practices [33], the Instagram API was completely deprecated [23].

An alternative with a more permissive API is *Flickr* [26], but we did not explore the possibility of using this smaller social network.

3.3.2 Defining a food entity

The set of Dialogflow predefined entities does not include one for food. This is in fact a nontrivial problem, because enumerating all possible food requires knowing about both "raw" ingredients, and commercial food which might be referred to as their brand name or with a specific product denomination. We therefore tried to handcraft our own food corpus, in the hope it would provide sufficient coverage.

To collect food entities, we first used the Open Food Facts database [30]. Besides being freely accessible, this option was selected because of the large number of entries, 374,259, the presence of generic food identifiers associated with commercial product for 59,383 of the entries, and a nutritional approved health rating on a A to F scale. Moreover, besides a raw data export, the service provides an experimental JSON API for online queries.

The raw database was exported as a MongoDB [19] object, in *bson* format. After having created an empty *openfooddata* table, using the *mongoimport* command we copied the contents of this object in the new database. Then, through the mongo console, we wrote a script to extract all food with a name in English.

```
var cursor = db.products.find(
    {$and: [
      {$or: [
          {``generic_name'' : {$ne: "", '$exists':true
            ↪ }},
          {``generic_name_en'' : {$ne: "", '$exists':
            ↪ true}}
        ]},
      {"countries" : {$regex: ``en|UK|United States|
         ↪ Canada''}
    ]}
)

var cursor = db.products.find({$and: [{$or: [{"" : {$ne:
   ↪ ", '$exists':true}},{"generic_name_en": {$ne: ", '
   ↪ $exists':true}}]}, {"countries" : {$regex: ``en|UK|
   ↪ United States|Canada''}}]})
```

```
while (myCursor.hasNext()) {
    printjson(myCursor.next());
}
```

Unfortunately, this produced only 795 food items which were indexed as being listed in English with a proper name, and manual testing for common foods resulted in several misses.
Subsequently, we found an online corpus of British raw food at LanguaL.org [79](1316 items), but it was also insufficient for common queries.

In the end, we decided to compromise and accept the risk of some false positives, deeming missing out on some uncommon foods more harmful. Our final approach was marking the food parameters as a *@sys.any* entity, which is equivalent to a wildcard match. This is fine for a more structured intent, where the user prefaces tier logged food with an action verb (“I ate”, “I had” etc.), but these are not sufficient to capture the full range of inputs, as some users will also just say the name of the food. For those type of formulations, we had to add an intent case which just matches entity *@sys.any*, but that obviously has the unwanted consequence of also catching utterances that match neither the intent nor the smalltalk module.

3.3.3 Classifying food

Just like our handcrafted leafy green vegetable set, we would have liked the capability of automatically recognise whenever food belonged to a certain category (meat, fish, nuts etc.), which we would have used to create a set of simple heuristics to detect food missing from the user’s diet.
The apparently simple task of associating the name of a food to a category is deceptively complex; even humans assign multiple categories to the same food (e.g. fruit or snack for *apple*), and thinking about a certain food as belonging to a certain category will influence their beliefs in regard to its nutritional properties [64].

Our naive attempt to classify food according to its category was to cluster it based on its nutritional values: ideally, similar kind of foods would have ended up being classified in the same clusters (”high in sugar”, ”high in proteins”, ”low in vitamins” etc.) and manual inspection of classified data could have been used to assign an intuitive category to each cluster.

The *k-means clustering* algorithm is used to group points in n-dimensional space into a predetermined *k* clusters, by iteratively computing the cluster each point belongs to, based on a distance metric, until cluster membership becomes stable.
While our vector space was 250-dimensional, corresponding to the number of distinct nutritional values identified by the USDA nutritional database, it is not trivial to determine the value of *k*. If we had had a distinctive number of food groups we wanted to obtain (like, for instance, the 10 categories identified by the Australian National Nutrition Survey of 1995 [85]), we could have used that as *k*, but obviously any kind of foods that belonged to a cluster we have not considered would have been incorrectly classified. Napoleon (2011)[84] describes an algorithm to both automatically select a value *k*, and to reduce the dimensionality of our data set. This allows us to reduce computation, by eliminating nutrients that do not contribute significantly to the classification, and is a necessary operation, because the size of our training set was too small compared to the dimension of each data point. Known in the literature as *curse of dimensionality*, it can lead to a variety of issues that would make any classification meaningless [73].
Once clusterings for the training set were calculated, any subsequent food the user logged would have been classified based on its distance from the calculated cluster centres.

We built a training dataset by fetching from the Nutritionix API the nutritional values of the 1316 foods in the LanguaL.org dataset, minus the 213 foods Nutritionix had no nutritional value on record for. The Nutritionix API returns a list of values of type (ID, quantity), where the ID Corresponds to nutritional values as identified by USDA. Because the free tier of the API is rate-limited to 200 requests per day, we wrote a script to break up the dataset into smaller chunks, download a portion of data every day, and recombine it. Having collected all our samples, we passed the data to a custom Node script to expand each food's value into a 250-dimensional vector. Then, using the *mljs* library [8], it would perform dimensionality reduction with PCA [56], find good starting cluster centres, and execute k-means clustering on the entire dataset through the *clusterfck* library [9]. The cluster centres would then be saved to a file, for the purpose of being used by the application to classify new food.

All attempts at classification, however, where disastrous. While we used the algorithm

[8] https://github.com/mljs
[9] https://github.com/tayden/clusterfck

from Ding, 2004 [56], to dynamically identify the number of cluster centres with the highest variance (a number of 7 for our dataset), our results were only marginally better than other clusterings. For any number of clusters we tried, most datapoints would group around 2 or 3 cluster centres, with one centre attracting around 800 values, a few hundreds for the other two, and all other centres having between 10 and no value. This result is corroborated by Kim et al. (2015)[70]'s finding that foods of similar origins do not necessary cluster together when grouping based on raw nutritional values. Their proposed network classification and metrics of nutritional fitness do offer a potential method to recommend complementary meals for the chatbot, but deriving the full network from the provided dataset would have required a significant amount of time, and each query would also require calculations that would take so long as to make a chatbot unresponsive.

Automatic classification of foods into distinct groups is still an open question. Besides nutrient counts, alternative classification methods could be the usage of word embeddings in recipes [37]. While this might be a good way to find if two foods are culinarily related, it does not necessarily satisfy the property of nutritional equivalence, which is what we would ultimately want Healthbot to do. Eftimov, 2017 [57] developed an automatic classification method for European standard food classification system FoodEx2, which lists not only composition values, but also chemical contaminants, food consumption and pesticide residuals. Their algorithm goes through a classified learning phase and a probabilistic natural language extraction phase for description, which achieves a good accuracy of 89%, but the resulting classification categories are too broad to be useful for our purposes.

Chapter 4

Evaluation

As with all software artefacts that have a user facing component, testing can be lead both on a technical level and on a usability level. Since the chatbot infrastructure contains many different components, it will be necessary to conduct some testing on the robustness of each.

4.1 Testing

Throughout the whole implementation, particular care was taken to follow software engineering best practices. All our code was regularly checked into a git version control system, with commits for any new logic features, and informative commit messages. Because JavaScript is a dynamic, weakly typed interpreted language, it lacks the checks for code correctness typical of a compiler. As a consequence, we run all code to find out if it works. To avoid deploying a broken commit to the server, we wrote a git pre-commit hook to test if the program crashes immediately and to pass a linter for catching any syntax error in the code:

```
npm start > /dev/null &
sleep 5
if ``eslint *.js`` && [[ -n `pidof -k node` ]] ; then
    echo "Pass linter and npm doesn't crash"
    exit 0
else
    pkill node
```

```
    exit 1
fi
```

While the combination of linting and running the program was a good tool to statically catch some errors, some bugs we encountered during development could have been avoided if we had had the convenience of a safer type system. It is possible to add type annotation onto Javascript code using the Typescript language [27]. There are also Dialogflow client APIs for a variety of additional programming languages besides Node. It would even have been possible to use languages without any official support, because the use of the client is not necessary to communicate with the platform (our own implementation only took advantage of a minimal number of services it offers): the only requirement to handle fulfilment is a webhook that accepts the correct kind of HTTP requests and provides a correct response in return.

It is also necessary to conduct some dynamic testing to ensure the interaction of all components is successful. The ecosystem for end to end bot testing is still very primitive, with some bot platforms offering their own bespoke testing utilities, and the only cross platform library being *Botium*[1] and its unit test frontend *testmybot* [2]. These libraries are inspired by and offer the semantics of general purpose behaviour-driven testing frameworks, allowing for easily composing test cases through a simple syntax. For example, one of our unit tests was simply a file containing the script we would expect from interacting with our chatbot:

```
Begin Conversation Test Case

#me
greetings

#bot
Greetings! What's your name?

#me
You can call me Lorenzo

#bot
```

[1] https://github.com/codeforequity-at/botium-core
[2] https://github.com/codeforequity-at/testmybot

```
Hi Lorenzo pleasure to meet you!
I am Healthbot. I will be your personal diet assistant
You can tell me what you are eating and how much, or send
    ↪ me a picture of your food, and I will record it so
    ↪ we can try to understand how you eat better!
After that, whenever you want to think back about what
    ↪ you have been eating, just ask me to tell you what
    ↪ you had on any date!
```

While *testbot* initially looked like a promising solution for establishing a routine of test-driven development, it soon was evident that the *Botium* hooks for Facebook were still not mature enough to be used in production.
Therefore, rather than having a collection of sample conversations we could feed *testmybot* to determine whether any new code change would break any of the responses we had been getting before, we had to resort to manually testing each new feature, by messaging the chatbot from a personal Facebook account, repeating the same script for each different functionality we had previously implemented. This would show very clearly if our modifications had broken an existing behaviour. In case of errors, most debugging information was printed to the Heroku server logs through the *console.log* JavaScript function.

Various analytics tools exist for chatbot usage, both built-in in the chatbot platforms (both Dialogflow and Facebook Messenger provide one, although the first is still in beta), and external, like *dashbot*[3]. While we considered instrumenting Healthbot with one of these, the ecosystem did not seem mature enough, and the limited scope of our prototype made manual examination sufficient.

During the evaluation phase, we used Dialogflow's training module to update the intents with additional sample utterances we observed from our test users. While this helped us to refine some entities to include expressions we had missed, the training module did not seem to significantly improve accuracy; it remains to be seen whether it will be more useful once it comes out of beta.

[3]dashbot.io

4.2 User trial

For our evaluation, we ran an experiment giving out the chatbot to 11 university students, all within ages of 20 to 25 and at least moderately physically active, to use for a week. As a control group, another 9 university students were prescribed to use the MyFitnessPal app for the same duration. All users were recruited through Facebook chat or in person, and all were given the OK to start the evaluation on the same day after having read and signed a consent form describing the experiment and the role they would have.
In retrospect, having a more gradual rollout might have helped with spotting the first bugs sooner, and giving us a chance to fix the underlying issues without compromising the platform for every other user. As it was, while we identified several issues and features that would have been immediately easy to add, we did not push most of the modification to avoid breaking existing users' workflows. Unfortunately Facebook does not allow to have a separate testing and production environments until the application goes through a first review process, which we could not afford to spend time going through.

4.2.1 Record keeping

Since this was our first usage of the chatbot outside our own testing, we expected to encounter a variety of bugs and phrasings that had never been encountered through manual testing. We set up a detailed logging function for all error cases, printing the user ID, to help us reconstruct the causes at a later stage through cross referencing. We could also access a complete record of all communication through the Facebook app console, as well as having a list of intents identified and how the parameters were matched from the Dialogflow agent. Although it could have been possible to keep a larger number of logs, for example for success cases, we thought it would make records illegible in the eventuality of having to go through debugging.

While having this much access gave us some great insight into what might be affecting faulty behaviours, it was very concerning how we could read the conversations in their entirety, and while Dialogflow allows to deactivate the logging, there was no way of doing that through Facebook. Even if there was, it would be trivially easy to still log everything through the server.

4.2.2 Methodology

To initiate the experiment, the chatbot users' Facebook profiles were added as testers through the Facebook developer console. They were then sent a link to the Healthbot's Facebook page, where they could press a clearly visible button to start chatting. This would open a chat window, where they had the option of pressing another button to get started before being taken through their first conversation. Users were given no indication on how to proceed, except for the chatbot's introductory message. Over the course of the evaluation, users sent us some questions (never through the chatbot) on what they could do with it.

The MyFitnessPal testers were asked to give feedback a week after the evaluation started; the bot testers were sent feedback forms after 9 days.
The surveys sent to both testers were built using the Google Forms online tool. Most questions were similar between the two questionnaires, with some variations when it came to input methods and displays dependent on the app. In compliance with the consent form, none of the questions were made compulsory, and the survey was made anonymous.

The questionnaire asked some background information on the participants, to establish levels of fitness and computer literacy, thoughts on nutrition and previous dietary and food tracking histories.
Testers were then asked their opinion on the usability, utility, pleasantness of the entire platform they were evaluating, as well as for each specific functionality, and if they had any feedback on things they would have liked to see. Some answers were multiple choices, checkboxes or Likert scales, but most were open text input to allow the participants to give a full explanation of the reasoning behind their answers.
The full survey and responses can be found in the Appendix.

4.2.3 Response

7 participants responded to the MyFitnessPal survey, and 9 to the chatbot survey. While some questions were answered by all the participants who took the survey, none of the open ended questions were answered by all, with some questions or even entire sections were ignored by more than half of the respondents.

Participants seemed to be distributed similarly across the two trials, with chatbot users

being slightly more proficient with computers, as well as being more aware of their fitness levels. Similar splits were evident in the proportions of participants who had dieted before, with around three quarters of participants citing a good current health or scepticism with established diet, and some chatbot users using laziness as a motivation. The minority of users who had dieted chose to do so because of environmental, athletic or health-related issues, but did not maintain dieting after reaching their goals, or because of commitment issues.
Among both groups, about half the participants consistently had 3 meals per day, with some having a variable number of meals and no participants consuming less than two; our chatbot users however in general snacked less than MyFitnessPal users (there is a possibility that these answers might have been influenced by the experiment, even if participants were encouraged to think about their behaviour before; the fact that MyFitnessPal presents snack as a distinctly separate category, and the chatbot doesn't, might have affected responses to this question).
About half of the participants reported having tracked their diet before, either keeping a food diary, memorising their meals, or (the majority of respondents) using MyFitnessPal; most of these previous trackers also kept a record of their snacks.

For unclear reasons, more than half of the respondents skipped the section about their dietary makeup, but for those who did fill it in, definitions of "balanced diet" varied significantly: while a majority named a variation of having a correct proportion of Protein, Carbohydrates and Fats, with some allowing for vitamins and minerals as well, others named calories as a main concern, reducing some unhealthy food groups and increasing others, or avoiding stressing about their diet and making sure to have what made them feel good. Only half of the respondents consider their diet to be balanced, including all those who planned their meals in advance, and most respondents tend to cook their own meals, eat out or do both things in equal measure.

MyFitnessPal testers found the app on average more useful than users of the chatbot, although the latter was rated as generally more pleasant to use. The food diary, the macronutrient breakdown graph and the remaining calorie counter were all generally considered clear and useful, with the graph giving the most pleasant feedback. For input, the majority of users preferred scanning the barcode of the meal they were having, although for some the kind of food they were eating factored into their choice, and people who tended to cook their own food preferred text entries. All users had some issues with finding the food they wanted using text entry, but no one complained about the

method being too slow; barcode scanning seemed to perform better, with only some users reporting difficulties identifying a barcode or matching the correct item in the database.
By contrast, chatbot users generally found the feedback useless, or insufficient. Chatting was highly preferred as an input method, although several participants did not think it understood their queries well enough, and some were annoyed by the prompts for size. Some users who took pictures for input found it not accurate enough, but the larger problem of this feature seemed to be people who were not aware of the functionality.

Retention rates were much higher for MyFitnessPal users, with the highest number of missed meals estimated to be 5, and some users logging all their meals; chatbot users, were much less active, with only one person logging almost every meal, and everyone else estimating having missed between 5 and 20. For both populations, the leading cause of having missed logging a meal was lack of time or forgetfulness, with some chatbot users finding input methods cumbersome or lack of interest because the feedback did not seem useful.
As a consequence, almost all chatbot testers did not log their meals on several days. Only half of the users reported having received a reminder the day after an inactive day, even though every user received at least one reminder (this might be explainable by the ambiguity of the question: a reminder was not sent out after every day of inactivity, and some users received one even if they had been active the day before). The reminders were generally found to be useful, and mostly made the users log their food on the day, and one user even expressed a desire to receive more frequent prompts to avoid forgetting meals.
MyFitnessPal also provided a reminder functionality, but it is off by default. All users who turned it on got a reminder, but it did not convince them to use the app after.

One stark difference in responses between chatbot and app users was on preference between noting their food records with absolute measurement (number of portions or unit of measurement plus numerical value). MyFitnessPal users overwhelmingly declared a preference for absolute value metrics, because of the need to calculate precise calorie counts that the app provides, and as a more reliable comparison method to standard recommended portion sizes. The majority of chatbot users instead indicated a preference for relative values, because they found easier not to have to constantly measure portions.

Despite the fact that the utility of a food diary comes from the ability to look back on previous meals, only a third of the chatbot users, and just over half the app users took advantage of this feature on a later date, and those who did reported the information presented to them to be accurate, but unhelpful; in fact, about half of the chatbot users and two thirds of app users do not think using the meal log has given them a better idea of how they eat.

Overall, most participants did not think that logging their food had helped them to eat better, although for many users that was because they already are happy with their diet. Those that registered a positive impact mentioned that having a better oversight on their food trends did prove helpful for them, and MyFitnessPal users specified sugar tracking and suggested recipes as useful features, although some comments also pointed out that the paid version of the app could have been more useful. However, two thirds of chatbot users found that they had become more "mindful" about their diet by using the chatbot, as opposed to less than half of the MyFitnessPal users.

Expectations for the chatbot were high for some users who were hoping for "*[a] good AI*" which would be talkative and give them active reminders and regular feedback; some were just looking for a more convenient way to log their food; but most participants did not expect much from it. Needless to say, the former group were disappointed by our implementation, with the natural language parsing of quantities, repetitive replies and image recognition capabilities being particularly frustrating. At the same time, users appreciated the ability to choose input method, and some found the chatbot's personality less annoying than they expected. MyFitnessPal testers also were expecting ease of use, a complete database, and a tool that would prompt small change in their behaviours by highlighting trends that were needed to be changed. Most of these were met by participants, although the majority of American commercial products in the database was deemed a problem.

When asked if they would continue to log their meals after the evaluation period, participants on both platforms were mostly uninterested, either because they did not find it useful enough, or because logging took too much time, and in the case of the chatbot, they perceived the product development as not being ready enough for regular usage. However, some users who seemed to have benefited from its usage were willing to continue interacting with it, or at least considered the possibility in case of future more rigorous dietingr. One MyFitnessPal user who had tracked their meals before was convinced to resume their paper food diary.

About half the MyFitnessPal users enabled fitness tracking functions, which seemed generally well received, although there were some concerns to how accurate their estimations were, and how useful it is to simply subtract exercise from calorie intake from a nutritional standpoint. Participants who did not use the feature were potentially interested, but the interface was not easy enough to understand, and there were perceived barriers to entry such as downloading a separate companion app or owning a smartwatch to better track calorie expenditure. A few chatbot users tried texting about their activity, but when they got no reply they did not make another attempt.

Testers of the app suggested they would have liked to have dedicated fruit and vegetable counters, automatic exercise calorie calculations and personalized recipe suggestions based on a specific ingredient or past meals and goals. For the chatbot, suggestions included pointing out the recommended amount of food, more reminders, especially around 5-a-day tracking, retroactively adding past meals, adding more variations to the automatic replies to make them less boring, and better onboarding functionality.
As part of the survey participants were also asked if they thought that the information they were uploading was being kept safe, and if they considered the issue important. Most participants responded were actually concerned about their dietary records being exposed, with some particularly worried of being judged because of their diet, while others did not think food records were a particularly sensitive topic. One users conceded that while data protection is important, anonymising dietary records could be used to benefit medical research organisations. Users of the chatbot generally considered their information to be secured, and while one participant specified “I know its developer takes security seriously”, another identified that platform issues were a problem: “I mean it’s on facebook so not really.” [4]. On the other hand, MyFitnessPal users were more worried about the platform, or unsure whether their information was safe or not. And with good reason: two days after the study completed, the app’s parent company *Under Armour* publicly announced it had been a victim in one of the largest ever leaks of personal user information [15].

[4] A very cognisant assesment, in retrospect, given the revelations transpired through the Cambridge Analytica scandal [33]

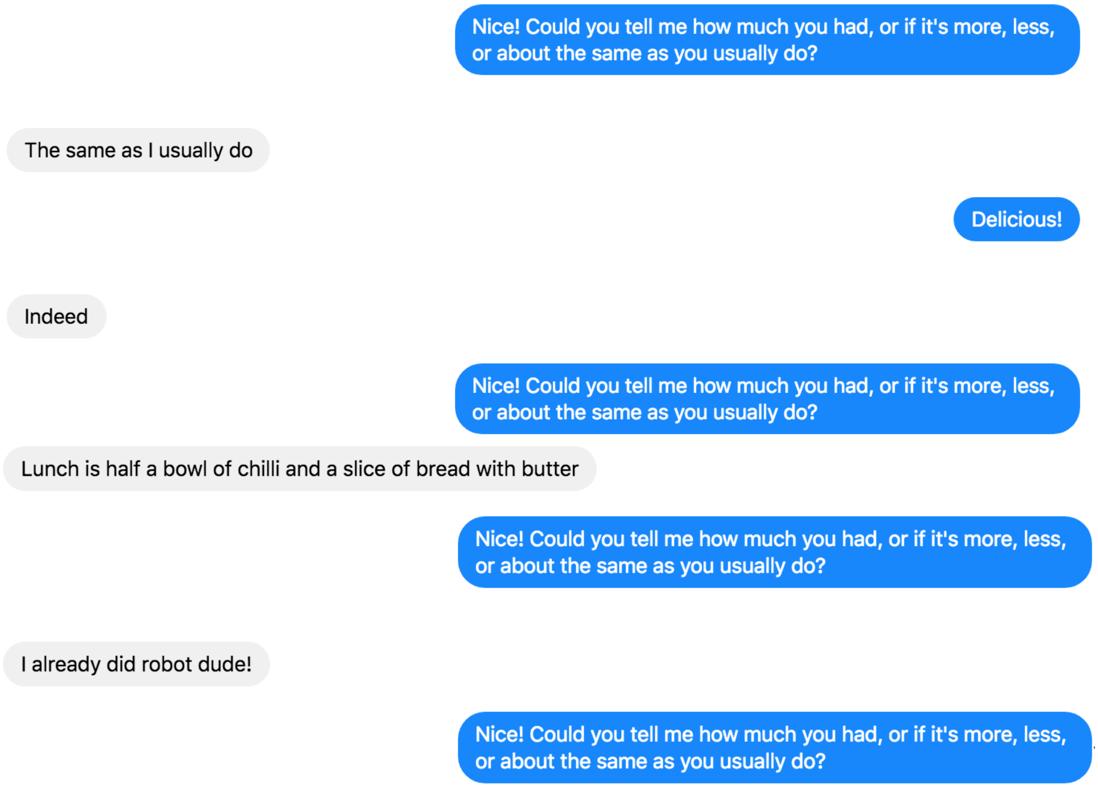
Chapter 5

Discussion

5.1 Chatbot performance

The nature of chatbot development is one of predicting what a user will say; this is already a hard job for any large software development effort, making it incredibly difficult for a single developer. Testing on our own account with phrasings we knew the chatbot would recognise was very different from deploying it in the wild and experiencing all the different ways participants in the study tried to express the same concepts.

Observing how the users were interacting with the chatbot made it clear that our design choice of minimising false negatives by using the *@sys.any* entity to capture food was not particularly successful. Users would often try to interact with the chatbot for non functional conversation, such as greeting it, thanking it about one of its encouragements to keep logging food, complaining about a mistake or simply acknowledging a previous message. In those cases, if the intent-triggered conversation had already been terminated, the new message would be incorrectly identified as food, without a quantity, and a clarification request message to specify the portion size would be fired. This also caused the intent recognition model to have several false positives, misidentifying stop words as food within a message that also contained valid food words. Fortunately, this kind of errors turned out to be less relevant, because users would already have to identify some portion sizes, so they would not notice any misidentification issues, and the recognised food was only stored in the database if matching with some record in the *Nutritionix* API.



But while the API did provide a safeguard against incorrect food identification, it also showed issues by misidentifying real food, either by catching only part of the word (cupcake was flagged as cake mix in one instance), or sometime inexplicably (Long island ice teas were stored in the database more than once, without the user having logged them). The ability to list several different food items within a single message, as well as the users' habit of splitting up different elements of the same meal between different messages, also caused difficulties in matching quantity to food, with Dialogflow's models simply not being advanced enough to catch the many subtleties of word ordering. This was another cause for the size clarification message to be presented after a quantity had already been given.

The *clarifai* API rarely conformed to our ideal parameter of a single identification over the 97% confidence threshold; however, among the top three choices, there was often a good candidate the user could choose to select. When confronted with no suitable choice, if the user tried to clarify what the food was, the chatbot appeared to recover gracefully, while actually logging the clarification as a new meal.

A considerable effect on how the chatbot performed with users was also due to operational issues. After obtaining consent to participate in the experiment, users were

provided with a URL they could open through the Facebook Messenger app or in the browser. Before it became clear that individual users had to be added as testers through the Facebook developer console, the first users who were given a URL did not receive any reply to their messages until they were added. While it was explained to them that this was not an issue with the chatbot but a temporary account management problem, it might have contributed to the perception that the chatbot was buggy. In fact, even after testing initiated correctly, there were some uncaught bugs that users could easily spot, which betray the fact that they were conversing with an automaton. For many of the chatbots' interactions, responses were selected randomly from a predefined list; over time, the user would exhaust all variants for a response, and since these were selected randomly sometimes would have the exact same responses given to them in quick succession. Sometimes, the database would fail silently when accessing the user name, so a message would be sent to the user containing "undefined" rather than their name.
Perhaps most grievously, an uncaught exception in the logic of *worker.js* caused the entire periodic reminder script to fail after having sent messages to the first user. Because the first user was our test account, we kept receiving reminders through the evaluation period, which caused us to believe everything was behaving correctly until the 6th day of the experiment. Since this was a very important feature we were interested in evaluating, we decided to push a fix in production, and to extend the trial by two more days to verify its effects. We deactivated the reminder for feedback after 3 days had passed, because with the change in schedule it would have fallen too close to the conclusion.
The fix was not perfect, causing the no log reminder to fire up for every user, even if they had just logged their food the night before. However, while this might have been annoying for these users, it proved highly effective, causing every single user to log at least a meal on the day and after, including those who had only tested it on the first day and given up after. The leafy greens reminder was also subjected to a similar bug, firing for every user just before the no log reminder. Unlike the latter, most users did not seem to react to the suggestion, and while some registered the message, it is unclear whether users failed to modify their behaviours because the greens reminder encourages a more difficult change, or simply because they only noticed one of the remainders. The second explanation is not unlikely, given the fact that our testers seem to ignore larger body of text while messaging, like the introductory welcome message which explained the chatbot's capabilities, causing users who were not given an explanation before the experiment to be unaware of some functionality.

Another functionality that was not presented to the users due to some small logic errors was the undereating reports. One of the condition for undereating was that if only two small items of food were inserted before 11 PM, we should flag an undereating option; this however should only have applied if the request was made for food on the current day, causing previous days' requests executed before 11 not to trigger. This issue might however be considered negligible for the purposes of our evaluation, since users rarely asked for previous days' logs, and those that did never fulfilled our conditions to trigger an overeating message.

5.2 Evaluation results

While our study provides us with some interesting qualitative data, it should be noted that we cannot conduct hard generalisations on our results, because of the tiny sample size. Our participants come from the uniform demographic of university students because of issues with the logistics of recruitment; while this enables us to compare data without taking into account how demographics affect replies, it also means we lack external validity, and we cannot understand how other groups will react, especially in vulnerable less tech-savy age groups or with major dietary issues. Further studies will be necessary to ascertain how usable the interface is for different demographics.

Responses to our survey confirmed that the chatbot is not ready for public release, and in its current incarnation provides little utility compared to app based food diaries; however, under some metrics the chatbot did provide a stronger performance, leaving open the possibility that with more work it might provide a suitable replacement.
More than one participant did note that interacting with the food log through conversation was, despite the frustrations with some faulty NLP, a benefit over MyFitnessPal, even wishing that the conversations could be more human like. In fact, while more users rated MyFitnessPal as being useful compared to the chatbot, the chatbot was rated as more pleasant to use than the app, despite the bugs.
Some evidence that the chatbot interface is easier to interact with than the MyFitnessPal menu can be glimpsed from the fact that a higher proportion of users did not log their meals because they were too busy or it would have been too much of a hassle. Indeed, there were complaints about the speed of entering a meal into the app being too slow.
Conversational interfaces also seem to influence users' thoughts on how food should

be measured in logs, with participants who were exposed to chat marking a preference for keeping measurement relative to previous meals, as opposed to MyFitnessPal users preferring absolute measurement. This can be attributed as a reflection on the functionality provided by the two systems: our chatbot does not provide precise calorie counting calculations, making it less important to have a precise metric on how much food has been eaten. This can make food logging easier, because the user does not need to worry about using a scale or counting portions, but only need to give the easy estimation on how much they have eaten compared to previous meals. Of course, having precise quantity would enable us to perfect data analysis on the backend, but having a relative measurement allows us already to build simple heuristics such as what we use to detect any new over or undereating trend in the users' habits, and there are some research projects looking into using machine learning to supplement missing health data [105].
Conversational reminder seem to be more effective than regular app notifications: every user of the chatbot logged a meal after receiving a reminder, while MyFitnessPal users who had turned on reminders did not act on them, or reported that they probably would not have been convinced by a reminder to log food. This of course might have been caused by the sparsity of reminders, since we never sent any in the first few days of the trial, and reminders to take actions outside the chatbot, like eating some leafy green vegetables, were not acted on. Nonetheless, users seemed to react favourably to the feature, and even requested having more frequent notifications to avoid skipping further meal logs.

The most important outcome from the survey that we did not observe during the previous evaluation trial was the issue of feature discoverability. Our solution of simply stating the chatbot functionalities in the opening message did not work, causing a significant number of participants not to be aware of features like picture logging or past logs requests. Users also reported not knowing about the analytics function the chatbot had, but those were intentionally left vague in the welcoming dialogue. The onboarding information was presented as a rapid succession of messages, which were also longer than usual, and the combination might have proven too much for the users who just scanned their content quickly. We also provided a helper message in case someone needed reminders on how to use the bot, but it was never invoked. Potential solutions to this issue include improving the initial dialogue, making it more interactive, breaking down the onboarding into a more gradual set up over the first day, or, taking a page from what most current Facebook Messenger bots do, using a persistent menu, which somewhat nullifying the benefits of having a conversational interface, and always presenting all features as quick reply options at the end of a previous conversation. This approach is highly recommended, as it provides user with a visual list of all input possibilities, as well as giving the chatbot a good understanding of what the user is intending to do, effectively resolving issues such as our chatbot's tendency of misclassifying an excessive amount of messages as food insertions. While we had been aware of this approach during development, we chose not to use it because in our opinion it broke the suspension of disbelief of having a conversation with an intelligent agent; in retrospect, it would have greatly improved our design, and it would be easier for user than being explicitly *trained* into using the correct formulation to match an intent.

Across both platform, we saw a common trend of users not being able to draw any utility from the logging because they considered themselves to already have a good grasp of their diet. As much as users may believe that is the case, the advantages for more conscious users will emerge through further data analytics over larger periods of times to identify long term trends, and better visualisation tools to display more useful information. In the case of the chatbot, since textual feedback is somewhat limited in the type of information we can display to the user, it would be beneficial to insert a webview within the Messenger app, which could enable us to display rich graphical content (the Forksy chatbot has a nice graphical overview for a user's food diary).

More adequate feedback could also be provided by taking into account users' goals and dietary preferences. It is meaningless to send a warning for having too little or too much food if we do not know whether the user is trying to lose weight or training to build muscle, and it might be harmful for users who already struggle with dieting to receive feedback their not having enough nutrients. Knowing whether a user is on a special diet, such as vegetarian or vegan, would also be beneficial, both for keeping track that users are getting all their nutrients, and to avoid making inappropriate recommendations for food that is missing from their diet.

Since our chatbot deals with people who may be in the vulnerable frame of mind of being insecure about their diet, we have to take extra care of what users are saying. Running some sentiment analysis on user input may avoid tragically inappropriate exchanges such as this, caused by the random selection of encouraging phrases in response to logging food, which besides making the user feel mocked, also betrays our

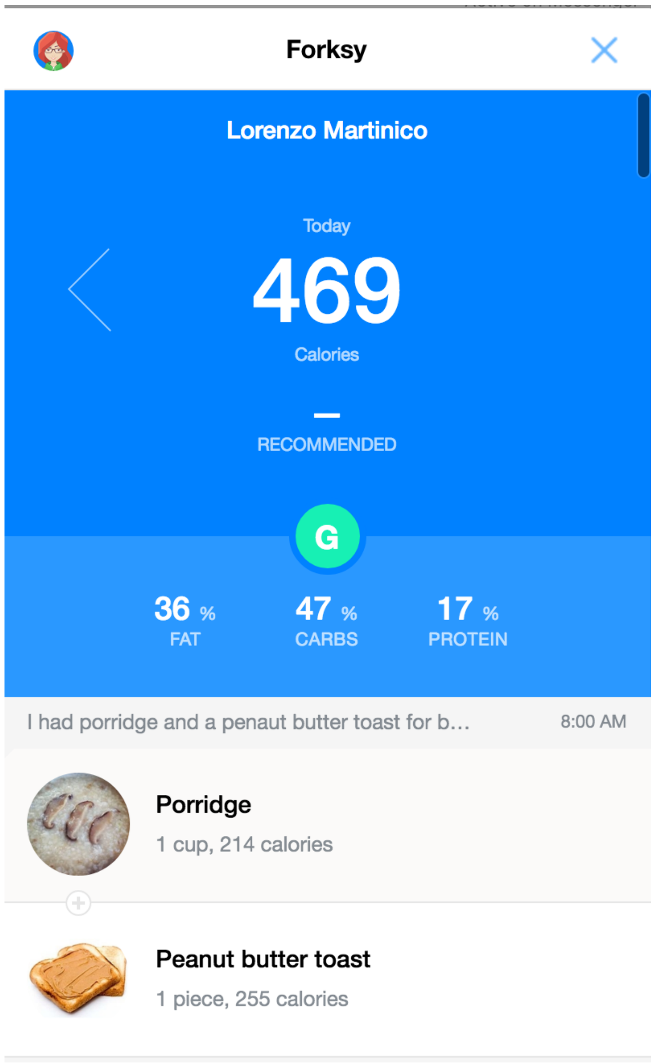
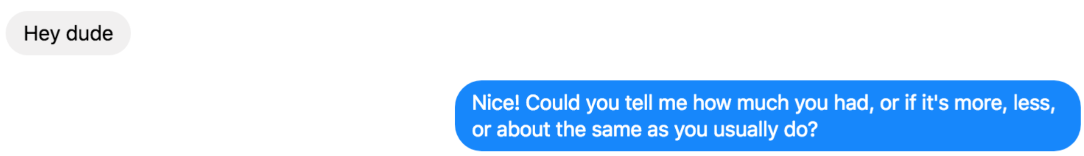


Figure 5.2: The Forksy food diary

lack of understanding of the input's context.



Some interesting feedback we received through the surveys was about users' mental model of how the chatbot worked (or rather, didn't). Users attributed undue relevance to the "seen" message icon on the Messenger application. While this would appear whenever a reply was sent, it would remain unmarked when the chatbot received a message it did not acknowledge. The lack of read receipt co-occurred in some occasions where some expected remarks were not sent, most likely due to the Heroku server not waking up in time, causing users to associate the two events. Similarly, after the erroneous reminder for lack of usage was sent overnight, participants reported the chatbot not being able to distinguish whether they were sleeping or putting off logging food. These assumptions can be powerful to leverage when designing our conversations, by exploiting users' misconceptions to cover up flaws in the dialogue system. More studies should be conducted in this area.

5.3 Future improvements

There are many directions in which a project like this could go, and while our implementation was a bit barebones and only offered the essential features, there are several things we thought of during development or received requests from users which we might have added to make for a more useful and complete nutritional assistant, but had to hold off from because of time constraints.

An immediate feature addition that was reported as being useful during evaluation, was the option of retroactively adding a meal to the previous day. We immediately added the feature, which was a simple fix of changing the date object in the MongoDB update code from statically being the current date to taking the intent's argument. We did not push the new feature however because we were concerned deploying it would affect current users.

To match our functionality with MyFitnessPal's, we should also allow easy scanning of commercial food packages through a barcode scanner. While we could enable scanning on submitted images, this could lead to negative user experience when testing with badly framed images, requiring several photographs of the food to be taken before a match would be identified. Instead, we could rely on a Facebook Messenger webview object to integrate a live barcode scanner[1], but it is uncertain how well it would perform.

People interact in different ways with modern messaging platforms, and our chatbot should be able to take advantage of these new forms of communications: emojis, stickers, gifs. There are tagsets to classify this rich content based on emotion expressed, so we could use this as a basic form of sentiment analysis for incoming message, and

[1] possibly integrating https://github.com/serratus/quaggaJS

to express personality when sending outgoing messages. We already have some hard-coded emojis as part of our script, but adding stickers and gifs to a generative model for text would make the chatbot feel more like a real entity, since the number of options is very large compared to the number of scripted responses we can write.

Some people tend to have a very regular diet, where they will eat the same meal for several days in a row. MyFitnessPal acknowledges this, and it provides an option to copy a meal from a previous day. This interaction pattern would fit naturally with a chatbot dialogue, where a user could simply say: "I had the same breakfast as yesterday", and the system should just replicate the same food as in the previous day. Of course, this would require enforcing a more strict separation between meals that we currently have, but we could guess it based on time of input, or by simply setting the meal type as a required parameter in the logging intent. Even simpler, although a bit less natural, would be asking the chatbot to set a recurring meal to be copied every day until manually interrupted.

A feature that was particularly appreciated by MyFitnessPal testers was recipe suggestions. It would be good if the chatbot could automatically suggest recipes, based on what nutritional requirements the user needs to fulfil, and maybe taking into account dietary history and preferences. In fact, with enough food data it might be possible to infer what kind of foods the user likes, and aggregating data across several users we might be able to provide a recommender system for discovering new recipes you might not know.

A nutritional assistant should also be able to give you some advice on specific food. It would be pretty simple to add an intent to query information from the *Nutritionix* API about the nutritional values of the food in question. However, the challenging aspect of such a feature would be displaying the information in a useful way. It could be possible to display nutritional content in grams for each macro and important micronutrients, but that would only benefit people who can understand what those nutritional values mean, and does not fit within the final aims of our chatbot. Perhaps contextualisation with past meals might provide useful, but we need to be able to classify what kind of food that is, which is still an unknown problem.

We should also be interested in tracking other factors beside the food eaten. Sleep, mood, level of stress and exercise are all massive factors that can affect and be affected by what we eat [46]. The chatbot should capture these factors, either by interfacing

with another Quantified Self platform, or by asking users questions at the start of the day.
The above features would mostly be relatively straightforward to implement, and we would have added them to our chatbot if we had had more development time available. We also envisioned several other possible features, which would be interesting, but would also require a lot more work to make possible.

An interesting component in diet tracking is social awareness, as discussed in the Background chapter. While the chatbot interface in itself does not provide any obvious advantage from this point of view, the network effect of having an agent that can interact with your contacts, be it your social circles in Facebook Messenger or your co-workers on Slack, can provide some interesting opportunities to leverage social pressure. The most immediate idea is having the chatbot set “challenges” to a group of users. This might take the form of a user invoking the chatbot in a group chat through the Facebook Chat Extensions [10], and setting a randomised “healthy eating” challenge (something along the lines of *eat 5 portion of vegetables every day for the next week*). The chatbot would then end ask for participants among the group chat members, and once everyone has accepted or declined the challenge, it would keep a score of how everyone is doing, publicly complimenting those who are on track and encouraging who has been lagging off to do better. To avoid polluting the group chat, whose purpose might not be just having fitness challenges, logging should happen within the regular direct message interaction with the chatbot, and shoutouts and encouragements should be limited to a small daily window.
The idea of using the chatbot has a motivator tool to set challenges against themselves could also be expanded to the context of an individual user. This would work very similarly to the group challenges, except the user would be competing against themselves, trying to surpass their previous goals or stick with a new healthy habit.

We would also like to go back to our original idea of integrating images from social networks. With Instagram being impossible to use (besides scraping public facing profiles through a crawler), Flickr seems like the best alternative. Syncing with a user’s photo collection allows us not only to save time from logging their new meals, and to capture a snapshot of historic data, but also leverage metadata that the social network makes available to us, such as image tags and the social graph of people interested in that picture. This would help strengthen the prediction for our food classification, but would also provide a basis to group related foods together.

Another helpful deception our chatbot could adapt, until an appropriate nutritional knowledge base and algorithms to understand an individual's diet can be developed, would be giving a trained nutritionist the possibility to examine data that has been submitted so far, and intervene to give appropriate advice. Facebook Messenger does expose a Handover protocol [11], designed for such a case where a human operator needs to takeover the chatbot's operations. We would also need to have some visualization tool, so that the human nutritionist might understand at a glance what the situation is and offer the necessary advice.

5.4 Chatbots and privacy

Throughout our implementation and evaluation phases, it became evident how little privacy a chatbot user can expect. Participants on our experiment, who come from a variety of academic backgrounds, mostly showed having some expectations of privacy, regardless of how much they were concerned about the confidentiality of this specific domain, and overall customers across Europe report greater concerns with how their personal information is being handled by private companies [8].

In our architecture, we had several stages where data was "leaked" to a third party: all communications between users and chatbot were completely accessible to Facebook, as the provider of the Messenger platform, and us, as well as any other potentially malicious administrators to the Facebook page the bot was linked to; all conversations were also forwarded to Google, to be analysed as part of their Dialogflow service; food eaten was stored on our MongoDB provider website; and some logs were leaked to the Heroku instance from our webhook code, which potentially could have stored a lot more information than it did. All of this data was associated with a personal identifier, or even the user's own name.
While ideally we should be able to trust these service providers with user information, revelations in the last few months have increased users' awareness, and concerns, with certain data brokerage and collection practices that are common among the tech giants, which leave data vulnerable to be exploited by malicious agents. And while the sensitivity of nutritional habits may not seem important, up until you consider how knowing about a major eating disorders might be valuable to insurance companies, using a chatbot in any other medical setting might be a cause of concern. Narrow rules in the United States govern how sensitive medical information can be dealt with (regulated

by the Health Insurance Portability and Accountability Act), and soon Europe will enforce data protection on all data generated by its own citizens (through the General Data Protection Regulation).

While most of the major messaging platforms nowadays provide an option for end-to-end encryption, especially thanks to the Signal Messaging protocol [52], which has been widely deployed through industry, no one provides facilities to encrypt chatbot conversations [35], except the Wire messaging app, which has a barebones chatbot API in alpha stage (but it is unclear whether it is actually being used beyond their initial demos [2]), and the Matrix protocol [3], which allows chatbot users within encrypted channels, but does not provide an officially documented API to develop them, which means there are about a dozen customly developed bots available.
While an API such as Wire's uses end-to-end encryption to protect messages from the platform provider, a user still has to trust the other end, whoever controls the server the chatbot runs on, not to steal their data or leak it to malicious third parties. It might be possible to mitigate this risk with a revised client-only chatbot architecture, where some basic parsing happens immediately on the client side, perhaps using historic chatbot technologies such as AIML [2], or some ondevice machine learning, which has become more feasible in the past few years [4]. This architecture would still include a server, whose only function would be storing the encrypted data, and conduct the relevant data analytics in an anonymous and confidential manner by using zero-knowledge proofs, an encryption scheme which enables a system to prove properties of a message without decrypting it. Of course, this would probably prevent more complex natural language processing which requires large scale data analysis, but maybe with proper conversation design it could be enough to maintain all important functionality, and any necessary data collection to improve on-device machine learning models might be conducted through privacy preserving techniques.

[2]https://github.com/wireapp/wire
[3]https://github.com/matrix-org

Chapter 6

Conclusion

The main goal of our project was creating an alternative interface to make it easier for health data to be collected and understood.

Our implementation had us explore the architecture of popular SaaS toolkit Dialogflow and its functionalities, working with a platform that has only been superficially documented to create a fully functioning agent. We evaluated some basic principles of chatbot dialogue design, various backend technologies to finally choose Heroku, an option that would grant us the most flexibility and ease of use; we also explored the space of various 3rd party tools, such as database technologies, nutritional information databases, image recognition for food objects and social networking APIs. While our original design included a vast breadth of features, having to deal with time limitations forced us to take into consideration what aspects of the interface would be truly fundamental, and prioritise what would take us to a minimum viable products. We highlighted some issues we run into along the way, from failing to obtain permission to use the Instagram API for our purposes, to collecting a sufficiently large dataset of food names. Particular effort was put into classifying food based on its nutritional value, but we failed to achieve even barely usable performance, due to the complexity of the problem domain.

Despite the many roadblocks, we delivered a functioning prototype, the first chatbot, to our knowledge, which combines texting and pictures as input for a dietary log. We also led the first comparative user trial between a food logging app and chatbot, with a small group of tester across a single demographic.
While our final release was still plagued by many bugs that affected usability, we were

able to provide a useful service to some users who gained some insights on their diet. We also achieved better results than MyFitnessPal with the control group in some areas, such as reminders and general pleasantness of use, and found out some interesting correlations between platform used and importance attributed to measurement units, as well as users' mental models and attitudes towards the confidentiality of nutritional data.

This leaves us with some promising expectations for future progress of the project, and we explored some possible extensions of various implementation difficulties. Although these features might improve how much users can do with our chatbot, and how it feels to interact with it, we will still have to find a definitive way to replace our ad-hoc heuristic with a fully scalable and complete knowledge base, which would require a more systematic approach in its design, as well as larger scale testing.
As widely as these areas could be explored in the next phase of our project, however, during our development phase, and from feedback we received during our evaluation, we realised that it would be impossible for a chatbot that deals with medical data without compromising user's trust. Therefore, we would like to redirect our enquiries into exploring the possibility of having a secure chatbot, either from an architectural point of view, or in relation to their compliance with the upcoming General Data Protection Regulation in Europe. We realise that seems to go against recent trends in industry, where increasing data collection to enable intelligent features has become the norm, but we recognise that chatbots can create an illusion of intimacy that might lead to sensitive information being shared more willingly, and users have a right to have their most personal data protected.

Bibliography

[1] App Fatigue Allowed Chatbots To Emerge for Customer Service Engagement. URL: https://tinyurl.com/y75t78ud [accessed on 2018-03-19].

[2] Artificial Intelligence Markup Language (AIML). URL: https://www.pandorabots.com/docs/aiml/aiml-basics.html [accessed on 2018-06-04].

[3] Bitesnap - Photo Food Journal. URL: https://getbitesnap.com/ [accessed on 2018-03-19].

[4] Building TensorFlow on Android — TensorFlow. URL: https://www.tensorflow.org/mobile/android{_}build [accessed on 2018-04-10].

[5] Calorie Mama AI - Food Image Recognition and Calorie Counter using Deep Learning. URL: http://www.caloriemama.ai/ [accessed on 2018-03-19].

[6] Clarifai food model. URL: https://www.clarifai.com/models/food-image-recognition-model-bd367be194cf45149e75f01d59f77ba7 [accessed on 2018-04-09].

[7] Cloud Functions - Event-driven Serverless Computing — Google Cloud Platform. URL: https://cloud.google.com/functions/ [accessed on 2018-03-18].

[8] ePrivacy: consultations show confidentiality of communications and the challenge of new technologies are key questions. URL: https://tinyurl.com/z4xxz7o [accessed on 2018-04-06].

[9] Express - Node.js web application framework. URL: https://expressjs.com/ [accessed on 2018-03-18].

[10] Facebook Messenger Docs - Chat Extensions. URL: https://developers.facebook.com/docs/messenger-platform/guides/chat-extensions [accessed on 2018-04-05].

[11] Facebook Messenger Docs - Handover Protocol. URL: https://developers.facebook.com/docs/messenger-platform/handover-protocol [accessed on 2018-04-09].

[12] Fit Google. URL: https://www.google.com/fit/ [accessed on 2018-03-19].

[13] Forksy a robot-nutritionist that helps increase your energy. URL: https://getforksy.com/ [accessed on 2018-03-28].

[14] Free Calorie Counter, Diet & Exercise Journal — MyFitnessPal.com. URL: https://www.myfitnesspal.com/ [accessed on 2018-03-19].

[15] Important Message Regarding MyFitnessPal Account Security. URL: https://content.myfitnesspal.com/security-information/notice.html [accessed on 2018-03-30].

[16] Messenger Platform 2017 — The Evolution of Why We Message. URL: https://messenger.fb.com/blog/a-look-back-on-messenger-platform-in-2017 [accessed on 2018-01-22].

[17] Messenger Platform at F8 — Facebook Newsroom. URL: https://newsroom.fb.com/news/2016/04/messenger-platform-at-f8/ [accessed on 2018-03-18].

[18] MobileCoach. URL: https://www.mobile-coach.eu/ [accessed on 2018-03-20].

[19] MongoDB. URL: https://www.mongodb.com/.

[20] NHS patients' data was illegally transferred to Google DeepMind. URL: https://news.sky.com/story/nhs-patient-data-given-to-google-illegally-10935315 [accessed on 2018-03-19].

[21] Nutrient Data Laboratory : USDA ARS. URL: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/

beltsville-human-nutrition-research-center/nutrient-data-laboratory/ [accessed on 2018-04-02].

[22] Personalized Nutrition Designed for Better Health & Weight Loss Habit. URL: https://habit.com/ [accessed on 2018-03-21].

[23] Platform Changelog. URL: https://www.instagram.com/developer/changelog/ [accessed on 2018-04-10].

[24] Programmable Web - List of public APIs filtererd by keyword Food. URL: https://www.programmableweb.com/category/food/apis?category=20048 [accessed on 2018-04-02].

[25] SmartPlate®. URL: https://www.getsmartplate.com/ [accessed on 2018-04-02].

[26] The App Garden. URL: https://www.flickr.com/services/apps/about/ [accessed on 2018-04-10].

[27] TypeScript - JavaScript that scales. URL: http://www.typescriptlang.org/ [accessed on 2018-04-02].

[28] Most popular messaging apps 2018 — Statista. URL: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ [accessed on 2018-03-19].

[29] Social media: worldwide penetration rate 2018 — Statistic. URL: https://www.statista.com/statistics/269615/social-network-penetration-by-region/ [accessed on 2018-03-22].

[30] Open Food Facts, 2015. URL: https://world.openfoodfacts.org/ [accessed on 2017-11-08].

[31] Create Convincing Computer Conversations. 2016. URL: https://www.pullstring.com/conversational-writing.

[32] Digital in 2018. Technical report, We Are Social, 2018. URL: https://www.slideshare.net/slideshow/embed{_}code/key/jALQjcxPj3k7N.

[33] Tim Adams. Facebook's week of shame: the Cambridge Analytica fallout — Technology — The Guardian, mar 2018. URL:

https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence.

[34] Ziad Ahmad, Marc Bosch, Nitin Khanna, Deborah A. Kerr, Carol J. Boushey, Fengqing Zhu, and Edward J. Delp. A Mobile Food Record For Integrated Dietary Assessment. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management - MADiMa '16*, pages 53–62, 2016. URL: http://dl.acm.org/citation.cfm?doid=2986035.2986038, doi:10.1145/2986035.2986038.

[35] A Alesanco, J Sancho, Y Gilaberte, E Abarca, and J Garc. EMBEC & NBC 2017. 65:185–188, 2018. URL: http://link.springer.com/10.1007/978-981-10-5122-7, doi:10.1007/978-981-10-5122-7.

[36] Emily Alonso. How This Winner-Took-All by Helping People Lose. 2015.

[37] Jaan Altosaar. Food2Vec. URL: https://github.com/altosaar/food2vec [accessed on 2018-04-05].

[38] S. A. Bingham, C. Gill, A. Welch, K. Day, A. Cassidy, K. T. Khaw, M. J. Sneyd, T. J. A. Key, L. Roe, and N. E. Day. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *British Journal of Nutrition*, 72(04):619, oct 1994. URL: http://www.journals.cambridge.org/abstract{_}S0007114594001650, doi:10.1079/BJN19940064.

[39] Kristie Bishopp. The Recommended Serving Sizes for Raw Leafy Vegetable, 2017. URL: https://www.livestrong.com/article/545636-the-recommended-serving-sizes-for-raw-leafy-vegetables/ [accessed on 2018-03-22].

[40] C. J. Boushey, M. Spoden, F. M. Zhu, E. J. Delp, and D. A. Kerr. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. *Proceedings of the Nutrition Society*, (December 2016):1–12, 2016. URL: http://www.journals.cambridge.org/abstract{_}S0029665116002913, doi:10.1017/S0029665116002913.

[41] Petter Bae Brandtzaeg and Asbjørn Følstad. Why People Use Chatbots. In Ioannis Kompatsiaris, Jonathan Cave, Anna Satsiou, Georg Carle, Antonella Passani, Efstratios Kontopoulos, Sotiris Diplaris, and Donald McMillan, edi-

tors, *Internet Science*, pages 377–392, Cham, 2017. Springer International Publishing.

[42] Petter Bae Brandtzaeg, Ida Maria Haugstveit, Marika Lüders, and Asbjørn Følstad. How should organizations adapt to youth civic engagement in social media? A lead user approach. *Interacting with Computers*, 28(5):664–679, 2016.

[43] Petter Bae Brandtzæg, Jan Heim, and Amela Karahasanović. Understanding the new digital divide - A typology of Internet users in Europe. *International Journal of Human Computer Studies*, 69(3):123–138, 2011. doi:10.1016/j.ijhcs.2010.11.004.

[44] Jacqueline Brixey, Rens Hoegen, Wei Lan, Joshua Rusow, Karan Singla, Xusen Yin, Ron Artstein, and Anton Leuski. SHIHbot : A Facebook chatbot for Sexual Health Information on HIV / AIDS. *Proceedings of 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, (August):370–373, 2017.

[45] John Brownlee. No What Is Zero UI? (And Why Is It Crucial To The Future Of Design?), 2015. URL: `https://www.fastcodesign.com/3048139/what-is-zero-ui-and-why-is-it-crucial-to-the-future-of-design` [accessed on 2018-04-08].

[46] Matthew P Buman and Shawn D Youngstedt. Physical Activity, Sleep, and Biobehavioral Synergies for Health. In *Sleep and Affect*, pages 321–337. Elsevier, 2015.

[47] Tim Byers. Food frequency dietary assessment: how bad is good enough? *American Journal of Epidemiology*, 154(12):1087–1088, 2001.

[48] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O ' Neill, Cherie Armour, and Michael Mctear. Towards a chatbot for digital counselling. *Mental Health Artificial Intelligence*, pages 1–7. URL: `http://hci2017.bcs.org/wp-content/uploads/BHCI{\_}2017{\_}paper{\_}110.pdf`.

[49] Michelle Clare Carter, Victoria Jane Burley, Camilla Nykjaer, and Janet Elizabeth Cade. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. *Journal of medical Internet research*, 15(4), 2013.

[50] Yu Chen and Yu Chen. Exploring social accountability for pervasive fitness apps Exploring Social Accountability for Pervasive Fitness Apps. (January), 2014.

[51] Yuchuan Chen, Chien Yeh Hsu, Li Liu, and Sherry Yang. Constructing a nutrition diagnosis expert system. *Expert Systems with Applications*, 39(2):2132–2156, 2012. URL: http://dx.doi.org/10.1016/j.eswa.2011.07.069, doi:10.1016/j.eswa.2011.07.069.

[52] Katriel Cohn-Gordon, Cas Cremers, Benjamin Dowling, Luke Garratt, and Douglas Stebila. A Formal Security Analysis of the Signal Messaging Protocol. *Proceedings - 2nd IEEE European Symposium on Security and Privacy, EuroS and P 2017*, (November):451–466, 2017. doi:10.1109/EuroSP.2017.27.

[53] Kate Crawford, Jessa Lingel, and Tero Karppi. Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies*, 18(4-5):479–496, 2015. doi:10.1177/1367549415584857.

[54] Robert Dale. The return of the chatbots. *Natural Language Engineering*, 22(5):811–817, 2016. arXiv:0910.0834, doi:10.1017/S1351324916000243.

[55] Siena F Davis, Marisa A Ellsworth, Hannah E Payne, Shelby M Hall, Joshua H West, and Amber L Nordhagen. Health Behavior Theory in Popular Calorie Counting Apps: A Content Analysis. *JMIR mHealth and uHealth*, 4(1):e19, mar 2016. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4795330/, doi:10.2196/mhealth.4177.

[56] C Ding and X He. K-means clustering via principal component analysis. *Proceedings of the twenty-first international conference on Machine learning*, Cl(2000):29, 2004. URL: papers://b6c7d293-c492-48a4-91d5-8fae456be1fa/Paper/p2738{\%}5Cnfile:///C:/Users/Serguei/OneDrive/Documents/Papers/K-meansclusteringviaprincipalcomponent.pdf, arXiv:arXiv:0711.0189v1, doi:10.1145/1015330.1015408.

[57] Tome Eftimov, Peter Korošec, and Barbara Koroušić Seljak. Standfood: Standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. *Nutrients*, 9(6), 2017. doi:10.3390/nu9060542.

[58] Bobbie Eicher, Lalith Polepeddi, and Ashok Goel. Jill Watson Doesn ' t Care if You ' re Pregnant : Grounding AI Ethics in Empirical Studies. 2016.

[59] Danielle Elmasri and Anthony Maeder. Brain Informatics. 7670:243–251, 2012. URL: `http://link.springer.com/10.1007/978-3-642-35139-6`, doi:10.1007/978-3-642-35139-6.

[60] Asbjørn Følstad and Petter Bae Brandtzæg. Chatbots and the new world of HCI. *Interactions*, 24(4):38–42, 2017. URL: `http://dl.acm.org/citation.cfm?doid=3115390.3085558`, doi:10.1145/3085558.

[61] Silvia Gabrielli, F B K Create-net, and Silvia Gabrielli. Addressing Challenges in Promoting Healthy Lifestyles : The AI-Chatbot Approach Addressing Challenges in Promoting Healthy Lifestyles : The AI-Chatbot Approach. (July), 2017.

[62] H. Green. Should foods or nutrients be the focus of guidelines to promote healthful eating? *Nutrition Bulletin*, 40(4):296–302, 2015. doi:10.1111/nbu.12175.

[63] Eric Gregori. Evaluation of Modern Tools for an OMSCS Advisor Chatbot. *School of Computer Science Graduate Student Publications, Georgia Tech*, 2017. URL: `https://smartech.gatech.edu/bitstream/handle/1853/58516/evaluation{\_}of{\_}modern{\_}tools{\_}for{\_}an{\_}omscs{\_}advisor{\_}chatbot{\%}281{\%}29.pdf?sequence=1{\&}isAllowed=y`.

[64] Brett K. Hayes, Hendy Kurniawan, and Ben R. Newell. Rich in vitamin C or just a convenient snack? Multiple-category reasoning with cross-classified foods. *Memory and Cognition*, 39(1):92–106, 2011. doi:10.3758/s13421-010-0022-7.

[65] L. Hebden, A. Cook, H. P. van der Ploeg, L. King, A. Bauman, and M. Allman-Farinelli. A mobile health intervention for weight management among young adults: A pilot randomised controlled trial. *Journal of Human Nutrition and Dietetics*, 27(4):322–332, 2014. doi:10.1111/jhn.12155.

[66] J A Higgins, A L LaSalle, P Zhaoxing, M Y Kasten, K N Bing, S E Ridzon, and T L Witten. Validation of photographic food records in children: are pictures really worth a thousand words? *European Journal Of Clinical Nutrition*, 63:1025,

mar 2009. URL: http://dx.doi.org/10.1038/ejcn.2009.12http://10.0.4.14/ejcn.2009.12.

[67] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Lu, and Rama Akkiraju. Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. 2018. URL: http://arxiv.org/abs/1803.02952, arXiv:1803.02952, doi:arXiv:1803.02952v1.

[68] Javier Couto. Building a Chatbot: analysis & limitations of modern platforms - Tryolabs Blog, 2017. URL: https://tryolabs.com/blog/2017/01/25/building-a-chatbot-analysis--limitations-of-modern-platforms/ [accessed on 2017-10-29].

[69] Lele Kang, Chuan Hoo Tan, and Leon J. Zhao. The Impact of Intra-Transaction Communication on Customer Purchase Behaviour in E-Commerce Context. In *24th Australasian Conference on Information Systems*, Melbourne, 2013.

[70] Seunghyeon Kim, Jaeyun Sung, Mathias Foo, Yong Su Jin, and Pan Jun Kim. Uncovering the nutritional landscape of food. *PLoS ONE*, 10(3):1–17, 2015. doi:10.1371/journal.pone.0118697.

[71] Jenny Stamos Kovacs. The Dos and Don'ts of Counting Calories, 2009. URL: https://www.webmd.com/diet/features/dos-donts-counting-calories{\#}1.

[72] Tobias Kowatsch, Dirk Volland, Iris Shih, R Dominik, K Florian, Filipe Barata, Andreas Filler, B Dirk, Katrin Heldt, and Pauline Gindrat. Designing the Digital Transformation. 10243:485–489, 2017. URL: http://link.springer.com/10.1007/978-3-319-59144-5, doi:10.1007/978-3-319-59144-5.

[73] Hans-Peter Kriegel, Peer Kröger, and Arthur Zimek. Clustering high-dimensional data. *ACM Transactions on Knowledge Discovery from Data*, 3(1):1–58, 2009. URL: http://portal.acm.org/citation.cfm?doid=1497577.1497578, doi:10.1145/1497577.1497578.

[74] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. In *International Conference on Smart Homes and Health Telematics*, pages 37–48. Springer International Publishing, 2016.

[75] Kim Ly, Min Zhao, and Dilip Soman. A Practitioner's Guide To Nudging. Technical report, Rotman School of Management, University of Toronto, 2013.

[76] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'Neill. How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. In Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler, editors, *Human-Computer Interaction - INTERACT 2017*, pages 441–459, Cham, 2017. Springer International Publishing.

[77] William Meisel. Specialized Digital Assistants and Bots Vendor Guide and Market Studies. Technical report, TMA Associates, 2016.

[78] Aniello Minutoloa, Massimo Espositoa, and Giuseppe De Pietroa. A Conversational Chatbot Based on Kowledge-Graphs for Factoid Medical Questions. In *New Trends in Intelligent Software Methodologies, Tools and Techniques: Proceedings of the 16th International Conference SoMeT_17*, volume 297, page 139. IOS Press, 2017.

[79] Anders Møller. LanguaL Home Page. dec 2007. URL: http://www.langual.org/.

[80] Susan V Morris. How many lost customers have you won back today? An aggressive approach to complaint handling in the hotel industry. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 1(1):86–92, 1988.

[81] Patel MS, Asch DA, Rosin R, and Et al. Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial. *Annals of Internal Medicine*, 164(6):385–394, 2016. URL: +http://dx.doi.org/10.7326/M15-1635, doi:10.7326/M15-1635.

[82] Patel MS, Benjamin EJ, Volpp KG, and Et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: The be fit randomized clinical trial. *JAMA Internal Medicine*, 177(11):1586–1593, 2017. URL: +http://dx.doi.org/10.1001/jamainternmed.2017.3458, doi:10.1001/jamainternmed.2017.3458.

[83] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang,

and Kevin Murphy. Im2Calories: Towards an automated mobile vision food diary. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter(December):1233–1241, 2015. doi:10.1109/ICCV.2015.146.

[84] D Napoleon. A New Method for Dimensionality Reduction using K- Means Clustering Algorithm for High Dimensional Data Set. 13(7):41–46, 2011.

[85] Australian Bureau of Statistics. National Nutrition Survey Selected Highlights Australia, 1995.

[86] Mitesh S. Patel, Luca Foschini, Gregory W. Kurtzman, Jingsan Zhu, Wenli Wang, Charles A.L. Rareshide, and Susan M. Zbikowski. Using wearable devices and smartphones to track physical activity: Initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. *Annals of Internal Medicine*, 167(10):755–757, 2017. doi:10.7326/M17-1495.

[87] Patrick, K., Griswold, G. W., Raab, and F. Health and the mobile phone. *American Journal of Preventive Medicine*, 35(2):177–181, 2008. URL: `http://apps.isiknowledge.com/full{\_}record.do?product=CABI{\&}search{\_}mode=Refine{\&}qid=49{\&}SID=3CEcJmNAPbH4nkMjbpO{\&}page=3{\&}doc=107`, doi:10.1016/j.amepre.2008.05.001.Health.

[88] Calie Pistorius. Developments in emerging digital health technologies. (1), 2017.

[89] James O. Prochaska, Wayne F. Velicer, Colleen Redding, Joseph S. Rossi, Michael Goldstein, Judith DePue, Geoffrey W. Greene, Susan R. Rossi, Xiaowu Sun, Joseph L. Fava, Robert Laforge, William Rakowski, and Brett A. Plummer. Stage-based expert systems to guide a population of primary care patients to quit smoking, eat healthier, prevent skin cancer, and receive regular mammograms. *Preventive Medicine*, 41(2):406–416, 2005. doi:10.1016/j.ypmed.2004.09.050.

[90] Amon Rapp and Federica Cena. Self-monitoring and Technology: Challenges and Open Issues in Personal Informatics. In Constantine Stephanidis and Margherita Antona, editors, *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice*, pages 613–622, Cham, 2014. Springer International Publishing.

[91] Jaclyn Rich, Hamed Haddadi, and Timothy M Hospedales. Towards Bottom-Up Analysis of Social Food Food in Social Media. (July 2015), 2016.

[92] Ellen Henrietta (Swallow) Richards and Louise Harding. Williams. The dietary computer. Explanatory pamphlet; the pamphlet containing tables of food composition, lists of prices, weights, and measures, selected recipes for the slips, directions for using the same., 1902.

[93] Verónica Rivera-Pelayo, Valentin Zacharias, Lars Müller, and Simone Braun. Applying quantified self approaches to support reflective learning. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, page 111, 2012. URL: http://dl.acm.org/citation.cfm?doid=2330601.2330631, doi:10.1145/2330601.2330631.

[94] Iulian V. Serban, Chinnadhurai Sankar, Mathieu Germain, Saizheng Zhang, Zhouhan Lin, Sandeep Subramanian, Taesup Kim, Michael Pieper, Sarath Chandar, Nan Rosemary Ke, Sai Rajeshwar, Alexandre de Brebisson, Jose M. R. Sotelo, Dendi Suhubdy, Vincent Michalski, Alexandre Nguyen, Joelle Pineau, and Yoshua Bengio. A Deep Reinforcement Learning Chatbot. (Nips):1–9, 2017. URL: http://arxiv.org/abs/1709.02349, arXiv:1709.02349.

[95] Sanket S Sharma and Munmun De Choudhury. Measuring and Characterizing Nutritional Information of Food and Ingestion Content in Instagram. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 115–116, New York, NY, USA, 2015. ACM. URL: http://doi.acm.org/10.1145/2740908.2742754, doi:10.1145/2740908.2742754.

[96] Amalia Suzianti, Rizky Puti Minanga, and Felisa Fitriani. Analysis of User Experience (UX) on Health-Tracker Mobile Apps. *International Journal of Computer Theory and Engineering*, 9(4):262–267, 2017. URL: http://www.ijcte.org/index.php?m=content{\&}c=index{\&}a=show{\&}catid=87{\&}id=1402, doi:10.7763/IJCTE.2017.V9.1148.

[97] Latanya Sweeney. Computational disclosure control: A Primer on Data Privacy Protection. *Computer Science*, Ph.D.:216, 2001.

[98] Adam Tanner. For Sale: Your Medical Records. *Scientific American*, 314(2):26–27, jan 2016. URL: http://www.nature.com/doifinder/10.1038/scientificamerican0216-26, doi:10.1038/scientificamerican0216-26.

[99] Ida Torres. Samsung testing out Bixby abiity to easily track calories - Android Community. URL: https://androidcommunity.com/samsung-testing-out-bixby-abiity-to-easily-track-calories-20180111/ [accessed on 2018-03-22].

[100] Gabrielle M. Turner-McGrievy, Michael W. Beets, Justin B. Moore, Andrew T. Kaczynski, Daheia J. Barr-Anderson, and Deborah F. Tate. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. *Journal of the American Medical Informatics Association*, 20(3):513–518, 2013. doi:10.1136/amiajnl-2012-001510.

[101] USDA Food and Nutrition Service. Implications of Restricting the Use of Food Stamp Benefits: Summary. 2007. URL: http://www.fns.usda.gov/sites/default/files/arra/FSPFoodRestrictions.pdf.

[102] Joseph Weizenbaum. ELIZAa computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[103] Henry H. Wilmer, Lauren E. Sherman, and Jason M. Chein. Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in Psychology*, 8(APR):1–16, 2017. doi:10.3389/fpsyg.2017.00605.

[104] Ryan Witt. Healthcare data: A hacker's jackpot — Healthcare IT News. URL: http://www.healthcareitnews.com/blog/healthcare-data-hackers-jackpot [accessed on 2018-03-19].

[105] Maria Wolters. "Missing " health data. URL: http://mariawolters.net/research-overview/missing-health-data/ [accessed on 2018-04-05].

Appendix A

Feedback forms

My Fitness Pal experiment feedback

This survey will be anonymous, so please try to answer to all questions in full and as honestly as you can.

1. **How old are you?**

2. **How much do you weight in Kilograms (estimate if unsure)?**

3. **What is your height in cm (estimate if unsure)?**

4. **Do you know what your BMI is? If so, write it down**

5. **How proficients are you with computers?**
   Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Expert user

6. **Are you on a diet, or have you attempted a diet in the last two years?**
   Mark only one oval.
   - ◯ Yes *Skip to question 8.*
   - ◯ No *Skip to question 7.*

Skip to question 10.

7. **Please explain why you haven't considered a diet in the last two years**

Skip to question 10.

Untitled Section

8. **Please briefly describe your diet and your motivations for dieting**

9. **If you are not currently dieting, why did you stop?**

Before the app

10. **How many meals do you usually have in a day (excluding snacks)?**
    Mark only one oval.
    - ◯ 1
    - ◯ 2
    - ◯ 3
    - ◯ 4
    - ◯ I don't have the same number of meals every day

11. **How often do you have snacks outside of meals?**
    Mark only one oval.
    - ◯ Very rarely
    - ◯ Once per week
    - ◯ Several times per week
    - ◯ Daily
    - ◯ More than once per day

12. **Have you ever tracked your diet before?**
    Mark only one oval.
    - ◯ Yes *Skip to question 13.*
    - ◯ No *Skip to question 21.*

Skip to question 15.

13. **Describe how you track or tracked your diet**

14. **Do you also track your snacks?**
    Mark only one oval.
    - ◯ Yes
    - ◯ No

Untitled Section

15. **What do you think is a "balanced" diet?**

16. **Do you believe your diet to be balanced?**
    Mark only one oval.
    - ◯ Yes
    - ◯ No
    - ◯ Other:

17. **Please estimate how many portions of fruits and vegetables you consume in a day on average**
    Mark only one oval.

1	2	3	4	5	6	7
◯	◯	◯	◯	◯	◯	◯

18. **Do you mostly cook your own meal, buy ready-made food, or eat out?**
    Mark only one oval.
    - ◯ I mostly cook my own meal
    - ◯ I mostly buy ready-made food
    - ◯ I mostly eat out
    - ◯ I do all three in equal measure

19. **How far in advance do you plan your meals?**
    Mark only one oval.
    - ◯ I don't need planning - I always eat the same thing
    - ◯ I don't plan at all, I just eat what I want at the moment
    - ◯ I plan all my meals for the day in the morning
    - ◯ I plan the next several days of meals when I am grocery shopping
    - ◯ I plan all the meals I will have in the next week/two weeks far in advance
    - ◯ Other:

20. **How often are your meal plans disrupted by an unexpected event (a missing ingredient, being invited to eat out)?**
    Mark only one oval.
    - ◯ Daily
    - ◯ A few times a week
    - ◯ A few times a month
    - ◯ Very rarely

Using the app

21. **On a scale from 1 to 5, how useful did you find the app?**
    Mark only one oval.

	1	2	3	4	5	
Not useful at all	◯	◯	◯	◯	◯	Extremely useful

22. **On a scale from 1 to 5, how pleasant did you find to use the app?**
    Mark only one oval.

	1	2	3	4	5	
Extremely annoying	◯	◯	◯	◯	◯	Extremely pleasant

23. **What did you think of the way the app visualized your information?**
    Check all that apply.

	Useful	Useless	Plesant	Unplesant	Too detailed	Not enough detail	Clear	Unclear
The Diary	☐	☐	☐	☐	☐	☐	☐	☐
Calorie Breakdown (Macronutrients)	☐	☐	☐	☐	☐	☐	☐	☐
Calories Left ticker	☐	☐	☐	☐	☐	☐	☐	☐

24. **Estimate how many of your meals during your trial you didn't log through the app**

25. **If you didn't log a meal, what was the reason?**
Check all that apply.

- [ ] I forgot
- [ ] I found typing/scanning the barcode for my meal cumbersome
- [ ] I stopped being interested in diet tracking
- [ ] I had a technical issue
- [ ] I didn't find the feedback I was getting useful
- [ ] I felt intimidated by watching the numbers in my tracker
- [ ] I was too busy
- [ ] Other: ____________

Untitled Section

26. **Was there a day where you didn't use the app at all?**
Mark only one oval.

- ( ) Yes - one day *After the last question in this section, skip to question 28.*
- ( ) Yes - several days *After the last question in this section, skip to question 28.*
- ( ) No *After the last question in this section, skip to question 31.*

27. **My Fitness Pal provides a reminder functionality; did you turn it on?**
Mark only one oval.

- ( ) Yes *Skip to question 28.*
- ( ) No *Skip to question 30.*
- ( ) It was already turned on *Skip to question 28.*

Skip to question 31.

Untitled Section

28. **Did you get a reminder to log your food?**
Mark only one oval.

- ( ) Yes *Skip to question 29.*
- ( ) No *Skip to question 30.*

Skip to question 31.

29. **Did getting a reminder convince you to use the app that day?**
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) I was going to use it anyway

Skip to question 31.

Untitled Section

30. **If the app had reminded you to use it, do you think it would have convinced you to use it?**
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) I was going to use it anyway

Untitled Section

31. **Which input method did you prefer?**
Mark only one oval.

- ( ) Texting
- ( ) Scanning the barcode (when possible)
- ( ) Either
- ( ) Depends on the food

32. **Did you have any issues using text logging?**
Check all that apply.

- [ ] It wasn't accurate enough
- [ ] It wasn't fast enough
- [ ] It couldn't find the exact brand / food I was having
- [ ] Other: ____________

33. **Did you have any issues using the barcode scanner?**
Check all that apply.

- [ ] It wasn't accurate enough
- [ ] It didn't understand what the barcode was
- [ ] It couldn't find the food in its database
- [ ] The camera didn't work
- [ ] Other: ____________

34. **How accurate do you think your portion /size estimates were?**

35. **Would you prefer recorded meal sizes to be absolute values (number of portions, weight) or relative values (based on previous meals, more, less or the same as usual)?**
Mark only one oval.

- ( ) Relative Values
- ( ) Absolute values

36. **Why?**

37. **Did you check back on what you had eaten on the day or on a previous day?**
Mark only one oval.

- ( ) Yes *Skip to question 40.*
- ( ) No

Skip to question 40.

Untitled Section

38. **When you checked back on your meal, did you find the information...**
Check all that apply.

- [ ] Accurate
- [ ] Useful
- [ ] Incomplete
- [ ] Unhelpful
- [ ] Badly formatted

39. **Having used the app, do you think you have a better idea of how much you are eating in a day?**
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) Other: ____________

Untitled Section

40. **Overall, did you find the app helps you with to your diet?**
Mark only one oval.

- ( ) Yes
- ( ) No

41. **Please explain**

42. **Has using the app made you more "mindful" of what you eat?**
Mark only one oval.

- ( ) Yes
- ( ) No

43. **What did you expect when you started to use the app?**

44. **Did using the app conform to your expectation? If not, why?**

45. **Do you trust the app to store your information securely? Do you think it's important?**

46. **Would you use the app past this evaluation period? Why, or why not?**

47. **Did you think any important features you would have liked were missing?**

48. **Did you use any of the fitness tracking features? Did you find them a useful complement to the diet tracker? Please describe your experience**

49. **Is there anything else that could be done to improve the experience of using the app?**

Chatbot feedback

This survey will be anonymous, so please try to answer to all questions in full and as honestly as you can. If possible, go back and check on your interaction with the chatbot.

1. **How old are you?**

2. **How much do you weight in Kilograms (estimate if unsure)?**

3. **What is your height in centimeters (estimate if unsure)?**

4. **Do you know what your BMI is? If so, write it down**

5. **How proficients are you with computers?**
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Expert user

6. **Are you on a diet, or have you attempted a diet in the last two years?**
Mark only one oval.
- ◯ Yes *Skip to question 8.*
- ◯ No *Skip to question 7.*

Skip to question 10.

7. **Please explain why you haven't considered a diet in the last two years**

Skip to question 10.

8. **Please briefly describe your diet and your motivations for dieting**

9. **If you are not currently dieting, why did you stop?**

Before the chatbot

10. **How many meals do you usually have in a day (excluding snacks)?**
Mark only one oval.
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ I don't have the same number of meals every day

11. **How often do you have snacks outside of meals?**
Mark only one oval.
- ◯ Very rarely
- ◯ Once per week
- ◯ Several times per week
- ◯ Daily
- ◯ More than once per day

12. **Have you ever tracked your diet before?**
Mark only one oval.
- ◯ Yes *Skip to question 13.*
- ◯ No *Skip to question 21.*

Skip to question 15.

13. **Describe how you track or tracked your diet**

14. **Do you also track your snacks?**
Mark only one oval.
- ◯ Yes
- ◯ No

Untitled Section

15. **What do you think is a "balanced" diet?**

16. **Do you believe your diet to be balanced?**
Mark only one oval.
- ◯ Yes
- ◯ No
- ◯ Other:

17. **Please estimate how many portions of fruits and vegetables you consume in a day on average**
Mark only one oval.

1	2	3	4	5	6	7
◯	◯	◯	◯	◯	◯	◯

18. **Do you mostly cook your own meal, buy ready-made food, or eat out?**
Mark only one oval.
- ◯ I mostly cook my own meal
- ◯ I mostly buy ready-made food
- ◯ I mostly eat out
- ◯ I do all three in equal measure

19. **How far in advance do you plan your meals?**
Mark only one oval.
- ◯ I don't need planning - I always eat the same thing
- ◯ I don't plan at all, I just eat what I want in the moment
- ◯ I plan all my meals for the day in the morning
- ◯ I plan the next several days of meals when I am grocery shopping
- ◯ I plan all the meals I will have in the next week/two weeks far in advance
- ◯ Other:

20. **How often are your meal plans disrupted by an unexpected event (a missing ingredient, being invited to eat out)?**
Mark only one oval.
- ◯ Daily
- ◯ A few times a week
- ◯ A few times a month
- ◯ Very rarely

Using the chatbot

21. **On a scale from 1 to 5, how useful did you find the chatbot?**
Mark only one oval.

	1	2	3	4	5	
Not useful at all	◯	◯	◯	◯	◯	Extremely useful

22. **On a scale from 1 to 5, how pleasant did you find to use the chatbot?**
Mark only one oval.

	1	2	3	4	5	
Extremely annoying	◯	◯	◯	◯	◯	Extremely pleasant

23. **What did you think of the chatbot's feedback on your diet?**
Check all that apply.
- ☐ I didn't get any feedback
- ☐ I found the feedback useful
- ☐ I found the feedback useless
- ☐ I didn't care about the kind of feedback I was getting from the chatbot
- ☐ I would have liked more feedback
- ☐ Option 6
- ☐ Other:

24. **Estimate how many of your meals during your trial you didn't log through the chatbot**

25. **If you didn't log a meal, what was the reason?**
Check all that apply.

- [ ] I forgot
- [ ] I found typing/photographing my meal cumbersome
- [ ] I stopped being interested in diet tracking
- [ ] I had a technical issue
- [ ] I didn't find the feedback I was getting useful
- [ ] I felt judged by the chatbot
- [ ] I was too busy
- [ ] Other: ______________________

26. **Was there a day where you didn't use the chatbot at all?**
Mark only one oval.

- ◯ Yes - one day *Skip to question 27.*
- ◯ Yes - several days *Skip to question 27.*
- ◯ No *Skip to question 31.*

Skip to question 31.

27. **Did you get a reminder the next day from the chatbot to log your food?**
Mark only one oval.

- ◯ Yes *Skip to question 28.*
- ◯ No *Skip to question 30.*

Skip to question 31.

28. **Did getting a reminder convince you to use the chatbot that day?**
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ I was going to use the chatbot anyway

29. **What did you think of the reminder(s) from the chatbot?**
Check all that apply.

- [ ] Useful
- [ ] Annoying
- [ ] Out of place
- [ ] Incorrect
- [ ] Other: ______________________

Skip to question 31.

Untitled Section

30. **If the chatbot had reminded you to use it, do you think it would have convinced you?**
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Other: ______________________

Interactive with the chatbot

31. **Which input method did you prefer?**
Mark only one oval.

- ◯ Texting
- ◯ Taking pictures
- ◯ Either
- ◯ Depends on the food

32. **Did you have any issues using text logging?**
Check all that apply.

- [ ] It wasn't accurate enough
- [ ] It didn't understand what I was saying
- [ ] I found the prompts for portions annoying
- [ ] Other: ______________________

33. **Did you have any issues using picture logging?**
Check all that apply.

- [ ] It wasn't accurate enough
- [ ] It didn't understand what the photographed food was
- [ ] I found the multiple choice options annoying
- [ ] Other: ______________________

34. **Do you prefer giving relative estimates of your food logging (more, less, same as usual) or precise estimates (2 cups, 500 grams?) Why?**

35. **Did you check back on what you had eaten on the day or on a previous day?**
Mark only one oval.

- ◯ Yes *Skip to question 38.*
- ◯ No

Skip to question 38.

36. **When you checked back on your meal, did you find the information...**
Check all that apply.

- [ ] Accurate
- [ ] Useful
- [ ] Incomplete
- [ ] Unhelpful
- [ ] Badly formatted

37. **Having used the chatbot, do you think you have a better idea of how much you are eating in a day?**
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Other: ______________________

Untitled Section

38. **Overall, did you find the chatbot helped you to eat well?**
Mark only one oval.

- ◯ Yes
- ◯ No

39. **Please explain**

40. **Has using the chatbot made you more "mindful" of what you eat?**
Mark only one oval.

- ◯ Yes
- ◯ No

41. **What did you expect when you started to use the chatbot?**

42. **Did using the chatbot conform to your expectation? If not, why?**

43. **Do you trust the chatbot to store your information securely? Do you think it's important?**

44. **Would you use the chatbot past this evaluation period? Why, or why not?**

45. **Did you think any important features you would have liked were missing?**

46. **Is there anything else that could be done to improve the experience of using the chatbot?**

Appendix B

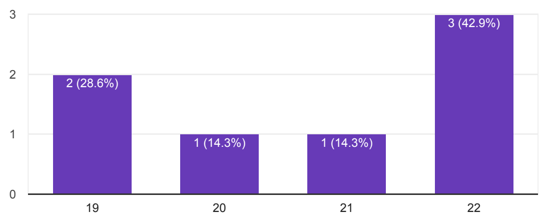
Aggregated responses

My Fitness Pal experiment feedback

7 responses

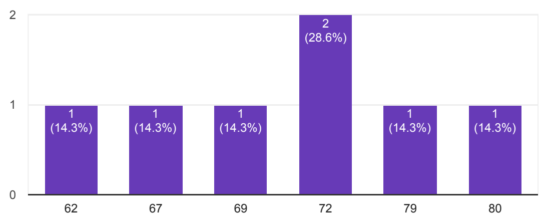
How old are you?

7 responses



How much do you weight in Kilograms (estimate if unsure)?

7 responses



What is your height in cm (estimate if unsure)?

7 responses

170

176

168

174

183

161

185

Do you know what your BMI is? If so, write it down

3 responses

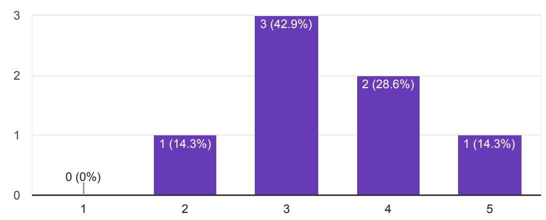
no

22.3

21.4

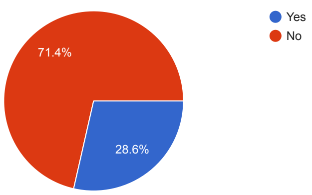
How proficients are you with computers?

7 responses



Are you on a diet, or have you attempted a diet in the last two years?

7 responses



Please explain why you haven't considered a diet in the last two years

5 responses

They don't work!

Dont feel dieting is useful: more a fan of increasing exercise as a means of slimming

I was happy with my weight

I think I am in a good shape and do not need a diet

I'm healthy and I don't think I need it

Untitled Section

Please briefly describe your diet and your motivations for dieting

2 responses

Trying to snack less and plan my meals. I also try to do more sport

Vegetarian: environmental concern (meat industry == really bad)

If you are not currently dieting, why did you stop?

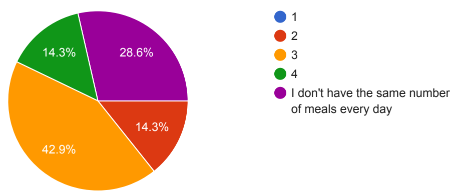
0 responses

No responses yet for this question.

Before the app

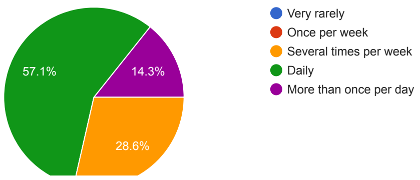
How many meals do you usually have in a day (excluding snacks)?

7 responses



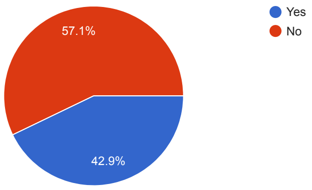
How often do you have snacks outside of meals?

7 responses



Have you ever tracked your diet before?

7 responses



Describe how you track or tracked your diet

3 responses

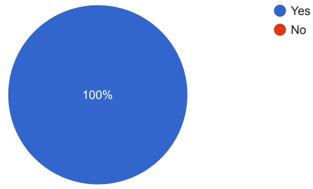
Using the myfitnesspal app.

i had a food journal

plan all meals

Do you also track your snacks?

3 responses



Untitled Section

What do you think is a "balanced" diet?

3 responses

The pie chart thing of proportions of what to eat every meal.

eat something from everything and not to leave one food out completely

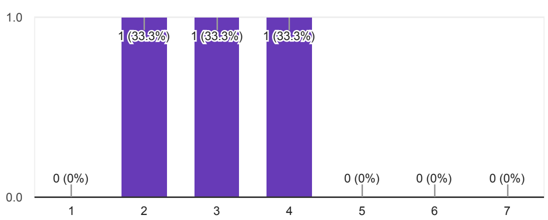
the right amount of protein, carbohydrates, fats and vitamins

Do you believe your diet to be balanced?

3 responses

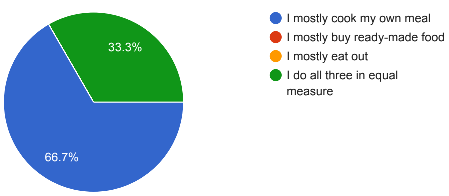
Please estimate how many portions of fruits and vegetables you consume in a day on average

3 responses



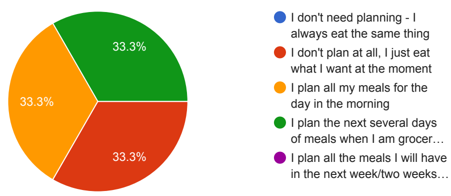
Do you mostly cook your own meal, buy ready-made food, or eat out?

3 responses



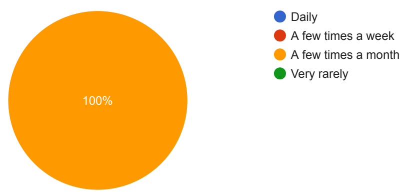
How far in advance do you plan your meals?

3 responses



How often are your meal plans disrupted by an unexpected event (a missing ingredient, being invited to eat out)?

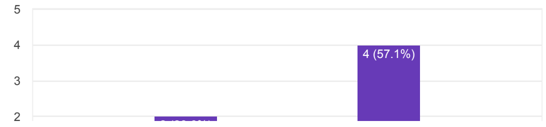
3 responses



Using the app

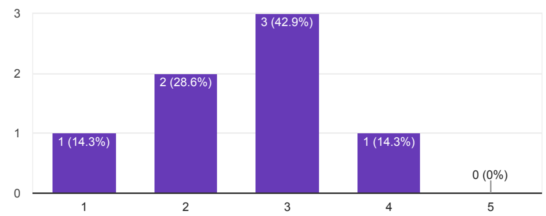
On a scale from 1 to 5, how useful did you find the app?

7 responses

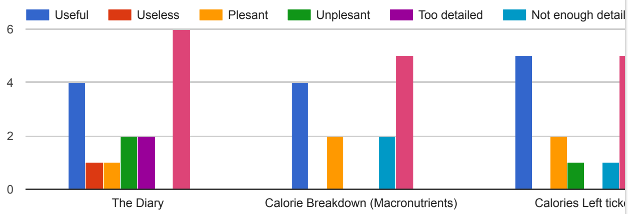


On a scale from 1 to 5, how pleasant did you find to use the app?

7 responses

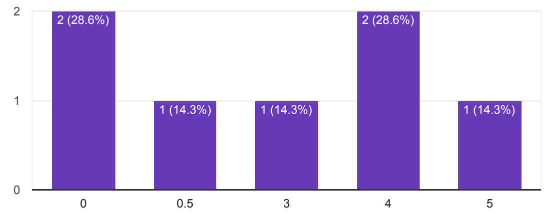


What did you think of the way the app visualized your information?



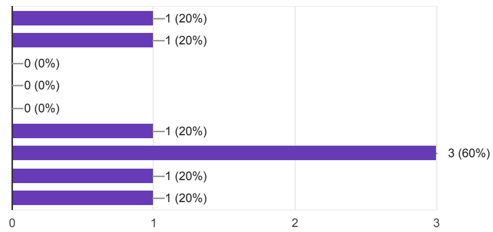
Estimate how many of your meals during your trial you didn't log through the app

7 responses



If you didn't log a meal, what was the reason?

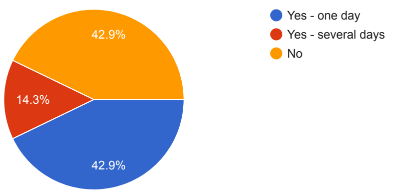
5 responses



Untitled Section

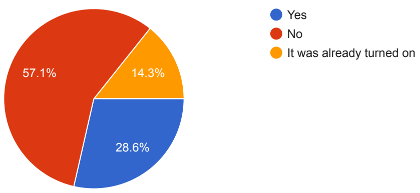
Was there a day where you didn't use the app at all?

7 responses



My Fitness Pal provides a reminder functionality; did you turn it on?

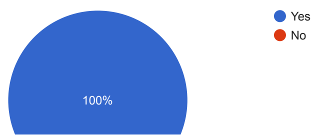
7 responses



Untitled Section

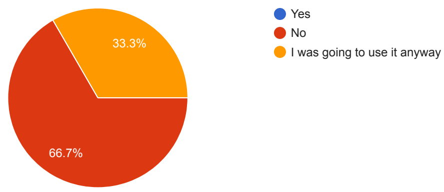
Did you get a reminder to log your food?

3 responses



Did getting a reminder convince you to use the app that day?

3 responses



Untitled Section

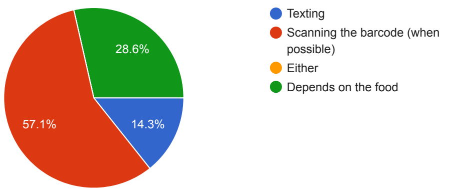
If the app had reminded you to use it, do you think it would have convinced you to use it?

4 responses

Untitled Section

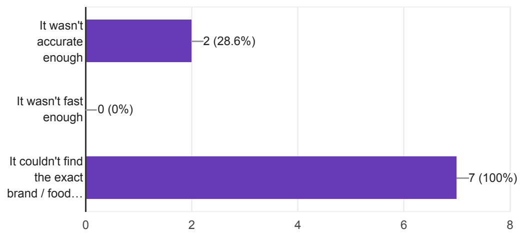
Which input method did you prefer?

7 responses



Did you have any issues using text logging?

7 responses



Did you have any issues using the barcode scanner?

5 responses

How accurate do you think your portion /size estimates were?

6 responses

Ok

Reasonably

not that good because i didn't have a scale

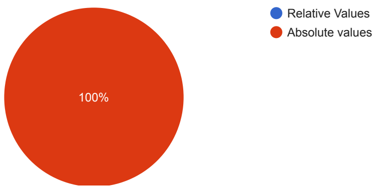
More or less accurate (within 5-10%)

I don't think they were accurate enough.

They were perfect for snacks & shop bought items but often buggy for imported recipes (there were frequent misreads of portions per recipe, perhaps due to inconsistent formatting between recipe websites).

Would you prefer recorded meal sizes to be absolute values (number of portions, weight) or relative values (based on previous meals, more, less or the same as usual)?

7 responses



Why?

7 responses

Can be compared to recommended more easily and you can weigth the portion difference.

Absolute is easier for me personally to process

Because than the calories and nutrition is more exact

It's more precise

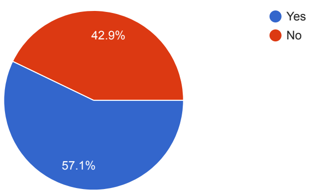
absolute values seem more accurate thsn relative values

It would be easier to calculate.

My portion sizes are inconsistent, it depends usually on how much I have snacked throughout the day.

Did you check back on what you had eaten on the day or on a previous day?

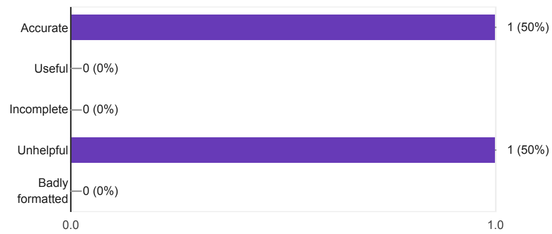
7 responses



Untitled Section

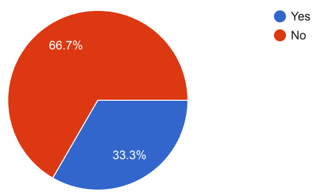
When you checked back on your meal, did you find the information...

2 responses



Having used the app, do you think you have a better idea of how much you are eating in a day?

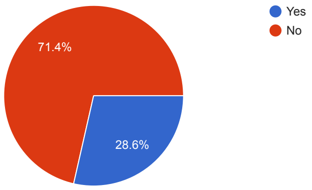
3 responses



Untitled Section

Overall, did you find the app helps you with to your diet?

7 responses



Please explain

7 responses

Said I was eating to little but I'm not..

Im content with what i eat calorie wise compared to my exercise and all the features i wanted to use were on pay extra

at the moment with out a scale it's not helping that much and to create new recipes was annoying. i already think i know how much i should eat

It made me aware of some of my worst habits and I tried some of the suggested recipes

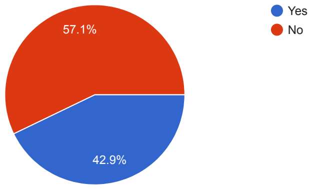
I think I have a good grip on my diet and nutritional requirements already

I didn't want to change it so I don't think I needed to monitor it. But if I tried to e.g. lose weight, I think it would be helpful.

It warned me of a surplus of sugar in my diet, which I plan to cut down on.

Has using the app made you more "mindful" of what you eat?

7 responses



What did you expect when you started to use the app?

7 responses

It to be easy to use!

Little change in personal behavior

track my calories and maybe help me to avoid snacks

It confirmed some of the fears I had about my eating habits

To have a better idea of how much of my nutritional requirements I am meetingg

That there wouldn't be all the food I eat in the database.

That it would warn me about how much sugar I was eating.

Did using the app conform to your expectation? If not, why?

6 responses

Yes (2)

No only american options appeared at the top of searches.

it did

To a certain extent, yes. It gives me a good idea of what I'm eating, but I believe it overestimates how much I need to need.

There was actually more than I expected.

Do you trust the app to store your information securely? Do you think it's important?

7 responses

No because they're American. Yes.

No, but there wasnt much i put on there that i thought was overly in need of being secured

i'm not sure and it think it imported even though i try not to say too much about me

Yes. I don't think it's important

I did not really consider the question of protection of my data, so I am unsure what the app actually does with my meal logs. I consider it important to not have my information stored securely because I don't want other people to judge me based on how and what I eat.

I don't know but I don't think it's important.

I think information in general should be securely held. However, I don't mind diet information being given to health services / research bodies. Poor diets are already putting stress on health services, especially in the US for diabetes & obesity. Perhaps anonymised dietary data could help understand the issue.

Would you use the app past this evaluation period? Why, or why not?

7 responses

No it takes to long to log meals.

No, it didnt really help me that mhch

probably not and start my own meal dairy again

Yes, I think it's going to help me eating in a more healthy and conscious way

I would not use it anymore because I don't think it's telling me anything new about my diet and so I think it's a bit of a useless effort to keep track of it.

I probably will when I decide I want to lose weight.

No, unless I change my diet significantly. If you don't change your diet, it will not tell you anything new. That said, once I have tried to find a way of cutting down on sugar I will revisit the app to check what progress I have made.

Did you think any important features you would have liked were missing?

5 responses

Fruit/ veg counter. Meal suggestions based on low carb/ high protein/ high fibre requirements.

no

No, I think the app does a good job

I'd like if the app could calculate how much calories I burn during exercise. I had to do that manually.

Custom recipes where you choose the ingredients from a list (similar to what you might see at a self-checkout for fruit & veg, it wouldn't be hard to do) so that you don't have to approximate what you ate by a recipe from a site that is kind of similar.

Did you use any of the fitness tracking features? Did you find them a useful complement to the diet tracker? Please describe your experience

6 responses

Was hard to use without downloading the other sister ap for exercise

no

I didn't really understand or spent much time figuring out how they worked

I let the app measure how many steps I walked every day. I think this is quite a useful feature to complement the diet tracker as it takes into account the fact that you might require more food after exercising.

Yes, I think it's very useful. But I'm not sure if it's accurate on a phone.

I used it, it was the worst part of the app. Very over simplified. It also compounded the issue with the calorie counter, by implying that a healthy diet would be achieved by eating as much as you want, as long as you can work off the equivalent amount of calories in the gym. This is an outdated mindset.

Is there anything else that could be done to improve the experience of using the app?

6 responses

I would have prefered more features (as added above) being free because it was too much tk pay for premium in comparison to the benefit i would have got from doing so.

easy way to make you own recipes

More personalized diet suggestions for foods and recipes to try based off of a person's individual requirements

I feel like the way that the diary is displayed is a bit too crammed even though I think that the things it shows are useful. Maybe there is an optically more pleasant way of visualising the differet meals during a day.

Maybe if I had a smartwatch it would improve recording the exercise.

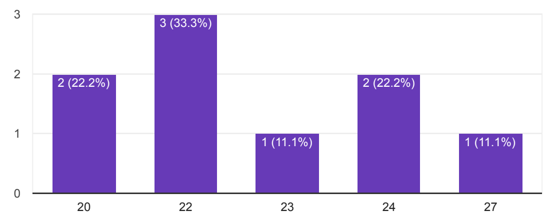
Not sure.

Chatbot feedback

9 responses

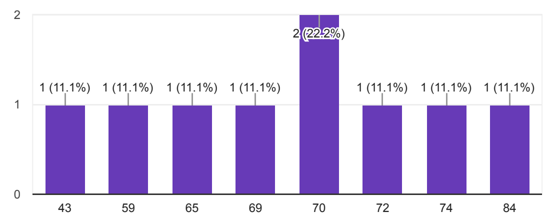
How old are you?

9 responses



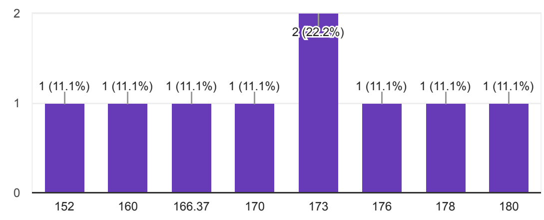
How much do you weight in Kilograms (estimate if unsure)?

9 responses



What is your height in centimeters (estimate if unsure)?

9 responses



Do you know what your BMI is? If so, write it down

5 responses

23.3

19.7

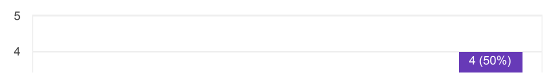
No

26.3

26

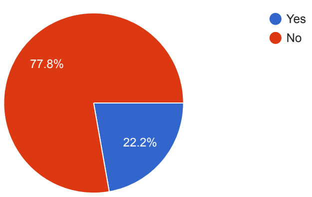
How proficients are you with computers?

8 responses



Are you on a diet, or have you attempted a diet in the last two years?

9 responses



Please explain why you haven't considered a diet in the last two years

7 responses

My diet is pretty balanced normally and I stay fit so don't need to lose weight. I'm pretty close to being underweight but not enough that I'd bother trying to gain weight.

Because I thought I was a reasonable weight and had a good excersize schedule.

I generally eat pretty healthily and I excericse a fair bit anyway so I don't feel like I need to make a particular effort to go out of my way to diet

bad at commitment

There was no need

Laziness

Because I eat well and in moderation and I exercise

Please briefly describe your diet and your motivations for dieting

2 responses

Making weight for taekwondo competitions

Want to be in a better shape

If you are not currently dieting, why did you stop?

2 responses

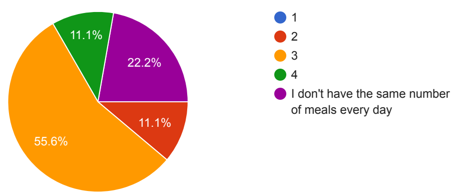
Dieting should only ever be a temporary thing. You need to be making lifestyle changes for long term results.

Couldn't keep up for longer than 3 months

Before the chatbot

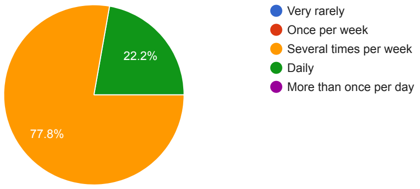
How many meals do you usually have in a day (excluding snacks)?

9 responses



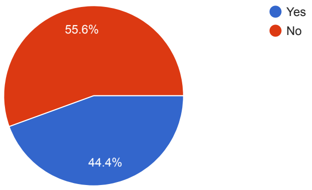
How often do you have snacks outside of meals?

9 responses



Have you ever tracked your diet before?

9 responses



Describe how you track or tracked your diet

4 responses

fitnesspal app

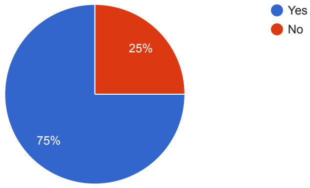
My Fitness Planner App

MyFitnessPal

In my head

Do you also track your snacks?

4 responses



Untitled Section

What do you think is a "balanced" diet?

4 responses

A diet in which you get all the vitamins, minerals and macronutrients to have healthy life. A diet that doesn't cause you stress and makes you feel good.

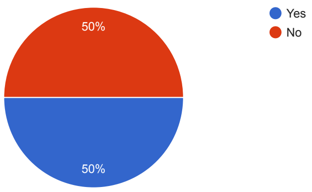
within calories allowance and with enough variety to provide the body with vitamins neutriants etc it needs.

More portions of vegetables, less sugar

Protein, carbs, fruit and veg, dessert

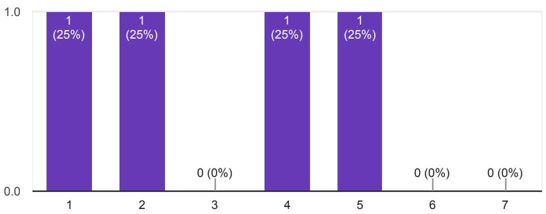
Do you believe your diet to be balanced?

4 responses



Please estimate how many portions of fruits and vegetables you consume in a day on average

4 responses

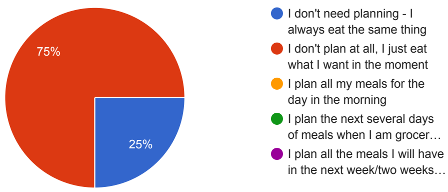


Do you mostly cook your own meal, buy ready-made food, or eat out?

4 responses

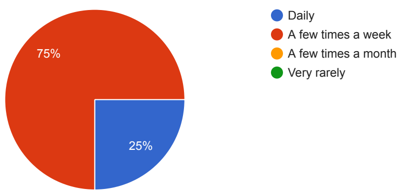
How far in advance do you plan your meals?

4 responses



How often are your meal plans disrupted by an unexpected event (a missing ingredient, being invited to eat out)?

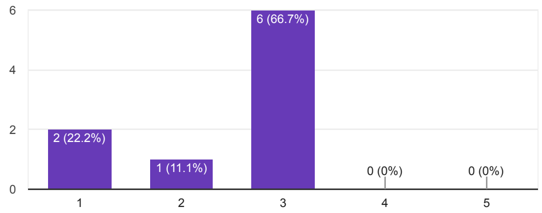
4 responses



Using the chatbot

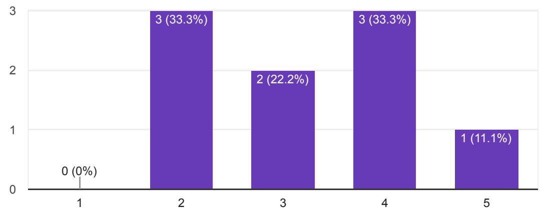
On a scale from 1 to 5, how useful did you find the chatbot?

9 responses



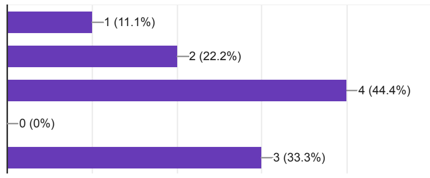
On a scale from 1 to 5, how pleasant did you find to use the chatbot?

9 responses



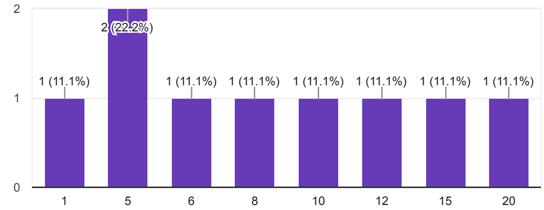
What did you think of the chatbot's feedback on your diet?

9 responses



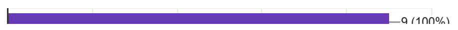
Estimate how many of your meals during your trial you didn't log through the chatbot

9 responses



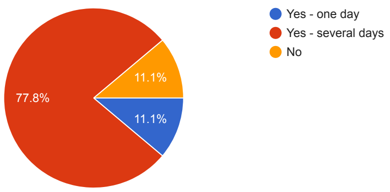
If you didn't log a meal, what was the reason?

9 responses



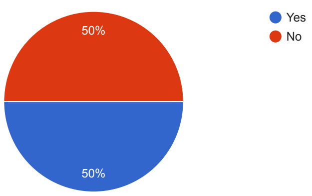
Was there a day where you didn't use the chatbot at all?

9 responses



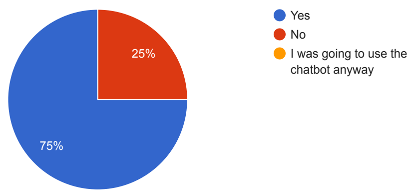
Did you get a reminder the next day from the chatbot to log your food?

8 responses



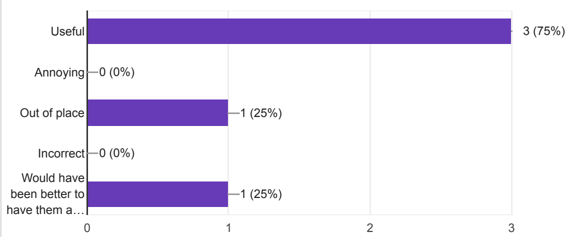
Did getting a reminder convince you to use the chatbot that day?

4 responses



What did you think of the reminder(s) from the chatbot?

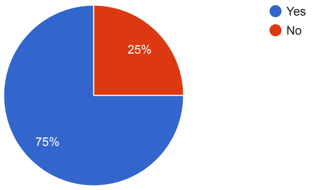
4 responses



Untitled Section

If the chatbot had reminded you to use it, do you think it would have convinced you?

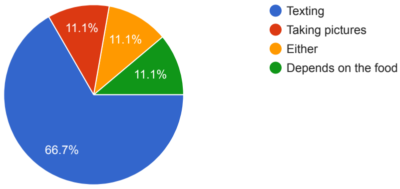
4 responses



Interactive with the chatbot

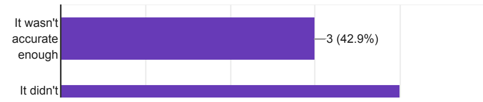
Which input method did you prefer?

9 responses



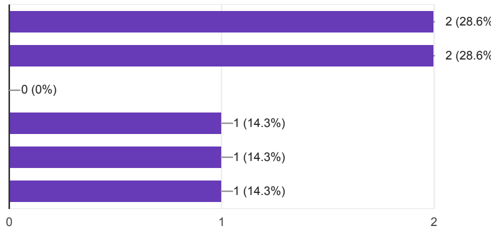
Did you have any issues using text logging?

7 responses



Did you have any issues using picture logging?

7 responses



Do you prefer giving relative estimates of your food logging (more, less, same as usual) or precise estimates (2 cups, 500 grams?) Why?

9 responses

relative. I don't measure out my food, I have more or less depending on how much I feel like eating.

relative its easier to measure

relative estimates

More, less and same is easier I feel as it doesn't require as much detail

I think if I'd been given any indication that healthbot wanted measurements in a particular unit I'd have preferred that.

relative: no need to measure, easier to write

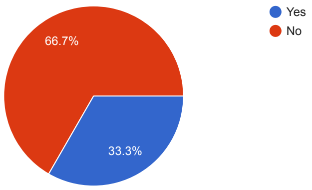
Yes, because more/less tells you nothing if you don't have a baseline

precise estimates

Relative

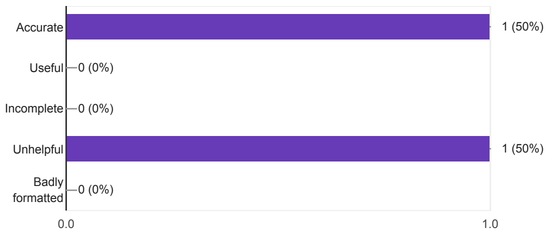
Did you check back on what you had eaten on the day or on a previous day?

9 responses



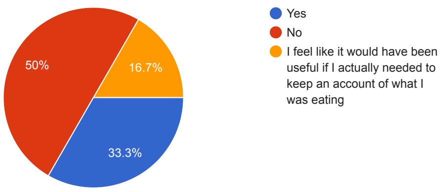
When you checked back on your meal, did you find the information...

2 responses



Having used the chatbot, do you think you have a better idea of how much you are eating in a day?

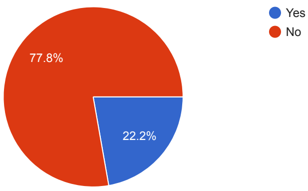
6 responses



Untitled Section

Overall, did you find the chatbot helped you to eat well?

9 responses



Please explain

7 responses

I'm a pretty healthy person, my diet is balanced enough as it is it allowed me to see how my diet is over all which is something which you don't see when you're not recording things

It didn't influence what I ate

I don't feel I received enough feedback to change my diet

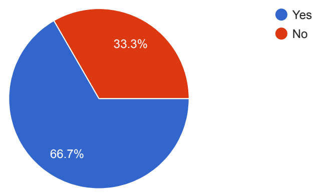
reflecting on my food choices helps make better diet decisions

I didn't even know that was what it was for

I ate a lot of junk this week lol

Has using the chatbot made you more "mindful" of what you eat?

9 responses



What did you expect when you started to use the chatbot?

8 responses

something which would just record what I ate

Not much really. I was hoping it to be more convenient than other apps.

I thought feedback might be more regular although it may have been due to my lack of interaction

A good AI

It would take more initiative, remind me to log things

No expectations

I expected the chatbot to be more talkative.

Not many expectations

Did using the chatbot conform to your expectation? If not, why?

7 responses

slightly. Thought it wouldn't give more suggestion although that could occur with more use

I wished it had better picture recognition

No. It wasn't very good at recognizing what information meant. It's like it needed slightly more word recognition, or to ask or the exact data it wanted from the user. It didn't understand quantities I gave, but it never asked for anything other than relative information which was often irrelevant. It also got confused when I slept and thought I just wasn't logging those 8 hours.

Was less annoying than it thought it might be

No, didn't have any

No, it gives repetitive replies that aren't that useful or customized to my input.

Didn't have any

Do you trust the chatbot to store your information securely? Do you think it's important?

9 responses

Honestly I don't care as long as its useful.

yes to both as food data could be very useful to some people

I do trust it. I do think it is important

I don't really care too much about storing information about what I eat. Generally I feel data protection is important

I mean it's on facebook so not really.

Yes. I know its developer takes security seriously

Yes, yes

No. Yes.

Eh, I don't know & yes

Would you use the chatbot past this evaluation period? Why, or why not?

9 responses

Yes, evaluation period was short I still haven't used some of the features and want to try them out.

yeah good to record what I'm eating

no, I believe it would need to be bug free first.

I would definitely try to use it for a while more

No. There's better services. My Fitness Pal being one of them. The only benefit is having a pseudo-human interaction, but that becomes frustrating when the bot is so limited in dialogue.

Not. I have good eating habits and I am mindful of my eating even without a chatbot

No, it didn't fill any purpose

No, I find logging my intake a waste of time.

No. I think about what I'm eating enough on my own

Did you think any important features you would have liked were missing?

6 responses

Having recommended nutrition amounts? Also tracking of the 5-a-day.

yes add in meals the day before.

It telling me what I had eaten and how much macronutrients I had for the day

better recognition of food and amounts, or, recognizing casual quantities like "a bowl" or "a mug" and asking for more detail.

reminders

It's harder to say what it should do rather then what it shouldn't.
To me it just felt like a logger which I might as well just written down my eating habits in a text file on my desktop and it would achieve the same as the chatbot.
It was too repetitive, not enough varians in phrases.

Is there anything else that could be done to improve the experience of using the chatbot?

8 responses

A full list of everything the chatbot could do being made available would have been nice.

maybe make it less like its a chat bot (even though thats it's name) more human makes it a nicer experience

The generic replies can get annoying/boring

It needs a lot of revision tbh, designers need to ask is it meeting it's purpose, is it worthwhile compared to existing products. Outside of design it needs better programming, users can't be trusted to know what input you expect from them you have to point them in the right direction. Limited responses and a generic tip about leafy greens isn't good enough.

filling this survey made me realise I didn't know about some functionality of it like taking pictures or looking back on what I ate; maybe the bot should introduce all this functionality in the beginning

Previous question

The chatbot is quite buggy. It called me Undefined. Sometimes it doesn't read my messages when I send from my laptop rather than my phone.

Better , more diverse feedback and responses from it