Nudging Healthy Choices: Leveraging LLM-Generated Hashtags and Explanations in Personalized Food Recommendations

Ayoub El Majjodi[1,*], Alain Starke[1,2], Christoph Trattner[1], Alessandro Petruzzelli[3] and Cataldo Musto[3]

[1]MediaFutures, University of Bergen, Bergen, Norway
[2]University of Amsterdam, Amsterdam, The Netherlands
[3]University of Bari Aldo Moro, Bari, Italy

Abstract

Making healthy recipe choices can be challenging for users, requiring time and knowledge to differentiate among various options. These choices are often generated by personalized recommender systems that account for individual preferences. One effective approach to encouraging healthier food choices is to intervene in how these choices are presented to users. In this paper, we explore the impact of nutritional food labels and evaluate the effectiveness of a Large Language Model (LLM) in generating high-quality explanations and hashtags to support users in making healthier food decisions. In an online experiment (N = 240), we designed a knowledge-based recommender system to generate personalized recipes for each user. Recipes were annotated with one of four intervention, a Multiple Traffic Light (MTL) nutrition label, LLM-generated explanations, LLM-generated hashtags, or no label (baseline). Our findings indicate that the interventions significantly enhanced users' ability to select healthier recipes. Additionally, we examined how different system components affected the overall user experience and how these components interacted with one another.

Keywords

Digital Food Nudges, Food recommender system, Explainability, Large language models, User-centric evaluation, Decision-making

CEUR-WS.org/Vol-4027/paper3.pdf

1. Introduction

People are increasingly relying on online platforms to find recipes. To reduce information overload, food recommender systems provide assistance through suggesting personalized food options within those platforms, by analyzing user preferences and ranking items accordingly. However, within the food domain, such systems may inadvertently promote unhealthy food choices [1, 2], as unhealthy recipes often receive the highest ratings, the most comments, frequent bookmarks, and the most attention overall [3]. This raises concerns about their impact on dietary habits and general health. Therefore, it is important to understand how recommenders can be designed to support various healthy eating goals, not only in terms of content, but also in terms of how that content is communicated to the end user.

Researchers have focused on addressing the health-related challenges of food recommender systems, proposing various solutions to mitigate these issues. One approach involves integrating users' caloric requirements by generating personalized menu recommendations based on user specific needs [4]. Another direction extends beyond calorie-based recommendations to provide tailored or personalized dish suggestions, that contain essential nutrients to address specific dietary concerns [5]. For example, knowledge-based recommender systems have demonstrated effectiveness in improving nutritional outcomes for elderly males by incorporating multiple user-related factors, such as dietary restrictions, nutritional needs, ingredient composition, preparation time, cost, and allergies [6].

12th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS'25), Prague, Czech Republic.
*Corresponding author.
ayoub.majjodi@uib.no (A. E. Majjodi); alain.starke@uib.no (A. Starke); christoph.trattner@uib.no (C. Trattner); alessandro.petruzzelli@uniba.it (A. Petruzzelli); cataldo.musto@uniba.it (C. Musto)
0000-0002-7478-5811 (A. E. Majjodi); 0000-0002-9873-8016 (A. Starke); 0000-0002-1193-0508 (C. Trattner); 0009-0008-2880-6715 (A. Petruzzelli); 0000-0001-6089-928X (C. Musto)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Nudges have been hailed as an approach to address health-related challenges in recommender systems. In doing so, the research focus has shifted from only focusing on algorithmic accuracy to a more comprehensive study of the recommender system as a whole, where the interface plays a significant role in determining user decision making [7]. Digital nudging principles have been recognized as an effective strategy for encouraging healthier choices within food recommendation domain [8]. Nudging represents a low-cost and easily deployable approach that specifically targets the presentation phase of recommender systems, such as emphasizing healthier options in decision-making scenarios [8]. Evidence suggests that the integration of nudging within food recommender systems not only increases the likelihood of users selecting healthier items [9] but also enhances the overall user experience [10], while simultaneously improving system effectiveness [11]. However, most nudging techniques are still generated using traditional statistical approaches [8]. While these methods have proven effective, there is a growing need to incorporate emerging trends and advancements to further optimize their impact and produce more personalized and dynamic nudges. [12].

There is a lot of potential for AI-based methods in addition to traditional recommender approaches, particularly with the rise of Large Language Models (LLMs). Recent research has explored different ways to leverage LLMs for recommendation tasks, which can be broadly categorized into *direct generation* and *enhanced adaptation* approaches. *Direct generation* methods rely on prompt engineering to elicit recommendations from LLMs without modifying the underlying model [13]. These approaches take advantage of the model's general knowledge and reasoning abilities. *Enhanced adaptation* methods, on the other hand, fine-tune smaller, open-source models for specific recommendation tasks. By tailoring the model to a particular domain, these approaches aim to improve relevance and personalization. Additionally, a particularly promising aspect of LLM-based recommendation systems is their ability to provide not only item suggestions but also meaningful, context-aware and personalized explanations for their recommendations [14, 15]. This capability represents a significant advancement over traditional recommender systems, offering users both recommendations and the reasoning behind them. For instance, LLM-generated explanation are highly appreciated by users as they help in the evaluation of recommended movie items [16]. However, within the food domain, the application of LLM-generated explanations remains underexplored [17, 18].

While the benefits of digital nudges in promoting healthier choices [7, 19] and the role of LLMs in enhancing recommender systems have been shown [20], this study investigates the use of LLMs for generating explanations and hashtag based nudges within a recipe-personalized environment. Specifically, it addresses the gap between LLM-driven nudges and user evaluation in the context of a personalized recipe recommender system. In an online user study, we evaluate the effectiveness of three different nudging (i.e, Multiple Traffic Light (MTL) nutrition label, LLM-generated Explanations, and LLM-generated Hashtags), strategies applied to a knowledge-based recipe recommender in encouraging healthier food choices. Furthermore, the study examines the interaction between various system components and their influence on user experience. We formulated the following research questions:

- **RQ1:** To what extent can LLM-generated explanations, hashtags, and nutritional food labels serve as nudges to promote healthy recipe choices within a personalized food recommender system?
- **RQ2:** How do users evaluate various aspects of recommender interfaces with food nudges, and how do these aspects relate to each other?

The remainder of this paper is structured as follows: Section 2 reviews related work on food recommender systems, digital nudges, and LLM-based explanations. Section 3 outlines the methodology, while Section 4 presents the results. Finally, Section 5 discusses the findings and future research directions..

2. Related Work

2.1. Explaining Recommendations

Explanations of recommendations serve as a bridge between complex algorithms and user understanding [17]. Integrating them into a system has shown to improve the overall user experience, such as by helping the user making better and more informed decisions [21]. In recent years, there has been growing academic interest in post-hoc explanations [22], which provide users with insight into the reasoning behind a recommended item. A survey conducted by Zhang et al. [23] demonstrated that explanations substantially enhance the perceived usefulness of recommender systems. Moreover, in the context of persuasiveness, Gkika et al. [24] has found that explanations can influence individuals to change their attitudes or adopt behaviors conducive to improved lifestyles, even when users initially exhibit a low intention toward the recommended items. This effect is achieved by emphasizing the benefits of consuming the recommended items within the presented explanations [21].

2.2. Recommendation explanations with LLMs

A Large Language Model (LLM) is an advanced natural language processing model capable of processing, understanding, and generating human-like text [25]. LLMs have been successfully applied across diverse domains, including content generation [26], language translation [27], and code completion [28], contributing to the creation of relevant information based on given tasks. To accomplish these tasks, such models must be instructed with prompts [29], which are specific text inputs designed to guide the models in generating the desired outputs in a natural language.

The current research on large language models (LLMs) for recommender systems remains in its early stages [20, 30]. A survey by Zihuai et al. [18] systematically reviews emerging advanced techniques for enhancing recommender systems using LLMs, including pre-training, fine-tuning, and prompting. LLMs provide high-quality representations of textual features and extensive coverage of external knowledge [31]. These capabilities can be leveraged in recommender systems by either integrating LLM embeddings into traditional recommendation algorithms or employing LLMs as standalone recommender systems [32].

The power of large language models (LLMs) lies in their ability to generate human-like language with high fluency and contextual understanding [33]. However, the generation of explanations for recommendations using LLMs remains scarce, as highlighted by Said et al. [33]. Using LLMs for post hoc explanations for recommender systems leads to more flexible and personalized natural language, that bring improved user engagement within movie domains [34]. Moreover, LLMs lead to more creative and coherent explanations, which lead to improved user engagement[16]. Post-hoc LLM-generated explanations for a movie recommender systems have shown greater user appreciation and effectiveness compared to traditional item feature-based explanations, positively impacting understandability, satisfaction, transparency, trust, efficiency, and persuasiveness [16].

A hashtag is a keyword or phrase preceded by a hash sign (#), commonly used on social media platforms to categorize and identify content related to specific topics [35]. Research highlights hashtags as essential tools for information discovery and content visibility [36]. While some studies have utilized keywords to explain recommendations [37], others, such as [38], demonstrate that keyword-based explanations, like hashtags, can enhance understandability and decision-making in movie recommender systems. However, to the best of our knowledge, no prior work has investigated the use of LLMs for generating hashtags as a means of providing recommendation explanations [17, 18].

2.3. Digital Food Nudges

Digital nudges involve techniques applied during the presentation phase of the recommender system process [8]. These techniques aim to support users in making informed decisions [39]. In food recommender systems, researchers have leveraged digital food nudges to design systems that promote healthier food choices and enhance user satisfaction [10, 9]. Nutritional information associated with recommended items, such as front-of-pack labels like multiple traffic light (MTL) labels and Nutri-Score labels, has been shown to encourage healthier food choices within personalized recommender systems [19, 10]. Similarly, associating attractive images with healthy food items and unattractive ones with unhealthy options has been effective in influencing users toward healthier selections [9].

Generating feature-based explanations (e.g., user or item-base) as food nudges has been demonstrated

Generate {six grammatically correct hashtags | three grammatically correct lines of explanations} to describe {recipe title}. To generate {the hashtags | explanations} emphasize the ingredients of the dish and their healthiness, {recipe fsa score} as FSA score, {calories} calories, fat fat, and {protien} protein. The hashtags should convince a user with a {user BMI level}, an eating goal to {usre eating goal}, {sleep hours} hours of sleep, {depression level}, and {physical activity} physical activity to {prepare | avoid preparing} this recipe.

Figure 1: Example of the used prompt template to generate recipe explanations and hashtags for a given user. The prompt varied based on the study condition (explanation or hashtags) and whether the recipe was healthy or not.

to positively influence users' willingness to make more informed and healthier food choices [40]. Moreover, compelling explanations about recommended recipes significantly increase the likelihood of users selecting healthier options [41]. However, the potential of using large language models (LLMs) to generate explanations that nudge users toward healthier food choices remains unexplored in the literature [18].

2.4. Contribution

Digital food nudges have been shown to effectively guide users toward healthier choices within recommender systems [42, 19]. However, the potential of large language models (LLMs) to enhance these nudges has remained largely unexplored. Building on the latest advances in recommender systems [8, 18], this study is the first to investigate how LLMs can generate short textual explanations and hashtags to nudge users toward healthier recipes. As discussed in our literature review (Section 2), while we do not introduce a new recommender algorithm, our work demonstrates how integrating LLM-driven content generation can strengthen the persuasive impact of existing systems. Specifically, guided by the user-centric evaluation framework of Knijnenburg et al. [43], we make the following contributions:

- **LLM-Based Nudge Generation:** We demonstrate, for the first time, how short textual explanations and hashtags produced by LLMs can effectively steer users toward healthier recipe choices in a personalized recipe recommender.
- **User-Centric Impact Assessment:** We employ Structural Equation Modeling (SEM) to evaluate how different system components LLM-driven content, interface design, and personalization, collectively influence user experience, behavioral intentions, and healthfulness of choices.

All data, system components, prompts, LLM API calls, and analytical methods used in this study are openly available at [44] to enable transparency and reproducibility.

3. Methodology

3.1. Dataset

We addressed our research questions using a dataset from the popular recipe platform Allrecipes.com. From a collection of 58,000 recipes, we extracted a stratified sample of 3,000 main dishes representing diverse food categories, including barbecue, fruits, vegetables, seafood, pasta, meat, and poultry. The sampling process ensured that all selected recipes had no missing data for relevant attributes, such as related to health. These attributes included salt, sugar, protein, fat, saturated fat, carbohydrate, fiber, sodium, magnesium, and serving size.

3.2. System Design and Procedure

Before beginning the experiment, participants were provided with a brief introduction to the study and were required to provide informed consent. The first phase involved completing questionnaires to collect personal information (e.g., age, gender, education level) and assess their food knowledge. The next step, consists of a user profile builder that required the user to fill out a form detailing their food preferences such as eating goals, sleep time, cooking experience and daily exercises. This information is then processed by a knowledge-based recommender system previously developed in our previous work [10]. The system generate personalized recommendations for both healthy and unhealthy recipes.

After the recommender algorithm generates personalized recipes for a given user, a feature extractor identifies key recipe features and combines them with user features to construct a prompt. The prompt is input to a large language model (LLM), which generates explanations and hashtags designed to nudge users toward healthier recipe choices. An example prompt is shown in Figure 1. The LLM is instructed to produce a single explanation and a set of hashtags tailored to the specific recipe and user. We utilized GPT-3.5 Turbo, which, as demonstrated in our previous work [45], is a powerful language generation model trained on rich of knowledge, capable of producing highly coherent, contextually relevant, and detailed responses [46]. Participants were presented with six recipes, including three healthy ones and

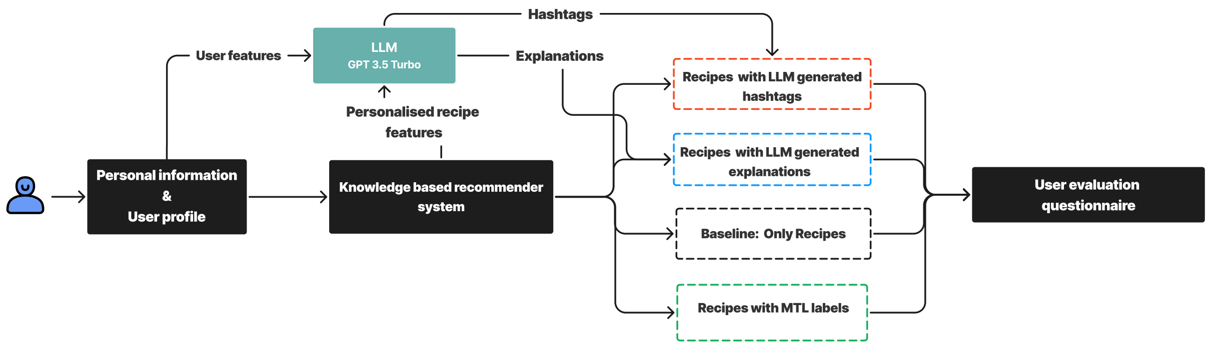

Figure 2: System design and user flow illustrating all steps to complete the study.

three unhealthy, then they asked to select the recipe they found most appealing and would consider trying at home. Participants were divided into groups corresponding to different study conditions.

The final phase involved evaluating the user experience using pre-validated questionnaires related to choice satisfaction, choice difficulty, perceived effort, understandability, and usability [43, 47, 16]. Figure 2 details the user flow and the system architecture of the online experiment.

3.3. Research Design

The recommender interface in this study was subject to 4 between-subjects between subjects conditions: noLabel, MTL label, LLM-based explanation, and LLM-based hashtags (cf. Figure 3). In each condition, participants were presented with six recipes, including three healthy options and three unhealthy ones. In the baseline condition, recipes were unannotated. In the MTL label condition, each recipe was accompanied by a corresponding Multiple Traffic Light (MTL) label. In the explanation condition, recipes were supplemented with LLM-generated textual explanations. In the hashtags condition, six hashtags were associated with each recipe.

3.4. Study Participants

Participants were recruited from the crowdsourcing platform Prolific. Upon successful completion of the study, each participant received GBP 1.50 as a compensation for an average participation time of 7–10 minutes. Eligibility criteria included a minimum approval rate of 95%, fluency in English, and the successful completion of at least 30 previous submissions. To determine an appropriate sample size,

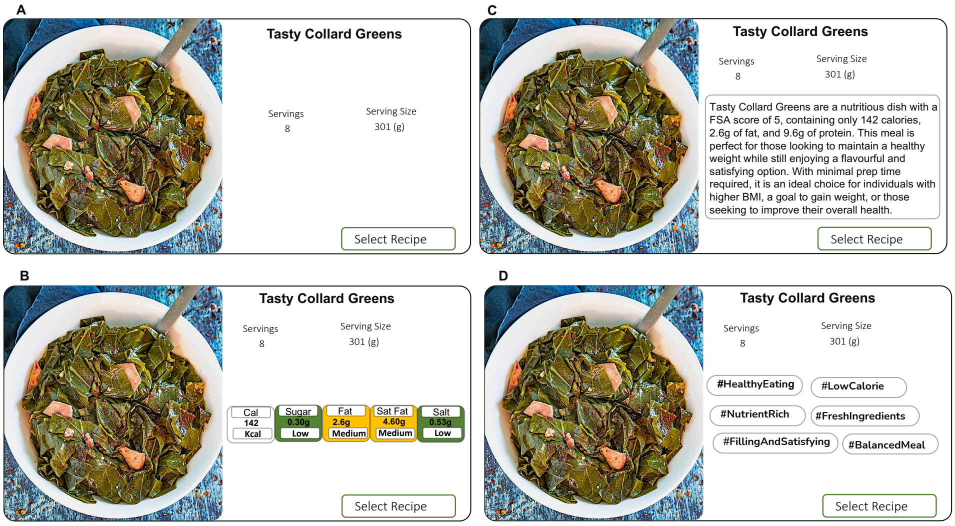

Figure 3: Examples of recipes: with No label (baseline) (A), or annotated with either a Multiple Traffic Light (MTL) label (B), a LLM-based explanation (C) or LLM-based hashtags (D).

we first approximated the research design by performing an 'A priori ANOVA: Fixed effects, special, main effects and interactions' test in G*Power 3.1.9.7, under 90% power, a medium effect size ($f = 0.25$) and a numerator of 1 (for ANOVAs that used dummy predictor variables). This resulted in a required sample size of $N = 171$. Since studies with Structural Equation Models (SEMs) have shown better results with sample sizes greater than 200 when involving multiple latent factors [48], we eventually recruited $N = 240$ participants. Among them, 34% were between 25 and 35 years old, 60% identified as female, and 97% had attained at least a high school degree.

3.5. Measures

Table 1
Results of the confirmatory factor analysis. Items were measured on 5-point Likert scales. Cronbach's Alpha is denoted by α, items in gray were omitted from analysis. Usability was analyzed separately for it was absent in the baseline condition.

Aspect	Item	Loading
Subjective Food Knowledge $AVE = .539$ $\alpha = .812$	Compared with an average person, I know a lot about healthy eating.	.697
	I think, I know enough about healthy eating to feel pretty confident when choosing a recipe.	
	I know a lot about food to evaluate the healthiness of a recipe.	.754
	I do not feel very knowledgeable about healthy eating.	-.751
Choice Satisfaction $AVE = .685$ $\alpha = .845$	I like the recipe I have chosen.	.823
	I think I will prepare the recipe I have chosen.	.825
	The chosen recipe fits my preference.	.877
	I would recommend the chosen recipe to others.	
Understandability $AVE = .526$ $\alpha = .718$	It was easy to understand the contents of a recipe.	
	I could easily understand why recommended recipes fitted my preferences.	.589
	I understood the healthiness of each recipe.	.825
Usability $\alpha = .718$	The {*explanations* | *labels* | *hashtags*} helped me to understand the healthiness of each recipe.	
	The {*explanations* | *labels* | *hashtags*} helped me to choose a recipe.	

3.5.1. Recipe healthiness

To assess the healthiness of the recipes in our dataset, we relied on the well-established and pre-validated FSA score, introduced by the British Food Agency [49, 50]. The metric evaluates the healthiness of each recipe based on four macronutrients: sugar, fat, saturated fat, and salt. The score ranges from 4, indicating the healthiest recipes, to 12, representing the least healthy. Details of the FSA computation method can be found in [9]. To distinguish between healthier and less healthy recipes within our recommendation set, we applied a healthiness threshold to partition the dataset into two groups: healthy and unhealthy. This approach allowed us to generate recommendations from both subsets, ensuring a balanced presentation. Each user received a personalized selection of six recipes, comprising three from the healthy set and three from the unhealthy set.

3.5.2. User evaluation metrics

To evaluate the users' nutritional knowledge levels, we employed the Subjective Food Knowledge (SFD) questionnaire, which was validated in prior studies [51, 52]. The SFD questionnaire comprised 4 items, that are rated on a five-point Likert scale.

To measure the user experience of participants, we based our analysis on the user-centric evaluation framework for recommender systems [43]. This framework facilitated the assessment of multiple metrics to evaluate user experience through pre-validated questionnaires across several domains of recommender systems. The framework includes measures such as choice satisfaction [53, 54], which assesses how satisfied users are with the overall experience and the quality of recommendations, choice difficulty [54], which evaluates the process of interacting with the system and making final decisions, and perceived effort [55], which measures the overall effort expended by users during their interaction with the system.

Additionally, we adopted validated metrics from the explainability literature on recommender systems to assess the system's understandability [22, 24, 16]. Finally, participants were asked to evaluate the experienced usability [56] of each nudging intervention (i.e., labels, explanations, and hashtags). Since the baseline condition was excluded, a separate analysis was conducted using Cronbach's alpha to assess the internal consistency of the usability questionnaire items.

To assess the validity of each questionnaire and the selected measures, we performed confirmatory factor analysis. Surprisingly, the perceived effort and choice difficulty questionnaires were excluded from the analysis as they did not meet internal consistency guidelines ($\alpha > .70$) or convergence validity criteria based on the average variance explained ($AVE > .50$). All other aspects, however, met the necessary standards. Table 1 describes the factor loadings and Cronbach's Alpha (α), showing that items with low loadings (indicated in grey) were excluded from the analysis.

3.6. Ethical Statement

This research adhered to the ethical guidelines of the University of Bergen and Norway's regulations for scientific research. The study was judged to meet the ethical standards of the university and therefore did not require a more extensive review, as it contained no misleading information, stressful tasks, or content that would likely provoke extreme emotions. All collected data were collected and processed anonymously to ensure participant confidentiality and privacy.

4. Results

We addressed both research questions by performing a Structural Equation Model (SEM) analysis. We examined the influence of different LLM-based interface nudges on recipe healthiness (RQ1) and user evaluation (RQ2) by constructing a path model, analyzing how different observed and evaluative system aspects related to each other, following the user evaluation framework by Knijnenburg and Willemsen [43]. Changes in objective system aspects (i.e., No Label, MTL label, LLM explanations, LLM

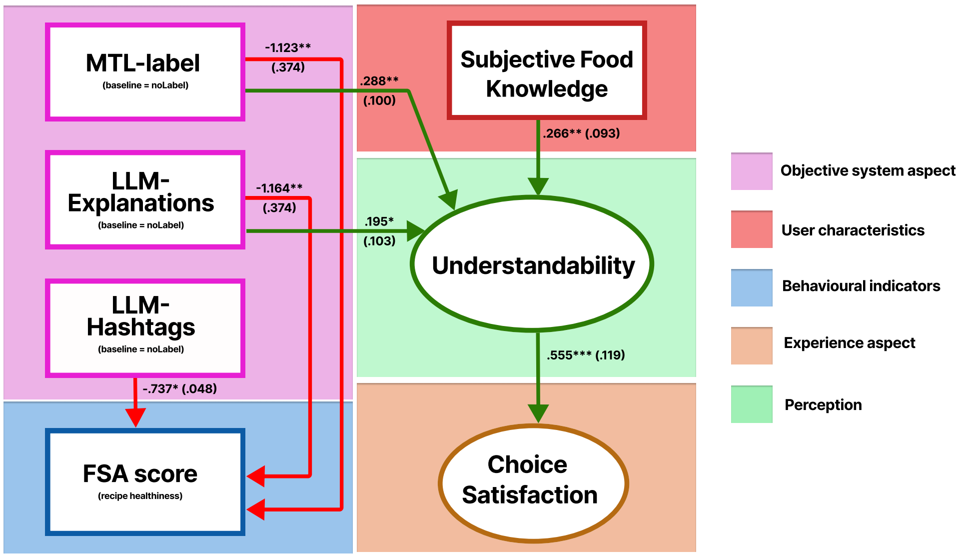

Figure 4: Structural Equation Model (SEM). Numbers on the arrows indicate β-coefficients, with standard errors shown in brackets. Effects between subjective constructs are standardized and represent correlations: green arrows denote positive correlations, while red arrows indicate negative correlations. Other effects represent regression coefficients. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

hashtags) were related to an observed factor (i.e. healthiness of the chosen recipe) and user experience factors: based on perception (i.e., understandability) and experience (i.e., choice satisfaction). In addition, we also examined the role of personal characteristics (i.e., subjective food knowledge).

In line with [43], we first specified a fully saturated model, from objective to experience aspects. Non-significant paths were then pruned, while modification indices were used for model optimization. The results of the final model are presented in Figure 4, which had a near optimal fit: $\chi^2(43)$, $p < .05$, $CFI = .991$, $TLI = .987$, $RMSEA = .023$, $90\% - CI$: [.000; .051]. The model met the guidelines for discriminant validity, as the correlations between latent constructs were smaller than the square root of each construct's Average Variance Extracted (AVE) (cf. [43]).

4.1. RQ1: Healthiness of recipe choices

Participants, on average, selected 20% more healthy recipes (FSA score < 7) when presented with recommendations incorporating nudging interventions (i.e., MTL labeling, LLM-based explanations, and LLM-based hashtags) compared to the baseline condition (i.e., no label). Figure 5 illustrates the distribution of user selections for healthy recipes and less healthy recipes across the various study conditions. The structural equation model (SEM) illustrated in Figure 4 reveals a significant path from the objective system aspects (i.e., MTL label, LLM-based explanation, and LLM-based hashtags) to the behavioral aspect (i.e., the healthiness of selected recipes, as measured by the FSA score). This indicates that the nudging interventions effectively promote healthier choices, with participants in the treatment conditions selecting recipes with lower FSA scores compared to the control condition.

Among the interventions, MTL labels exhibited the strongest influence on FSA scores ($coef. = -1.123, p < 0.001$), followed by LLM-based explanations ($coef. = -1.164, p < 0.001$) and hashtags ($coef. = -0.750, p < 0.05$). These findings highlight the effectiveness of the interventions in promoting healthier choices compared to the baseline condition. To test for differences between non-baseline condition, we performed a one-way ANOVA on the chosen FSA Score with the conditions as a predictor: $F(3, 236) = 4.69$, $p < 0.01$, followed by a post-hoc Tukey HSD test. This, however, revealed no significant differences in effectiveness among the different nudging conditions and only confirmed the significant differences between the no-label baseline and different nudging in terms of recipe healthiness (i.e., chosen FSA score).

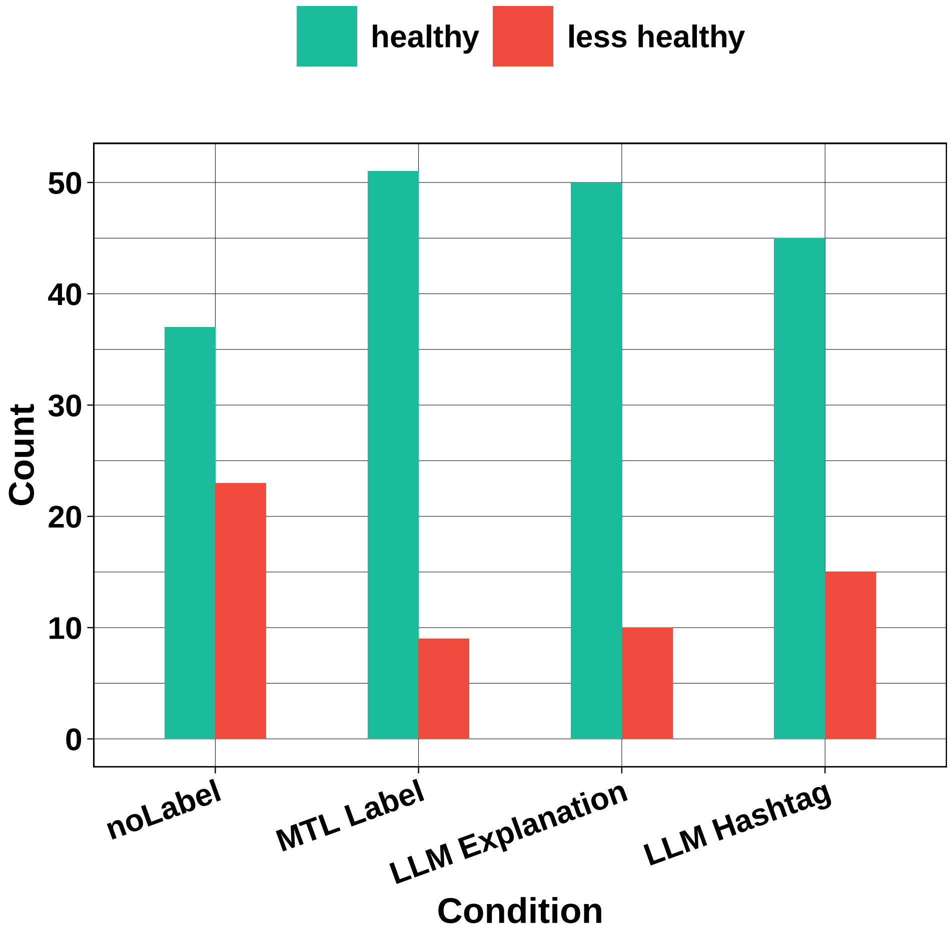

Figure 5: Healthy and less healthy recipe choices across the various study conditions.

4.2. RQ2: User Evaluation

4.2.1. User experience

The second main path depicted in Figure 4 stemmed from the objective system aspects towards the user perception. Two types of nudging interventions, MTL labels and LLM-generated explanations, positively affected the understandability of recommendations. Specifically, the MTL intervention improved user understandability compared to the no-label baseline ($coef. = .288, p < 0.01$). Similarly, the addition of LLM-generated explanations had a significant positive effect on user perceptions, enhancing understandability ($coef. = .195, p < 0.05$). In contrast, the hashtag intervention did not show a significant impact on user perceptions. Figure 6 presents the understandability score across different conditions.

This path extends further into user experience, as understandability significantly influences choice satisfaction ($coef. = .555, p < 0.01$). Thus, the understandability perception factor mediates the relationship between objective system aspects (i.e., MTL label, LLM explanations) and user experience outcomes (i.e., choice satisfaction). Figure 7 further illustrates the variation in choice satisfaction as a function of user understandability across the different study conditions. The understandability levels are computed as the mean understandability for each respective condition. The final path in Figure 4 demonstrates that the perception aspect also mediates the relationship between user characteristics and user experience. Specifically, subjective food knowledge (SFD) positively affected understandability ($coef. = 0.266$, $p < 0.01$), which subsequently affected choice satisfaction. This indicates that SFD enhances user understandability, which in turn leads to higher choice satisfaction.

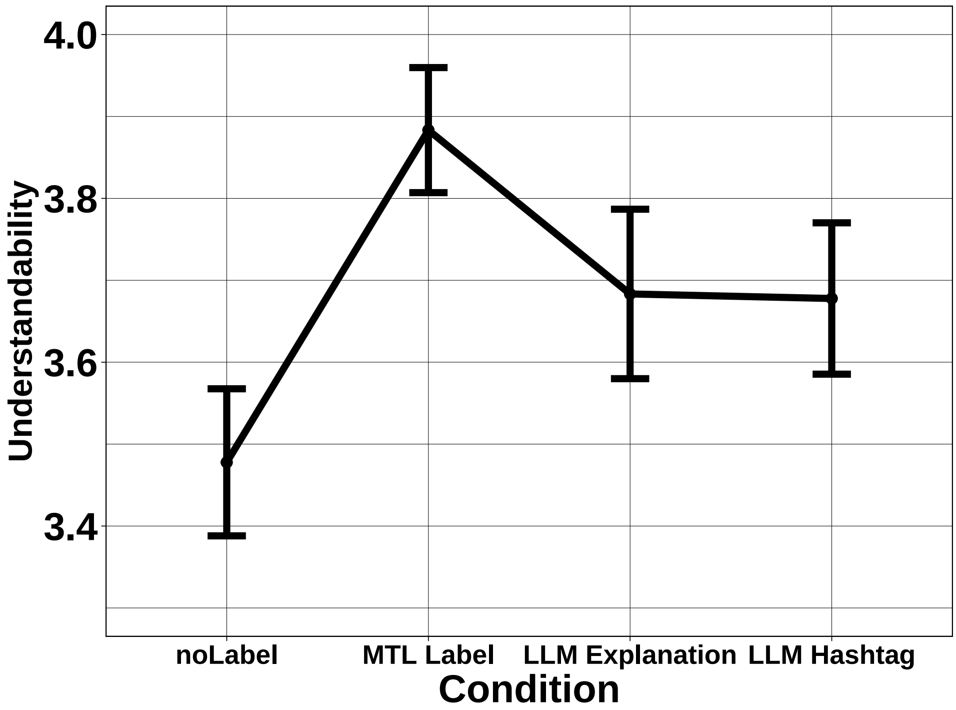

Figure 6: Mean scores for the understandability, the perception aspects across conditions. Errors bars represent 1 S.E.

Finally, a test of indirect effects checked whether the described paths from the objective system aspects and SFD to choice satisfaction were fully mediated by understandability. The test confirmed that the paths from both SFD ($coef. = 0.145$, $p < 0.01$) and MTL-label ($coef. = 0.160$, $p < 0.05$) were significantly mediated by understandability, while those from the other two conditions were not significant ($0.1 < p < 0.05$). This suggested that, compared to the no-label baseline, changes in choice satisfaction can be explained by changes in understandability, which stemmed from the presence of MTL labels.

4.2.2. Usability evaluation

Finally, we evaluated the perceived usability across the different nudging interventions. The results of a one-way ANOVA revealed a significant effect of the interventions on usability scores: $F(2, 177) = 14.29$, $p < 0.0001$. To identify specific differences between conditions, we conducted a post-hoc Tukey HSD test. The MTL labels demonstrated the highest mean usability score $(M = 4.08, SD = 0.70)$ followed by LLM explanations $(M = 3.79, SD = 0.80)$ with no significant difference between the two interventions. In contrast, LLM hashtags resulted in significantly lower usability scores $(M = 3.18, SD = 1.23)$, compared to both MTL labels and LLM explanations. Figure 8 visualizes the variations in usability scores across the intervention conditions.

5. Discussion

This study has examined the benefits of LLM-based descriptive nudges for a food recommender system. To date, most nudging strategies have been static [8], usually not adapted to a user's input. In the context of recommender systems, both academic and industry studies have concentrated on the effectiveness of digital nudges in recommender systems, particularly in enhancing the visibility of the most relevant items, either for the user or the recommender system provider [55]. The food recommender domain has garnered significant attention due to the complexity of food choices and their direct impact on the overall health [57].

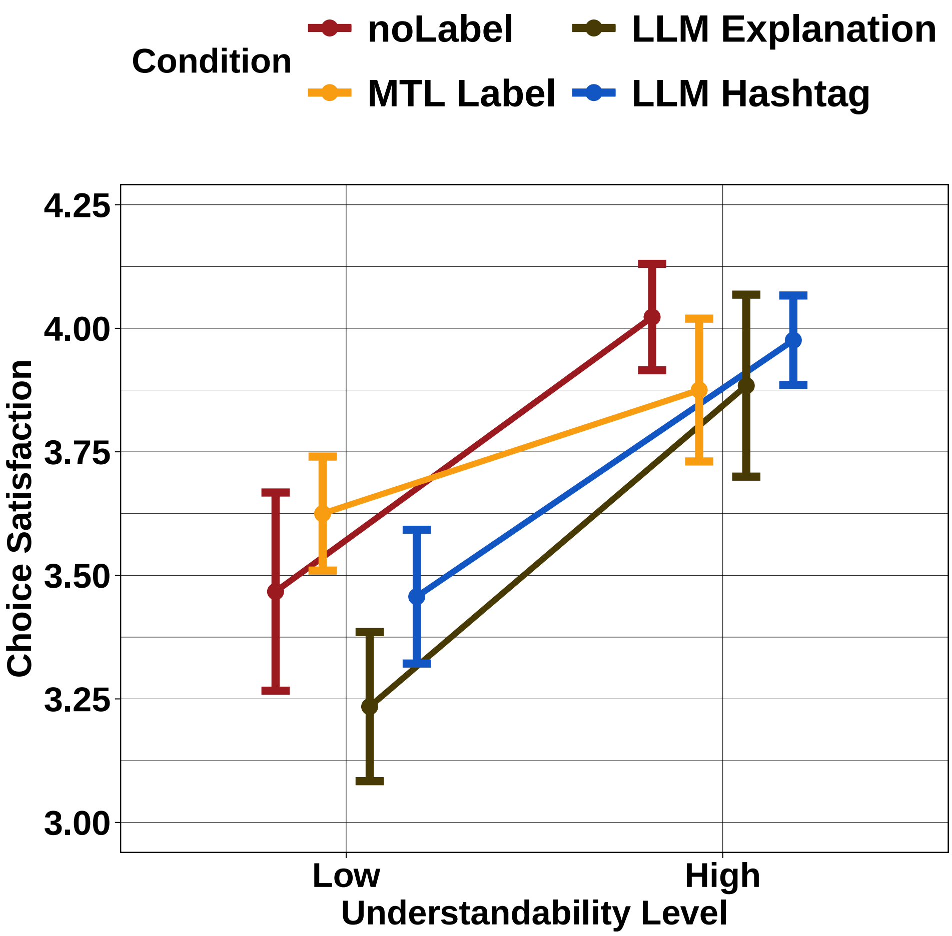

Figure 7: Mean scores for the choice satisfaction given the level of user understandability aspects across conditions. Errors bars represent 1 S.E.

This study has investigated a new direction for explanations and nudges in food recommender systems. We have compared static nudges (in this case: MTL labels) with more dynamic, based on user input nudges (in this case: LLM-based explanations and hashtags), where each strategy aimed to promote healthier food consumption, through supporting healthier choices in the interface as well as enhancing the user's experience.

The use of LLMs to generate explanations and hashtags, proved to be effective in leading users to more healthy food choice compared to the no-nudge condition, leveraging their ability to generate human-like language, and flexible, and personalized nudges within recommender systems (RQ1). This finding aligns with various studies that support the use of LLMs as persuasive tools, especially in the context of health and food-related domains [58]. Similarly, MTL labels as nudge to support user to make healthier choices. This reinforces the idea that current online recommendations can often lead to unhealthy choices [59, 1].

Based on our findings, it can be concluded that LLM-based techniques, when compared to traditional nudging strategies (e.g., MTL labels), perform comparably within the context of our study and supporting healthier choices. However, the finding of El Majjodi et al. [19], suggest that MTL labels has lead to less healthier choices within personalized recipes environment. Consequently, further research is required to examine the comparative effectiveness of these nudging strategies across different settings, recommendation domains, and methodologies, with particular focus on the underlying mechanisms that drive their impact. [8, 60].

Our path model has revealed a significant interaction between various aspects of the recommender system (RQ2), underscoring the importance of user-centric evaluation, particularly with regard to its

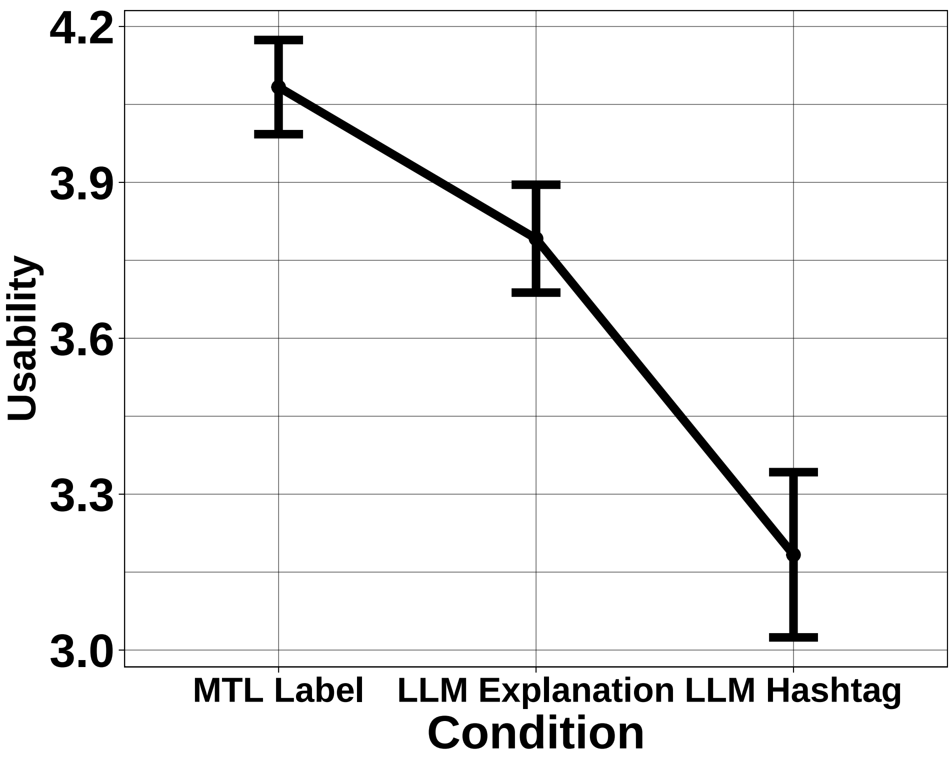

Figure 8: Mean usability score across study condition. Errors bars represent 1 S.E.

impact on user experience [43]. Our findings suggest that the nudging intervention, through MTL labels and LLM explanations, significantly enhances user understanding of the generated recipes, which in turn influences choice satisfaction, a crucial aspect of the overall user experience. This supports the finding that nutritional labels not only significantly impact user choices but also influence overall user behaviors in offline domains [61]. Additionally, it highlights that content presentation is crucial in enhancing how users experience the system.

The use of LLMs for generating personalized explanations has been shown to enhance user experience, particularly by increasing choice satisfaction through improved understandability, whereas LLM-generated hashtags demonstrated no significant effect. These findings highlight the critical role of clarity in LLM-generated explanations in facilitating users' comprehension of recommended recipes. This, in turn, leads to significantly better user experience compared to the baseline condition, where no explanations are provided. The enhanced understandability of these explanations may be attributed to the flexibility of the generation approach, enabling the LLM to adapt more effectively to recommended items and user preferences by leveraging its vast general knowledge. This adaptability likely contributes to more satisfying recommender systems [16]. Furthermore, by aligning explanations with user preferences, LLMs not only improve user satisfaction but also promote healthier food choices. This aligns with the findings of [40], which demonstrated that explanations generated using traditional computational approaches also promoted healthier choices. However, the explanations were not personalized and only involved comparisons between two recommended recipes.

Another key contribution of the path model is its emphasis on the relationship between user characteristics, perception, and overall user experience. The findings suggest that users with greater food knowledge exhibit a higher understanding of nudge interventions (i.e., MTL labels and LLM-generated explanations), which positively influences choice satisfaction. The significance of user knowledge has been well established not only in the recommender systems literature [62, 10] but also in psychological research [63].

Finally, our findings demonstrate the effectiveness of both MTL labels and LLM-based explanations in enhancing usability, as reflected in higher usability scores. This also explains the lack of significant effect of LLM-generated hashtags on perception and user experience. MTL labels directly communicate a recipe's nutritional values, while LLM-generated explanations have the advantage of emphasizing user preferences, thereby improving usability by helping users better evaluate the recommended items. This effect has also been demonstrated in domains such as movies and news [16, 64]. Although hashtags can attract user attention [65], they may be less engaging due to their limited coherence and constructive value [66]. However, this topic warrants further exploration.

An interesting avenue for future research arises from our findings. First, we aim to explore a range of user evaluation metrics for both MTL labels and LLM-generated explanations, including transparency, trust, persuasiveness, and efficiency. Additionally, we seek to investigate the effectiveness of these nudges in conjunction with other recommender approaches beyond knowledge-based systems, such as collaborative, content-based, and deep learning-based methods.

Our contribution paves the way for exploring the use of LLMs to generate personalized nudges within recommender systems, with the potential to enhance adaptability, user engagement, and to promote behavioral change. Furthermore, we highlight a crucial yet underexplored area: the long-term effects of LLM-generated explanations and nutritional labels in influencing users' dietary habits, lifestyles, and sustained behaviors.

Acknowledgments

This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT to: Grammar and spelling check, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content. It's also important to note that the author did not use generative AI to generate totally new text.

References

[1] C. Trattner, D. Elsweiler, S. Howard, Estimating the healthiness of internet recipes: a cross-sectional study, Frontiers in public health 5 (2017) 16.

[2] C. Trattner, D. Parra, D. Elsweiler, Monitoring obesity prevalence in the united states through bookmarking activities in online food portals, PloS one 12 (2017) e0179144.

[3] R. Raghunathan, R. W. Naylor, W. D. Hoyer, The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products, Journal of Marketing 70 (2006) 170–184.

[4] T. N. Trang Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, Journal of Intelligent Information Systems 50 (2018) 501–526.

[5] T. Ueta, M. Iwakami, T. Ito, A recipe recommendation system based on automatic nutrition information extraction, in: Knowledge Science, Engineering and Management: 5th International Conference, KSEM 2011, Irvine, CA, USA, December 12-14, 2011. Proceedings 5, Springer, 2011, pp. 79–90.

[6] J. Aberg, Dealing with malnutrition: A meal planning system for elderly., in: AAAI spring symposium: argumentation for consumers of healthcare, 2006, pp. 1–7.

[7] M. Jesse, D. Jannach, B. Gula, Digital nudging for online food choices, Frontiers in psychology 12 (2021) 729589.

[8] M. Jesse, D. Jannach, Digital nudging with recommender systems: Survey and future directions, Computers in Human Behavior Reports 3 (2021) 100052.

[9] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, Frontiers in Artificial Intelligence 4 (2021) 621743.
[10] A. El Majjodi, A. D. Starke, M. Elahi, C. Trattner, et al., The interplay between food knowledge, nudges, and preference elicitation methods determines the evaluation of a recipe recommender system., in: IntRS@ RecSys, 2023, pp. 1–18.
[11] G. Castiglia, A. El Majjodi, A. D. Starke, F. Narducci, Y. Deldjoo, F. Calò, Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels, in: Fourth Knowledge-aware and Conversational Recommender Systems Workshop (KaRS), 2022.
[12] S. Dalecke, R. Karlsen, Designing dynamic and personalized nudges, in: Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics, 2020, pp. 139–148.
[13] L. Wang, E.-P. Lim, Zero-Shot Next-Item Recommendation using Large Pretrained Language Models, 2023. URL: http://arxiv.org/abs/2304.03153, arXiv:2304.03153 [cs].
[14] Q. Ma, X. Ren, C. Huang, Xrec: Large language models for explainable recommendation, 2024. URL: https://arxiv.org/abs/2406.02377. arXiv:2406.02377.
[15] L. Li, Y. Zhang, L. Chen, Personalized prompt learning for explainable recommendation 41 (2023). URL: https://doi.org/10.1145/3580488. doi:10.1145/3580488.
[16] S. Lubos, T. N. T. Tran, A. Felfernig, S. Polat Erdeniz, V.-M. Le, Llm-generated explanations for recommender systems, in: Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 276–285.
[17] A. Said, A review of llm-based explanations in recommender systems, arXiv preprint arXiv:2411.19576 (2024).
[18] Z. Zhao, W. Fan, J. Li, Y. Liu, X. Mei, Y. Wang, Z. Wen, F. Wang, X. Zhao, J. Tang, et al., Recommender systems in the era of large language models (llms), arXiv preprint arXiv:2307.02046 (2023).
[19] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: Proceedings of the 30th ACM conference on user modeling, adaptation and personalization, 2022, pp. 48–56.
[20] Q. Wang, J. Li, S. Wang, Q. Xing, R. Niu, H. Kong, R. Li, G. Long, Y. Chang, C. Zhang, Towards next-generation llm-based recommender systems: A survey and beyond, arXiv preprint arXiv:2410.19744 (2024).
[21] S. Lubos, T. N. T. Tran, S. P. Erdeniz, M. E. Mansi, A. Felfernig, M. Wundara, G. Leitner, Concentrating on the impact: Consequence-based explanations in recommender systems, arXiv preprint arXiv:2308.16708 (2023).
[22] G. Peake, J. Wang, Explanation mining: Post hoc interpretability of latent factor models for recommendation systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2060–2069.
[23] M. Zanker, D. Ninaus, Knowledgeable explanations for recommender systems, in: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 1, IEEE, 2010, pp. 657–660.
[24] S. Gkika, G. Lekakos, The persuasive role of explanations in recommender systems., in: BCSS@ PERSUASIVE, 2014, pp. 59–68.
[25] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, A. Mian, A comprehensive overview of large language models, arXiv preprint arXiv:2307.06435 (2023).
[26] S. Moore, R. Tong, A. Singh, Z. Liu, X. Hu, Y. Lu, J. Liang, C. Cao, H. Khosravi, P. Denny, et al., Empowering education with llms-the next-gen interface and content generation, in: International Conference on Artificial Intelligence in Education, Springer, 2023, pp. 32–37.
[27] H. Huang, S. Wu, X. Liang, B. Wang, Y. Shi, P. Wu, M. Yang, T. Zhao, Towards making the most of llm for translation quality estimation, in: CCF International Conference on Natural Language Processing and Chinese Computing, Springer, 2023, pp. 375–386.
[28] D. Nam, A. Macvean, V. Hellendoorn, B. Vasilescu, B. Myers, Using an llm to help with code understanding, in: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, 2024, pp. 1–13.
[29] L. Beurer-Kellner, M. Fischer, M. Vechev, Prompting is programming: A query language for large

language models, Proceedings of the ACM on Programming Languages 7 (2023) 1946–1969.
[30] J. Lin, X. Dai, Y. Xi, W. Liu, B. Chen, H. Zhang, Y. Liu, C. Wu, X. Li, C. Zhu, et al., How can recommender systems benefit from large language models: A survey, arXiv preprint arXiv:2306.05817 (2023).
[31] 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, World Wide Web 27 (2024) 60.
[32] J. Liu, C. Liu, P. Zhou, R. Lv, K. Zhou, Y. Zhang, Is chatgpt a good recommender? a preliminary study, arXiv preprint arXiv:2304.10149 (2023).
[33] S. Shahriar, K. Hayawi, Let's have a chat! a conversation with chatgpt: Technology, applications, and limitations, arXiv preprint arXiv:2302.13817 (2023).
[34] Í. Silva, L. Marinho, A. Said, M. C. Willemsen, Leveraging chatgpt for automated human-centered explanations in recommender systems, in: Proceedings of the 29th International Conference on Intelligent User Interfaces, 2024, pp. 597–608.
[35] A. Erz, B. Marder, E. Osadchaya, Hashtags: Motivational drivers, their use, and differences between influencers and followers, Computers in Human Behavior 89 (2018) 48–60.
[36] A. Laucuka, Communicative functions of hashtags, Economics and Culture 15 (2018) 56–62.
[37] A. Dattolo, F. Ferrara, C. Tasso, The role of tags for recommendation: a survey, in: 3rd International Conference on Human System Interaction, IEEE, 2010, pp. 548–555.
[38] J. Vig, S. Sen, J. Riedl, Tagsplanations: explaining recommendations using tags, in: Proceedings of the 14th international conference on Intelligent user interfaces, 2009, pp. 47–56.
[39] R. H. Thaler, C. R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness, Yale University Press, New Haven, CT, 2008.
[40] A. D. Starke, C. Musto, A. Rapp, G. Semeraro, C. Trattner, "tell me why": using natural language justifications in a recipe recommender system to support healthier food choices, User Modeling and User-Adapted Interaction 34 (2024) 407–440.
[41] M. Rostami, V. Farahi, K. Berahmand, S. Forouzandeh, S. Ahmadian, M. Oussalah, A novel explainable and health-aware food recommender system., KDIR 2022 (2022) 208–215.
[42] S. Forberger, L. A. Reisch, P. van Gorp, C. Stahl, L. Christianson, J. Halimi, K. K. De Santis, L. Malisoux, T. de Magistris, T. Bohn, 'let me recommend…': use of digital nudges or recommender systems for overweight and obesity prevention—a scoping review protocol, BMJ open 14 (2024) e080644.
[43] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: Recommender systems handbook, Springer, 2015, pp. 309–352.
[44] A. E. Majjodi, Umap_25_llms_hash: Scalable umap for 25 llms with hashing and visualization, https://github.com/ayoubGL/UMAP_25_LLMS_Hash, 2025. GitHub repository.
[45] A. Petruzzelli, C. Musto, L. Laraspata, I. Rinaldi, M. de Gemmis, P. Lops, G. Semeraro, Instructing and prompting large language models for explainable cross-domain recommendations, in: Proceedings of the 18th ACM Conference on Recommender Systems, 2024, pp. 298–308.
[46] J. Ye, X. Chen, N. Xu, C. Zu, Z. Shao, S. Liu, Y. Cui, Z. Zhou, C. Gong, Y. Shen, et al., A comprehensive capability analysis of gpt-3 and gpt-3.5 series models, arXiv preprint arXiv:2303.10420 (2023).
[47] N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, in: 2007 IEEE 23rd international conference on data engineering workshop, IEEE, 2007, pp. 801–810.
[48] T. A. Kyriazos, et al., Applied psychometrics: sample size and sample power considerations in factor analysis (efa, cfa) and sem in general, Psychology 9 (2018) 2207.
[49] Department of Health and Social Care UK, Front of Pack nutrition labelling guidance, 2016. URL: https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance.
[50] M. Rayner, P. Scarborough, A. Boxer, L. Stockley, Nutrient profiles: development of final model, London: Food Standards Agency (2005).
[51] L. R. Flynn, R. E. Goldsmith, A short, reliable measure of subjective knowledge, Journal of business research 46 (1999) 57–66.
[52] Z. Pieniak, J. Aertsens, W. Verbeke, Subjective and objective knowledge as determinants of organic vegetables consumption, Food quality and preference 21 (2010) 581–588.

[53] A. Starke, M. Willemsen, C. Snijders, Promoting energy-efficient behavior by depicting social norms in a recommender interface, ACM Transactions on Interactive Intelligent Systems (TiiS) 11 (2021) 1–32.
[54] M. C. Willemsen, M. P. Graus, B. P. Knijnenburg, Understanding the role of latent feature diversification on choice difficulty and satisfaction, User Modeling and User-Adapted Interaction 26 (2016) 347–389.
[55] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.
[56] N. Tintarev, J. Masthoff, Explaining recommendations: Design and evaluation, in: Recommender systems handbook, Springer, 2015, pp. 353–382.
[57] H. Schäfer, S. Hors-Fraile, R. P. Karumur, A. Calero Valdez, A. Said, H. Torkamaan, T. Ulmer, C. Trattner, Towards health (aware) recommender systems, in: Proceedings of the 2017 international conference on digital health, 2017, pp. 157–161.
[58] I. Papastratis, D. Konstantinidis, P. Daras, K. Dimitropoulos, Ai nutrition recommendation using a deep generative model and chatgpt, Scientific Reports 14 (2024) 14620.
[59] C. Trattner, D. Elsweiler, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: Proceedings of the 26th international conference on world wide web, 2017, pp. 489–498.
[60] K. Bergram, M. Djokovic, V. Bezençon, A. Holzer, The digital landscape of nudging: A systematic literature review of empirical research on digital nudges, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–16.
[61] K. L. Hawley, C. A. Roberto, M. A. Bragg, P. J. Liu, M. B. Schwartz, K. D. Brownell, The science on front-of-package food labels, Public health nutrition 16 (2013) 430–439.
[62] M. F. Ghori, A. Dehpanah, J. Gemmell, H. Qahri-Saremi, B. Mobasher, How does the user's knowledge of the recommender influence their behavior?, arXiv preprint arXiv:2109.00982 (2021).
[63] C. Dietrich, Decision making: Factors that influence decision making, heuristics used, and decision outcomes, Inquiries Journal 2 (2010).
[64] Y. Luo, M. Cheng, H. Zhang, J. Lu, E. Chen, Unlocking the potential of large language models for explainable recommendations, in: International Conference on Database Systems for Advanced Applications, Springer, 2024, pp. 286–303.
[65] P. Chakrabarti, E. Malvi, S. Bansal, N. Kumar, Hashtag recommendation for enhancing the popularity of social media posts, Social Network Analysis and Mining 13 (2023) 21.
[66] O. Litvinova, F. B. Matin, M. Matin, B. Zima-Kulisiewicz, C. Tomasik, B. N. Siddiquea, J. Stoyanov, A. G. Atanasov, H. Willschke, Patient safety discourse in a pandemic: a twitter hashtag analysis study on# patientsafety, Frontiers in public health 11 (2023) 1268730.