Applied Soft Computing Journal 178 (2025) 113234

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

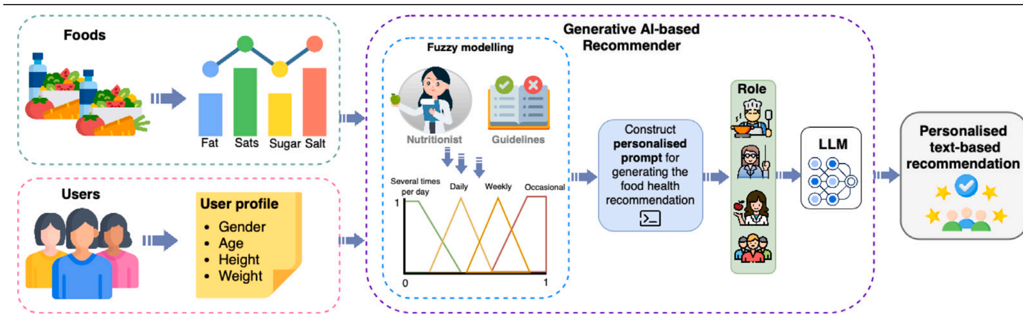
Personalised healthy food text recommendations through fuzzy linguistic variables: A generative AI-based approach

Andrea Morales-Garzón [a,*], Ana María Rojas-Carvajal [b], Roberto Morcillo-Jimenez [a], Maria J. Martin-Bautista [a], Karel Gutiérrez-Batista [a]

[a] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center for Information and Communication Technologies), University of Granada, Spain

[b] Department of Nursery, Faculty of Health Sciences, University of Granada, Spain

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Food recommendation
Generative models
Prompt engineering
Natural language processing
Fuzzy logic

ABSTRACT

Nutrition and healthy eating habits are fundamental for the global population. Nowadays, there is an increasing tendency to consume less healthy recipes and a low general knowledge of nutrition. In these terms, generative AI arises as a potential tool for health-aware food recommendations, especially when improving communication with the user. This study presents a pipeline to enrich prompts with fuzzy modelling to increase the quality of textual recommendations. We apply our pipeline to generate a personalised frequency of food consumption, considering both nutritional and individual profiles. This is an essential task for increasing the health-conscious recommendation systems. We conducted extensive experimentation across different roles and prompt strategies. We evaluated the quality of the text and the nutritional rigour of the text responses. Our results show that enriching prompts with fuzzy modelling of the nutritional information of the foods significantly improves the quality of the prompt responses.

1. Introduction

Nowadays, unhealthy diets and inadequate nutrition constitute significant risk factors for various diseases worldwide. According to a 2019 study by The Lancet, one in five deaths globally is linked to poor nutrition, encompassing conditions such as cardiovascular diseases (e.g., heart attacks and strokes), certain cancers, and diabetes.[1] For instance, cardiovascular diseases represent the leading cause of death worldwide, accounting for over half of all global mortality, claiming approximately 17.9 million lives annually.[2] Additionally, the

* Corresponding author.
E-mail address: amoralesg@decsai.ugr.es (A. Morales-Garzón).

[1] www.who.int/news-room/fact-sheets/detail/malnutrition

[2] www.who.int/health-topics/hypertension/cardiovascular-diseases

https://doi.org/10.1016/j.asoc.2025.113234
Received 16 July 2024; Received in revised form 10 April 2025; Accepted 26 April 2025
Available online 12 May 2025

World Health Organisation (WHO) estimates that the global prevalence of overweight adults, defined as those with a Body Mass Index (BMI) of 25 or higher, stands at 38.9%. These statistics underscore the critical need to establish rigorous nutrition standards within computational systems applied to nutrition, dietary advice, and assessment.

Here arises health-aware food recommendation systems [1] as a computational approach to help individuals integrate healthy habits into their daily routine. The accessibility and online availability of large language models (LLMs) [2,3], such as ChatGPT, has empowered individuals to use these resources for queries concerning recommendations regarding food, cooking, and nutrition [4]. They have also recently attracted considerable attention in recommendation systems, presenting a promising tool for enhancing their performance and increasing the communication of recommendation systems with the users [5].

Integrating generative AI in food recommendation systems opens a new paradigm in personalised recommendation. There is growing interest in incorporating LLMs into conversational recommendation systems, providing personalised explanations to users, and increasing their user-friendly features. However, this communication may lack exactness, clarity, engagement, or trustworthiness, opening new scenarios regarding user interaction with the system. Prompting generative general language models for nutritional recommendations can lead to severe health issues for the population [6]. We highlight the lack of nutritional rigour, absence of nutritional vision in training and limited context of the individuals [7], and problems derived from gender bias [8]. In the case of nutrition, it is especially relevant that personalised recommendations are gender-sensitive, as this is a differentiating feature in nutritional dietary assessment.

Since LLMs have cold-start capabilities in zero-shot scenarios, providing reliable responses is challenging. In these terms, knowledge-enhanced prompt learning arises as an alternative for textual-based recommendations, mainly used in conversational recommendation systems [9], as well as retrieval augmented generation for improving the prompt content [10]. In nutritional recommendations, relevant data comes from diverse sources: nutrition information, individual profiles, expert guidelines, and health standards [11]. Incorporating all the necessary information into prompts increases the difficulty of achieving high-level performance in the response. Fuzzy modelling techniques have been demonstrated to model complex relationships and integrate expert knowledge [12]. Motivated by this, this study proposes a novel generative AI-based approach that enriches prompts based on fuzzy modelling of foods' nutrition profiles, nutritional advice, and healthy standards.[3] We specifically apply this approach to generate personalised textual recommendations regarding food frequency of consumption based on their composition and individual needs. This task is significant in healthy-aware recommendation systems since they provide educational explanations that help the user understand the nutritional profile of foods while enhancing healthy diet adherence and long-term healthy habits.

To the best of our knowledge, this is the first attempt at applying fuzzy logic to enrich user interaction in recommendation scenarios. The contributions of this study are as follows:

1. We present a generative AI-based approach to provide textual recommendations regarding the frequency of consumption of foods based on health standards and nutritionist advice. It considers the individuals' profile and the nutritional information of foods to provide tailored textual recommendations for adequate food consumption.
2. We propose a methodology for integrating fuzzy linguistic variables into LLM-based food recommendations and dietary assessment pipelines, with the potential for further textual recommendation tasks.
3. We conduct an extensive study of textual recommendation considering prompting roles, text quality, nutritional rigour, and robustness of prompting strategies using the LlaMA3 model. Our research demonstrates that integrating fuzzy modelling in generative AI prompts enhances nutritional rigour and improves model behaviour.

[3] The code and materials for this paper will be publicly available on GitHub upon acceptance.

2. Related work

The term *food computing* refers to technological applications related to food [13,14]. These sorts of applications range from food recommendation systems and diet management [11,15] to recipes adaptation [16–18].

In [19], the authors analyse the healthiness of Internet-Sourced Recipes. They state that only a tiny percentage of the recipes on Allrecipes.com can be assumed healthy, considering the standard guidelines proposed by WHO and FSA, and even recipes in the *healthy recipes* category can be deceptive. This illustrates the level of unhealthy recipes that feed data-driven recommendations, as well as the tendencies of consumption of the population.

Food recommendation allows suggesting nutritional recommendations to users at food, recipe and diet levels. They are based on different criteria such as user preferences [20], dietary restrictions and nutritional requirements [21], and cultural impact [22]. By leveraging data such as user reviews, ingredient lists, cooking methods, and nutritional content, recommendation systems aim to provide personalised and relevant recipe suggestions that align with the user's needs [23].

Recently, significant advances have been made in food recommendation systems, aiming to predict users' preferences and guide their choices based on predetermined criteria [24,25]. The studies above fail to recommend healthy food as they mainly focus on users' preferences.

2.1. Health-aware food recommendation systems

We find several approaches in the literature striving to design healthy-aware food recommendation systems. That is the case in [23], where the authors summarise the state-of-the-art recent approaches of healthy-aware food recommendation systems. In the same way, the study depicted in [26] provides an overview of recommendation systems in the healthy food domain, covering both individual and group recommendation approaches and discussing research-related challenges.

According to [27], a framework (namely NutRec) is presented for predicting relevant ingredients and their amounts. The proposal creates a healthy pseudo-recipe that searches the dataset for the most similar healthy recipes to improve the healthiness of the recommended recipes without requiring any pre-computed nutritional information for the recipes. In [25], the study explored whether a recommendation system that integrates healthy and personalised suggestions can influence individuals to choose healthier recipes than they typically would. It is achieved by incorporating a healthy bias into the recommendation algorithm and displaying a healthy tag on recipe cards.

Another interesting work, [20], introduces a food recommendation approach to generate personalised daily meal plans tailored to users' nutritional needs and preferences. The approach incorporates an AHPSort-based pre-filtering stage for excluding inappropriate foods from the recommendations. Finally, the approach generates a menu to maximise user preferences while ensuring nutritional requirements. A new model named the Healthy and Time-Aware Food Recommendation System (HTFRS) is presented in [15]. The HTFRS model outperformed other state-of-the-art food recommendation systems by considering food ingredients and user ratings and incorporating a novel time-aware similarity metric to capture changes in user preferences over time. Market2Dish [28] presents a significant initiative for health-aware food recommendations. The system aims to map ingredients available in the market to user-preferred dishes, thereby facilitating a recipe retrieval task that suggests health-conscious options to users.

In [29], the authors propose an affective computing-based meal recommendation and menu planning system that considers a person's emotional state towards different foods, their nutritional value, and the individual's calorie requirements to provide personalised meal recommendations and full-day menu plans. Another interesting work is presented in [30], proposing a novel framework for personalised healthy-aware meal planning using generative AI models, paying attention to inpatient clinical dietetics. The authors state the fact that nutritionists play a crucial role in observing patient adherence and preferences, contributing to a continuous learning health system cycle for the personalised meal planning system.

Our approach is food-level, focusing on providing users with the recommended frequency of consumption for requested ingredients. However, the studies detailed above do not focus on the need to model the frequency of ingredient consumption despite their importance for balanced and long-term habits. Accurately modelling this frequency is crucial in food recommendation, as it determines the appropriate selection of ingredients in recommended recipes and diet plans. The flexibility of our approach allows for its integration into both existing and future recipe recommendation systems.

Additionally, these systems still need to be improved: (1) visualisation and explanation of healthy-aware food recommendations have only been studied superficially, and (2) overlooking user characteristics like weight, age, allergies, etc. Lately, LLMs have been used for solving general downstream NLP tasks, demonstrating being a suitable tool for recommendation systems [31] in different domains like dialogue agents [32], news [33], and nutrition recommendation [34]. Table 1 includes the most relevant approaches aligned to our proposal reviewed in this section, highlighting their pros, cons, and peculiar features.

2.2. Open challenges in health-aware food recommendation

After reviewing the related literature, we observed that although health-aware food recommendation systems have made notable progress, many still rely on rigid nutrient limits that often fail to adjust to real-world food variations and the reality of food recommendations. These systems may offer impractical suggestions that miss users' dietary goals. We propose a fuzzy modelling approach that treats nutrition and health targets as flexible rather than binary, better capturing the trade-offs in dietary recommendations. Grounded in official dietary guidelines, our model supports realistic, personalised recommendations while maintaining compliance with health standards. This adaptive framework enhances system credibility and supports transitioning from generic food suggestions to health-supportive dietary guidance.

Most existing food recommendation systems — including many health-aware variants — generate context-independent outputs. They rarely consider dynamic factors such as the user's current health status or evolving dietary needs. This static treatment limits their capacity to deliver timely, relevant suggestions in everyday settings. Our approach addresses this gap by including LLMs to encode contextual metadata into prompt structures. These prompts guide the generation of recipe descriptions and explanations that adapt to the user's current situation. This context awareness enhances both the personalisation and practicality of the system outputs, allowing it to respond to individual needs more precisely.

A recurring limitation in health-focused recommendation systems is the lack of interpretability of the recommendations. Users are often left without a clear understanding of why a particular food item was suggested, which is especially problematic in domains involving health decisions. Without accessible, evidence-based justifications, users may be unable to assess whether recommendations suit their specific goals, reducing trust and engagement. Our approach provides natural language explanations generated through prompting LLMs and explicitly grounded in nutritional knowledge. This is intended to support informed food choices and foster user confidence in recommendation scenarios.

Table 2 summarises these gaps, as well as how our approach addresses them and the intended impact of our contributions. Therefore, our proposal not only seeks to provide healthy-aware recommendations considering features like nutrition information, users' preferences, expert guidelines, and health standards but also aims to incorporate more healthy-aware nutritional information into prompts for enhancing LLMs-based healthy-aware food recommendations. As far as we know, this is the first approach using fuzzy logic, prompt engineering, and LLMs to increase the performance of LLMs and the quality/understandability of healthy-aware recipe recommendations. In particular, we study healthy-aware text recommendations based on food frequency of consumption, taking into account prompting roles, text quality, nutritional rigour, and the robustness of prompting strategies.

3. Methodology

As stated before, the main objective of our proposal is to develop a personalised fuzzy-based healthy-aware text recommendations system for requested ingredients based on food frequency of consumption. In order to estimate the frequency of food consumption, the system uses individual profiles, food nutritional content, and established dietary guidelines (e.g., WHO standards). We define the problem as follows:

Let U represent an individual, characterised by their attributes $U = \{w_i, h_i, a_i\}$ where w_i is weight (kg), h_i is height (cm) and a_i is age (years). Let $F = \{f_1, f_2, \dots, f_n\}$ represent the set of requested foods. Each food item f_j is characterised by a set of four nutrient values $f_j = \{sugar_j, fat_j, saturates_j, salt_j\}$ where $sugar_j$ represents the quantity of sugar, fat_j represents the quantity of fat, $saturates_j$ represents the quantity of saturates and $salt_j$ represents the quantity of salt of f_j. Firstly, the system computes the Basal Metabolic Rate (BMR) for each user using the Harris–Benedict Equation (HBE) (see Eqs. (1), (2) and (3)). Secondly, we define four fuzzy sets $FS^U_{f_j} = \{FS_{sugar}, FS_{fat}, FS_{saturates}, FS_{salt}\}$, one for each nutrient to fuzzify its amount, for the food f_j and the user U, using three fuzzy linguistic variables as $\mathcal{L}^U_{f_j} = \{$*low*, *medium*, or *high*$\}$, based on threshold values derived from nutrition experts guidelines. Finally, we define a set of fuzzy rules that allow us to compute the fuzzy score of consumption $S(f_j, U)$ for a food f_j and an individual U in the inference stage.

Note that the fuzzy linguistic variables $\mathcal{L}^I_{f_j}$ are integrated into a prompt, which is used to generate personalised user-friendly textual recommendations based on the individual's profile and food characteristics that align with WHO nutritional guidelines.

Fig. 1 illustrates the workflow for building the generative AI-based recommendations. Our approach parts from three primary resources, the individual profile (needed to adjust the food frequency of consumption to the individual), food and nutrition data (typically obtained from food datasets), and nutrition standards (such as WHO or nutritionists' advice). We use this information to model the nutritional amounts of the foods, taking into account four nutrients (fat, saturated fats, sugar, and salt) and three fuzzy linguistic input variables to represent the amounts (low, high, and medium). This modelling is not a one-size-fits-all approach but a personalised solution for tailored individual needs and body composition. It allows us to obtain a unique score for quantifying food's recommended frequency of consumption for a given individual, ensuring that each individual's unique requirements are met. Then, the fuzzy variables are considered in the prompt for generating personalised nutrition information for the food for final users.

We enrich the prompt with fuzzy modelling to provide user-friendly and personalised textual recommendations on food nutritional profiles. These recommendations are based on the computed scores and the

Table 1
Most related approaches summarising their pros, cons, and peculiarities.

Article	Pros	Cons	Peculiarities
[20] Toledo et al. (2019)	System with a pre-filtering stage to remove unsuitable foods and an optimisation stage to create meal plans that align with user preferences and nutritional needs while avoiding recently eaten foods. A case study shows its effectiveness, excluding 32 foods for overweight and 40 for diabetic users.	It lacks integration of long-term food logs to ensure balanced intake over time. Additionally, the evaluation was conducted on synthetic data, highlighting the need for future validation with real users.	General framework for daily meal recommendations that incorporates the simultaneous management of nutritional- and preference-aware information.
[25] Pecune et al. (2020)	It explores how adding a health bias and healthy tags to recipe recommendations affects user choices. It develops three recommender systems (health-focused, preference-based, and hybrid) to understand how they influence healthier decision-making.	Limited recommendation diversity, lacks a conversational interface for capturing user preferences, and no focus on users' healthy eating interests to tailor recommendations.	It explores if a healthy bias and health tags in recipe recommenders influence people's decision-making when selecting recipes.
[27] Chen et al. (2020)	It outperforms a random baseline in WHO scores, highlighting the value of considering ingredient amounts, not just types. Training on diverse and extensive data proved more beneficial than limiting it to healthy recipes.	Potential improvements include ensemble ingredient models, broader dietary goal support, and user rating integration for benchmarking.	The system recommends healthier recipes by generating a healthy pseudo-recipe and then retrieving a similar healthy recipe to recommend.
[28] Wang et al. (2021)	System with robust performance in user health profiling, outperforming baseline methods by extracting health-related information from social media. It also achieves superior results compared to existing methods across multiple evaluation metrics.	More advanced systems could profile users' health, offer food-level recommendations prioritising health protection, and better integrate healthy diet knowledge.	Personalised recommendation scheme that maps market food products to the healthy dishes eaten at home by profiling user health and recommending healthy recipes.
[24] Gao et al. (2022)	System that outperforms state-of-the-art food recommendation methods by effectively leveraging ingredient–ingredient, ingredient–recipe, and recipe–user relationships.	It ignores key factors like diversity. It only considers internal data among foods, recipes, and users.	Using a sum aggregator in the ingredient graph yields better results than a concatenation aggregator.
[29] Islam et al. (2023)	It provides meal recommendations and full-day menus by incorporating users' feelings, excitement, and preferences, distinguishing it from traditional food recommendation systems. The hierarchical ensemble model accurately predicts affective states	Further exploration of collaborative filtering would be beneficial. Using sequential deep learning models could enhance the accuracy of detecting food-related emotions.	Affective computing-based meal recommendation and menu planning that considers nutritional requirements and users' emotions and feelings towards different foods.
[15] Rostami et al. (2023)	The system outperforms state-of-the-art food recommender systems by combining time-aware collaborative filtering with ingredient-based rating prediction. The proposal also adds the time factor into user similarity calculations, which led to an average performance improvement of 14.96%.	Main limitations: dataset bias, limited cultural diversity in evaluation, lack of user-specific data (e.g., age, medical history), and an overfocus on accuracy, risking overfitting.	Healthy- and time-aware recommender system that integrates user preferences, ingredients, and nutrition data to provide personalised recommendations to guide users towards healthier eating habits.
[30] Kopitar et al. (2024)	Advanced personalisation with generative AI; menu visualisation; integrated feedback.	Depends on high-quality clinical data, complex hospital integration, and ethical considerations.	Novel application in clinical nutrition; emphasis on patient participation; multidisciplinary approach.
[31] Xu et al. (2024)	Framework for using LLMs in recommendation tasks. It analyses factors influencing recommendation, e.g., model architecture, context length, prompting approaches, or user interest modelling, among others.	Inefficient inference and challenges in formulating appropriate prompts enable LLMs to fully understand recommendation tasks.	Pioneering use of LLMs in recommender systems.
[34] Papastratis et al. (2024)	It improves the explainability and accuracy of diet recommendations by modelling user profiles in a descriptive latent space and aligning suggestions with nutritional guidelines. ChatGPT enriches the database with culturally equivalent meals, enhancing variety and precision.	It focuses on a specific set of cuisines and dietary preferences, overlooking a broader range of cuisines and dietary needs, such as vegan or gluten-free options. It lacks evaluation with real users.	Diet recommendation system that leverages deep generative networks and LLMs like ChatGPT to provide personalised, accurate, and nutritious weekly plans that align with nutritional guidelines.

individual's profile. Starting from the individual's profile, we calculate their daily energy intake based on height, weight, and gender. This calculation refers to BMR, the body's energy needs to maintain vital functions at rest, such as breathing, blood circulation, and body temperature regulation [35]. The HBE provides an estimation for calculating the BMR of a specific individual. Let w_i be the weight in kilogrammes of the individual i, h_i their height in centimetres and a_i their age in years. Eqs. (1) and (2) illustrate the calculation used for obtaining BMR for male and female individuals based on the Harris–Benedict equation. To consider inclusivity, we also compute a gender-independent version of BMR. This ensures that our nutritional recommendations are representative across all genders and can be used in non-specified gender use cases. For this, we calculate the individual energy intake using the average values of the male and female BMR estimations, as shown in Eq. (3). This approach not only caters to a wider audience but also reflects our understanding of the diverse nutritional needs of

Table 2
Summary of key challenges in health-aware food recommendation.

Challenge	Our approach	Impact
1. Rigid nutritional thresholds	Fuzzy modelling of nutrients and goals	Enables flexible, realistic dietary suggestions
2. Static recommendations	Context-enriched LLM prompts	Personalises outputs based on evolving user context
3. Low interpretability	LLM-generated, guideline-based explanations	Enhances user trust and decision-making clarity

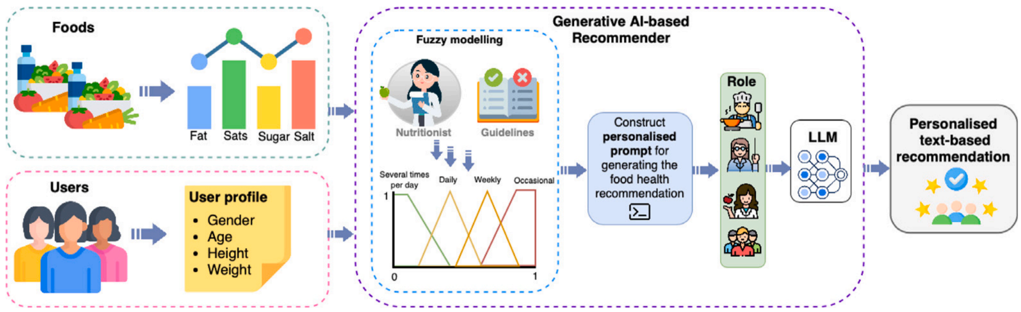


Fig. 1. Workflow of the proposed approach.

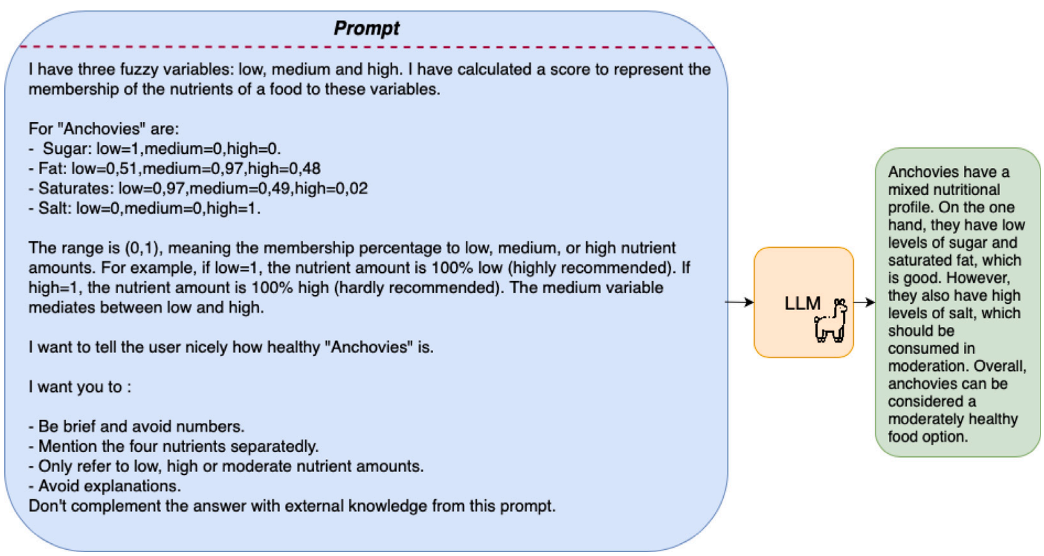


Fig. 2. Prompt for generating recommendations for the fuzzy strategy. We provide the model with the values of the input fuzzy variables for a given food (anchovies, in this case).

individuals.

$$\mathrm{BMR}_{\mathrm{male}}(i) = 88.362 + (13.397 \times w_i) + (4.799 \times h_i) - (5.677 \times a_i) \tag{1}$$

$$\mathrm{BMR}_{\mathrm{female}}(i) = 447.593 + (9.247 \times w_i) + (3.098 \times h_i) - (4.330 \times a_i) \tag{2}$$

$$\mathrm{BMR}_{\mathrm{avg}}(i) = \frac{\mathrm{BMR}_{\mathrm{male}}(i) + \mathrm{BMR}_{\mathrm{female}}(i)}{2} \tag{3}$$

We use BMR to estimate the maximal quantity of a nutrient that should be consumed per day. A gram of carbohydrates provides 4 kilocalories, a gram of protein provides 4 kilocalories, and a gram of fat provides 9 kilocalories. Since the OMS provides the recommended maximal amounts for each nutrient as specific percentages of total energy intake, we can calculate the values for an individual in specific. This allows us to personalise the fuzzy modelling to the individual profiles and provide accurate textual recommendations. We detail the modelling in Section 4.

3.1. Prompt strategy

Fig. 2 shows the prompt used with a specific food. Specifically, we illustrate the prompt for “Anchovies”, after enriching it with the information obtained from the fuzzy linguistic variables. As shown in the figure, we include the membership value for each fuzzy variable and explain the model and the meaning of the membership values. Additionally, we instruct the model on specific requirements to fulfil. Specifically, we ask the model to avoid numbers (so it is easy to understand individuals) and mention each of the four ingredients using low, high or moderate amounts. The latter is included to comply with the nutritional rigour that experts expect from the response. Also, we expressly indicate that the model cannot include external knowledge so that we can control the experiments and the trustworthiness of the prompt response.

In this case, we have chosen an ingredient with a low value of sugar and fats and a high amount of salt. Regarding the saturated facts, while belonging to the low interval, they have a high value. In the rest of the paper, we refer to this prompt strategy as the *fuzzy* prompt.

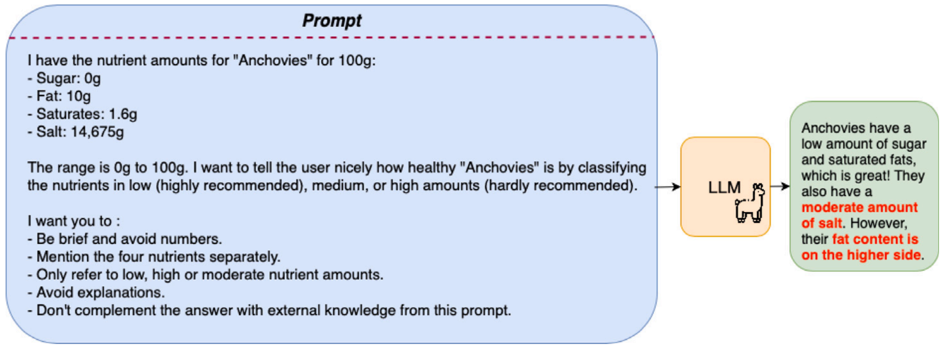


Fig. 3. *Crisp* version of the prompt. We only provide the original nutrient amounts and indicate that the system should consider low, medium, and high amounts.

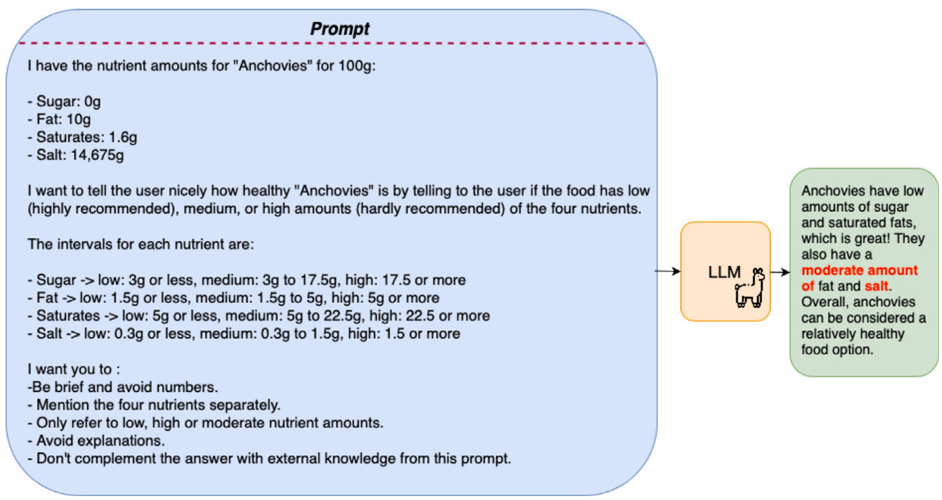


Fig. 4. *Crisp2fuzzy* version of the prompt. We provide the nutrient amounts and the intervals to classify them into low, medium, and high quantities and ask the model to infer the linguistic variable with the highest membership so that it can be used in the response.

To analyse the efficacy of the fuzzy prompt, we incorporate two additional prompts to compare their performance. First, we include a *crisp* prompt, in which we provide the crisp nutrition information of the food and ask the model to provide a nutrition-based explanation in terms of low, medium, and high amounts. Fig. 3 shows an example of the crisp prompt for Anchovies. Second, we include a *crisp2fuzzy* prompt, which has a strategy in the middle of the crisp and the fuzzy prompts. Fig. 4 shows an example of this strategy with the same ingredient, Anchovies. In this case, we provide the model with the nutrition details of the food and the intervals and ask the model to generate the corresponding linguistic variables for each nutrient, i.e., we ask the model to assign the membership to the linguistic variables instead of providing it with the fuzzification.

As highlighted in Figs. 3 and 4, both examples show poor behaviour in comparison to the fuzzy prompt. In Section 5, we test this hypothesis by analysing the performance of the three strategies to illustrate the large language model behaviour for providing nutrition information about foods.

3.2. Role prompting

We have considered four different roles to style the prompt responses in order to study their performance for providing easy-to-understand and correct explanations to the user. With this, we aim to study nuanced variations in the provided explanations to optimise the prompting performance. We hypothesise that the responses provided for the different roles may differ in vocabulary, style, or correctness, which is important to optimise the prompts. Specifically, we have considered four roles to provide context and guide the explanation generation, detailed as follows:

- **Chef**: "You are an expert chef with large knowledge of nutrition".
- **Professor of nutrition**: "You are a professor of nutrition".
- **A person with no experience in nutrition**: "You are a person with no expertise in nutrition".
- **Nutritionist**: "You are an experienced nutritionist".

We indicate the role when prompting the LLM. Fig. 5 shows a more detailed workflow for generating the fuzzy enriched textual recommendations.

4. Fuzzy linguistic modelling

A linguistic variable has values expressed in words or sentences in a natural or artificial language [36]. This concept offers an approximate representation of ideas that can be described using quantitative terms, leading to the field known as fuzzy logic. In the dietary field, linguistic variables offer us a representation strategy for quantifying healthy consumption of food ingredients. We can use them to model a fuzzy system that considers the food nutrition composition to quantify the recommended consumption.

Multiple health charities and organisations have established standards for a healthy and balanced diet [37]. In this case, we focus on those nutrients that foods must not exceed to be considered healthy and nutritious. Specifically, we have used a total of four nutrients that, consumed in high amounts, are directly related to diet-related diseases: fat, saturated fats, sugar and salt. Table 3 illustrates the guidelines proposed by the WHO and European Union associations for these nutrients. The first three columns of Table 3 —Low, Medium and High — specify the amounts for considering a specific nutrient's low,

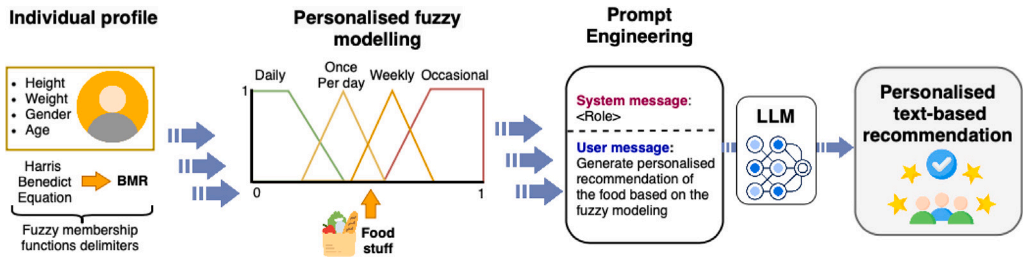


Fig. 5. Workflow for enriching the prompts using fuzzy modelling.

Table 3
Guide for choosing healthy foods provided by Heart UK, the UK's cholesterol charity. It states recommended quantities of specific nutrients per each 100 g of food. The last three columns correspond to the daily recommended amounts by the WHO for a male, female, or non-binary gender individual of healthy body weight consuming about 1500, 2000 and 2500 kilocalories per day.

Per 100 g of food	Low	Medium	High	Max per day (1500 kcal)	Max per day (2000 kcal)	Max per day (2500 kcal)
Fat	<3 g	3–17.5 g	>17.5 g	50 g	66 g	83 g
Saturates	<1.5 g	1.5–5 g	>5 g	16 g	22 g	27 g
Sugars	<5 g	5–22.5 g	>22.5 g	37.5 g	50 g	62.5 g
Salt	<0.3 g	0.3–1.5 g	>1.5 g	5 g	5 g	5 g

medium, and high presence per 100 g of food. Therefore, we have 12 linguistic variables as input, three for each one of the four nutrients. These quantities depend on the proportion of nutrition amount per 100 g of food and do not rely on individual body composition or activity level.

However, it is insufficient to consider these ranges to understand how healthy or unhealthy an ingredient is, especially in the High amount interval. The WHO guidelines for a healthy diet also indicate the daily amount the population must not overcome for specific nutrients.[4] They allow us to distinguish those quite unhealthy foods since their nutrition composition surpasses the daily recommendations stated by health organisations. According to the WHO, the daily amount of salt should be less than 5 g daily. The intake of fats should be less than 30% of total energy intake, and the intake of saturated fats should be less than 10%. The daily amount of free sugar should be less than 10% of total energy intake. The three latter depend on the individual's daily energy intake, which is calculated by using the age, gender, weight, and height using the BMR Equations [38], detailed in Section 3. Thus, the fuzzy modelling can be personalised for each individual by taking their energy intake.

Therefore, this information must be controlled when recommending food based on their nutrition composition and health profile. The last three columns of Table 3 show the value of the maximum daily amount of these nutrients for individuals whose daily energy intake is 1500, 2000, and 2500 kcal, respectively (independently of the individual gender).

We have used these variables to create three linguistic variables that allow us to create a fuzzy system to model a healthiness score for foods. We use an additional output linguistic variable called *Frequency of consumption* for this. We have modelled this variable based on the recommended frequency of consumption of foods by nutrition and healthcare experts (daily, once per day, weekly, and occasional). After defuzzification, it represents a score of 0 and 1, from less to more frequency of consumption.

[4] Healthy Diet: www.who.int/news-room/fact-sheets/detail/healthy-diet

4.1. Fuzzy linguistic variables

For designing the fuzzy system, we have created three input linguistic variables for each nutrient: low, medium, and high, following the guidelines in Table 3. However, the nutrition domain requires considering the specific amount inside the intervals. For example, while 3.1 g and 15 g are both medium amounts for fat, the former is healthier since it is very close to low values. Our modelling must be aware of this to model a precise healthiness score.

Given a nutrient $n \in \{\text{fat}, \text{saturates}, \text{sugar}, \text{salt}\}$, the amount of nutrient x falls in the interval $I := [0, 100]$. Then, we split I into three disjoint subintervals denoted as $I_{\text{low}} := [0, c^n_{\text{low}}]$, $I_{\text{medium}} := [c^n_{\text{low}}, c^n_{\text{high}}]$ and $I^n_{\text{high}} := [c^n_{\text{high}}, 100]$. These subintervals represent the low, medium and high areas for the healthy amount of each ingredient. We define $c^n_{\text{medium}} := \frac{c^n_{\text{low}} + c^n_{\text{high}}}{2}$, and c^n_{max} as the maximum quantity of the nutrient n that should be consumed per day. Then, we define the membership function for the fuzzy variables for each nutrient n (see Fig. 6) as follows:

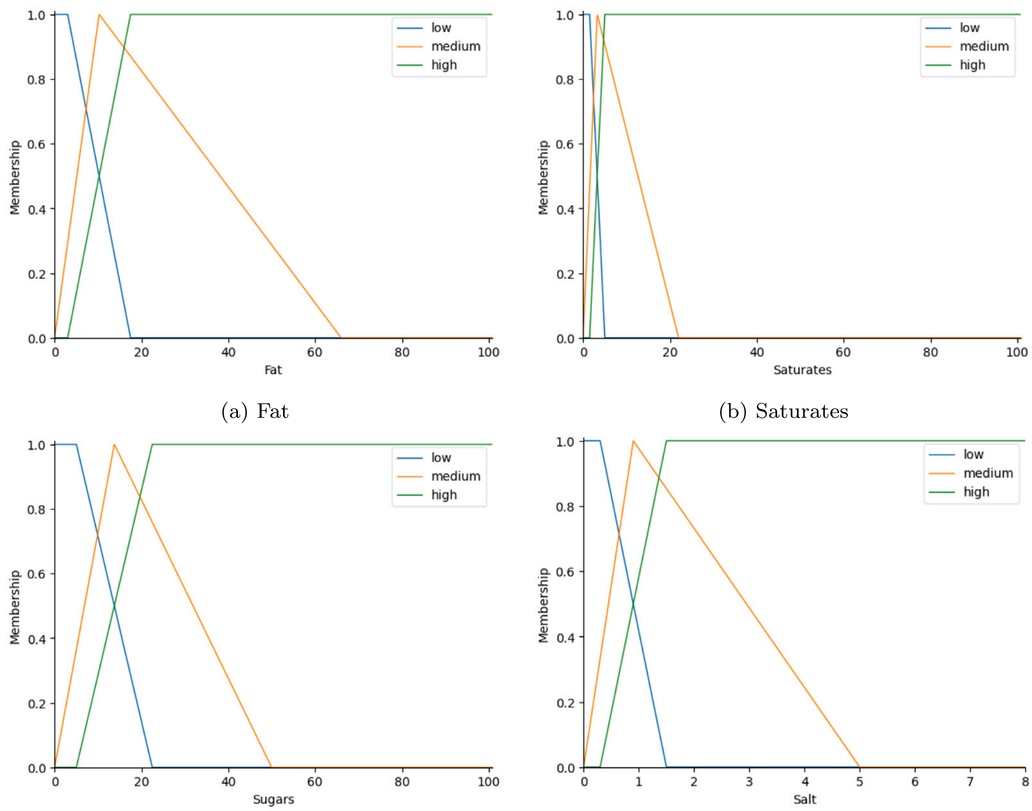
4.1.1. The low-[nutrient] variable

To quantify the amount of a given nutrient in a food item and determine its suitability based on dietary recommendations, we use fuzzy logic to categorise nutrient quantity as low, medium, or high. Specifically, we define a trapezoidal membership function $\mu^n_{\text{low}}(x)$ to represent the degree to which a nutrient amount x belongs to the *low* label. This function assigns a membership value in the range $\mu^n_{\text{low}} : I \to [0, 1]$ based on predefined threshold values c^n_{low} and c^n_{high}. The function is defined as follows:

$$\mu^n_{\text{low}}(x; c^n_{\text{low}}, c^n_{\text{high}}) = \begin{cases} 1 & ; \quad x < c^n_{\text{low}} \\ 0 & ; \quad x > c^n_{\text{high}} \\ \frac{c^n_{\text{high}} - x}{c^n_{\text{high}} - c^n_{\text{low}}} & ; \quad c^n_{\text{low}} \le x \le c^n_{\text{high}} \end{cases} \tag{4}$$

where:

- For nutrient values below c^n_{low}, the food is fully considered to have a *low* amount of the nutrient.



(a) Fat (b) Saturates

(c) Sugars (d) Salt

Fig. 6. Fuzzy variables for fat, saturates, sugar, and salt amounts per 100 g of food. Note that in (d), the x-axis has been reduced for better visibility in this illustration. In this case, the membership function remains stable until value 100 (data domain).

- For values above c^n_{high}, the food is no longer considered to belong to *low* label.
- For values in between c^n_{low} and c^n_{high}, the membership function decreases linearly, reflecting a gradual transition of the nutrient amount.

4.1.2. The high-[nutrient] variable

For *high* label, we also use trapezoidal membership function $\mu^n_{\text{high}}(x)$ to represent the degree to which a nutrient amount x belongs to the *high* label. This function assigns a membership value in the range $\mu^n_{\text{high}} : I \to [0,1]$ based on predefined threshold values c^n_{low} and c^n_{high}. The function is defined as follows:

$\mu^n_{\text{high}} : I \to [0,1]$ defined as follows:

$$\mu^n_{\text{high}}(x; c^n_{\text{low}}, c^n_{\text{high}}) = \begin{cases} 1 & ; \quad x > c^n_{\text{high}} \\ 0 & ; \quad x < c^n_{\text{low}} \\ \frac{x - c^n_{\text{low}}}{c^n_{\text{high}} - c^n_{\text{low}}} & ; \quad c^n_{\text{low}} \le x \le c^n_{\text{high}} \end{cases} \tag{5}$$

where:

- For nutrient values above c^n_{high}, the food is fully considered to have a *high* amount of the nutrient.
- For values below c^n_{low}, the food is no longer considered to belong to *high* label.
- For values in between c^n_{low} and c^n_{high}, the membership function decreases linearly.

4.1.3. The medium-[nutrient] variable

We use a triangular membership function $\mu^n_{\text{medium}} : I \to [0,1]$. The main reason is due to the specificities of the nutrition domain. While the amount of a specific nutrient can be medium, this value can be closer to low or high values. With this function, we have modelled that medium values that are very close to both low and high values are not considered equals. It is defined as follows:

$$\mu^n_{\text{medium}}(x; c^n_{\text{medium}}, c^n_{\text{max}}) = \begin{cases} 0 & ; \quad x \ge c^n_{max} \\ \frac{x}{c^n_{\text{medium}}} & ; \quad x \le c^n_{\text{medium}} \\ \frac{c^n_{\text{max}} - x}{c^n_{\text{max}} - c^n_{\text{medium}}} & ; \quad c^n_{\text{medium}} < x < c^n_{\text{max}} \end{cases} \tag{6}$$

This fuzzification allows us to relate to unhealthy foods with sugar, salt, fat, and saturates that are higher than the daily maximum recommended amounts. Additionally, the *medium* variable allows us to measure if it is closer to low or high amounts.

4.1.4. Frequency of consumption

We have utilised an output variable called "consumption" to calculate a numerical score, which helps us determine the ideal frequency of consuming a particular food item. Based on the recommendations of nutrition experts, we have established four linguistic variables that represent this frequency: "daily" (for foods that could be consumed multiple times a day), "once per day", "weekly", and "occasional". We have used triangular membership functions for modelling weekly and once-per-day variables and trapezoidal functions for occasional and daily (see Fig. 7).

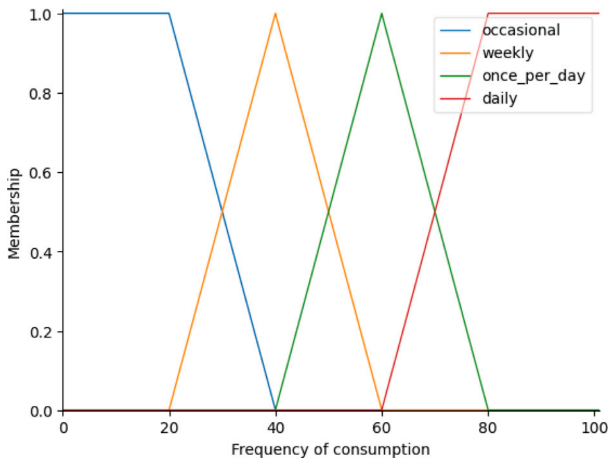


Fig. 7. Modelling of the frequency of consumption variable.

4.2. Fuzzy rules based on expert criteria

The four nutrients were correlated to build the rules of the system. This allows us to model the frequency of food use based on its nutritional composition. These rules were decided following two main principles: (1) the higher the amount of these nutrients, the lower should be the frequency of intake, and (2) the more nutrients with high amounts, the lower should be the frequency of intake.

We have defined a total of 15 fuzzy rules to model the behaviour of the *Frequency of consumption* output variable. First, we have modelled the situation with the highest value for the output variable. This is the case when the food has low amounts of fat, saturated fats, sugar and salt:

{Fat=low, Saturates=low, Sugar=low, Salt=low} → {Frequency of Consumption=daily}

We also consider that food can be consumed daily if it has a low amount of at least three of the four nutrients and a medium amount for the remaining one. One example is as follows:

{Fat=low, Saturates=low, Sugar=low, Salt=medium} → {Frequency of Consumption=daily}

We have also modelled scenarios where following expert advice should lead to lower consumption. Specifically, we have incorporated rules to ensure that foods containing two of these nutrients above low ranges are limited to once per day. For example, this is illustrated by the following rule:

{Fat=low, Saturates=medium, Sugar=medium, Salt=low} → {Frequency of consumption= once per day}

We force a weekly consumption of those foods in which at least three nutrients are in the medium interval, and the fourth is not high. For example:

{Fat=medium, Saturates=medium, Sugar=medium, Salt=medium} → Frequency of consumption=weekly}

In addition, the experts indicate that food should be consumed occasionally if at least one of the four nutrients is high:

{Fat=high, Saturates=low, Sugar=low, Salt=low} → {Frequency of Consumption=occasional}

Finally, the worst scenario is if the four variables have a high amount of the four nutrients. In this case, the frequency of consumption is also occasional:

{Fat=high, Saturates=high, Sugar=high, Salt=high} → {Frequency of Consumption=occasional}

We have modelled the system using the functionalities implemented in the *skfuzzy* library in Python. This library provides tools to create and manipulate fuzzy sets and define membership functions. It also allows us to construct the fuzzy control system by aggregating fuzzy rules to obtain a final crisp value for the recommended frequency of food consumption.

5. Experimentation set-up

Data. We have used two nutritional datasets to ensure diverse and representative coverage of foods across different contexts. The first one is the CoFID dataset,[5] maintained by the Public Health Agency (PHE) of the Department of Health and Social Care in England, with nutritional information of 2889 food items. This dataset contains many raw materials in the food industry, such as agricultural products such as grains, fruits, and vegetables, and animal-derived products such as milk or meat. Additionally, it contains numerous non-branded processed foods that the population often acquires in the supermarket. Therefore, this allows us to test our approach with different levels of food processing products commonly used by the population for cooking activities. The second one is the BEDCA dataset,[6] developed by the Spanish Agency for Food Safety and Nutrition (AESAN) in collaboration with academic and governmental institutions. BEDCA includes a total of 940 food items commonly consumed in Spain, covering both traditional ingredients and processed items typical of the Mediterranean diet. By incorporating BEDCA, we aim to complement CoFID with a culturally distinct perspective, enabling a broader evaluation of the model in terms of regional dietary patterns. For both datasets, we used the food description column as textual input for the prompt since it contains the name of the food. Preprocessing was applied to this column to remove non-alphanumeric characters and symbols and to standardise capitalisation. This preprocessing was consistent across all datasets used in the experimentation. From CoFID, we have used the *Proximates* table to obtain information regarding fats, saturates and sugar. Since measurement units in CoFID follow dietary standards, no conversion was needed. Additionally, we used the *Inorganics* table to obtain the salt content of the foods. However, CoFID provides sodium information instead of salt. As WHO guidelines are stated in terms of salt, and it is a more manageable nutrient for the population, we converted sodium to salt for our modelling. According to the CoFID database documentation, we multiplied these values by 2.5 to obtain the corresponding salt amount in grams.

Language model. We have used LLaMA3, an extension of LLaMA2 large language model [3] to generate the prompts, specifically the version with 8 billion parameters.[7] The maximum number of tokens to generate in the output has been set to 1024 and top_p to default, i.e., 0.9. It defines the size of the candidate word set (higher values of top_p means that the model considers a more significant number of possible words). As shown in Fig. 2, we have firmly delimited the output of the prompt, so we use this parameter to allow more expressivity. We run all the experiments with temperature values of 0.0, 0.6 and 0.9. This parameter models the randomness of the response and enables us to test our approach's robustness.

Fuzzy modelling. We have computed the fuzzy modelling for an individual of healthy body weight consuming approximately 2000 kilocalories daily. This average intake is recommended for maintaining weight and supporting normal physiological functions.

Prompt configuration. We generate the food textual recommendation for each food detailed in the CoFID and BEDCA databases. For this, we run the LLaMA3 model with three prompt strategies (fuzzy, crisp, and crisp2fuzzy) for the four roles (nutritionist, professor, chef, and non-expert person). In total, we produced 12 prompt experiments that we analysed regarding quality, similarity, and nutritional rigour.

[5] https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid

[6] https://www.bedca.net/

[7] meta-llama/Meta-Llama-3-8B-Instruct

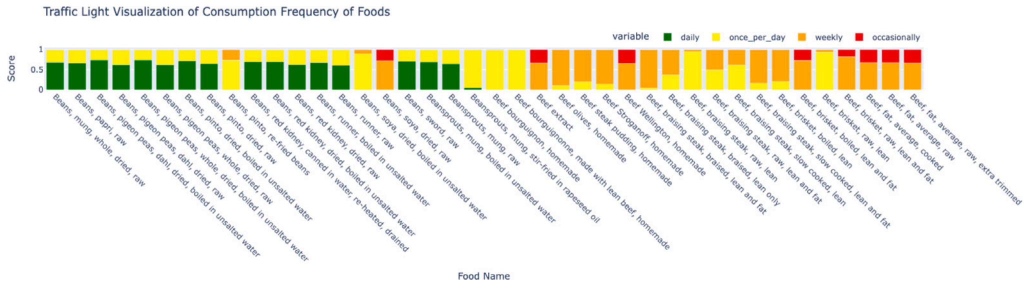


Fig. 8. A visualisation of the frequency of consumption score for the CoFID dataset.

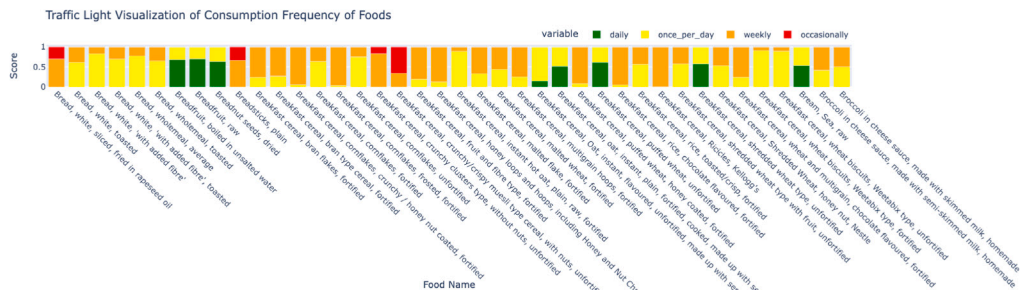


Fig. 9. A visualisation of the frequency of consumption score for the CoFID dataset.

6. Results and discussion

6.1. Fuzzy score for modelling frequency of consumption

Fig. 8 shows a visualisation of the score for selecting foods from the CoFID dataset. It comprises the membership value to each fuzzy variable for a food. If one variable does not appear for a food, the membership value to this fuzzy variable is zero. As can be appreciated in the figure, unprocessed foods have a more significant membership to *daily* and *once per day* variables, e.g., raw beans are mostly recommended to be eaten several times per day (if wish) because of their highly recommended properties. However, the higher the processing degree of the foods, the higher the membership value proportionally for weekly and occasional variables. For example, "Beansprouts, mung, raw" has a better nutritional profile than "Beansprouts, mung, stir-fried in rapeseed oil" since the latter contains fried processing, which is less recommended for a healthy lifestyle.

Fig. 9 shows another selection of foods, in this case, focused on bread and cereals. It shows consistent results; for example, the first two foods show disparities in the score because of the processing of the foods. In the case of the cereals (in the middle of the figure), their composition affects the membership to the variables. For example, those enriched with more ingredients show higher membership to occasional intake, thus being an indirect advantage of this score modelling.

Due to the fuzzy modelling we have designed, obtaining the frequency of consumption score is transparent. This allows us to systematically quantify and interpret the varying degrees of frequency in consumer behaviour. By assigning linguistic variables such as "low", "medium", and "high" to specific ranges of data and utilising fuzzy logic rules, we can derive a comprehensive frequency score. This approach enhances the interpretability and accuracy of our consumption analysis, ensuring that the scoring process remains clear and justifiable. Fig. 10 shows this casuistic for the food "Pate, Liver".

6.2. Generative AI-based recommendations

We have studied the performance of the prompt strategies and roles by analysing the similarity among the prompt responses. These metrics provide quantitative measures of how well two text responses are aligned. However, they do not fully capture the nuanced quality of the text or the specific nature and goals of the prompt responses for our domain-specific task. To address this, we have complemented these metrics by studying the quality of the prompt responses tailored to our task.

Our work analyses free-text outputs from language models in response to nutrition-related prompts. In this open-ended generation context, the lack of predefined correct answers or ranking criteria makes standard evaluation metrics such as precision, recall, or normalised discounted cumulative gain (NDCG) challenging to apply. We have followed a threefold approach to evaluate the quality of the prompt responses. First, we assessed the quality based on the role and prompt strategy, evaluating the AI-generated responses for clarity, subjectivity and polarity. Second, we have analysed the nutritional accuracy of the prompt responses. With the latter, we aim to capture the factual correctness, safety, and appropriateness of model-generated dietary recommendations. This metric functions similarly to accuracy, as it assesses the validity of the nutritional information provided — rated as low, medium, or high — based on comparisons with original values tailored to each user. This aligns more closely with hallucination detection and factual consistency verification than traditional information retrieval metrics. Third, we have conducted a human-based evaluation, where users reviewed a subset of responses to validate our automated assessments and provide qualitative insights into the perceived usefulness and reliability of the recommendations.

6.2.1. Similarity study among the prompts

We have computed three similarity metrics to compare the responses based on the prompt strategy and roles. SacreBLEU [39] is a version of the BLEU metric widely used in text generation and machine translation tasks. It measures the n-gram overlap between two texts regarding exactness and fidelity. Precisely, this metric assesses how many n-grams in the generated text match another text, typically the reference. ROUGE-L [40] is a metric that measures the longest common sub-sequence between two texts. It is focused on recall since its goal is to reflect the fluency and coherence of the generated text. BertScore [41] utilises pre-trained BERT embeddings to evaluate the semantic similarity between the generated and reference texts.

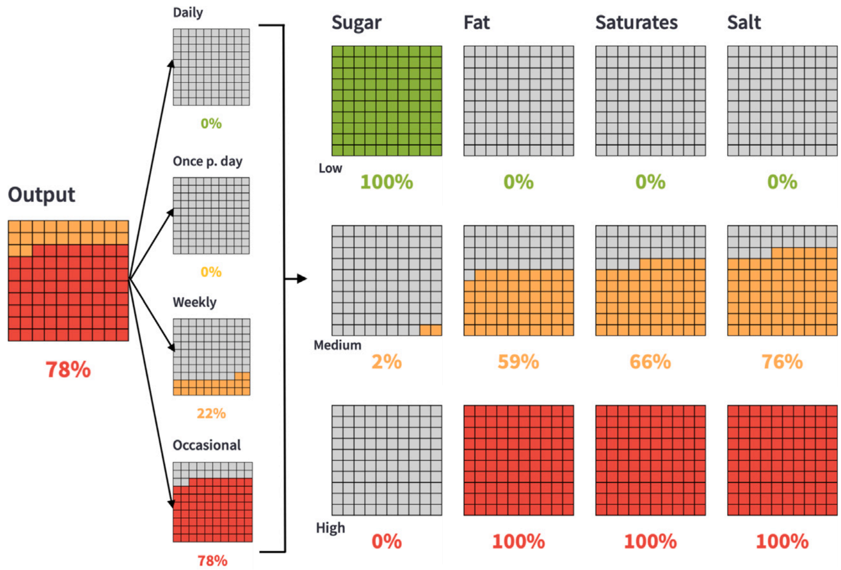


Fig. 10. Interpretation of the Frequency of Consumption score. The score (on the left) is constructed using the membership degree assigned to the frequency of consumption variables. This degree is determined based on the levels of sugar, fat, saturates, and salt present in the food items (low, medium, and high).

Table 4
Average similarity results for both datasets (CoFID and BEDCA) grouped by role.

Role A	Role B	CoFID			BEDCA		
		BERT Score	Sacre BLEU	ROUGE-L	BERT Score	Sacre BLEU	ROUGE-L
chef	nutri	**76.015**	**28.733**	43.464	74.722	31.864	44.314
chef	prof	75.647	27.751	**43.466**	74.585	30.877	43.683
nutri	prof	75.264	27.766	43.056	**76.597**	35.059	47.054
nutri	chef	74.357	24.593	40.772	75.230	34.441	46.391
nonexpert	nutri	73.612	22.699	39.135	75.500	**36.563**	**48.199**
chef	nonexpert	73.031	22.760	38.825	73.364	31.281	42.438
prof	nutri	72.304	21.468	36.560	75.597	34.302	47.059
nonexpert	chef	72.300	20.060	36.967	74.447	33.622	46.248
nonexpert	prof	72.273	19.666	37.334	74.679	34.085	46.979
prof	chef	72.020	20.826	36.338	74.595	32.901	45.636
nutri	nonexpert	70.957	19.266	35.171	74.873	33.535	45.117
prof	nonexpert	70.409	18.380	33.849	75.148	33.917	47.398

Unlike SacreBLEU and ROUGE-L, which focus on exact word matching, BertScore is designed to capture deeper contextual relationships between words, making it a valuable complement to the latter metrics.

We computed similarity metrics for each pair of strategy/role combinations, such as fuzzy/chef versus crisp/nutritionist responses. By aggregating the similarities grouping by role, we derived the average similarity scores presented in Table 4. Each row represents the average similarity for each pair of roles. The rows in the table are ordered in descending order based on their BERTScore values. Across both datasets, the non-expert role tends to produce outputs that are less similar to the reference, indicating weaker performance. In CoFID, non-experts consistently show lower performance across all metrics, with particularly low BERTScore values highlighting poor semantic similarity. In Bedca, the difference in BERTScore between roles is smaller, but the non-experts still perform worse on lexical metrics like BLEU and ROUGE.

To measure the clarity of the prompt responses, we studied the text readability, taking into account three scores that consider different dimensions of clarity and readability. Let W be the total number of words in the text and S and L the total number of sentences and syllables in the text, respectively. We have used the Simple Measure of Gobbledygook (SMOG) index, as seen in Eq. (7), a well-known readability measure that measures how easily a text is read. It quantifies the years of education needed to understand a specific text. This standard measure is used in public health to measure the understandability of health-related texts for the general public. Let P be the number of polysyllabic words. The Smog grade score is defined as follows:

$$\text{SMOG index} = 1.0430\sqrt{30\left(\frac{P}{S}\right)} + 3.1291. \tag{7}$$

As shown in Eq. (8), we have also included the Dale-Chall readability score, which allows us to measure readability in terms of word familiarity, which is a relevant quality in our task since text recommendations have to be easy to understand by the population. This metric compares the words in the text to a list of 3000 common words understood by 4th-grade students. Let DW be the number of complex words (not found on the previously mentioned list of familiar words). This score is defined as follows:

$$\text{Dale-Chall Score} = 0.1579\left(\frac{DW}{W} \times 100\right) + 0.0496\left(\frac{W}{S}\right). \tag{8}$$

The SMOG index and Dale-Chall metrics show higher average values for the fuzzy strategy, indicating high levels of readability and understandability in the text (see Figs. 11 and 12). Regarding role performance for the fuzzy strategy, the non-expert role shows lower performance in both scores, while the texts generated with the professor role show higher performance.

Additionally, we analyse the polarity and subjectivity of each prompt response using the functionalities provided by the Python library Textblob. The polarity measure ranges from −1 to 1. Values closer

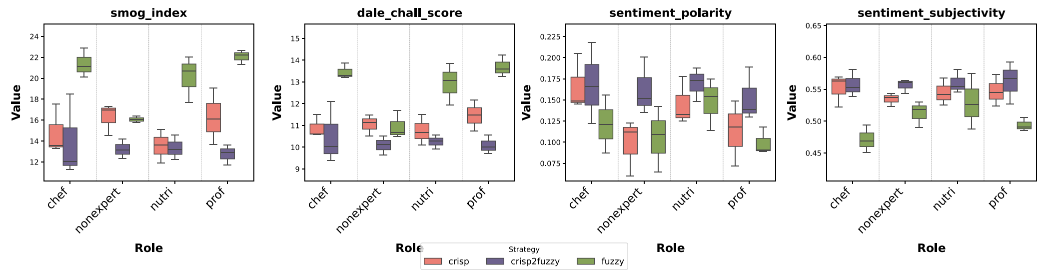


Fig. 11. Comparison of the text quality metrics detailed in Table 5 (CoFID dataset).

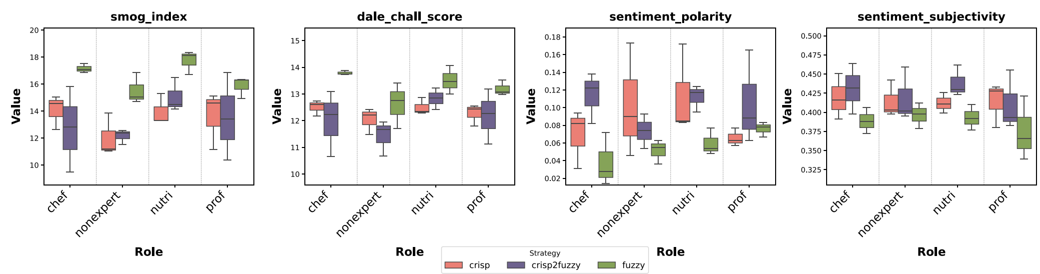


Fig. 12. Comparison of the text quality metrics detailed in Table 5 (BEDCA dataset).

to −1 refer to negative sentiments, 0 to neutral and 1 to positive. In our case, nutritional recommendations should remain neutral because maintaining a balanced perspective is crucial for providing reliable and unbiased information to users. Likewise, nutritional information should be as less subjective as possible since the information that users receive must not lead to misunderstandings. As shown in Figs. 11 and 12, using the fuzzy strategy achieves greater results for polarity and subjectivity for both datasets.

Table 5 shows the quantitative results for the quality check of all strategy/role combinations across different temperature values. On average, the responses obtained with the fuzzy strategy show better behaviour than crisp and crisp2fuzzy. Using a high-temperature value allows us to increase the text quality overall. In general, increasing the temperature generally helps achieve higher maximum scores across both datasets, whereas temperature 0 reaches fewer performance peaks.

6.2.2. Nutritional rigour

One key challenge in large language models is tackling hallucinations, inconsistencies and vague vocabulary in the prompt responses. In nutrition applications, experts demand the information these models provide to be exact, meticulous and rigorous. Motivated by this, we have analysed the model responses for all the possible strategy/role combinations. Our analysis involves a comprehensive examination of the text to determine if the model has correctly assigned the nutrition variable to each nutrient (low, medium, high). With this measure, we convey the concept of a strict and precise evaluation or practice regarding nutrition, which is essential for providing high-quality food recommendations. The results of this analysis, as shown in Table 6, reveal the mean accuracy for each nutrient across all the pairs of strategy/role combinations for the prompt. From the results, we extract five main findings:

1. Fig. 13 shows a visual comparison of nutritional rigour performance for the CoFID dataset, considering the standard deviation of different temperature values. It illustrates that the performance is considerably higher when enriching the prompt with fuzzy variable information. Providing the nutrition quantities (crisp) or the intervals for the variable membership (crisp2fuzzy). A similar pattern is observed in BEDCA (Fig. 14), where the fuzzy strategy also achieves the highest performance.
2. The responses with greater nutritional accuracy are provided by using the role of a professor of nutrition. Fig. 15 illustrates the results obtained for each role in the CoFID dataset, showing the standard deviation of the three temperature values. The roles of non-experts and nutritionists follow, and in most cases, the role of a chef produces the least accurate results. In BEDCA (Fig. 16), this patterns also repeat, achieving the role of professor the better results for the fuzzy strategy (the one with better performance). For the other strategies, the performance of the professor role remains high, but the differences across roles are less pronounced.
3. In crisp and crisp2fuzzy strategies, we provide the nutrition information and the low, medium, and high intervals. Even though the prompt fails to infer precisely this information in more than half of the foods for both datasets.
4. Enriching large language models with crisp nutrition amounts is not enough; they need external knowledge to decide the health of foods. Thus, using this information solely is inappropriate for this task, but it could be greatly improved with fuzzy modelling enrichment.
5. After replicating the study with different temperature values, the conclusions remained consistent, demonstrating the robustness of the results across varying levels of output expressiveness and creativity.
6. Overall, combining a fuzzy strategy and a professor role achieves higher performance for all the experiments.

6.2.3. Human-based evaluation

We have conducted user-centred evaluations to assess the effectiveness of the generated explanations in real-world scenarios. To this end, we designed an annotation platform where users are presented with a variety of food items and their corresponding textual recommendations. They are then asked a series of questions about them. For each explanation, users perform two annotation tasks:

1. **Usability study: evaluating how well users understand and trust the generated explanations.** In this task, users are shown several food options accompanied by text recommendations. For each explanation, they assess: **clarity** (is the explanation easy

Table 5
A comparison of SMOG, Dale–Chall (D–C), subjectivity (subj.), and polarity (pol.) values across different temperature (T) strategy/role configuration prompts.

T	Strategy	Role	CoFID				Bedca			
			SMOG	D-C	Pol.	Subj.	SMOG	D-C	Pol.	Subj.
0.0	crisp	chef	13.576	10.618	0.205	0.569	15.028	12.608	0.082	0.451
	crisp	nonexpert	14.53	10.525	0.112	0.543	13.871	12.427	0.090	**0.403**
	crisp	nutri	13.617	10.677	0.178	0.542	15.315	12.873	0.085	0.411
	crisp	prof	13.654	10.745	0.149	0.545	14.592	12.558	0.063	0.433
	fuzzy	chef	20.112	13.209	0.087	**0.469**	17.526	13.887	0.072	0.406
	fuzzy	nonexpert	16.039	10.492	**0.065**	0.49	14.727	11.705	0.063	0.412
	fuzzy	nutri	17.688	11.942	0.154	0.575	**18.117**	**14.072**[a]	**0.048**	0.392
	fuzzy	prof	**21.3**	**13.247**	0.091	0.485	16.327	13.528	0.083	0.421
	crisp2fuzzy	chef	11.301	9.382	0.166	0.553	15.840	13.099	0.122	0.464
	crisp2fuzzy	nonexpert	13.126	9.649	0.135	0.564	11.537	10.681	0.093	0.459
	crisp2fuzzy	nutri	12.244	10.277	0.173	0.554	16.481	13.222	0.117	0.462
	crisp2fuzzy	prof	11.69	9.712	0.139	0.593	16.854	13.179	0.088	0.455
0.6	crisp	chef	13.303	10.583	0.145	0.563	12.626	12.179	0.031	0.416
	crisp	nonexpert	16.98	11.474	0.123	0.537	11.038	11.480	0.046	0.398
	crisp	nutri	11.885	10.113	0.125	0.568	13.316	12.337	0.083	0.399
	crisp	prof	16.116	11.47	**0.072**	0.524	11.161	11.808	0.077	0.380
	fuzzy	chef	**22.89**[a]	**13.871**	0.156	**0.494**	17.086	**13.739**	0.014	0.388
	fuzzy	nonexpert	15.756	10.672	0.109	0.53	15.040	12.762	0.055	0.398
	fuzzy	nutri	20.697	13.056	0.175	0.526	**18.334**[a]	13.463	0.054	0.377
	fuzzy	prof	22.213	13.589	0.118	0.506	14.928	12.981	0.067	**0.339**[a]
	crisp2fuzzy	chef	18.495	12.092	0.218	0.581	12.815	12.223	**0.138**	0.432
	crisp2fuzzy	nonexpert	14.217	10.122	0.201	0.561	12.373	11.942	0.054	0.395
	crisp2fuzzy	nutri	13.21	9.926	0.188	0.581	14.500	12.855	0.095	0.430
	crisp2fuzzy	prof	13.619	10.566	0.189	0.567	10.368	11.132	0.165	0.393
0.9	crisp	chef	17.559	11.515	0.149	0.522	14.554	12.729	0.094	0.391
	crisp	nonexpert	17.316	11.138	**0.06**[a]	0.523	11.186	12.222	0.173	0.442
	crisp	nutri	15.099	11.507	0.133	0.525	13.309	12.296	0.172	0.426
	crisp	prof	19.08	12.164	0.118	0.573	15.106	12.446	0.057	0.428
	fuzzy	chef	21.113	13.296	0.121	**0.451**[a]	**16.850**	**13.756**	0.028	0.372
	fuzzy	nonexpert	16.408	11.688	0.142	0.518	16.843	13.412	**0.036**[a]	0.379
	fuzzy	nutri	22.026	13.84	0.114	0.488	16.702	12.999	0.077	0.410
	fuzzy	prof	**22.677**	**14.236**[a]	0.089	0.491	16.299	13.079	0.078	**0.366**
	crisp2fuzzy	chef	12.033	10.026	0.122	0.539	9.481	10.661	0.082	0.398
	crisp2fuzzy	nonexpert	12.348	10.51	0.152	0.543	12.561	11.671	0.074	0.402
	crisp2fuzzy	nutri	14.601	10.559	0.148	0.546	14.169	12.422	0.124	0.423
	crisp2fuzzy	prof	12.898	10.004	0.13	0.527	13.397	12.273	0.063	0.383

[a] Indicates the maximum value for each nutrient.

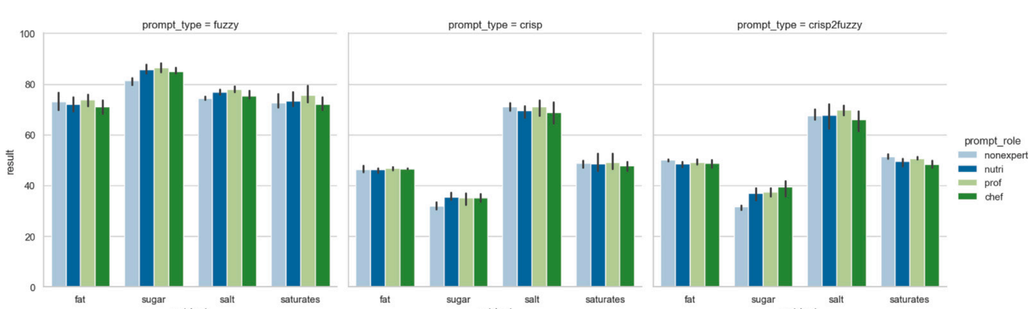


Fig. 13. Nutritional rigour for each strategy used in the prompt (CoFID dataset).

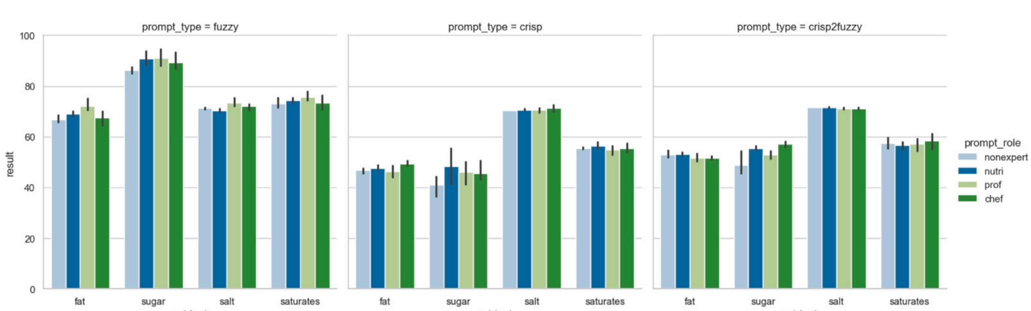


Fig. 14. Nutritional rigour for each strategy used in the prompt (BEDCA dataset).

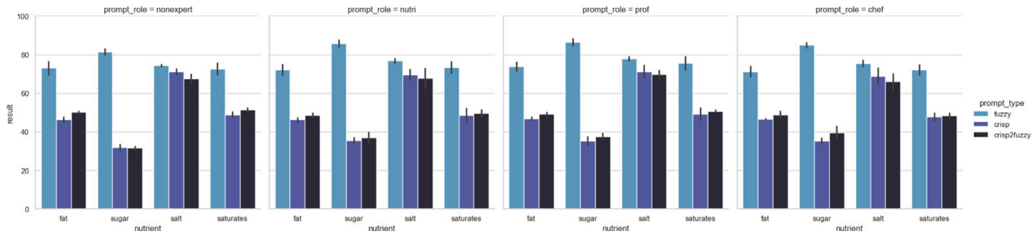


Fig. 15. Nutritional rigour for each role used in the prompt (CoFID dataset).

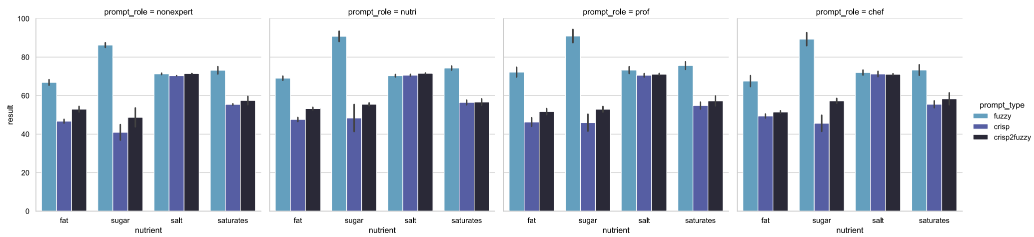


Fig. 16. Nutritional rigour for each role used in the prompt (BEDCA dataset).

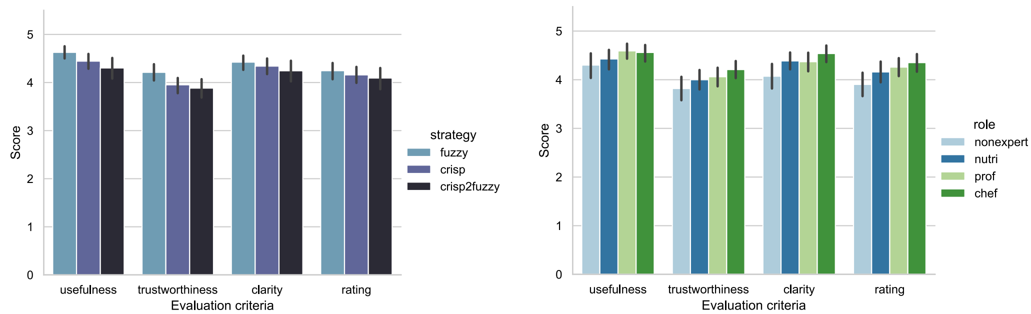
Table 6
Nutritional rigour of the prompt responses. The highest value for each temperature is indicated in bold.

T	Strategy	Role	CoFID				BEDCA			
			Fat	Salt	Saturates	Sugar	Fat	Salt	Saturates	Sugar
0.0	crisp	chef	48.944	74.939	61.405	36.726	48.504	71.230	55.138	43.204
	crisp	nonexpert	46.798	74.005	57.009	31.533	45.620	70.374	55.359	36.408
	crisp	nutri	49.083	73.763	62.375	36.345	47.009	70.267	55.359	41.262
	crisp	prof	48.079	74.939	61.648	34.199	44.124	70.909	52.928	41.262
	crisp2fuzzy	chef	50.952	70.786	55.59	42.887	51.068	71.444	55.028	58.252
	crisp2fuzzy	nonexpert	50.64	71.305	54.517	32.503	54.594	71.444	55.359	46.117
	crisp2fuzzy	nutri	50.156	73.451	55.729	38.56	52.564	71.337	54.807	55.340
	crisp2fuzzy	prof	49.117	72.966	56.525	38.56	53.419	71.016	54.365	52.913
	fuzzy	chef	74.628	78.089	79.128	87.504	69.979	72.727	76.243	93.204
	fuzzy	nonexpert	**76.982**[a]	74.87	78.816	80.305	68.483	71.658	75.249	86.408
	fuzzy	nutri	75.32	78.712	79.232	88.058	70.085	70.267	75.470	93.689
	fuzzy	prof	76.393	**79.993**[a]	**81.931**[a]	**88.75**[a]	**75.107**[a]	**75.294**[a]	**77.901**[a]	**94.660**[a]
0.6	crisp	chef	49.706	72.136	58.671	40.256	50.534	72.620	53.923	43.204
	crisp	nonexpert	47.456	72.724	57.909	34.683	47.436	70.374	55.801	42.233
	crisp	nutri	49.083	72.62	58.221	39.875	47.009	71.123	56.243	48.544
	crisp	prof	48.702	73.451	56.629	38.802	46.154	71.230	55.138	46.602
	crisp2fuzzy	chef	51.229	69.436	56.386	42.506	51.175	70.909	58.785	55.825
	crisp2fuzzy	nonexpert	51.575	67.774	56.421	33.16	51.816	71.444	57.348	45.631
	crisp2fuzzy	nutri	50.848	70.682	56.906	40.879	53.205	71.765	57.348	54.854
	crisp2fuzzy	prof	49.637	71.028	55.348	40.118	50.321	71.551	59.116	54.369
	fuzzy	chef	72.205	75.32	75.632	85.601	68.269	**72.727**	72.818	87.864
	fuzzy	nonexpert	73.139	75.389	75.32	83.42	66.346	71.337	72.707	87.379
	fuzzy	nutri	72.586	76.67	76.22	86.397	69.124	71.016	73.370	**90.291**
	fuzzy	prof	**74.801**	**78.366**	**79.439**	**87.574**	**70.833**	72.513	**74.586**	**90.291**
0.9	crisp	chef	52.302	70.855	58.186	40.775	49.252	69.840	57.459	50.485
	crisp	nonexpert	48.564	70.855	53.375	32.364	47.115	70.267	55.138	44.175
	crisp	nutri	52.198	74.178	58.152	41.606	48.825	70.374	57.790	55.340
	crisp	prof	48.771	69.747	53.617	37.245	48.611	69.519	56.464	50.000
	crisp2fuzzy	chef	54.067	68.986	58.325	42.575	52.244	70.802	61.215	57.767
	crisp2fuzzy	nonexpert	51.194	67.774	54.794	30.876	52.457	71.444	59.669	54.369
	crisp2fuzzy	nutri	54.586	70.128	57.944	41.087	53.953	71.551	57.901	56.311
	crisp2fuzzy	prof	51.436	69.263	54.863	36.691	51.389	70.695	58.343	51.456
	fuzzy	chef	68.778	75.666	76.67	84.666	64.423	70.481	70.829	86.893
	fuzzy	nonexpert	70.059	75.666	75.286	82.485	65.705	70.909	71.492	84.951
	fuzzy	nutri	69.643	77.085	75.147	84.354	67.949	69.733	74.144	**88.350**
	fuzzy	prof	**72.24**	**78.643**	**77.812**	**85.497**	**70.620**	**71.979**	**74.365**	87.864

[a] Refers to the maximal value for each nutrient.

to understand?), **usefulness** (does it provide relevant dietary insights?), and **trustworthiness** (does it seem reliable and scientifically accurate?). We included an additional question, in which annotators have to provide a **general rating** for the whole explanation. Users rate each of these aspects on a scale from 1 (poor) to 5 (excellent). This task helps us evaluate the usability and perceived quality of the explanations from the user's perspective.

2. **Impact study: analysing if the explanations influence user choices.** In this task, users report how frequently they think a given food must be consumed in a healthy diet (several times per day, daily, weekly, or occasionally) both before and after reading



(a) Results grouped by strategy

(b) Results grouped by role

Fig. 17. Human evaluation scores by strategy and role. It contains the average results of the annotators for usefulness, trustworthiness, clarity, and general rating of the explanations.

the explanation. Then, we measure changes in their responses to assess the potential impact of the explanation on their awareness of food healthiness.

We randomly selected the food items to annotate, ensuring equal representation across all frequency-of-consumption categories. We used the recommendations generated with temperature 0, as these showed the best performance in our prior evaluations. A total of 15 items were chosen from each dataset and tested across all combinations of strategy and role, resulting in 330 distinct annotations from users.

15 annotators participated in the study. The age range was 22 to 64, and the gender distribution of the annotators was 9 male vs 6 female, with 0 identifying as non-binary or preferring not to disclose. In terms of expertise, 9 considered themselves beginners, 5 intermediate, and 1 advanced.

Fig. 17 presents the results from the usability study. Overall, the average scores were high, exceeding 4 in all categories: clarity (4.35/5), usefulness (4.48/5), trustworthiness (4.1/5), and overall rating (4.2/5). These results demonstrate the strong performance of our approach, regardless of the strategy and role used to generate the explanations. When analysing the human annotation scores by strategy (see Fig. 17(a)), the fuzzy strategy consistently achieves the highest scores, although the differences are relatively small. In contrast, the crisp and crisp2fuzzy strategies show smaller variations, especially in the overall rating, where the average difference is minimal. Focusing on the role dimension (see Fig. 17(b)), the results align with our previous findings: the non-expert role received the lowest scores across all dimensions. In comparison, the explanations generated by the chef and professor roles were rated higher by annotators, although again, the differences were slight

To evaluate the impact of the explanations on users' nutritional awareness, we compared their responses about consumption frequency before and after reading the explanations. For this analysis, we treated the responses as ordinal data with the following ordered categories: occasional < weekly < once-a-day < daily. We mapped these categories to integers to compute the absolute difference between each user response and the reference value (i.e., the label assigned by the fuzzy modelling). Then, we checked if the post-reading explanation responses were close to the reference values. Results showed an improvement of 14.85% in the post-explanation responses. We confirmed with a Wilcoxon signed-rank test that this reduction in error was statistically significant ($W = 1208.0, p = 0.019$), supporting the hypothesis that the explanations positively influenced users' dietary decision-making.

6.3. Computational cost and scalability

Running time. Our experiments were run on two NVIDIA RTX A5000 GPUs (24 GB VRAM each), with CUDA version 12.6 and driver version 560.35.03. During inference, each GPU operated at moderate utilisation (23%–32%) and used between 8.4–10.2 GB of memory. In practice, without using batching and computing one instance at a time to generate its explanation, we observed a running time of approximately 4–5 s per explanation, depending on the final length of the generated text.

Scalability of the approach. The computational complexity of prompting LLaMA is $O(n^2 \cdot d)$. Since it is transformer-based, its inference time is defined by the self-attention mechanism, which scales with $O(n^2 \cdot d)$, where n is the sequence length and d is the embedding dimension (for LLaMA-7B, $d = 4096$). Therefore, the computational cost is dominated by n. In our case, we use a maximum sequence length of $n = 1024$, although on average, the number of tokens per recommendation does not exceed 140 (see Fig. 18).

6.4. Ablation study

From a general perspective, this section analyses the different variants we can use with our proposed method. Specifically, the text readability and nutritional rigour are assessed on both datasets.

As one of our main contributions is the use of fuzzy linguistic variables for improving personalised healthy food text recommendations utilising generative AI, we have decided to analyse the three prompt strategies presented in this work (crisp, crisp2fuzzy and fuzzy), with three different values of model temperature and taking into account the mean for all supported roles (nutritionist, professor, chef, and non-expert person). With this, we aim to study the results when no modelling is introduced in the prompt (crisp), when our modelling approach is used (fuzzy), and when the crisp value and the intervals used in the modelling are provided (crisp2fuzzy, a middle ground between the previous strategies).

Table 7 showcases that the proposed approach, based on fuzzy linguistic variables (*fuzzy*), achieves the best performance for almost all the analysed measures. This entails that the fuzzy-based strategy improves the proposal from an interpretability and understandability perspective. We must highlight that this strategy outperforms the other two for all the measures used to assess the nutritional rigour, ensuring better health-aware recommendations.

6.5. Explainability, interpretability, and other considerations

Our approach offers textual recommendations guided by experts, enhancing large language models with fuzzy nutrition modelling of foods. Our goal is to mitigate inaccuracies, which are crucial when deploying general artificial intelligence models for specialised applications, such as those in nutrition. In our case, we have enriched the prompt with a personalised fuzzy modelling of the health information of the ingredients, showing a superior performance in terms of nutritional accuracy. With our approach, we contribute towards trustworthy applications of AI-generated models, which are beneficial for informed decision-making.

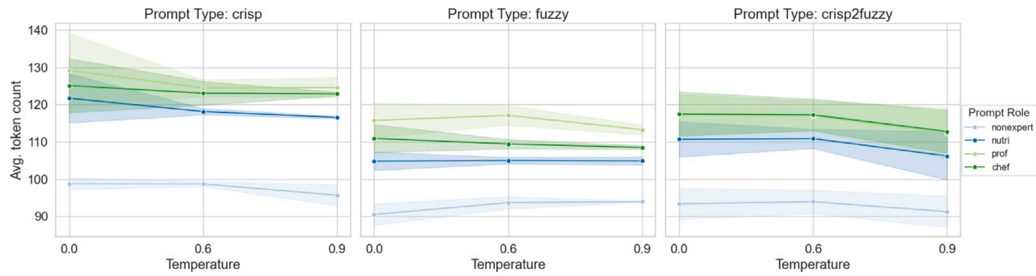


Fig. 18. Average token count per strategy and role across temperature values, aggregated over both datasets (CoFID and BEDCA). Note that in the crisp2fuzzy strategy (on the right), the roles of the professor and chef overlap.

Table 7
Averaged SMOG, Dale–Chall (D–C), polarity (Pol.), subjectivity (Subj.), and nutritional values grouped by strategy. The highest value per metric and temperature is highlighted in bold with an asterisk (*).

Dataset	Measure	T = 0.0			T = 0.6			T = 0.9		
		crisp	crisp2fuzzy	fuzzy	crisp	crisp2fuzzy	fuzzy	crisp	crisp2fuzzy	fuzzy
CoFID	SMOG	13.844	12.090	**18.785***	14.571	14.885	**20.389***	17.763	12.970	**20.556***
	D–C	10.641	9.755	**12.222***	10.910	10.676	**12.797***	11.581	10.275	**13.265***
	Pol.	**0.161***	0.153	0.099	0.116	**0.199***	0.139	0.115	**0.138***	0.117
	Subj.	0.550	**0.566***	0.505	0.548	**0.572***	0.514	0.536	**0.539***	0.487
	Fat	48.226	50.216	**75.831***	48.737	50.822	**73.183***	50.959	52.821	**70.180***
	Salt	74.412	72.627	**77.893***	72.733	69.730	**76.436***	71.409	69.038	**76.268***
	Saturates	60.609	55.590	**79.277***	57.858	56.765	**76.423***	55.833	56.982	**76.229***
	Sugar	34.701	38.628	**86.654***	38.404	39.668	**85.748***	38.998	37.807	**84.250***
BEDCA	SMOG	14.702	15.178	**16.674***	12.035	12.514	**16.847***	13.539	12.902	**16.674***
	D–C	12.616	12.545	**13.298***	11.951	12.038	**13.486***	12.423	11.757	**13.561***
	Pol.	0.080	**0.105***	0.066	0.059	**0.113***	0.048	**0.124***	0.086	0.055
	Subj.	0.424	**0.460***	0.408	0.398	**0.412***	0.375	**0.422***	0.402	0.382
	Fat	46.819	52.911	**70.389***	47.283	51.629	**68.143***	48.451	52.511	**67.173***
	Salt	70.620	71.310	**72.137***	71.338	71.417	**71.648***	69.500	71.123	**71.281***
	Saturates	54.946	54.890	**76.716***	55.801	58.649	**73.370***	56.713	59.232	**72.708***
	Sugar	40.534	53.156	**91.740***	45.146	52.170	**88.456***	50.000	55.976	**86.515***

However, the modelling proposed in this study does not incorporate historical user data. This implies that a daily diet should not consist solely of combinations of ingredients meant for occasional consumption but should rather maintain a balance of various available foods. The approach presented in this paper does not consider prior user information to adjust remaining daily or weekly calorie intake when providing text recommendations. These textual recommendations provide the general population with personalised explanations regarding the frequency of food consumption in their daily lives. Integrating cumulative energy intake with textual recommendations remains a future prospect for our approach in larger dietary assessment systems.

Employing a prompt-based strategy to generate textual recommendations enhances explainability. This approach provides transparency into the model information processing and output generation, as the prompt aids in understanding the rationale behind its decisions, thereby promoting trust in its outcomes. Additionally, using fuzzy modelling to enrich the prompt further enhances transparency and interpretability in the procedure, as previously stated in Fig. 10. Additionally, our modelling is not dependent on gender data. We adapt the modelling to male, female, and non-binary individuals; the latter also includes cases in which they prefer not to indicate gender, ensuring privacy and intimacy.

The application detailed in this study is essential in two ways. First, it helps to build educational nutritional dialogue tasks which provide accurate and user-friendly information in natural language. In this way, users can access personalised health information that can be incorporated into healthy cooking activities. Second, due to our approach's low complexity and generalisation capabilities, it is easy to integrate into existing platforms, unlike knowledge-based methods. This is crucial in food computing due to the added complexity of integrating knowledge from different origins, data structures, and scopes [11].

We can apply our approach to building broader health-aware applications such as recipe and diet health-aware recommendation systems, domain-focused chatbot applications, or fine-grained recipe analysis while considering trustworthiness and interpretability. Note that the methodology we focus on is not dependent on the specific rules used for fuzzy reasoning. These rules can be reformulated to address the particular constraints of the recommender system and can be dynamically updated to adapt to user needs, considering required intake adjustments due to dietary updates. Nutritional food databases typically share a standardised structure, encompassing, at a minimum, textual descriptions of the foods and their corresponding dietary information. Consequently, adapting the database to a different nutritional resource becomes effortless, ensuring the proposed model geographic interoperability.

Additionally, the approach detailed in this study applies to other fields of application that involve modelling based on expert guidance. In particular, fuzzy modelling, context-enriched prompting, and text outputs can be extended to domains such as healthcare assistance or financial advice, where recommendations align with established guidelines while adapting to individual profiles. For example, in healthcare, similar methods could support personalised treatment suggestions that apply clinical thresholds but remain flexible to the patient's conditions. In financial assessment systems, fuzzy constraints could represent varying risk tolerances or investment goals, while natural language explanations enhance user understanding and trust.

Limitations and challenges. While this study proposes a flexible and context-aware system for health-oriented food recommendation, several limitations must be acknowledged. First, the nutritional analysis currently focuses on a core set of macronutrients and critical components—namely fat, saturated fat, sugar, and salt. This selection reflects practical and communicative considerations: these nutrients are commonly used in public health labelling schemes, consistently available across food databases, and easily interpretable by a general audience. However, this focus does not encompass other dietary factors such as fibre, vitamins or minerals, which also play essential roles in overall dietary quality and health outcomes. One of the challenges in incorporating these additional elements lies in the limited and often inconsistent availability of detailed micronutrient information in existing food composition datasets. Nonetheless, our proposed framework is designed to be modular and extensible. As nutritional databases become more complete and as public literacy around broader nutritional components increases, future iterations of the system can integrate these additional nutrients without requiring changes to the core methodology.

One limitation of this approach is that explanations rely on the ingredients available in the database, as it is required for extracting nutritional information. In recommendation scenarios, where ingredient databases with nutritional information are typically used to build recommended menus, this is not a major issue. However, in other applications, it is important to consider that some ingredients may be excluded from the fuzzy modelling process, preventing the generation of explanations for them. This limitation also extends to missing values in the database, where appropriate data imputation strategies must be applied to address the issue.

With our approach, the explanations are personalised in terms of content but not in terms of style. Although we have included several metrics to ensure the quality of the prompts in terms of understandability and appropriateness for users, further refinements could enhance the personalisation of the prompt style. Tailoring the style of explanations could make them more engaging and better suited to individual user preferences and specific culinary contexts.

Considerations on nutritional data and model reliability. Regarding nutritional advice, we follow the WHO dietary guidelines, which are internationally recognised, broadly validated, and designed to support healthy eating across diverse populations. For food composition data, we rely on two official and well-established sources — CoFID and BEDCA. While we acknowledge that such databases may include minor inaccuracies due to regional variability, food processing differences, or measurement techniques, they are widely used in public health and nutrition research and provide a solid foundation for our modelling. While the fuzzy modelling in our approach adds flexibility, we are aware it may introduce subtle biases through rule design or threshold tuning. The most critical risk in this context would be to incorrectly label a food item or recommendation as healthy when it is not. To minimise this risk, we apply strict criteria for the "high" nutritional rating, aligning closely with WHO thresholds to ensure that our recommendations remain conservative, thus avoiding overestimating healthiness.

7. Conclusions and future work

The fuzzy linguistic variables enrich the prompt with fuzzy reasoning, enhancing the model performance. We demonstrate that incorporating this reasoning leads to more accurate and reliable results than providing the model with explicit values and intervals. Our findings indicate promising and reliable outcomes, opening a new avenue for including fuzzy modelling of linguistic variables to enrich the input of large language models for healthy-aware food recommendations.

We plan to extend this study to consider more nutrients, especially health benefits, such as vitamins. Additionally, we plan to integrate this approach in a recipe-level recommendation approach to study the role of this modelling when computing recommendations to generate dietary plans.

CRediT authorship contribution statement

Andrea Morales-Garzón: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Ana María Rojas-Carvajal:** Writing – review & editing, Visualization, Validation, Methodology. **Roberto Morcillo-Jimenez:** Writing – review & editing, Methodology, Investigation. **Maria J. Martin-Bautista:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Karel Gutiérrez-Batista:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially supported by the Grant PID2021-123960OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe. It was also funded by "Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía" through a pre-doctoral fellowship program (Grant Ref. PREDOC_00298). In addition, this research has been partially supported by the project CITIC-2024-06, funded by the Research Center for Information and Communication technologies of the University of Granada. Funding for open access charge: Universidad de Granada/CBUA.

Data availability

Data will be made available on request.

References

[1] T.N.T. Tran, A. Felfernig, C. Trattner, A. Holzinger, Recommender systems in the healthcare domain: state-of-the-art and research issues, J. Intell. Inf. Syst. 57 (1) (2021) 171–201.

[2] A.Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D.S. Chaplot, D.d.l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, et al., Mistral 7B, 2023, arXiv preprint arXiv:2310.06825.

[3] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, 2023, arXiv preprint arXiv:2307.09288.

[4] M.B. Garcia, ChatGPT as a virtual dietitian: Exploring its potential as a tool for improving nutrition knowledge, Appl. Syst. Innov. 6 (5) (2023) 96.

[5] L. Wu, Z. Zheng, Z. Qiu, H. Wang, H. Gu, T. Shen, C. Qin, C. Zhu, H. Zhu, Q. Liu, et al., A survey on large language models for recommendation, 2023, arXiv preprint arXiv:2305.19860.

[6] P. Niszczota, I. Rybicka, The credibility of dietary advice formulated by ChatGPT: robo-diets for people with food allergies, Nutrition 112 (2023) 112076.

[7] S. Arslan, Exploring the potential of chat GPT in personalized obesity treatment, Ann. Biomed. Eng. 51 (9) (2023) 1887–1888.

[8] H. Kotek, R. Dockum, D. Sun, Gender bias and stereotypes in large language models, in: Proceedings of the ACM Collective Intelligence Conference, 2023, pp. 12–24.

[9] X. Wang, K. Zhou, J.-R. Wen, W.X. Zhao, Towards unified conversational recommender systems via knowledge-enhanced prompt learning, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1929–1937.

[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Adv. Neural Inf. Process. Syst. 33 (2020) 9459–9474.

[11] A. Morales-Garzón, K. Gutiérrez-Batista, M.J. Martin-Bautista, Link prediction in food heterogeneous graphs for personalised recipe recommendation based on user interactions and dietary restrictions, Computing (2023) 1–23.

[12] C. Fernandez-Basso, K. Gutiérrez-Batista, R. Morcillo-Jiménez, M.-A. Vila, M.J. Martin-Bautista, A fuzzy-based medical system for pattern mining in a distributed environment: Application to diagnostic and co-morbidity, Appl. Soft Comput. 122 (2022) 108870.

[13] C. Harper, M. Siller, OpenAG: a globally distributed network of food computing, IEEE Pervasive Comput. 14 (4) (2015) 24–27.

[14] W. Min, S. Jiang, L. Liu, Y. Rui, R. Jain, A survey on food computing, ACM Comput. Surv. 52 (5) (2019) 1–36.

[15] M. Rostami, V. Farrahi, S. Ahmadian, S.M.J. Jalali, M. Oussalah, A novel healthy and time-aware food recommender system using attributed community detection, Expert Syst. Appl. 221 (2023) 119719.

[16] A. Morales-Garzón, J. Gómez-Romero, M.J. Martin-Bautista, A word embedding-based method for unsupervised adaptation of cooking recipes, IEEE Access 9 (2021) 27389–27404.

[17] A. Morales-Garzón, J. Gómez-Romero, M.J. Martín-Bautista, Contextual sentence embeddings for obtaining food recipe versions, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2022, pp. 306–316.

[18] C. Pellegrini, E. Özsoy, M. Wintergerst, G. Groh, Exploiting food embeddings for ingredient substitution, in: HEALTHINF, 2021, pp. 67–77.
[19] C. Trattner, D. Elsweiler, Food recommender systems: Important contributions, challenges and future research directions, 2017, arXiv:1711.02760.
[20] R.Y. Toledo, A.A. Alzahrani, L. Martinez, A food recommender system considering nutritional information and user preferences, IEEE Access 7 (2019) 96695–96711.
[21] A.A. Shari, N.A. Pajar, N. Sabri, M.R.M. Noordin, F.M.I. Zainudin, A.S. Shari, A. Ahmad, Mobile application of food recommendation for allergy baby using rule-based technique, in: 2019 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS, IEEE, 2019, pp. 273–278.
[22] K.-J. Kim, C.-H. Chung, Tell me what you eat, and i will tell you where you come from: A data science approach for global recipe data on the web, IEEE Access 4 (2016) 8199–8211.
[23] C. Trattner, D. Elsweiler, Food recommender systems: Important contributions, challenges and future research directions, 2017, ArXiv abs/1711.02760, URL: https://api.semanticscholar.org/CorpusID:23192718.
[24] X. Gao, F. Feng, H. Huang, X.-L. Mao, T. Lan, Z. Chi, Food recommendation with graph convolutional network, Inform. Sci. 584 (2022) 170–183.
[25] F. Pecune, L. Callebert, S. Marsella, A recommender system for healthy and personalized recipes recommendations, in: HealthRecSys@ RecSys, 2020, pp. 15–20.
[26] T.N. Trang Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, J. Intell. Inf. Syst. 50 (2018) 501–526.
[27] M. Chen, X. Jia, E. Gorbonos, C.T. Hoang, X. Yu, Y. Liu, Eating healthier: Exploring nutrition information for healthier recipe recommendation, Inf. Process. Manage. 57 (6) (2020) 102051.
[28] W. Wang, L.-Y. Duan, H. Jiang, P. Jing, X. Song, L. Nie, Market2Dish: Health-aware food recommendation, ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 17 (1) (2021) 1–19.
[29] T. Islam, A.R. Joyita, M.G.R. Alam, M. Mehedi Hassan, M.R. Hassan, R. Gravina, Human-behavior-based personalized meal recommendation and menu planning social system, IEEE Trans. Comput. Soc. Syst. 10 (4) (2023) 2099–2110, http://dx.doi.org/10.1109/TCSS.2022.3213506.
[30] L. Kopitar, G. Stiglic, L. Bedrac, J. Bian, Personalized meal planning in inpatient clinical dietetics using generative artificial intelligence: System description, in: 2024 IEEE 12th International Conference on Healthcare Informatics, ICHI, 2024, pp. 326–331, http://dx.doi.org/10.1109/ICHI61247.2024.00049.
[31] L. Xu, J. Zhang, B. Li, J. Wang, M. Cai, W.X. Zhao, J.-R. Wen, Prompting large language models for recommender systems: A comprehensive framework and empirical analysis, 2024, ArXiv abs/2401.04997, URL: https://api.semanticscholar.org/CorpusID:266902921.
[32] M. Shanahan, K. McDonell, L. Reynolds, Role play with large language models, Nature 623 (7987) (2023) 493–498.
[33] X. Li, Y. Zhang, E.C. Malthouse, Prompt-based generative news recommendation (PGNR): Accuracy and controllability, in: European Conference on Information Retrieval, 2024, URL: https://api.semanticscholar.org/CorpusID:268692881.
[34] I. Papastratis, D. Konstantinidis, P. Daras, K. Dimitropoulos, AI nutrition recommendation using a deep generative model and ChatGPT., Sci. Rep. 14 (1) (2024) 14620, URL: https://api.semanticscholar.org/CorpusID:270736117.
[35] C. Henry, Basal metabolic rate studies in humans: measurement and development of new equations, Public Heal. Nutr. 8 (7a) (2005) 1133–1152.
[36] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, Inform. Sci. 8 (3) (1975) 199–249.
[37] World Health Organization, et al., Healthy Diet, Technical Report, World Health Organization. Regional Office for the Eastern Mediterranean, 2019.
[38] E. Pavlidou, S.K. Papadopoulou, K. Seroglou, C. Giaginis, Revised harris–benedict equation: new human resting metabolic rate equation, Metabolites 13 (2) (2023) 189.
[39] M. Post, A call for clarity in reporting BLEU scores, 2018, arXiv preprint arXiv:1804.08771.
[40] C.-Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Text Summarization Branches Out, 2004, pp. 74–81.
[41] T. Zhang, V. Kishore, F. Wu, K.Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with bert, 2019, arXiv preprint arXiv:1904.09675.