Something Similar:
Developing Human-Centered Recommendations to Support "in-the-moment" Meal Choices for Health Goal Attainment

Pooja M. Desai

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

© 2025

Pooja M. Desai

All Rights Reserved

Abstract

Something Similar: Developing Human-Centered Recommendations to Support "in-the-moment" Meal Choices for Health Goal Attainment

Pooja M. Desai

Chronic disease remains one of the most pressing health challenges today and nutrition plays a central role in its management. Yet effective nutrition management is difficult for individuals to sustain. Existing informatics tools tend to focus on planning and tracking, but provide little support for real-time decisions — the very moments when people are most likely to abandon their goals. This dissertation addresses that gap by exploring similarity-based recommendations as a way to help people make meal choices that align with their health goals while still fitting into their existing plans and in-the-moment contexts, when decisions actually occur.

Aim 1 investigated how people reason about meal similarity and substitutability, and whether computational methods could reflect that reasoning. Human participants rated meal pairs and created a gold-standard dataset. Findings showed that similarity judgments drew on multiple dimensions, including form, cuisine, macronutrients, and ingredients, and that meals judged as more similar were also considered more substitutable, though context shaped these substitutability perceptions. Computational methods, especially LLM-based approaches, aligned closely with human judgments, suggesting that these methods can generate interpretable and human-aligned similarity assessments.

Aim 2 characterized user needs and assessed the feasibility of similarity-based recommendations in everyday contexts. Findings showed that meal choices were shaped by both personal priorities (taste, cost, preparation time) and the constraints of the situational setting. In open settings, people balanced multiple priorities with the help of existing tools. However, in constrained settings, they made on-the-spot compromises and found similarity-based recommendations particularly useful. Participants also emphasized the importance of autonomy and control at the system, algorithm, and moment-to-moment levels, highlighting the need for flexible, context-aware systems that adapt to evolving goals and constraints.

Finally, Aim 3 explored how large language models (LLMs) could enrich meal data and support conversational recommendation. Study 1 showed that LLMs could generate structured tags from free-text data, creating an enriched corpus that supported tailored recommendations. Study 2 used participatory design to identify preferences for short, user-led dialogues that allowed refinement along attributes such as flavor, texture, macronutrients, and effort while maintaining a supportive tone. Study 3 tested two conversational prototypes and found that participants qualitatively preferred the Dimension-Elicitation Chatbot, valuing its structured, collaborative style. At the same time, technical breakdowns underscored the fragility of current systems and the need for more adaptive and generative approaches.

Together, these studies contribute a new approach to meal recommendation that is grounded in human reasoning, demonstrates the feasibility of using free-text data as both input and corpus, and highlights the unique value of similarity-based methods for in-the-moment nutrition support. More broadly, this dissertation advances human-centered AI by showing how technical methods, human priorities, and ethical safeguards can come together to create decision-support systems that are interpretable, actionable, and designed for real-world use.

Table of Contents

Acknowledgments ix

Dedication xi

Chapter 1: Introduction 1

1.1 Dissertation Aims 3

1.2 Dissertation at a Glance 5

1.2.1 Aim 1 — Investigate the technical feasibility of human-reasoning aligned computational approaches for assessing meal similarity. 5

1.2.2 Aim 2 — Characterize user needs and assess the feasibility of similarity-based meal recommendations in everyday decision settings. 7

1.2.3 Aim 3 — Explore the use of large language models (LLMs) to elicit users on in-the-moment priorities and accordingly tailor meal recommendations. . 11

1.3 Contributions 15

1.4 Limitations 17

Chapter 2: Review of Related Literature 19

2.1 Challenges with Nutrition Self-Management 19

2.2 Theoretical Perspective: Human-Centered AI (HCAI) 21

2.3 Core Informatics Technologies for Nutrition Management 22

2.3.1 Personal Informatics (PI) Technologies 23

2.3.2 Health Coaching Systems 25

2.3.3 Recommendation Systems 26

2.4 User-Generated Nutrition Data 28

2.5 Development of a Free-Text 3k Meal Dataset 30

2.6 Takeaways: Towards More Human-Centered Nutrition Technologies 32

Chapter 3: Aim 1 — Investigate the technical feasibility of human-reasoning aligned computational approaches for assessing meal similarity 34

3.1 Related Work 35

3.2 Methods 36

3.2.1 RQ 1.1: Understanding Human Perceptions of Meal Similarity and Substitutability 36

3.2.2 RQ 1.2: Computational Approaches for Assessing Meal Similarity 39

3.3 Results 42

3.3.1 RQ 1.1: Understanding Human Perceptions of Meal Similarity and Substitutability 42

3.3.2 RQ 1.2: Computational Approaches for Meal Similarity 47

3.4 Discussion 50

3.5 Key Takeaways 53

Chapter 4: Aim 2 — Characterize user needs and assess the feasibility of similarity-based meal recommendations in everyday decision settings 57

4.1 Methods 58

4.2 Results 62

4.2.1 Participant Background and Experience 62

4.2.2 RQ 2.1: Existing Meal Selection Practices and Variation Across Situational Settings 64

4.2.3 RQ 2.2: Preference for Similarity-Based Recommendations Across Decision-making Context 67

4.2.4 RQ 2.3: Additional Design Considerations for Meal Recommendation Systems 70

4.3 Discussion 73

4.4 Key Takeaways 79

Chapter 5: Aim 3 — Explore the use of large language models (LLMs) to elicit users on in-the-moment priorities and tailor meal recommendations 81

5.1 Study 1 – Using LLMs to Generate Meaningful and Human-Aligned Tags for Meal Dimensions 83

5.1.1 Related Work 84

5.1.2 Methods 85

5.1.3 Results 88

5.1.4 Discussion 90

5.2 Study 2 — Participatory Design of AI-Based Conversational Agent for Meal Support 93

5.2.1 Related Work 93

5.2.2 Methods 94

5.2.3 Results 100

5.2.4 Discussion 111

5.3 Study 3 — Exploring User attitudes towards a Dimension-Elicitation Conversational Meal Recommendation System 114

5.3.1 Prototype Development Process 115

5.3.2 Methods 115

5.3.3 Results 118

5.3.4 Discussion 125

5.4 Key Takeaways 129

Chapter 6: Conclusions and Future Work 131

6.1 Summary of Dissertation 131

6.2 Contributions 133

6.3 Limitations 135

6.4 Future work 138

6.5 Conclusion 142

References 143

Appendix A: Dataset Information 161

A.1 Health Goals 161

A.2 Dataset Details 161

Appendix B: Aim 1 Supplementary Materials 164

B.1 RQ 1.2: Clustering Details and Coherence 164

Appendix C: Aim 2 Supplementary Materials 166

C.1 User Interview Guide 166

Appendix D: Aim 3 Supplementary Materials 168

D.1 RQ 3.1: Validating the use of LLMs for Meal Dimension Tagging 168

D.2 RQ 3.2: Participatory Design of Conversational Interactions for in-the-moment Preference Elicitation 170

D.2.1 Part 1 — Participatory Design Interview Guide 170

D.2.2 Part 2 — Participatory Design Activity Tasks 170

D.3 RQ 3.3: Users Attitudes Towards an LLM-based Dimension Elicitation Chatbot . . 173

D.3.1 Chatbot Prototype Prompts 173

List of Figures

1.1 Conceptual Approach for Similarity-Based Recommendation 4

2.1 Socioecological model for nutrition 20

2.2 Human-Technology-Ethics Framework for human centered AI 23

2.3 Mobile App for Meal Data Collection 31

3.1 Example of a pair of meal descriptions from the meal-pair dataset 37

3.2 RQ 1.2 Study Design 42

3.3 Key Identi ed Dimensions of Meal Similarity 45

3.4 Dimensions for Similairty Assessment and Association with Substitutability 55

3.5 Comparison of Computational Similarity Approach Against Gold-Standard Ratings 56

4.1 Example of Similarity Recommendation Scores and User-Facing Prototype 61

4.2 RQ 2.1 Results: Top Meal Priorities Across Users 65

4.3 RQ 2.2 Preference for Similarity-Based Recommendations in Open and Constrained Settings 68

5.1 Word clouds of annotated meal dataset by dimension. 91

5.2 Proposed "This or That" Conversation to Elicit Momentary Preferences 96

5.3 Roleplay Activity Sample Artifact 98

5.4 Feedback on Proposed "This or That" Probe 99

5.5 Example dialogs from the Baseline and DE Chatbot. 116

6.1 Proposed multi-agent system architecture for conversational meal recommendation. 140

A.1 The number of meals described across each of the 84 users in the nal dataset. The red line indicates the threshold of 50% of the meals in the corpus 162

B.1 Visualization of Clustering Approach in 2D 165

D.1 Crowdworking Task for validating LLM generated labels for meal dimensions . . . 168

D.2 Dream Based Brainstorming Activity sample artifact 171

List of Tables

3.1 Comparison of Meal Attributes Discussed by Humans v. LLMs 50

4.1 Demographics of participants in the Aim 2 user study (N=15) 63

4.2 Comparison of Features in Open and Constrained Settings 71

4.3 Summary of Requested Features and Their Proportions 72

5.1 Candidate Prompt Structure for LLM Tagging Task 87

5.2 Meal Dimension De nitions for Prompting 88

5.3 Model x Prompt Performance for LLM Tagging 89

5.4 Participant Demographics, Self-Ef cacy, and Health Literacy 101

5.5 Participant Experience with and attitudes towards AI 103

5.6 Study 3 Participant Demographics 119

5.7 Task Satisfaction Ratings for Baseline vs. DE-Chatbot 120

5.8 System Usability Ratings for Baseline vs. DE-Chatbot 121

A.1 Nutrition Goals and Meal Criteria 163

B.1 Cluster Descriptions with Example Meals 164

D.1 Tag Performance — Tags Match Dimension Described 169

D.2 Tag Performance — Tags Match Meal Described 169

Acknowledgments

I am grateful for the many people who supported me on this journey. This document uses "I" a lot, but the truth is none of this work was done alone. I carry such deep gratitude for everyone who has shaped this process.

First of all, a hearty thanks to my committee for their support throughout the process.
To my advisor, Lena Mamykina: You rst welcomed me into your lab (and into informatics) when I was 19. Thank you for guiding me through the past decade, for indulging my curiosity, and for our deeper conversations about science and the world that have shaped how I think.
To Sue Bakken: I admired you long before we ever met. Thank you for your boldness, for your leadership, and for being a champion for the Justice Informatics group. I have learned so much from watching how you navigate your research.
To Noémie Elhadad: Thank you for your feedback, for ideating with me, and for pushing me to make this dissertation stronger. I admire you in so many ways and am grateful to keep learning from your example.
To my external committee members, Pedja Klasnja and Heather Cole-Lewis: I have followed your work for years and I am so thankful that you joined my committee. Thank you for your time, insight, and support.

To the participants who shared their time and re ections: human-centered research is impossible without you. Thank you for making this work real.

To the ARCH Lab members who came before me — especially Silis, Maria, and Elliot — thank you for adopting me as a “lab little sister” and for continuing to offer perspective and advice. To the ARCH Lab students that follow, especially Kayla Schiffer-Kane and Muying Li: you are amazing, and I am always here for you in return.

To the Justice Informatics Collaborative — Oliver, Chris, Apara, and Sal — I have learned so much from you and am proud of the work we have built together.

To the students of DBMI: thank you for the community we built here. You kept me grounded and made this journey joyful. In particular thank you to Harry, Courtney, Rebecca, Vinny, JB, Tara, Lauren, Adit, Young, and so many others. A special shoutout to my work-wife, Adrienne: without you, I would have been even more of a mess.

To my Wolfpack (Ashna, Meredith, Re, Elizabeth, and Nicole) and my motley crew from middle-school (Alexi, Amelia, Simone, and Rachel): thank you for believing in me long before I knew this path was possible. I am so grateful for the way you have championed me and my dreams, in all their forms.

And nally...my family. Thank you for fostering my curiosity and always, always having my back. To my parents: thank you for answering my endless “but why?” questions at bedtime, for peer-reviewing my stuffed animal experiments, and for encouraging me to get up every time I stumbled. To my baby brother: you are my biggest cheerleader and constantly push me to see things from new perspectives. I love you to pieces. And to Josie — my puppy love, my sunshine girl. You can't read this (obviously), but you bring immeasurable light to my life. It's true what they say: you rescued me as much as I rescued you. Pup cups when I get home!

I am deeply grateful to the NLM and NIDDK for funding this work. Objective, rigorous, and equitable science matters. Fund it and the diverse researchers behind it.

Thank you, all of you, for walking this path with me.

Dedication

To my family – Mama, Baba, Aman, and Josie – you are my world. Everything I am is because of you. Thank you, thank you, thank you, thank you...

Chapter 1: Introduction

Chronic diseases are among the most pressing and widespread health challenges affecting one in three adults and having an estimated cost of $47 trillion by 2030 [1, 2]. It is well understood that nutrition plays a central role in managing these conditions, yet effective nutrition management tends to be extremely dif cult for individuals to navigate. One reason is that while general guidelines for healthy eating are available, physiological responses to food are highly individual making it dif cult to identify personally meaningful nutritional targets [3, 4, 5]. Even when someone has clear nutrition goals, aligning daily meals with these targets requires them to have high levels of nutritional literacy, persistence, and sustained motivation [6, 7, 8].

Furthermore, people do not make meal choices in a vacuum: in addition to considerations around nutritional goals, people must navigate other factors such as competing priorities, other preferences or restrictions, and be able to adapt their plans to dynamic contexts [9].

The ubiquity of smartphones and digital technologies offer new opportunities to leverage personal data to support individuals as they make nutrition choices [10, 11, 12, 13, 14]. Within informatics and HCI, a range of technologies have been explored to support nutrition self-management [5, 12, 15, 16, 17, 18]

One type of solution are personal informatics (PI) technologies which help individuals aggregate meal data and give them a way to re ect on their collected meal records [17, 19]. Both research prototypes and commercial PI technologies (e.g., MyFitnessPal, LoseIt!) have demonstrated the utility of these tools in promoting awareness of existing behaviors through tracking meals and learning through re ecting on past meal records [20, 21, 22, 23, 24]. However, while PI tools support the collection of and re ection on personal data, they offer little support in translating awareness of patterns into speci c actions to help them meet their health goals in the future [25, 26, 27].

Somewhat addressing this limitation, digital health coaching systems aim to help users identify health goals to focus on and provide reminders and nudges to encourage them to stay on track [28, 29]. These systems have shown promise in helping motivate users and support them with accountability as they navigate making nutrition changes [30, 29]. While support setting higher-level goals is important, users still struggle with how to break down and operationalize these health goals in the routines of their daily life. They seek more tailored in-the-moment guidance to help them better make needed adjustments in-the-moment [29, 30, 31, 32].

A third class of technologies, Recommendation Systems (RecSys) can help users discover new meals that align with their preferences [16, 33, 34]. Broadly, these systems focus on supporting exploration and novelty by generating personalized suggestions based on factors such as past choices or stated taste preferences. However, most nutrition-focused RecSys give little attention to factors beyond preference — such as health goals, dietary constraints, or contextual considerations like current meal plans or available ingredients [35].

In essence, existing informatics tools for nutrition support long-term planning or post-hoc analyses but offer little support for real-time action. This gap — supporting individuals as they make food choices in-the-moment — is a critical and underexplored area. Prior work has shown that these in-the-moment decisions, often made when individuals are constrained by considerations like hunger, social settings or pre-existing plans (e.g., leftovers, a dinner reservation), are when people are most likely to abandon their health goals [36]. In these moments, even if someone knows what they should eat to meet their health goals, following through is dif cult without guidance that is both actionable and realistic.

A growing body of work in informatics and human-computer interaction discusses Human-Centered AI (HCAI) as an approach to address these kinds of gaps. Fundamentally this approach focuses on building technologies aligned with how people actually engage in real-world decision-making with the intention of supporting user control, responsible use, and thoughtful consideration of how the system will impact users and the world if integrated in their daily lives [37, 38, 39]. Xu et al. discuss HCAI as a balance of three factors that all in uence on another: technology, human

needs, and ethics/responsible use [39].

Keeping this approach in mind enables new ways of exploring how to support individuals with nutrition management in-the-moment. Using this framework equips us to better understand important questions, such as: How can we help people make meal choices on the spot that respect their preferences and existing plans while better aligning with their health goals? The diverse socioecological factors in uencing meal decisions highlight the need for human-centered nutrition technologies tailored to users' unique decision-making contexts [15, 18, 9].

1.1 Dissertation Aims

This dissertation proposes a novel approach to answering this question: similarity-based meal recommendations. Rather than asking users to make dramatic changes to their plans or consider unfamiliar options, this method starts from a user's initial meal choice and recommends similar alternatives that better align with their goals. As a result, individuals make meal choices in-the-moment that align with their health goals while remaining consistent with their existing plans and preferences.

Though similarity-based recommendation had been explored in other domains such as e-commerce, it had not been examined in the context of nutrition. This approach differs from traditional nutrition recommendation approaches by anchoring recommendations in users existing plans and health goals. As is expanded upon in Chapter 2, traditional recommendation approaches focus on generating suggestions aligned with data around users previous meal choices or just their health goals. These approaches tend to be data expensive, requiring highly labeled metadata and considerable information about users' preferences that can result in a delay in making quality recommendations until suf cient user data is gathered (cold start problem). Additionally, they do little to account for the lived experience of users who often have existing meal plans that they are unwilling to abandon entirely.

Instead, by anchoring recommendations in what the user already plans to eat, this system offers a more practical and persona-aligned path toward making healthier choices. By narrowing the

Figure 1.1: Conceptual Approach for Similarity-Based Recommendation: Rather that suggesting something that requires users to abandon existing plans this approach focuses on recommending adjustments to the existing meal plan that helps meals better align with existing plans and health goals.

distance between intention and action, this approach offers a new kind of real-time decision support — one that does not require extensive long-term data or user pro ling but instead adapts in-the-moment to the meal at hand. Crucially, this approach can offer support exactly when people need it most: in-the-moment of choice, when users are most likely to struggle and when timely guidance can have the greatest impact.

However, such an approach for meal recommendations raises several open research questions at the intersection computational modeling, user practices, user-centered design and ultimately HCAI. First, it is necessary to understand what similarity means in the context of meals, both from the perspective of human perception and computational approaches. Second, there are questions about how similarity-based recommendations can be generated and presented in ways that are useful and actionable for users. Third, there are questions about how to ensure technical solutions are human-centered and can facilitate in-the-moment decision-making in a manner that supports individual agency, minimizes cognitive burden, and adapts to different decision-making environments in which people make meal choices.

This dissertation explored these questions towards the design of a conversational similarity-based meal recommendation system aimed at helping individuals make in-the-moment dietary

decisions that align with their health goals.

Aim 1 investigated how humans think about meal similarity and explores the technical feasibility of computational methods for assessing meal similarity. Details of this aim are found in Chapter 3.

Aim 2 characterized user needs and examined the practicality of implementing similarity-based recommendations in everyday food decision-making, through a lab feasibility study. Details of this aim are found in Chapter 4.

Aim 3 explored how LLMs can be leveraged to deliver dynamic and personalized similarity-based recommendations in-the-moment. Details of this aim are found in Chapter 5.

Lessons from these aims advance knowledge about how informatics solutions can better support healthy meal choices in-the-moment and inform considerations for building HCAI for nutrition that better meets and balances the human, technical, and ethical considerations.

1.2 Dissertation at a Glance

The following section includes an overview of the main objectives, research questions, methods and key ndings from each aim of the dissertation. The subsequent chapters of this dissertation expand on the motivations, methods, and ndings of each aim in more detail.

1.2.1 Aim 1 — Investigate the technical feasibility of human-reasoning aligned computational approaches for assessing meal similarity.

Objective: This aim investigated what attributes humans consider when determining how similar meals are and how this relates to their perceptions of how interchangeable two meals are for each other. Then, it explored the alignment of different computational approaches for determining meal similarity with human reasoning.

Research Questions Below are the core research questions that guide Aim 1.

'& 1•1: What attributes of meals inform human perceptions of meal simi-

larity and substitutability?

'& $_{1 \bullet 2}$: Which computational similarity approaches align with human perceptions of meal similarity?

Study Methods & Analysis: For Aim 1, a two-part mixed methods study with crowdworkers was conducted to understand how humans consider meal similarity, and which computational approaches align with these perceptions.

The rst part of the study (RQ 1.1) characterized human perceptions of meal similarity and identi ed key attributes that inform these judgments. Participants rated pairs of meal descriptions and provided explanations of their perceived similarity. This generated a gold-standard dataset capturing human reasoning about meal similarity and substitutability.

The second part of the study (RQ 1.2) evaluated how well different computational methods could replicate human assessments of similarity, providing insight into the technical feasibility of aligning algorithmic recommendations with human reasoning. This evaluation focused on three different computational approaches (cosine similarity, k-means clustering, and large language models) to determine the similarity between free-text meal descriptions similarity. To evaluate these different approaches, the gold standard dataset generated in RQ 1.1 was used to measure their alignment with human perceptions of meal similarity.

Primary Findings: Meals pairs in the study were rated by 14 unique crowd workers. In RQ 1.1, primary ndings showed that people considered several attributes of meals that inform perceptions of similarity. Most frequently they referenced: meal form, meal cuisine, macronutrient composition, and core components. Meals perceived as more similar were also regarded as more substitutable for one another ($r = 0.63$, $p < 0.001$). However, participants also noted substitutability as a complex construct often impacted by situational constraints. Ultimately, these ndings suggest that similar

meals might be good candidates for actionable recommendations of alternative meals to substitute.

Next, RQ 1.2 explored three computational approaches for assessing the similarity between two free-text meal descriptions: cosine similarity, k-means clustering, and LLMs (GPT-4o). Cosine similarity and LLM-based approaches showed the strongest alignment with human ratings ($r=0.68$ and $r=0.66$, respectively, both $p < 0.001$), with LLMs exhibiting the fewest discrepancies between predicted and human similarity judgments. K-means clustering performed poorly in this context ($r=0.26$, $p=0.09$) and produced the highest rate of inconsistencies. Overall, the results demonstrated the technical feasibility of leveraging computational approaches, particularly LLMs, to approximate human reasoning about meal similarity, laying the foundation for human-centered meal recommendation systems.

Key Takeaways: Aim 1 demonstrated that human perceptions of meal similarity are shaped by multiple dimensions, including meal form, cuisine, macronutrient composition, and core ingredients. Meals judged as more similar were also considered more substitutable, though situational factors also in uenced this relationship. Among the computational approaches tested, cosine similarity and LLM-based methods showed the strongest alignment with human judgments, with LLMs producing the fewest discrepancies. These ndings show computational approaches are capable of making assessments of meal similarity that are comparable to those of humans, and align with human reasoning.

1.2.2 Aim 2 — Characterize user needs and assess the feasibility of similarity-based meal recommendations in everyday decision settings.

Objective: Characterize user needs and examine the practicality of implementing similarity-based recommendations in everyday food decision-making, through a lab

feasibility study.

Research Questions Below are the core research questions that guide Aim 2.

'& 2•1: What are individuals' existing practices for meal selection, and how do these differ between open and constrained decision-making settings?

'& 2•2: How do users perceive the usefulness of similarity-based recommendations for supporting meal decision-making in everyday decision settings?

- Hypothesis: In constrained settings users prefer recommendation that incorporate meals similar to their original plan

'& 2•3: What features do users value in nutrition-management technologies and how do these preferences vary across decision settings?

Study Methods & Analysis: This aim examined how situational constraints in uence meal choices and the design of nutrition recommendation systems that are both useful and actionable. Participants (N=15) were recruited using Craigslist and Columbia University's RecruitMe platform. They completed 60-minute virtual sessions combining surveys, semi-structured interviews, scenario-based recommendation tasks, and participatory design exercises. RQs 2.1 - 2.3 were addressed through activities within this single study.

To address RQ 2.1, participants rst completed surveys assessing the importance of factors such as cost, taste, preparation time, social context, and other in uences on meal selection, and reported the meals they had eaten the previous day. Semi-structured interviews explored their experiences and reasoning around meal choices. Survey responses were analyzed to identify general patterns in priorities, while interview transcripts were thematically coded to contextualize these patterns and capture nuances in decision-making across different situations.

For RQ 2.2, participants engaged with two meal decision scenarios: an open setting (grocery store) and a constrained setting (cafeteria). In each scenario, recommendations were generated based on cosine similarity to participants' previous meals, presenting two similar and two dissimilar options. Participants reviewed the recommendations, selected a preferred option, and explained their rationale. Analysis combined quantitative measures of preference for similarity with qualitative coding of participants' reasoning to understand how situational constraints shaped actionable choices.

Finally, RQ 2.3 explored participants' expectations for decision-support technologies. Using design probes and participatory exercises, participants indicated which features and types of explanations would be most helpful in open and constrained settings. Artifacts from these sessions were coded to quantify feature preferences, while qualitative re ections provided insight into user expectations for control, explanation style, and interaction modalities.

Primary Findings: A total of 15 participants took part in the study. This study explored how people's meal choices are in uenced by their priorities and setting, and implications for the design of nutrition recommendation systems that account not only for individuals' preferences but also for the context in which they make nutritional decisions. Taking a mixed methods approach, the ndings shed light on how decision settings shape decision-making priorities, and requirements of useful decision-support technologies.

RQ 2.1 explored what factors participants considered when making meal choices across different settings. Despite considerable variability in individual priorities, there were some common considerations that emerged including cost, taste, and meal preparation time. In open settings individuals focused on balancing priorities such as health and budget, assisted by technologies like mobile nutrition tracking and planning apps. However, in constrained settings they tended to abandon one priority for another, such as ignoring health goals to choose something budget friendly. Additionally, they rarely

used technologies in these settings to support nutrition decision-making.

Then, RQ 2.2 evaluated user-perceptions of a new approach for meal recommendation that makes suggestions based on similarity to a users' historical meal records in open settings and current meal plan in constrained settings. Participants signi - cantly preferred similarity-based meal recommendations in constrained setting as the recommendations closely aligned with their existing meal plans and were considered more practically actionable. However “similarity” emerged as a complex concept, with users considering many different dimensions, including meal form, cuisine, ingredient preparation, and nutritional value when considering what constituted a similar recommendation that was also actionable.

Finally, RQ 2.3 expanded on users' needs for technologies to support meal decision-making in both open and constrained settings. While they valued explanations that clari ed how recommendations met health goals, they were less interested in understanding how the algorithm generated those recommendations. Additionally they wanted more control at three levels: the system, the algorithm, and the output. At the system level they emphasized being able to input stable dimensions (e.g., allergies, health goals, and dietary restrictions) and turn on or off features of interest (e.g., explanations); these aspects would remain consistent across. However, they also emphasized wanting dynamic control where they could interact with the system in-the-moment to specify which dimensions of similarity the algorithm should use to generate recommendations, re ne options, and control how recommendations were presented.

Key Takeaways: Aim 2 revealed that meal choices are shaped not only by individual priorities, such as cost, taste, and preparation time, but also by the context in which decisions occur. In open settings, participants balanced multiple priorities with support from existing nutrition technologies, whereas in constrained settings they made on-the-spot compromises and rarely used technology. Similarity-based meal recommendations were particularly valued in constrained settings, as they aligned closely

with users' existing meal plans and were considered more actionable, though the specific dimensions shaping perceptions of similarity — such as meal form, cuisine, ingredients, or nutritional value — varied across constrained settings in ways that were not always predictable. Participants also emphasized the importance of control and customization in recommendation systems, favoring explanations of how suggestions met health goals over detailed algorithmic transparency.

These findings show that human-centered AI for nutrition needs to account for both people's stable priorities and the shifting constraints of everyday settings. Building in mechanisms to support user control and dynamically adjust around dimensions in real-time can make in-the-moment similarity based recommendation genuinely useful and aligned with real-world decision-making.

1.2.3 Aim 3 — Explore the use of large language models (LLMs) to elicit users on in-the-moment priorities and accordingly tailor meal recommendations.

As a brief summary, Aims 1 and2 highlighted the feasibility and utility of similarity-based meal recommendations. Findings also identified common dimensions of meal similarity, including: meal form, meal cuisine, main macronutrients, and key ingredients. Additionally, Aim 2 highlighted that the most relevant dimension of similarity varies by the situational setting; users sought control to dynamically filter and refine recommendations in-the-moment along these axes of similarity.

Objective: Building from Aims 1 and 2, this aim investigated how LLMs can be leveraged to dynamically elicit which meal dimensions similarity-based recommendation should attune to in-the-moment through three studies. RQ 3.1 focused on approaches to enrich an existing dataset of free-text meal descriptions with LLM generated tags along the four meal similarity dimensions (meal form, meal cuisine, main macronutrients, and key ingredients). Enriching the meal database used for recommendation with

these tags allows users to dynamically lter and re ne recommendation in-the-moment along these different dimensions. Next, RQ 3.2 used a participatory design approach and explored how user seek to engage with AI and would want to have conversational interactions in-the-moment to inform dynamic meal recommendations. Finally, RQ 3.3, explored users attitudes towards a lab prototype of an LLM-based conversational system for similarity-based meal recommendation in-the-moment.

Research Questions Below are the core research questions that guide Aim 3.

'& 3•1: How effectively can large language models generate semantically meaningful and diverse tags from meal descriptions in an unsupervised manner?

'& 3•2: How can LLMs be used to elicit in-the-moment meal preferences to support recommendation?

'& 3•3: What are users' experiences and perceptions of interacting with an LLM-based chatbot for in-the-moment meal recommendation?

Study Methods & Analysis: For this aim, three separate studies were conducted to understand how LLMs could be used to support dynamic in-the-moment recommendation support.

Study 1 (RQ 3.1) evaluated the ability of LLMs enrich a corpus of free-text meal records along four dimensions of meal similarity: meal form, meal cuisine, main macronutrients, and key ingredients. This study explored several candidate LLM models (ChatGPT, Llama, and Gemini) and prompting approaches (zero-shot, zero-shot with de nitions, few-shot, few-shot with de nitions). Each model x prompt combination was used to generate tags for each meal dimension for a small subset of meals in the corpus. Tags were validated by human crowdworkers to assess alignment with the meal and dimension to determine the optimal tagging approach. The most accurate approach was used to tag a larger meal library.

Study 2 (RQ 3.2) explored the design of interactive conversations using LLMs to elicit users' momentary preferences and contextual constraints. A total of 5 participants took part in the study. Participatory design sessions and scenario-based exercises were used to understand how users wanted recommendations to be re ned in in-the-moment, what constraints should guide the system, and how the system should balance user control with guidance.

Study 3 (RQ 3.3) examined users attitudes toward a lab prototypes of two LLM-based conversational system for similarity-based meal recommendation in-the-moment. The rst prototype was a Baseline-Agent that took a users health goal, meal plan and dietary restrictions as input, assessed if the input meal met the users health goal, and if not retrieved similar recommendations that did meet the goal from an expert reviewed database of meal records. The second prototype was a Dimension-Elicitation-Agent which did the same things as the Baseline Agent, but also asked users a few short questions about what meal dimensions mattered to them in that moment to tailor similarity-based to their in-the-moment preferences. A total of 8 participants took part in the study. Participants used each prototype to get recommendations in two simulated scenarios and re ected on their experiences using each agent through surveys and qualitative interviews.

Primary Findings: For this aim we conducted three separate studies to understand how LLMs could be used to support dynamic in-the-moment recommendation support.

In Study 1 (RQ 3.1), LLMs and human annotation were used to label meals across multiple dimensions. Overall Llama and the Few-Shot + De nition emerged as the optimal model–prompting combination. Tags generated through this approach had an accuracy of 96% or higher across all four dimensions: form, cuisine, macronutrients, and key ingredients. This approach was used to tag a larger meal library which serves as catalog for meal similarity recommendation in subsequent studies.

Then, Study 2 (RQ 3.2) used participatory design sessions (N=5) to explore how users expect meal recommendation conversations to progress, and to identify dialogue structures that feel natural and supportive in real-world decision-making. Text-based interaction was strongly preferred for its exibility and discretion, with limited interest in voice or image input. Findings elicited certain design requirements for these conversational interactions: (1) Conversations should short (minute) and follow a natural, context-aware progression, beginning with an initial suggestion and allowing re nement based on attributes such as avor, texture, macronutrients, effort, and setting; (2) Conversations should be user-led and provide opportunities for the user to control the direction of the interaction based on their needs; (3) They emphasized that AI should maintain a non-judgmental, supportive, and informative tone, while respecting privacy and using personal data only when relevant.

Finally, in Study 3 (RQ 3.3) I conducted a user evaluation of two chatbot prototypes: a Dimension-Elicitation Chatbot that let users re ne along recommendations along the four similarity dimension and a Baseline Chatbot that provided recommendations without re nement. A total of 8 participants evaluated the chatbots in randomized order. Qualitative feedback showed a preference for the Dimension-Elicitation Chatbot. They described its structured style as clearer, more purposeful, and better suited to quick meal decisions. The targeted follow-up questions not only helped them think through options but also made the process feel collaborative, as if they were working with the chatbot to reach a recommendation. Personalization through the dimension-elicitation conversations helped make suggestions feel relevant and engaging, but this strength was undermined by technical breakdowns such as context loss and misalignment, which sometimes frustrated users. Finally users discussed that while retrieval-based recommendations felt authentic they were also limiting, pointing to a need for more adaptive and generative recommendations to better support participants in future iterations.

Key Takeaways: Aim 3 demonstrates that LLMs can generate meaningful tags for free-text meal data and support interactive, similarity-based recommendations that respond to both momentary and stable user needs. Users preferred short, user-led dialogues that unfold in a context-aware sequence, allowing re nement based on avor, texture, macronutrients, effort, and setting, while maintaining a non-judgmental and supportive tone. In the prototype evaluation, participants qualitatively favored the Dimension-Elicitation Chatbot, describing its structured, targeted follow-up questions as clearer, more collaborative, and better aligned with quick decision-making. At the same time, technical limitations such as context loss and rigid retrieval highlighted the fragility of current LLM systems and the need for more adaptive, generative approaches.

These ndings show that LLMs can advance human-centered AI for nutrition by structuring meal data and enabling conversations that feel collaborative and user-driven. Technical approaches need to prioritize interactions that adapt to momentary priorities and remain short, supportive, and user-led. These ndings also underscored the importance of safeguards around privacy, discretion, and tone to build trust, and highlighted the need for generative approaches that can sustain long-term use without compromising safety.

1.3 Contributions

This thesis advances the eld of biomedical informatics by addressing a critical and underexplored challenge in nutrition management: supporting individuals in making real-time meal decisions that align with their health goals while respecting their preferences, plans, and constraints. This work used a human-centered AI approach that prioritizes on-the-spot decision support, breaking from existing technologies that focus on tracking, general exploration, or long-term planning. By anchoring the design of recommendations in a deep understanding of how people make food decisions in practice, this work offers a shift towards more human-centered intelligent systems

grounded in users' meal practices and designed to support them in the real world.

The signi cance of this work extends beyond nutrition to broader applications in personalized health informatics, decision-support systems, and human-AI interaction demonstrating how AI can be designed to align with human cognitive processes, preferences, and constraints in real-world settings. The contributions of this dissertation include:

1. Development of a Novel Similarity-Based Recommendation Method: This dissertation introduces a new approach to supporting meal choices through similarity-based recommendations. The method begins with what a person already intends to eat and suggests adjacent alternatives that better align with their nutritional goals. Few existing recommender systems focus on these in-the-moment decisions, even though this is when support is often most needed. By building from a user's existing choice rather than replacing it, this approach respects personal autonomy, situational constraints, and context. It also accommodates the multifaceted nature of food decision-making, where choices are in uenced by social settings, available options, and immediate practical needs.

2. Use of Free-Text Data for Meal Recommendation: This thesis also explores the feasibility of using unstructured, patient-generated meal data as a source for generating actionable, real-time dietary recommendations. Prior work has focused on using curated libraries of recipe data to deliver recommendations. The approach used here opens new opportunities for harnessing free-text inputs to understand users' existing eating patterns and support health goal-aligned adjustments.

3. Support for "In-the-Moment" Nutrition Decisions: This dissertation also focuses on how intelligent systems can support meal planning in-the-moment, bridging the gap between user intention and health-aligned action at the time of decision. Few informatics technologies for nutrition are intended for use in-the-moment, instead focusing on supporting re ection, goal setting and discovery. By focusing on helping individuals make health-aligned meal adjustments in-the-moment this approach can support individuals when they are most vulnerable

to faltering in their self-management.

4. Human-Centered Approach to AI-Driven Recommendation: Finally, this work builds on a growing body of work within informatics around building human-centered AI systems that center user practices, autonomy, and control in the development of AI systems. This dissertation uses this approach to guide research questions and the development of this system and explores the utility of such human-centered technologies.

1.4 Limitations

While this dissertation provides a foundational exploration into similarity-based, conversational nutrition RecSys, several limitations must be acknowledged.

First, the effectiveness of the similarity-based recommendation approach that relies on drawing recommendations from existing databases and catalogs is contingent on the diversity and granularity of such catalogs. If the dataset is limited, it may constrain the system's ability to generate recommendations that are meaningfully similar to the user's reference meal, potentially diminishing its usefulness.

Second, the scope of this work is bounded by a speci c set of nutrition health goals, which are based on general healthy eating principles. These may not be appropriate or suf cient for individuals managing speci c health conditions or with different dietary needs. As such, the system's applicability across various clinical contexts may be limited without further adaptation.

Third, although the dissertation addresses a real-world challenge, much of the evaluation is conducted through small, lab-based user studies. This means that several practical challenges — such as integration into users' daily routines, technical robustness in deployment, and long-term engagement — are not directly assessed, but are crucial areas for future exploration. While the ndings offer promising insights into the acceptability and potential impact of this meal similarity approach, future work is needed to validate these ndings with a larger, more diverse participant pool in real-world environments.

Fourth, this work focuses primarily on individual-level decision-making. However, nutrition

choices are shaped by broader structural factors, including social, political, cultural, and geographic in uences. While this research attempts to incorporate some of these contextual attributes, it does not fully address how systemic barriers may affect users' ability to act on recommendations. Deeper exploration of these dimensions is needed to ensure that future systems are equitable and inclusive.

Chapter 2: Review of Related Literature

This chapter reviews related work in health, informatics, and human-computer interaction that informs the questions explored in this thesis. Speci cally, the subsections below discuss challenges of nutrition self-management, introduce HCAI and frame the work of this dissertation through this lens, review core technologies for nutrition support, and examine opportunities to leverage user-generated data, and nally conclude with an overview of related prior work from the lab that informs this dissertation.

2.1 Challenges with Nutrition Self-Management

Food is a cornerstone of human society and is deeply intertwined with personal and cultural identity. It also plays a critical role in supporting health and well-being. Robust evidence has shown the importance of nutrition in chronic disease management, which impacts nearly 50% of individuals in the United States [1, 2]. Studies have shown that eating a healthy diet reduces the likelihood of becoming overweight and developing chronic conditions such as cardiovascular disease and diabetes, and is linked to reduced mortality for cancer survivors [40, 41].

Although making meal choices is a routine part of daily life, managing nutrition is often complex and demanding [42]. Individuals must navigate nutrition changes while balancing preferences with long-term goals, responding to evolving life circumstances, and making food decisions in the context of social, cultural, and environmental factors.

The Socioecological Model of Nutrition (see Figure 2.1) offers a valuable lens for understanding the complexity of food decision-making. It situates individual behaviors within a broader network of in uences, including personal preferences, family and social dynamics, physical surroundings (e.g., food availability and access), as well as broader policy and economic structures

(e.g., SNAP and EBT programs) [11, 18, 9, 36]. Keeping in mind this layered ecosystem, several key barriers to healthy eating have been identi ed.

Figure 2.1: Socioecological model for nutrition adapts the more general Socioecological Model to highlight the many different individual, social, and structural layers that impact meal choices. [9]

First, physiological responses to food are highly individualized, making it dif cult to de ne meaningful nutritional goals at the personal level [3, 4, 43].

Second, even with clear goals, aligning daily meals with nutrition targets requires persistence, nutritional literacy, and sustained motivation [6, 18, 44]. Studies show that a signi cant portion of the US population have low levels of nutritional literacy, which includes accurately estimating the nutritional content and portions of speci c meals [45, 46].

Third, making and sustaining healthy choices is cognitively demanding. Dual process decision making theories discuss two modes of thinking: Type 1 decisions are "fast" and habitual, while Type 2 decisions are "slow" and deliberative [47]. As food decisions are often habitual, shifting toward intentional and health-aligned choices demands signi cant cognitive effort. Even well-prepared individuals must sometimes make meal decisions on the spot. Past work has shown that even when individuals understand what choices are healthy, they still fail at the point of decision-making due to cognitive interference while buying groceries or ordering meals at restaurants [48, 49, 50].

Finally, even when individuals adopt healthier eating patterns, these patterns can be dif cult to sustain over time. Nutrition impacts the body gradually, requiring individuals to consistently make healthy meal choices amidst the chaos of daily life in order to observe changes. Prior work has shown that many individuals who start healthy diets, fail to sustain these positive changes, falling back onto their old habits [42, 48].

Simply put: knowing what to eat is not the same as being able to act on that knowledge consistently. This complexity underscores the need for nutrition technologies that are sensitive to users' lived experiences and responsive to the real-world contexts in which meal decisions are made [11, 9]. Technologies should better support nutrition decisions on the spot, at the exact moment they are made, whether that's ordering at a restaurant, buying groceries, or opening the fridge after a long day [33, 34, 51].

2.2 Theoretical Perspective: Human-Centered AI (HCAI)

Within interactive systems spaces there is a growing focus on building HCAI systems. Broadly, HCAI prioritizes human agency in AI-driven interactions by designing technologies that support and augment human skills, inform decision-making, and ultimately preserve user autonomy and control [52, 38, 37, 39, 53]. Though de nitions vary, many also consider the broader societal context in which these systems operate, including ethical considerations like privacy, equity, and governance [52, 38, 37, 39].

The importance of human centered solutions has been explored across several domains but bear particular importance which building AI solutions for healthcare settings, where decisions often carry high personal and societal stakes and can directly impact well-being [53, 54, 55]. Andersen et al. describe ve areas for how HCAI can be applied to improve health technologies [53]. These include: (1) alignment with human values, where AI systems are evaluated based on how well they re ect human and societal priorities; (2) design process, which centers users throughout development; (3) interaction, which includes both usability and the long-term dynamics between people and AI; (4) sociotechnical system, recognizing that healthcare happens through networks of people, tools, and work ows; and (5) implementation, which considers the practical realities of deploying AI in real-world clinical practice.

To operationalize this de nition, Xu et al. describe a three pronged Human-Ethics-Technology Framework for HCAI which emphasizes the importance of (1) ethically aligned design (2) technology that re ects human intelligence and retains human autonomy, and (3) useful and usable for decision-making in practice [39]. This framework is illustrated in Figure 2.2. Though broad, and not speci c to health-related technologies, this framework provides a structure through which to evaluate existing solutions and guide the development of new ones.

The research in this thesis drew on this three-pronged framework of HCAI as a lens to evaluate current tools that aim to support everyday meal decisions and inform the research questions and methods. This framework helps to develop new approaches that inform without overriding and not only respect but align with users' lived experiences and personal goals when making meal choices.

2.3 Core Informatics Technologies for Nutrition Management

Technologies for supporting eating practices has been a vibrant research area within HCI. Previous research in this community has advocated for a holistic approach to supporting nutrition that acknowledge food and eating as a and multifaceted and personally meaningful practice [11]. We highlight major types of interactive of technologies for nutrition management within HCI: (1) personal coaching systems, (2) personal informatics technologies, and (3) nutrition recommendation

Figure 2.2: Human-Technology-Ethics Framework for human centered AI by Xu et al. [39]

systems. The "Human-Technology-Ethics" HCAI framework serves as a lens to discuss the gaps and opportunities for these solutions [39].

2.3.1 Personal Informatics (PI) Technologies

A rst group of solutions are PI technologies examine ways individuals engage with personal data in order to increase their self-knowledge and awareness, and to inform their future actions [14, 15, 18]. These systems assist self-management by supporting data collection and data re ection and have been used in a variety of settings including personal nance and health [19, 56, 17].

Within the context of food and nutrition, PI tools have been explored to help individuals explore eating patterns [21, 57], identify speci c foods triggering health areups [22,24,59], encourage goal-oriented eating [23, 58, 59], and facilitate patient-provider communication around health behaviors [22, 60, 61]. For example, Foodprint helps users identify foods that may trigger IBS [22]. Users can easily log meal data through free-text, images, and can record how meals made

them feel, providing a simple way for users aggregate this data. The app helps them re ect on collected data to compare ingredients in meals that made them well and poorly as well as share logged meals with their healthcare providers.

Increasingly PI systems have leveraged machine learning to help users engage with their collected data. Simple computational methods have been used to help individuals identify underlying patterns in previously collected data that can inform their choices [57, 62, 63]. For example, HealthMashups enabled users to track many different kinds of data related to their health and wellbeing such as food, sleep, and physical activity [57]. Using regression to explore associations between collected data, the system could surface insights such as "You tend to consume more calories when you sleep less" for users to consider.

Additionally, machine learning has been leveraged to learn from tracked data and forecast future health responses [23, 64]. For example, Glucoracle is a mobile app for individuals with Type 2 Diabetes that used logged data about meal macronutrients and blood glucose to generate personalized blood glucose forecasts [23]. Users could enter a planned meal, and using their estimation of carbohydrates a personalized morel would predict its post-meal blood glucose levels. As a result, they could glean at a glance the predicted blood glucose impact of their meal and make in-the-moment adjustments they deemed appropriate.

While these technologies have been shown to facilitate re ection and learning from past data, from an HCAI perspective these systems do have some limitations.

From the human lens: while these systems help surface patterns users often struggle to identify how these patterns should translate into speci c actions to improve their health [23, 57]. Instead they seek technologies make more explicit suggestions for how they can modify their behavior [23, 57, 65, 64].

From the technical lens: users often nd the process of data collection to be incredibly burdensome often leading to lapses in tracking [21, 56]. As a result, data is often collected in ts and spurts, and this missingness can bias what patterns are observed and insights drawn from this data.

Finally, from the ethical lens: these technologies often require a high level of health literacy to

make sense of and act on the patterns they surface without additional guidance [66, 67]. This means that these tools are inequitable, often bene ting individuals with high levels of health literacy and access to support resources disproportionately.

2.3.2 Health Coaching Systems

Second, health coaching systems which aim to help users set and reach health goals with regular check-ins and support [68, 28, 69]. Human health coaching which often a standard part of chronic disease management, especially around nutrition [70, 69, 71]. Past systems in HCI have ranged in the degree of automation including technologies to facilitate human-led coaching, shared human-agent coaching, and fully automated agent coaching [68, 37, 55, 69, 71, 30].

Several systems have explored digital health coaching in the context of chronic disease management users can engage in goal setting, action planning, reminders to stay on track and question and answer support. For example, Platano is a mobile application that enables individuals with Type 2 Diabetes to choose health goals to work on [72]. While this application does not offer tailored nutritional guidance for meeting these goals, it helps users select a number of different health goals they can use to orient their self-management practices. Other solutions have explored coaching paradigms such as brief action planning to help individuals choose health goals to focus on and identify speci c action plans they can implement to meet these goals [30, 29].

Additionally, solutions such as Paola, a digital coach developed by Maher and colleagues are using LLMs to support adherence to physical activity goals and speci c dietary guidelines. This system provides users with nutritional education, weekly accountability check-ins and offers 24/7 Q&A support through an LLM-based chatbot [31].

These tools have shown promise in supporting self-management and accountability through checking in with users and keeping them goal oriented, however individuals still struggle to translate goals to concrete action that can be implemented in the moment [30]. As a result, from an HCAI lens coaching technologies have some limitations that can be addressed by future solutions.

From the human lens: while these technologies can support self-management by helping indi-

viduals engage in goalsetting this may not be suf cient to engender sustained change in behaviors and habits. In the moment, individuals re ect being uncertain to how apply goals and action plans in ways concordant with the constraints of their routines or speci c situational environment. There is an opportunity to make these systems even more human centered by providing support tailored for on the spot decision aid.

From the technical lens: there are data challenges that impact ef cacy of these systems. open questions around how these systems should facilitate data collection and augment coaching through personalized goals and computation. Additionally, there are questions about how to measure engagement and learning with these systems.

From an ethical lens: prior work has raised important safety concerns with using digital agents for health activities that rely on rapport as an important component in helping individuals change their behaviors. Particularly with regard to LLM-based coaches, there are questions around how to balance personalization with clear guardrails so these systems can support users' safety and appropriately.

2.3.3 Recommendation Systems

Finally, Recommendation Systems (RecSys) are designed to help individuals make decisions by offering concrete, personalized suggestions [73, 74, 75]. These systems are widely used in commercial domains, such as entertainment and e-commerce, where they generate suggestions based on similarity to user preferences, historic behaviors, or search terms. RecSys have been developed in a variety of different formats including web pages and mobile apps, with a growing interest in leveraging LLMs to elicit needs and deliver recommendations conversationally [76, 77, 78].

There are several approaches to generating recommendations. A common approach is content-based systems which recommend items that share metadata with a user's past preferences (e.g., movies of the genre a user historically likes). Another approach is collaborative ltering where recommendations are generated based on what similar users have enjoyed in the past (e.g., if user1 and user2

like similar movies, the algorithm will recommend movies user2 has seen and liked to user1) [73, 74].

RecSys have been increasingly applied to health-related domains, including chronic disease management, mental health, and nutrition [79, 80, 81, 82, 33]. Within nutrition, most existing systems focus on helping users discover new recipes aligned with their preferences or to nd ingredient substitutions based on dietary restrictions (e.g., vegan alternatives) [83, 84]. These systems have demonstrated utility in helping people nd interesting meals and ingredient alternatives [73, 52]. However, they often fall short in supporting the complex, multidimensional needs of users, such as meeting speci c health goals, aligning to cultural food practices, or navigating socioeconomic constraints [70, 30, 52, 73].

For example, many nutrition RecSys are built using large recipe datasets with the intention of recommending recipes to users [16, 33, 34, 85, 36, 51, 62]. However, individuals often need meal support outside of traditional meal preparation settings, such as when they are dining out with friends, partaking in an of ce catered lunch or grabbing a quick snack to tide them over. In these more real-world settings, recipe-based recommendations often fall short as they overlook important situational and health-related factors, including individual health goals, dietary needs, and resource limitations related to time, money, and food access [55, 78, 70, 30].

Within informatics, there is a growing interest in developing point-of-action RecSys that can provide support to users on the spot. These systems generally consider temporal, spatial, and social factors to generate more relevant suggestions. For example, such systems might recommend different activities for a "weekend in New York with kids" than for a "bachelorette party weekend in New York" [86, 74, 87]. While such systems have been explored in domains like travel, tourism, and entertainment [86, 86, 88], they are underexplored in nutrition and health [15, 89]. This is a missed opportunity, as chronic disease management and everyday eating decisions are often shaped by highly contextual factors, including routines, social settings, and availability of food [36]. A meal RecSys that operates "on-the-spot" could address this gap, especially as these spontaneous settings are often where individuals most struggle to adhere to their health goals.

From a human-centered AI perspective, effective RecSys must move beyond static user pro les and engage with users' evolving preferences, goals, and constraints.

From the human lens: literature emphasizes that preferences are dynamic and in uenced by the situational setting, mood, and current circumstances [11, 36]. In a review of human-centered RecSys, Konstan and Terven illustrate how even a highly accurate movie recommendation algorithm could fail if it does not account for context (e.g., someone who just nished a long movie series may not want to watch another lengthy lm, even if it's within their preferred genre) [51]. Nutrition decisions are similarly grounded in current needs and plans, yet few systems attempt to capture and help individuals re ne their existing plans [34, 33, 35].

From a technical lens: traditional recommendation approaches tend to be data expensive. These systems often suffer from the “cold start problem” where they struggle to make high quality recommendations until they gather suf cient user data [74, 75]. Additionally, existing systems often evaluate performance through metrics such as the novelty of recommendations [75, 90, 91]. However, in real-world decision-making, novelty is not always desirable. For a user trying to meet a nutrition goal, the most helpful meal suggestion may be one that aligns closely with an existing plan: something familiar and feasible rather than surprising or unfamiliar.

From an ethical lens: nutrition decisions are nuanced and deeply rooted in personal, social, and environmental contexts. This makes accurately inferring user preferences without misalignment dif cult. For example, someone who brought a homemade lunch to work may be constrained in what changes are feasible and can likely more easily specify the parameters for what recommendations are actionable than can be sensed by the system. Few current systems leverage data explicitly contributed by users, respecting autonomy and fostering transparency in how recommendations are generated [34].

2.4 User-Generated Nutrition Data

User generated data plays a critical role in personal health management, particularly in domains like nutrition where behaviors are frequent, contextual, and deeply individualized. Unlike clinical

data, nutrition data is generated in daily life — through meal choices, grocery lists, dining habits, and re ections on food [11, 92]. In the context of nutrition, where behavioral change often requires sustained engagement and support, leveraging patient-generated data can bridge the gap between general dietary guidelines and the complex realities of everyday eating. Prior work has explored many forms of collecting meal data from users, each with its own tradeoffs.

Several solutions have supported collection of meal data through capturing meal images [93, 94, 22, 95]. For example, Yang et al. describe YumMe which uses image analysis to recommend a meal from the crowdsourced database [94]. This method shows promise for using crowdsourced meal images to generating useful recommendations. At the same time work has shown that taking photographs of meals in social contexts can be stigmatized practice for individuals in particular communities, age, and gender groups making it not accessible to everyone [88]. Also, data collected from users "in the wild" is often noisy and includes partial ingredients, extraneous items in images (e.g., feet, pets, other people) which make generating recommendations more challenging.

Other systems have explored generating recommendations based on structured meal logs collected by users through popular commercial apps such as LoseIt! [96, 97]. These studies have shown the promise of using the similarity between meal title and ingredient vectors to identify pairs of related meals, the data used is relatively well structured with ingredients entered from an established database. While highly structured data is incredibly computationally useful, such detailed logging of data is often burdensome to users leading to lapses and inconsistencies tracking, particularly among individuals who do not identify as "self-quanti ers" [98, 25].

As an alternative, free-text meal data is a much lower burden mode of data collection and allows users maximum exibility for data entry [23]. While noisier and less standardized, these entries may better re ect how people naturally think and talk about food. Yet despite the ubiquity of free-text food data across apps and digital platforms, no prior work has systematically explored how such unstructured inputs can support meaningful meal recommendations. This gap points to an informatics opportunity to, despite its lack of structure, leverage the information within free-text meal data through computational approaches and create more inclusive, real-world nutrition

support tools.

2.5 Development of a Free-Text 3k Meal Dataset

As discussed in the preceding sections, existing nutrition RecSys generally rely on structured recipe databases or detailed ingredient logs, which, while computationally convenient, do not re- ect how individuals actually record or think about food. Designing systems that respond to users' in-the-moment decisions requires data that captures the variability, messiness, and contextual richness of real-world eating behavior. To address this need for more contextually rich, real-world data to inform human-centered meal recommendation systems, prior work in our lab has amassed a dataset that captures how people naturally describe and re ect on their meals in everyday life.

Prior work in the lab has explored a variety of solutions for how individuals use technology to support goal setting and nutrition self management [30, 72, 29]. These prior studies have focused on mobile and text-based interactions for supporting nutrition management. Though the speci c interactive system differed from study to study, as a part of all of these studies participants chose a health goals to focus on from an expert developed list, and used a progressive web application to capture meals with an image and free-text description (as shown in Figure 2.3).

Through these studies, the lab amassed a dataset of just over 3,000 meal records (henceforth referred to as the 3k Meal Dataset) captured in english by participants “in the wild.” The dataset consists of meals recorded by 84 unique individuals living in the New York City metropolitan area. These participants represented a diverse set of ages, genders, and health goals, and were primarily recruited through community-based studies focused on diabetes and general nutrition self-management. Participant contributed between 1 and 286 meals (mean=37.8, SD=55.62) over a period of several months, resulting in longitudinal data that captures changes in eating habits, routines, and goal adherence over time. Additional details about the full dataset and goal evaluation criteria are included in supplementary materials A.

Each record in the dataset includes:

- Meal title: a short free-text phrase written by participants (e.g., “Egg sandwich and coffee,”

Figure 2.3: Mobile interface through which participants could capture meal information. This enabled individuals to capture a meal title, descriptions, and a meal image.

"Salmon rice bowl," "Late-night snack – chips and hummus").

- Meal description: a longer free-text eld where participants elaborated on preparation details, portion sizes, or contextual factors (e.g., "Fried egg on a roll with avocado, and latte" or "salmon, soy sauce, brown rice, carrots, edamame").

- Health goal: one of eight expert-developed nutrition goals selected by participants at study enrollment (e.g., "reduce added sugar," "balance carbs," "increase vegetables").

- Goal alignment label: a binary expert assessment ("aligned" or "not aligned") determined by registered dietitians using a uni ed clinical protocol developed in collaboration with nutrition scientists. Dietitians underwent calibration sessions with an expert RD to ensure consistent interpretation of goals and to minimize inter-rater variability. Additional details about the speci c criteria used to determine if meals met each goal is available in Supplementary Materials A.

In addition to the primary elds, the dataset includes metadata such as timestamps, study source identi ers, and anonymized participant IDs, which enable temporal and within-subject analyses. Collectively, these data elements provide a unique opportunity to study real-world meal representations at the intersection of free-text natural language, goal-oriented health behaviors, and expert nutritional interpretation.

Unlike traditional recipe datasets, which focus on ingredients and nutrient composition, the 3k Meal Dataset captures the language of everyday eating around ow people articulate their meals, contextualize choices, and frame success or deviation relative to their own goals. This makes it particularly valuable for developing and evaluating algorithms that must interpret free-text inputs, detect implicit nutritional signals, and generate context-sensitive recommendations.

Additional details about dataset structure and labeling criteria are provided in Supplementary Materials A.

2.6 Takeaways: Towards More Human-Centered Nutrition Technologies

Together this body of prior work highlights both the promise and the persistent limitations of technologies designed to support nutrition management. While advances in recommendation algorithms and data collection have enabled increasingly personalized nutrition support, most existing systems remain disconnected from the lived realities of everyday eating.

Viewed through the lens of HCAI, three interrelated challenges emerge. First, from a human needs perspective, current systems provide limited assistance at the point of decision when individuals are choosing what to eat in dynamic, real-world contexts. Second, from a technological perspective, many existing approaches are data-hungry, requiring extensive input or historical data before becoming useful. This reliance on large, structured datasets makes them poorly suited for spontaneous, real-world use. Third, from an ethical and experiential perspective, most systems give users little agency over how recommendations are generated or adapted, leaving limited room for personal goals, situational constraints, and evolving priorities.

Addressing these challenges requires rethinking how nutrition data are represented and how

recommendations are produced. This dissertation responds by exploring how user-generated free-text meal can serve as a foundation for more human-centered, real-time meal recommendations. Speci cally, it uses the 3k meal dataset to investigate how such data can be leveraged to generate in-the-moment, contextually appropriate meal suggestions that both respect existing plans and empower users to maintain control over their decisions.

Chapter 3: Aim 1 — Investigate the technical feasibility of human-reasoning aligned computational approaches for assessing meal similarity

Nutrition management is central to human health and well-being, yet individuals often struggle to follow nutrition guidelines in everyday meal decisions. Informatics approaches can support meal choices by identifying healthier, similar alternatives to meals individuals already eat. However, little is known about how people assess meal similarity and substitutability, or how well computational approaches align with these human judgments. This understanding is an important rst step towards developing similarity-based recommendations that can help individuals nd a similar alternative to their meal plan that is more aligned with their health goals.

To address this gap, Aim 1 of the thesis explores human and computational approaches for assessing meal similarity and focuses on the following research questions:

- '& $_{1 \bullet 1}$: What attributes of meals inform human perceptions of meal similarity and substitutability?
- '& $_{1 \bullet 2}$: Which computational similarity approaches align with human perceptions of meal similarity?

RQ 1.1 focused on characterizing what attributes people consider when assessing how similar two meals are (what makes two meals alike), and how substitutable they are (what makes two meals exchangeable for each other). RQ 1.2 explored three computational approaches for assessing meal similarity and evaluated how they aligned with human perceptions of meal similarity. Together, these research questions provide a rst step to a scalable, human-aligned approach for identifying meals similar to one another that can inform the recommendation of similar meals.

3.1 Related Work

The idea of similarity and substitution-based recommendations has been explored in commercial recommender systems. For example, free-text user reviews have been used to learn substitutable relationships between products [99, 100, 101]. While this work highlights the potential of substitution as a recommendation mechanism, it has rarely been examined in nutrition contexts. Moreover, prior work has not addressed which attributes inform substitutability judgments or how these may differ from general similarity. This distinction is particularly important in meal planning, where a meal might be considered similar to another but not necessarily exchangeable in practice.

Within nutrition research, similarity has been used as the basis for recommending recipes and images. Each of these studies, however, made assumptions about a single attribute driving similarity. For instance, some focused on macronutrients [97], others on visual similarity [94], and others on shared ingredients [83, 96]. While these approaches demonstrate ways to operationalize similarity, they leave open important questions about how to compute similarity for free-form meal descriptions (as opposed to curated recipes) and, critically, whether these computed measures align with how humans perceive similarity between meals.

Food recommendation research has also explored how individuals themselves conceptualize similarity. For example, van Pinxteren et al. asked participants to group curated meal recipes in a card sort exercise, revealing dimensions such as meal form, meal heaviness, cuisine, and main carbohydrates [102]. These ndings illustrate that people rely on multiple, sometimes overlapping attributes when judging similarity. However, it remains unclear how these aspects of similarity translate to meal recommendations when individuals already have a meal plan in mind. Moreover, little is known about which dimensions, if any, predict a person's willingness to substitute a similar meal for their original plan.

A range of computational approaches have been applied to the problem of identifying similar meals, including cosine similarity on embeddings, clustering-based methods, and more recent

applications of large language models (LLMs) [97, 102, 89]. These approaches provide scalable ways to compare meals, but they have not been systematically evaluated against human judgments of meal similarity or substitutability. As such, it is unknown whether they capture the same attributes people actually use in decision-making.

3.2 Methods

A mixed-methods study was used to address the research questions in turn. The speci c methods are detailed below.

3.2.1 RQ 1.1: Understanding Human Perceptions of Meal Similarity and Substitutability

The goal of RQ 1.1 was to investigate how people perceived the similarity between meals and which meal features shaped these perceptions. The outcome was a gold-standard dataset containing pairs of meals, each accompanied by a human similarity rating and an explanation of the assessment.

Data Sources: As this aim focused on how users perceived the similarity between different meals and the alignment of computational approaches with user perceptions, a subset of the previously described 3k Meal Dataset was used for analysis. Details about the larger dataset and its attributes are available in Section 2.5 and in the Supplementary Materials A.

To generate the subset of data used in this study, 50 pairs of meals were selected from the 3k Meal Dataset. Meals were chosen somewhat at random: the rst two meals were selected at random, after which one of those meals was carried into the next set and its pair selected at random, and so forth. Having at least two comparisons involving each unique meal allowed each meal to serve as an anchor for exploring differences relative to other meals.

In total, this new dataset (termed the Meal Pairs Dataset) contained 50 pairs of meal descriptions comprising 59 unique meals.

Participant Recruitment: Crowdworkers from Amazon Mechanical Turk evaluated the meal pairs dataset, with three different crowdworkers assessing each pair. All crowdworkers were uent

Figure 3.1: Example of a pair of meal descriptions from the meal-pair dataset

in English and help an “AMT master” status, indicating high performance on previous tasks. A total of 14 unique crowdworkers participated in the evaluation; each crowdworker evaluated an average of 10.9 meal pairs (standard deviation 14.9).

Evaluation Task: In the evaluation task, crowdworkers were shown a pair of meal descriptions and asked to assess the meal pair on the four criteria detailed below. An example of the evaluation completed by crowdworkers for each meal is available in Supplementary Materials, Appendix B.

1. Overall Meal Similarity (Likert & free text): Raters were shown a pair of meal descriptions and asked: "Rate how similar the meals are. Explain your reasoning for the choice." Similarity was rated on a 5-point Likert scale ranging from "extremely different" to "extremely similar". Raters provided an explanation for their rating.

2. Dimensions of Similarity (choose multiple): Raters were shown a list of eight meal attributes identi ed from prior work by Pinxtreen et al. and asked to select all that in uenced their assessment of meal similarity [102]. These attributes included: meal form (e.g., soup), cuisine, carbohydrate or starch element (e.g., rice), meat or protein element (e.g., sh), preparation technique (e.g., fried), preparation time, heaviness (e.g., light meal), and healthfulness. Raters were able to revise their free-text reasoning about the similarity rating they assigned to incorporate explanations related to these attributes.

3. Meal Substitutability (Likert scale & free text): To understand willingness to substitute one meal for another, raters were asked: "If meal 1 is not available, how likely are you to substi-

tute meal 2 in its place?". Substitutability was rated on a 5-point Likert scale from extremely unlikely to extremely likely. Raters explained their selection in a free-text eld.

4. Dimensions of Meal Substitutability (choose multiple & free text): Raters were shown the same eight meal attributes as described above for similarity, and asked to select all that in uenced their assessment of meal substitutability. They could select multiple attributes, and were instructed to expand on the reasons for their selections in a free-text eld.

Data Analysis: As the goal of RQ 1.1 was to understand what dimensions of meals inform human perceptions of similarity, the approach to analysis consisted of the following steps.

1. Meal Similarity Assessment: For each meal pair, percent agreement was used to assess reliability in the meal similarity rating of each meal pair. Each meal pair received a similarity rating from three independent raters, and needed an agreement from 2 of 3 raters to be considered a reliable assessment. For meal pairs that passed this inter-rater agreement threshold, the average similarity score was computed as the ground truth. Meals not passing this threshold were agged for exploratory qualitative analysis to explore discrepancies in raters assessments.

   The human ratings of similarity generated through the process outlined above served as a gold standard dataset of meal similarity ratings that was used in RQ 1.2 to evaluate how well different computational approaches for assessing meal similarity aligned with human perceptions.

2. Dimensions of Meal Similarity: To identify which meal features most commonly informed similarity judgments, a frequency analysis of structured meal attributes was conducted. Qualitative open coding of free-text explanations was also conducted to understand how participants described similarity and surface any additional dimensions raters considered.

3. Meal Substitutability Assessment: As with the similarity assessment, inter-rater agreement was used to determine reliability, with a threshold requiring at least 2 of 3 raters to agree

on substitutability. For meal pairs meeting this threshold, the average substitutability score across raters was computed as the ground truth.

To examine the relationship between perceived similarity and substitutability, the correlation between these two variables was calculated.

4. Dimensions of Meal Substitutability: Finally, to understand what dimensions of the meals most commonly informed substitutability assessments frequency analysis of structured meal attributes was conducted.Qualitative open coding of free-text explanations was also conducted to understand how participants described substitutability and surface any additional dimensions raters considered.

Output – A Gold-Standard Meal Similarity Dataset: With the human assessments of meal similarity serving as a gold-standard evaluation, the meal-pairs dataset was expanded to include the human similarity and substitutability ratings for each pair. This expanded dataset serves as a gold-standard dataset that can be used for comparing other approaches for assessing similarity of meal pairs to human assessments, as is explored in RQ 1.2.

3.2.2 RQ 1.2: Computational Approaches for Assessing Meal Similarity

This RQ assessed three computational approaches (cosine similarity, k-means clustering, and large language models) to assess the similarity between free-text meal descriptions. To evaluate these approaches, the Gold-Standard Meal Similarity Dataset generated in RQ 1.1 to measure alignment with human perceptions of meal similarity was used as a benchmark.

Data Sources

This analysis used the Gold-Standard Meal Similarity Dataset generated as an outcome of RQ 1.1. This dataset adds human similarity ratings that serve as a gold-standard similarity assessment to evaluate how well each of the computational approaches align with human perceptions.

Data Pre-Processing:

To computationally represent each meal, the meal title and description elds were concatenated and encoded. The raw free-text data underwent several pre-processing steps to standardize it for use. First, spelling correction was applied using the National Cancer Institute's Diet History Questionnaire as a vocabulary of standard food-related terms and spellings [103]. Stop words (e.g., “a,” “the”), digits, special characters, measures (e.g., “tbsp”), and volume units (e.g., “oz,”) were removed. The resulting tokens, representing the meal title and ingredients, were concatenated and encoded using TF-IDF (term-frequency-inverse-document-frequency) to represent each meal as a single vector. By down-weighting tokens frequently used across the corpus, TF-IDF captured the rarity of a token describing one meal relative to others [104].

Computational Approaches:

To assess similarity three main approaches were explored: (1) cosine similarity, (2) k-means clustering, and (3) LLM-based assessment. Each of these approaches have been used in prior work to identify closely related items for recommendation. The operationalization for each computational approach is detailed in the section below.

1. Cosine Similarity: One common approach for assessing the similarity between two pieces of text is to use the dot product or cosine similarity. This method calculates the angle between the vectors representing two meals to assess how similar the meals are [105]. Output values rage from 0 (no similarity) to 1 (identical).

2. K-Means Clustering: Another common approach for similarity assessment is to use k-means clustering to group the dataset into k clusters with each cluster consisting of similar meals. After using the steps outlined above to pre-process the free-text meal data, I clustered meals. To determine the optimal number of clusters, a combination of the elbow method and silhouette coef cient, arriving at 28 clusters. Despite relatively low silhouette scores for individual clusters (ranging from 0 to 0.3) was used, which indicate potential issues with cluster separa-

bility, qualitative inspection showed strong thematic coherence within clusters. For instance, clusters often grouped meals that shared prominent ingredients or common cuisine types. Additional details on cluster construction is provided in the Supplementary Materials B.1.

To quantify the similarity between two meals, the normalized distance between the centroids of their assigned clusters was calculated (this meant that smaller value indicates greater similarity). To align this metric with those used in other approaches, it was transformed by computing 1 – normalized_distance, so that values closer to 1 re ect higher cluster-based similarity and would align on the same scale as similarity ratings from other methods.

3. Zero-Shot Large Language Model: I used Columbia's HIPAA-compliant OpenAI API to query GPT-4o-mini in a zero-shot setting [106]. For each meal pair, raw meal descriptions were passed directly to the model along with the same prompt used in the human rating task (noted in the box below). The model returns a Likert similarity rating and a text explanation.

LLM Meal Similarity Prompt:
(Meal 1 Description), (Meal 2 Description). Rate how similar the meals are. Options: 0 (extremely different), 1 (somewhat different), 2 (neutral), 3 (somewhat similar), 4 (extremely similar). Explain your reasoning for the choice.

Data Analysis:

To assess the performance of each computational approach, computational similarity assessments were compared to the human gold standard assessment generated during RQ 1.1.

1. Correlation with Gold Standard: Each of the computational approaches outlined was used to calculate a similarity rating for the 50 meal pairs in the meal pair dataset. These ratings were then compared with the human gold-standard ratings (generated in RQ 1.1) to evaluate how closely they aligned with human judgment. A spearman correlation was used to compare the similarity ratings from each computational approach to the human gold standard (this analysis was conducted using R).

2. Qualitative Exploration of Discrepant Cases: To better contextualize these outcomes I explored discrepancies between human and computation assessments for each of these approaches. As LLMs are able to output a natural language reasoning for the assessment, cross-case analysis was used to qualitatively compare the LLM reasoning to human reasoning for assessing the similarity of meal pairs. Qualitative analysis was conducted using nVivo coding software.

Figure 3.2: A graphical description of RQ 1.2 which evaluates candidate computational approaches for assessing meal similarity to determine which best align with human gold-standard assessments.

3.3 Results

The following section details the results for Aim 1. RQ 1.1 investigated the attributes of meals that shape human reasoning about similarity and substitutability, while RQ 1.2 evaluated the extent to which cosine similarity, clustering, and LLM-based approaches align with these perceptions.

3.3.1 RQ 1.1: Understanding Human Perceptions of Meal Similarity and Substitutability

The study rst explored how humans assess the similarity and substitutability of randomly selected meal pairs. Building on prior work [102], 14 crowdworkers completed a survey with structured and free-text elds to evaluate 50 pairs of meals on their similarity and substitutability. A mixed-methods evaluation was used to identify the key dimensions people consider when making

similarity judgments. Finally, the correlation between perceived similarity and substitutability was calculated.

Human Perceptions of Meal Similarity

When considering meal similarity, rater agreement was high, with 88% (44 out of 50) of meals showing consensus between at least two of the three raters. The mixed-methods ndings re ected four dimensions participants commonly considered when comparing meal pairs: (1) meal form; (2) main macronutrient; (3) meal cuisine; and (4) core ingredients. The rst three dimensions were consistent with prior ndings by van Pinxtreen et al. and were re ected in both participants' quantitative survey responses and their qualitative re ections [102]. Participants emphasized core ingredients as a distinct fourth dimension, which emerged primarily through their free-text responses.

While these were not all the dimensions identi ed, these four dimensions appeared most frequently and were consistently referenced across a variety of meal pairs and by multiple raters across in both qualitative and quantitative re ections. A deeper description of these four dimensions is below and an expanded table with descriptions and examples of all the dimensions identi- ed though analysis is available in the supplementary materials.

Meal Form: Participants noted paying attention to the meal form for the majority of meals, (69%). Qualitatively, this meant comparing if meals were soups, sandwiches, wraps, omelets, cereals or another meal form as a key point of similarity. Participants also noted that meal form often had implications of when that type of meal should be consumed and could inform which meal was substitutable for one another. For example, although the composition of cereal is liquid like a soup forms like cereal and omelets are often associated with breakfast time whereas soups are less commonly considered as breakfast foods. Finally, meals of shared form also often shared other characteristics such as how they were prepared and how they were typically eaten (e.g. warmed up, handheld).

"Both meals are omelets, Meal 1 is a simple cheese omelet, whereas Meal 2 includes

> additional ingredients like onions, pepper, and chorizo, making it somewhat different in composition and likely in avor pro le compared to Meal 1." - Pair 35 "Both meals are curries" - Pair 38

Main Macronutrient: Participants also described paying particular attention to the macronutrient composition of meals (61%). In particular they remarked on the protein and carbohydrate content, often discussing these components together although they were separated in the van Pinxtreen survey.

> "Both meals include a protein component (meat in Meal 1 and chicken nuggets in Meal 2) and a carbohydrate (noodles in Meal 1 and toast in Meal 2), but they differ in their speci c ingredients and overall composition." – Pair 9

Meal Cuisine: In many cases participants also indicated similarity by cuisine. Qualitatively, this meant considering similarities in meals "culinary style" and the corresponding avor pro le, even when the speci c ingredient components (such as the vegetables or proteins choices) differed. Certain cuisines were also associated with certain meal forms and meal preparation styles (e.g. a stir-fry being associated with Asian cuisine).

> "Both meals include vegetables and share an Asian culinary style, but Meal 2 includes a wider variety of components such as dumplings, broccoli, tofu, and white rice, making it more diverse and complex compared to the simpler stir-fry vegetables in Meal 1" – Pair 16

Core Ingredients: When thinking about key ingredients, participants differentiated between main components and accessory or side components of the meal. They perceived meals with similar based components to be alike, even if speci c sides or condiments change the avor pro le. As part of this consideration of key ingredients they also discuss different dietary pro les (e.g., vegetarian, vegan, plant-based vs. meat-based) as the inclusion or omission of key ingredients (e.g., butter, wheat) informed how similar the meals were.

"Both meals are omelets and include cheese. The main difference is that Meal 2 includes smashed plantains and butter, which adds some variation to the basic cheese omelet but keeps the core form and primary ingredient (eggs and cheese) quite similar" – Pair 35

Figure 3.3: The four main dimensions of meal similarity identi ed through the mixed-methods ndings are meal form, meal macronutrients, meal cuisine, and key ingredients.

Human Perceptions of Meal Substitutability

When considering meal substitutability, rater agreement was also high with 84% (42 out of 50) of meals showing consensus between at least two of the three raters. There was a signi cant positive correlation between users ratings of the similarity between meal pairs and their substitutability ($r = 0.63$, $p<0.001$), indicating that meal pairs rated as more similar were also considered more exchangeable for one another (Figure 1).

Participants highlighted many of the same dimensions when considering substitutability as they did when assessing meal similarity. For example, meals sharing the same forms, cuisines, and main macronutrients were perceived as likely to be satisfying substitutions for one another. However, when these elements were misaligned, the two meals were perceived as less exchangeable.

For example, one participant re ected on an “Extremely Likely” rating for the substitutability between two meals: “The meal form and cuisine are the same, both are stir-fries. The main difference is the added chicken in Meal 2, which changes protein content, heaviness, and slightly affects healthfulness. Overall, the similarity in ingredients and prep method makes it an easy substitute.” – Pair 21

Participants highlighted the importance of shared or similar core ingredients as particularly important when evaluating one meal as a suitable substitute for another. Across their qualitative re ections they often distinguished between components of the meal they considered central to the “meal experience”, as opposed to accessory components of the meal that held less importance. For example, one participant re ected:

> “I chose "Extremely Likely" because the main components of both meals: a roti, curry shrimp, and curry chickpeas, are exactly the same. The only difference is the drink, which doesn't signi cantly impact the overall meal experience. Unless there's a spe-ci c reason to avoid pomegranate juice, this substitution is a near-perfect match.” – Pair 43

This sentiment was echoed in the qualitative re ections on substitutability. At the same time, participants acknowledged that substitutability can be a more nuanced concept than similarity and impacted by momentary and personal factors including mood, contextual needs/preferences, and dietary restrictions.

One rater re ected on substituting a "vegetable soup" (meal 1) with "dumplings, broccoli & beef stir fry and rice" (meal 2):

> “If I were looking for a warm, liquid-based meal, I might not want to substitute [meal 2 for meal 1], but if I were open to a more lling, diverse meal with different textures I would” – Pair 30

> Another re ected on substituting a "ham and cheese sandwich" (meal 1) with an "Italian sandwich" (meal 2): “I am somewhat likely [to substitute] because both meals are sandwiches featuring ham, so they are similar in form and avor. . . if I were craving a sandwich or needed a quick substitute, it would work but it's not a perfect match nutritionally” – Pair 32

3.3.2 RQ 1.2: Computational Approaches for Meal Similarity

To evaluate how the three computational approaches aligned with human perceptions of similarity, correlation coef cients and discrepancies (de ned as greater than 1 standard deviation of difference between the human and computational ratings) between human and computational assessments were reviewed. These ndings are summarized in Figure 3.5. For each method, the gold standard meal similarity rating (Y axis) was plotted against the ratings generated for each computational approach (X-axis). The gray line indicates where x=y (i.e., where human and computational ratings would be the same). Points above the gray line are meal pairs humans rated more similar than the computational approach rated them. To the right of each plot the correlation coef cient and summary of discrepant cases.

Similarity Approach 1 – Cosine Similarity:

First, I examined how well cosine similarity aligned with human judgments of meal similarity. I found a signi cant positive correlation ($r=0.68$, $p<0.001$) between human and computer similarity ratings, suggesting strong agreement. However, there were discrepancies of greater than one standard deviation between human and computer similarity ratings for 30% of the meal pairs. Notably, in 85% of these discrepant cases (11 out of 13), the human gold-standard rating was higher than the cosine similarity score for the same pair indicating cosine similarity may underestimate similarity compared to human raters.

These discrepancies suggest that human raters may consider attributes not fully explicitly captured by the text descriptions of meals. For example, human raters often account for attributes such as cuisine and macronutrients when considering what makes two meals similar. However, given that cosine similarity is based purely on textual representation of the meal description, such elements may be absent in the vector space model and inferred by human raters.

Similarity Approach 2 – K-Means Clustering:

Next, alignment between human similarity judgments and similarity scores derived from k-means clustering was explored. Here, the correlation was weaker and not statistically signi cant ($r = 0.26$, $p = 0.09$), suggesting limited agreement. Discrepancies were more common, with 55% of meal pair ratings differing by at least one standard deviation between the human and clustering assessments. In contrast to cosine similarity, most of these discrepancies (87%, or 21 of 24 cases) re ected instances where the clustering method overestimated similarity relative to human ratings.

These ndings suggest that while clustering captured broad groupings of similar meals, it lacked the granularity and nuance evident in human assessments. The low silhouette scores further highlight the challenge of forcing high-dimensional, overlapping data into discrete, spherical clusters. As a result, k-means clustering may serve as a blunt estimator of similarity that overlooks ner distinctions perceived by human raters.

Similarity Approach 3 – Zero-Shot LLM:

When comparing human similarity ratings with those generated by a zero-shot LLM, a significant positive correlation ($r = 0.66$, $p < 0.001$) was observed, indicating strong alignment between the two. Discrepancies were relatively rare, occurring in only 11% of rated meal pairs. In most of these cases (80%, 4 out of 5 cases), the human gold-standard rating was higher than the LLM's rating of the same meal pair. One advantage of an LLM approach for similarity assessment, is the model's ability to generate an explanation of reasoning in natural language. This enabled us to directly compare how humans and LLMs “articulate” their reasoning for meal similarity assessments, particularly for discrepant cases.

Overall, both human and the LLM explanations referenced similar meal dimensions when evaluating meal similarity. Frequently discussed dimensions across both groups included included core ingredients, meal cuisine, and macronutrient composition.

However, in further examination of the 5 discrepant cases, I observed that human raters tended to place greater emphasis on dimensions such as meal form and preparation technique. Partici-

pants highlighted that preparation was often tied to dimensions like the meals form, with meals of the same form being prepared in the same way even though ingredients and avors might differ. One rater explained: “Both meals are soups, which makes them similar in form and potentially preparation technique. However, they differ in their main ingredients. Meal 1 focuses on vegetables, while Meal 2 includes shrimp and noodles, adding a protein and starch component that Meal 1 lacks.” – Pair 42

By contrast these dimensions were largely absent from the LLM's explanations. Instead, LLM explanations appeared to focus more on other aspects of the meal, such as shared avors, when assessing the similarity between meals.

Exploring Differences between Human and LLM Similarity Reasoning

Beyond examining individual discrepancies between human and LLM similarity judgments, I also took a broader view to compare the attributes each emphasized when reasoning about meal similarity. To do this, directed thematic coding using the known meal attributes from the human assessments was conducted for human and LLM similarity assessments for all meal pairs, with open coding to capture any additional dimensions identi ed.

Table 3.1 summarizes the attributes identi ed in human re ections of meal similarity and in LLM-generated explanations. Several attributes appeared with similar frequency across both sources, including ingredients, cuisine, food type, preparation technique, and preparation time.

However, notable differences also emerged. Human raters placed greater emphasis on meal form (e.g., salads, sandwiches, soups) and meal composition (the balance between "main" and "side components"). In fact, as discussed in RQ 1.1, meal form was among the most frequently cited and most important attributes in human assessments of both similarity and substitutability.

By contrast, LLM explanations more often referenced avor pro le (e.g., “umami” or “sweet”) and texture (e.g., “crunchy” or “soft”), attributes that participants mentioned far less often. LLMs also occasionally drew on broader cultural or contextual concepts, such as describing items as “comfort foods” or as “meals found in a deli” which though directly tied to human life were not commonly noted in human re ections of meal similarity.

Meal Attribute Code	A: GPT	B: Human
1: Cuisine	9	12
2: Dietary Plan	2	7
3: Flavor	24	12
4: Food Type	9	12
5: Healthfulness	3	2
6: Heaviness	5	2
7: Ingredients	33	42
8: Macronutrients — Carbs	6	5
9: Macronutrients — Protein	10	15
10: Meal Composition	13	20
11: Meal Form	19	28
12: Meal Type	8	9
13: Portion Size	0	1
14: Prep Technique	16	18
15: Prep Time	6	9
16: Texture	9	3

Table 3.1: Comparison of attributes referenced in meal similarity reasoning by humans and the LLM. Values indicate the number of meal pairs (N = 50) in which each attribute was mentioned during similarity assessments.

3.4 Discussion

This mixed-methods study examined (1) how people reason about meal similarity and substitutability, and (2) how well different computational methods capture those judgments. Findings offer guidance for designing more human-aligned tools for nutrition support.

First, the study explored how humans think about what attributes make meals similar and good substitutes for one another. Participants emphasized several dimensions that informed their perceptions of meal similarity, the most common being meal form, cuisine, macronutrient composition, and core ingredients. These dimensions appeared repeatedly across meal types and among diverse raters, underscoring their salience in shaping perceived similarity. While the rst three dimensions align with prior work by Pixtreen et al., core ingredients emerged in our study as a distinct consideration. Participants viewed certain ingredients as central to a meal's identity — meals sharing these core components were often seen as similar, even when they differed in form or preparation (e.g., an omelet and an egg salad sandwich both center around eggs). Additionally, meals

considered similar were also considered more substitutable with each other ($r = 0.63$, $p < 0.001$). However, participants also noted that contextual nuance beyond dimensions of similarity, including mood, availability, and situational preferences in uenced whether a similar meal would be seen as a viable substitute in the moment (e.g., craving a warm soup vs. a more lling meal).

Second, the study sought to explore how different computational approaches for assessing meal similarity from free-text meal data aligned with human perceptions. Cosine similarity ($r=0.68$, $p<0.001$) and a zero-shot LLM approaches ($r=0.66$, $p<0.001$) had signi cant positive correlations with human similarity ratings, indicating strong alignment. However, cosine similarity had a higher rate of discrepancies, likely because of semantic features inferred by human raters like cuisine or meal form not being directly represented in the meal text. Conversely, the zero-shot LLM approach had fewer discrepancies and appeared to consider many of the same dimensions that humans emphasize such as cuisine, macronutrient composition, and core ingredients even when they were not directly represented in the text. Cross-case comparisons between the LLM and human reasoning for similarity assessments also showed humans also drew on cues such as meal form and preparation style, which LLMs sometimes overlooked, suggesting an opportunity for improvement through re ned or few-shot prompting.

In contrast to these approaches, similarity estimates derived from k-means clustering were weakly correlated with human judgments ($r=0.26$, $p=0.09$) with signi cant discrepancies in over half of the half of the meal pairs. These ndings suggest that while clustering captured broad groupings of similar meals, it lacked the granularity and nuance evident in human assessments. The low silhouette scores further highlight the challenge of forcing high-dimensional, overlapping data into discrete, spherical clusters. As a result, k-means clustering in this case may serve as a blunt estimator of similarity that overlooks ner distinctions perceived by human raters.

Our ndings suggest that computational methods, particularly LLMs, can approximate human judgments of meal similarity and, in some cases, align with the reasoning behind them. This opens promising opportunities for nutrition self-management. First, similarity-based tools could help individuals re ect on their existing dietary data to uncover patterns and inform healthier changes.

Second, they can support real-time decision-making by suggesting alternatives that align with a person's current plans and goals. Unlike systems that focus on novelty or historic trends, similarity-based methods offer moment-to-moment relevance. Future research is needed to assess whether such similarity-based recommendations are perceived as useful, particularly in health-constrained or time-sensitive contexts.

Opportunities for Re nement and Limitations

Improving computational similarity models will require addressing gaps between human perception and what's captured in free-text meal data. Future work could incorporate additional meta-data, such as food ontologies or LLM-augmented descriptors, to surface missing attributes like form or cooking method. Multimodal approaches such as combining text with images may further improve alignment with human judgments, enabling more holistic representations of meals. For example, Yang et al.'s YumMe system used meal images to compute similarity, suggesting value in combining visual and textual cues [94].

This study also revealed challenges posed by the variability of free-text meal descriptions. Meals were logged at different levels of granularity (e.g., “green salad” vs.”),“and similar meals were described using diverse terms (e.g., “porridge,” “hot cereal,” “oatmeal”). These discrepancies limit computational precision. Future work could explore lexical normalization and structured knowledge (e.g., food ontologies) to standardize descriptions and improve consistency across entries [107, 108].

Meal Similarity for in-the-moment Meal Recommendations

Generating meal recommendations using meal similarity has several advantages over existing content-based, and collaborative ltering techniques.

First, the majority of contemporary recommendation systems approaches are constrained by a high data demand. Standard recommends approaches require extensive meal logs accompanied by meta-data (content based systems) or by user preference ratings (collaborative ltering sys-

tems) to model preferences and generate quality recommendations [109, 74]. Furthermore, these systems suffer from the “cold start” problem and are unable to generate useful recommendations at the beginning of user engagement when little personal data is collected. When deployed at scale, these recommendation systems (particularly non-commercial systems) are often limited by sporadic tracking and rating behaviors and fall short of the data needed to generate quality recommendations [16, 74]. By contrast, recommendations generated based on vector similarity are well suited to limited data availability as they require no additional metadata labels nor have a minimum threshold of data required for the recommendation system to perform. The lack of this “cold start problem” may make this approach to generating recommendations may be more scalable “in-the-wild”, where there are often sporadic usage patterns and gaps of lapsed tracking, particularly among economically disadvantaged communities [56].

Second, recommendations based on meal similarity to existing plans may be more aligned with practical ways people make sustainable health decisions. Traditional nutrition recommendation systems prioritize novelty and diversity to “break up the funk of weeknight dinners” and offer recipe suggestions that align with preferences [33, 110]. However, contextual constraints, for example factors like time, money, and ingredient availability, often limit the usefulness of recommendations generated through these approaches [87, 111, 65, 88]. Instead, literature on successful chronic disease management emphasizes small, sustained adjustments over time [112]. Recommendations generated through meal similarity are uniquely suited to this task: if combined with computational assessment of meal-goal alignment, this approach could help to arrive at recommendations that are similar to an individuals' past meals, but are more aligned with their nutritional goals. However, further studies are needed to understand whether recommendations generated through this approach are in fact easier to follow and more accessible to a broader audience.

3.5 Key Takeaways

Aim 1 investigated what attributes humans consider when determining the level of similarity between two different meals, and how this relates to their perceptions of how interchangeable two

meals are for each other. Then, it explored the alignment of different computational approaches (cosine similarity, k-means clustering, and LLMs) for assessing meal similarity aligned with human reasoning.

Findings for RQ 1.1 demonstrated that human perceptions of meal similarity are shaped by multiple attributes, including meal form, cuisine, macronutrient composition, and core ingredients. Meals judged as more similar were also considered more substitutable, though situational factors also in uenced this relationship.

For RQ 1.2, three computational approaches were evaluated. Among the computational approaches tested, cosine similarity and LLM-based methods showed the strongest alignment with human judgments, with LLMs producing the fewest discrepancies. These ndings suggest that computational approaches such as LLMs can reliably generate human-aligned similarity assessments of meal pairs. This provides a foundation for approaches that leverage similarity assessments to identify actionable meal recommendations that respect both similarity and substitutability.

Implications for Human-Centered Nutrition Technologies: Aim 1 underscores that human-centered AI for nutrition must re ect how people actually reason about similar and substitutable meals, drawing on attributes like form, cuisine, macronutrient composition, and core ingredients. Meals perceived as more similar were also seen as more substitutable, though factors such as mood and availability shaped whether a substitute felt viable. From a technical perspective, LLM-based methods most closely aligned with this human reasoning, outperforming cosine similarity and clustering.

Together, these ndings point to approaches for building nutrition technologies that are interpretable and aligned with human-reasoning.

(a)

(b)

Figure 3.4: (a) Summary of the top dimensions considered by users when assessing similarity from the taxonomy by van Pinxtreen et al. (b) Correlation between ratings of the meal pair's similarity (x-axis) and how substitutable (y-axis) the meals in the pair are for each other. Meals considered more similar are also considered more substitutable

Figure 3.5: Comparison of the three computational similarity approaches (x-axes) and gold-standard human ratings (y-axes). Points above the gray line are meal pairs humans rated more similar than the computational approach rated them.

Chapter 4: Aim 2 — Characterize user needs and assess the feasibility of similarity-based meal recommendations in everyday decision settings

Aim 1 explored human perceptions of meal similarity and identi ed four key dimensions: form, cuisine, macronutrients, and key ingredients. I also explore the technical feasibility of different computational approaches to identify similar meals, aligned with human judgment. Next, Aim 2 investigates how this computational assessment of similarity can be used to generate similarity-based recommendations. These recommendations can help people navigate making decisions around meals in their daily life, with particular attention to the settings in which they do so: from open-ended situations (e.g., cooking at home) to constrained ones (e.g., ordering from a xed menu).

This study addresses Aim 2 of the thesis. Here, I ask the following research questions:

'& $_{2.1}$: What are individuals' existing practices for meal selection, and how do these differ between open and constrained decision-making settings?

'& $_{2.2}$: How do users perceive the usefulness of similarity-based recommendations for supporting meal decision-making in everyday decision settings?

- Hypothesis: In constrained settings, users choose meals similar to their original plan.

'& $_{2.3}$: What features do users value in nutrition-management technologies and how do these preferences vary across decision settings?

I conducted a lab-based paraticipatory design study to invesigate these questions (N=15). RQ 2.1 sought to qualitatively understand how users priorities for meal selection in differed across situational settings and their existing practices for using technology to support making meal choices

in these different settings. Then, RQ 2.2 explored user attitudes toward similarity-based meal recommendations in these different settings. Finally, RQ 2.3 gathered feedback about features users nd valuable in nutrition-management tools to re ne the design of such a recommendation system.

This aim establishes a foundation for designing systems that are aligned with how humans make meal decisions and their values, to align with human work ows to augment their decision-making and informs the design of a system to enable human-driven decision-making.

4.1 Methods

Informed by prior work in personal informatics and RecSys for health and nutrition, this study focused on understanding what factors impacted individuals' nutritional decisions and how situational constraints in uenced meal choices. Table 1 summarized the central research questions that guided our study and analysis. This study used a mixed-methods approach and relied on a combination of qualitative and quantitative methods to answer each of these questions [113]. For clarity, I organized the following methods section by the research question; within each sub-section I detailed the study procedures, data elements, and analytical approaches used to address that question.

Participant Recruitment: Participants from various backgrounds were recruited using a combination of online recruitment platforms such as Craigslist, Amazon Mechanical Turk (AMT), and a participant recruitment website hosted by Columbia University. Participants self-assess eligibility using the following criteria: aged between 18 and 75 years old, no cognitive impairments, English uency, and have access to the internet. Research sessions are conducted virtually via Zoom and participants are compensated with a $15 gift card to thank them for their time.

Study Methods: Participants engaged in a 60-minute virtual research session which consists of a series of activities including surveys, semi-structured interviews, a participatory design task. The study activities ask participants to consider two different decision-making settings: an open setting (e.g., grocery store) where they had maximal exibility and choice, and a constrained setting (e.g., a cafeteria) where their choices are limited to a menu.

RQ 2.1 Experimental Design

To answer the rst research question participants completed a survey about their priorities for meal selection and had them expand on responses through a semi-structured qualitative interview. The following study components are related to RQ 2.1.

- Survey on Background & Meal Priorities: Prior to the study session participants respond to questions about their background, health experiences, and re ect on different factors that most inform when, what, and how much they choose to eat. They are asked to rate a list of factors identi ed as in uential for meal selection and health management in prior literature from not important (1) to very important (5). Factors include cost, taste, preparation time, planned activities, location, time, social setting, culture and religion [9, 36, 114]. Participants report the meals eaten the day before.

- Qualitative Interview on Experiences: The study session begins with a semi-structured interview where participants share their existing practices choosing meals, priorities for meal selection, and their prior experiences with health, technology in general, and health technology in particular.

Data Analysis: This research question focused on factors individuals prioritize when making meal choices and on ways these factors differ in situations with different constraints. Patterns from the background survey were analyzed to summarize trends and user priorities. Considerations were further explored through discussion in the research sessions to contextualize the quantitative results. One session transcript was reviewed jointly by the study team to align coding schemes, after which the remaining interviews were coded independently by me. The study team met periodically to discuss emerging codes and to resolve discrepancies through discussion.

RQ 2.2 Experimental Design

Next, participants were presented with two different situations for making nutritional choices an open setting (shopping in a grocery store) and a constrained setting (ordering a meal in a cafe-

teria). The following study components were related to RQ 2.2.

- Goal Selection: Participants were asked to select a goal to work on for the study session. They were given two nutrition goals (“Make half my meal non-starchy vegetables” and “Choose lean proteins”) and instructed to select the goal most aligned with their personal aspirations. During the study, participants only saw recommendations that met the chosen goal.

- Scenarios & Initial Plan Selection: Participants were presented with two meal decision-making scenarios: an open-ended grocery store setting and a constrained cafeteria setting (order randomized). In the open-ended setting (grocery store), participants' self-reported past meals were used to identify similar meals, while in the constrained (cafeteria) scenario, they were asked to choose from a menu.

- Recommendation Generation: Recommended meals were generated from the 3k Meal Dataset described in Aim 1. Cosine similarity was used to identify meals in the dataset similar to the user's initial plan. In each scenario, participants were presented with four recommendations: two similar (cosine similarity > 0.3) and two dissimilar (cosine similarity = 0) to the reference meal.

- Recommendation Selection & Discussion: In each scenario, participants reviewed the four recommendations generated in the prior step and re ected on them in a think-aloud process. They were then asked to indicate which recommendation they were most interested in following in that scenario and to explain their reasoning.

Data Analysis:

- Hypothesis: I hypothesized that in open-ended settings, individuals would make more diverse choices that were less dependent on their previous meals. In contrast, in the constrained setting, individuals were expected to prefer meal recommendations that were highly similar to the meal they selected from the restaurant menu.

(a)

(b) An example of recommended meals as shown to users in feasibility study.

Figure 4.1: (a) Example of reference meal and cosine similarity ratings for four meals from the goal-aligned meal library; two similar to the reference meal (>0.3) and two dissimilar (<0.3) (b) An example of recommended meals as shown to users in feasibility study

- Evaluation Metrics: A McNemar test was used to compare the proportion of preferred recommendations with high similarity scores in the open versus constrained settings. As participants also described their rationale for their choices, thematic analysis was conducted to examine qualitative factors in uencing recommendation preferences in each scenario.

RQ 2.3 Experimental Design

To answer RQ 2.3, participants were rst asked to discuss what technologies they would have found useful for making meal choices in each of the two settings (open and constrained). The following study components were related to RQ 2.3.

- Design Probe Selection: Participants were presented with seven design features identi ed from prior HCI literature on nutrition support [22, 23, 72, 94, 115, 95]. These included multiple formats for meal recommendations (e.g., text titles, descriptions, user photographs, stock photographs), two types of explanations (how a meal met a health goal and what data were used), and additional features such as recipes. In each setting, participants endorsed features of interest and explained their reasoning.

- Tool Brainstorming: Participants were then asked to propose additional features or tools they

would have wanted for decision-making in open and constrained settings. A member of the research team constructed a smartphone app interface by dragging features into a template in real time, and participants provided feedback throughout this process.

Data Analysis Plan: Artifacts generated through the participatory design activity were coded with a rubric to quantify the design probes endorsed and the features brainstormed for nutrition management technologies in each setting. The frequency of desired features in both open and constrained settings, as well as additional desired features, was reported. These patterns were further contextualized with qualitative re ections from participants as they thought aloud during the activity.

4.2 Results

A total of 15 participants engaged in the study. The results begin with a discussion of participants' backgrounds and their experiences with health management and technology, followed by a mixed methods integration of quantitative and qualitative insights to address each research question in turn.

4.2.1 Participant Background and Experience

Participants came from diverse backgrounds which informed their experiences and attitudes. Participants ages ranged from 19 to 75 years of age; most identi ed as women (73%). Just under half of the participants identi ed as white (47%), with other participants who self-identi ed as American Indian/Alaska Native, Asian, and Black. 20% of participants identi ed as Hispanic or Latino. 27% of participants self-identi ed as not having suf cient nancial resources to feel comfortable. Several participants also self-identi ed as having their eating practices informed by religious beliefs (e.g., keeping Kosher, Halal), cultural cuisine, or pre-existing health conditions (e.g., pre-diabetes, Pott's Disease).

When asked about their past experiences with health and nutrition all participants expressed some familiarity with nutrition concepts and awareness that nutrition is tied to health. However,

Category	N (%)
Age	
Mean (SD)	35.9 (18.6)
Range	19 - 75 y.o.
Gender	
Woman	10 (67%)
Man	4 (27%)
Another Term: Transgender Man	1 (6%)
Education	
Some College or less	5 (33%)
BA or higher	10 (67%)
Race	
American Indian or Alaska Native	1 (6%)
Asian	3 (20%)
Black or African American	4 (27%)
White	7 (47%)
Ethnicity	
Hispanic/Latinx	3 (20%)
Chronic Condition (any)	
Yes	4 (27%)

Table 4.1: Demographics of participants in the Aim 2 user study (N=15)

participants varied in how they engaged with nutrition in daily life. Some participants engaged with nutrition in a relatively relaxed fashion: they had basic goals for healthful eating but did not regularly invest much time and effort in thinking about food and nutrition. Others were extremely nutrition-oriented and had well-articulated perspectives on nutrition and diet. These participants fell into two distinct groups: metric-oriented and intentionally-distanced. Metric-oriented participants described setting targeted health goals such as losing weight or building muscle mass, and regularly engaged in practices like intensive macronutrient tracking and speci c diet regimens to attain their goals. Conversely, intentionally-distanced participants mindfully disengaged with practices like intensive tracking, nding a focus on metrics to negatively impact their mental and physical health. They voiced wariness about skewed nutritional narratives, diet-culture, and label-

ing certain foods as “not healthy”. These participants were often younger and several disclosed previous experience with disordered eating (either personally or in their family) which informed their intentional approach to engaging with nutrition.

> “I have a history of eating disorders... there's a lot of focus for me on working out and on food. I was restricting but was also obsessive about nutrients; I would be eating like a very small amount of food but was very conscious of how much was protein, how much was carb... now, I have a Fitbit and I'm very conscious to not track what I'm eating on that... I basically can't look at nutrition facts... but I will look at trends with my exercise” – P09

All participants described using technology in their daily lives with most using technology to support their health such as using YouTube videos and tracking apps for food, movement or menstrual activity to support health management. They noted barriers to using tracking technologies, including challenges around burdensome tracking, lapses in tracking and triggers for negative mindsets around nutrition.

4.2.2 RQ 2.1: Existing Meal Selection Practices and Variation Across Situational Settings

Findings show users juggle multiple priorities and constraints when making meal decisions; when time and resource constrained higher level priorities such as health goals often fall by the wayside to optimize for other factors such as taste, cost, and convenience.

Exploring individual patterns more deeply, I observed different subgroups of participants based on priorities (Figure 4.2b). For example, when choosing what to eat, one subset (yellow, 20%) placed health goals as a top priority but de-prioritized cost and taste; conversely, a second (orange, 20%) prioritized cost and taste but de-prioritized health goals. Finally, a third (pink, 60%) did not clearly identify any top priorities, instead rating several factors all as moderately important and qualitatively described the effortful process of nding balance between them when making meal decisions.

(a) (b)

Figure 4.2: (a) Top priorities for meal selection as ranked by users in the pre-study survey(b) Individual rankings for priorities that impact what users choose to eat; three subgroups of participants are underlined on the bottom in yellow, pink, and orange.

Beyond individual variation priorities, participants re ected that priorities were often entwined with the decision-setting in which the choice was made; certain factors were considered more important in open settings like grocery shopping or meal preparation, and others more important in constrained settings such as eating out. I observed the following qualitative themes:

Theme 1: Juggling long-term values with short-term priorities

Participants described open settings (e.g., grocery shopping, meal preparation) as environments where they had a variety of choice in available food options, time to consider options, and exibility to tailor options to personal needs. As a result, in these settings participants described quite consistently being able to make meal choices that effectively balanced their general priorities. For example, P20 described the process of actively weighing priorities such as taste, cost, and preparation plans (consistent with the factors highlighted in the survey) when considering which item to buy:

> "I shop in store because I like being able to price compare and compare it with what I want. I might see in my recipe I need salsa. Then I can just look at the [store] shelf and see what kind of salsa they have." – P20

By contrast, participants described constrained settings (e.g., eating out) as more limited, re-

quiring them to adapt on-the-spot based what options and actions were available to them. Consequently, they struggled to balance all of their priorities often needing to either negotiate which priority was most important or overlook their priorities altogether to make a decision.

Theme 2: Meal decision-making becomes “automatic” when constrained

Participants also discussed when constrained, particularly by time, they often defaulted to established habits that were convenient and required little cognitive effort. They raised several examples where they struggled to balance taste and cost with convenience, particularly in settings where they were time constrained such as running between meetings, studying, or exhausted after a long week at work. In these cases they described often simply settling for instant or ordered foods that could be conveniently obtained over meals that satis ed their self-identi ed taste, health, and cost preferences but required more effort to procure.

One participant described: "[When] I'm busy with exams, especially nal exams, deadlines or projects, I don't care about [food. I'll take bread and eat it as it is or a banana. I don't want to spend more than one minute in the kitchen." – P02

Theme 3: Limited technical support for in-the-moment decisions

All participants described using technology to support their health, such as watching cooking videos on YouTube and food, movement, or menstrual activity tracking apps. Within open settings, participants noted actively using several technologies to aid meal decision-making by helping to outsource cognitive burden and facilitate re ection and organization. For example, general purpose mobile apps such as “Notes” or store-speci c apps were used to organize shopping lists and aggregate discount relevant coupons to use at checkout. Participants also regularly used technology to assist priority-aligned meal discovery, such as using food recommendation apps or websites to discover recipes that enabled them to balance high-level priorities such as nutrition goals, budget, taste, and preparation time.

“I use the grocery store apps for my grocery list and I'm a member of and I'm in the

> rewards program, too. I use the grocery store app for the rewards and to put my list on and to keep track of coupons and all that... there's a website that I'll go to, and they have a bunch of recipes designed to be cheap. I can go and lter what ingredient they have, or they have some that are like especially easy recipes" – P20

Several participants described using generative AI to support non-health related tasks, with one person tried generative technologies to support nutrition and meal planning.

> "My friend typed into ChatGPT a sample menu. I ended up not using it just because it sounded really in depth and it had like seven meals planned out for us but I think if I'm ever looking for something really speci c... or when I'm really stuck on a recipe... I think it would maybe help." – P21

However, participants seldom described using technology to aid meal decision-making on the spot, and particularly in constrained settings. The technologies did use solely enabled instantaneous action, such as apps to place an order for food but did not help them connect to or balance their other priorities such as health goals. As a result, in constrained settings participants had to rely on their own judgment to consider possible adjustments, try and weigh priorities, and navigate cognitively taxing decisions on the spot.

4.2.3 RQ 2.2: Preference for Similarity-Based Recommendations Across Decision-making Context

The second research question aimed to assess the usefulness of similarity-based meal recommendations in supporting in-the-moment action in open and constrained decision settings. While the small sample size of this study precludes me from making strong statements, I report on the trends identi ed with statistical analysis.

Findings from the McNemar test align with our hypothesis, a greater proportion of participants preferred similarity-based recommendations, especially when constrained ($\chi^2 = 6.125$ – ? =

•013). In open settings individuals were neutral about receiving similarity-driven recommendations nding both novelty and similarity-driven recommendations equally useful. However, in constrained settings these same participants overwhelmingly preferred similarity-driven recommendations (Figure 4.3b). Given the small sample size, these results should be interpreted as indicative trends rather than de nitive conclusions.

(a) (b)

Figure 4.3: (a) Overall preferences for similarity-based recommendations in open (grocery store scenario) and constrained (cafeteria scenario) settings (b) Within user variation in preference for similarity-based recommendations in open v. constrained settings.

Qualitatively, participants described that they used recommendations differently when in open settings compared to constrained settings, impacting their situational preferences.

In open settings (e.g., the grocery store) participants used recommendations as a source of inspiration. Re ecting this, participants were interested in recommendations expanded the diversity of their repertoire and had novel ingredients, or familiar ingredients in novel combinations. In this setting participants were also willing to consider recommendations for meals not fully aligned with their preferences and constraints. Participants were open to make changes such as small ingredient swaps (often around dietary preferences), adding or removing sides, and slight changes to form of the meal so the result would be more aligned with their personal preferences.

> “I was already inclined towards carrots, yellow rice, onions, and chicken because it has vegetables in it and chicken in it. I might replace the rice.” – P07

By contrast, in constrained settings, participants overwhelmingly sought recommendations that were closely aligned with meals they had already planned to eat. They regarded recommenda-

tions as an instruction for what they should eat, and sought recommendations that were speci c, grounded, and implementable in the moment. As a result, they were particularly interested in similarity-based recommendations that shared core attributes with what they were already planning on eating and only required small adjustments to their existing plans. One participant described their logic:

> “When I already have a meal in mind, I'm thinking about that meal. It's unlikely that I'm going to go past a certain level of similarity” – P20

However, the concept of similarity was found to be complex; participants considered many different dimensions for what made a recommended meal similar or different from their initial meal choice. I enumerate and further describe some of these dimensions below.

1. Similarity in Meal Form: Most commonly, participants discussed wanting a meal of the same form to what they already planned to eat, particularly if they were in the constrained setting. For example, they preferred to swap one sandwich for another, even it included different ingredients. Not surprisingly, these considerations were based on both taste, functionality, and the convenience of ndings a substitute that was similar to the originally chosen meal. For example, participants remarked on waiting in a line for a particular type of food (“salad line/bar”, “sandwich station”) making alternatives that shared the same form easier to act on. One participant noted:

   > "I'm picturing the stations in my head and thinking about which station I would go to. Pasta is a different station, and sandwich is the same station [as my initial choice]. But that's also usually how I pick my food" – P20

2. Similarity in Meal Cuisine: Participants also considered the similarity of the meal cuisine when considering recommended alternatives. They articulated that a given cuisine often had many meals with similar avors to what they were in the “mood” to eat. On a practical note, they also pointed out restaurants were often cuisine-focused, making recommendations of similar cuisines not only appealing but more actionable.

3. Similarity in Meal Ingredients: Somewhat unsurprisingly and aligned with the rationale for similarity-based recommendations, participants considered the similarity of ingredients between their plan and alternatives they were willing to consider. However, participants' overall ingredient preferences also played a role – while they gravitated towards recommendations with similar ingredients to their initial plan, the inclusion of ingredients they did not like made those options less appealing. In open settings, participants were still willing to choose those recommendations and adjust; however, in constrained settings this often deterred them from the option altogether. On a practical note, they highlighted that preparation style made certain kinds of ingredient modi cations more feasible in the moment. One participant who was lactose intolerant noted:

   > "I technically can't eat either of these things [containing cheese], but it's easier to take cheese off a cold sandwich than a hot one." - P20

4. Similarity in Macronutrient Composition: Another consideration participants described was the macronutrient content of the recommendation. This was particularly the case in constrained settings where participants were generally considering the recommendations as an alternative to their existing plan. They discussed ensuring the meal recommendation selected had a similar nutritional pro le both quantitatively (for example equally protein rich) as well as qualitatively (the meal was equally heavy or lling) to their initial plan. One participant described:

   > "I would prefer the chicken marinara sandwich or Dagwood [sandwich over] the fruits...if I am hungry fruits are not going to quench that requirement." – P07

4.2.4 RQ 2.3: Additional Design Considerations for Meal Recommendation Systems

The third research question aimed to understand additional users' needs for technologies to support meal decision-making in both open and constrained settings, beyond focusing on the mech-

anisms for generating recommendations themselves. Through capturing participants responses to design probes that mirrored features commonly found in technologies throughout the literature and a guided brainstorming task I were able to better understand universally desired and setting-speci c features and capabilities.

With regards to the 7 design probes of features commonly found in nutrition technologies, some feature preferences were agnostic to the setting and others were decision-setting dependent (Table 4.2).

Feature from Literature	Open N (proportion)	Constrained N (proportion)
Text Meal Recommendation (title)	6 (0.40)	5 (0.33)
Text Meal Recommendation (title)	11 (0.73)	13 (0.87)
Meal Picture Recommendation from another user	2 (0.13)	4 (0.27)
Meal Picture Recommendation from a professional	9 (0.60)	9 (0.60)
Explanation: How recommendation meets health goals	9 (0.60)	7 (0.47)
Explanation: How recommendation was generated	3 (0.20)	3 (0.20)
Additional Feature: Recipe	12 (0.80)	8 (0.53)
Additional Feature: Other	12 (0.80)	13 (0.87)

Table 4.2: Comparison of Features in Open and Constrained Settings

Across all settings users liked receiving recommendations in a text-based format that had a short meal title and description. The addition of professional-quality pictures of recommended meals was generally well regarded as a source of inspiration; though in constrained settings individuals noted photos from other users helped them understand what they “were getting into” when choosing a meal recommendation.

Other features like recipes were more context-dependent and were more desired in open settings where individuals had the time and exibility to consider them in the moment. Interestingly, while users were moderately interested in seeing explanations that made clear why speci c recommended meals met identi ed health goals, they were uninterested in explanations that attempted to open the black box of how the recommendation algorithm functioned. One participant described:

> "I prefer the [explanations of the] goal because each individual is different... targeting the speci city of each person is more important [than] why the system picked this

particular meal for you" – P02

“Other” Feature Requested	N (proportion)
Control Feature Visibility	9 (0.60)
Ingredient Substitutions	8 (0.53)
Cost Information	8 (0.53)
Macronutrient Details	7 (0.47)
Location Data	4 (0.27)
Preparation Time	4 (0.27)
User Review Ratings	4 (0.27)
Dietary Restriction Labels	3 (0.20)

Table 4.3: Summary of Requested Features and Their Proportions

Additionally, participants suggested features not included in the original design probes that would be useful across all decision settings. In particular, they suggested features that lent users more control over the tool at three main levels to ultimate provide more user-centered decision support: (1) control over the overall system, (2) control over the algorithm mechanism, and (3) control over the in-the-moment output.

At the level of the overall system, users wanted to be able to provide direct input about the overarching unchangeable assumptions the technology was making about them. They wanted ways to provide direct input around food allergies or dietary restrictions. They also wanted ways to turn on or off certain features across the system, for example turning off a display of macronutrient information if it was triggering. As part of this exibility at the system level users discussed how their own needs evolved with time and experience; while features such as explanations of why recommendations meet goals was useful when they were a novice, they wanted to be able to turn off these features as they gained experience. One participant elaborated:

> "I feel like maybe the ability to turn on or off [explanations] based on what your goal was... if it was a goal that was less self-evident... [or] if you felt like it was something that you were still learning about" – P20

At the level of the algorithm, participants expressed interest in being able to modulate how the algorithm identi ed recommendations. They expressed enthusiasm at the idea of being able to

adjust in real-time the degree of similarity between their original meal plan and recommendations. They were also interested in being able to specify the dimension of similarity they were interested in to better “tune” the recommendations generated by the algorithm to meet their needs.

Finally, at the level of in-the-moment output, participants desired control over how recommendation results were displayed and the ability to manipulate the presentation of results in real-time to aid decision support. They suggested additional controls like lters that enabled them to sort recommendations by other data elements that aligned with their preferences such as cost, reviews or ratings to ascertain taste, or within recommendation ingredient substitutions. This gave them ne-grained control to actually integrate (rather than overlook) their higher priorities even in constrained settings.

4.3 Discussion

This aim explored how people's meal choices are in uenced by their priorities and setting, and implications for the design of nutrition recommendation systems that account not only for individuals' preferences but also for the settings in which they make nutritional decisions. Using a mixed methods approach consisting of surveys, qualitative interviews, and a participatory design activity this study generated a number of insights that have signi cant implications for the design of future human-centered nutritional recommender systems.

First, the study showed that meal choices are shaped not only by people's stable priorities—like cost, taste, preparation time, and health goals—but also by the context in which those choices happen. In open settings, participants described weighing and balancing multiple priorities, sometimes supported by nutrition technologies, while in constrained settings they made quick trade-offs and rarely turned to technology.

Second, this study demonstrated the utility of similarity-based recommendation particularly in constrained settings. Similarity-based recommendations were especially valued because they aligned with participants' existing meal plans and felt more doable in the moment. At the same time, what counted as `similar' was not straightforward. Participants noted that the attributes of

similarity (form, cuisine, ingredients, and macronutrients) that mattered in-the-moment shifted depending on the context in which they were making the choice.

Finally, participants also highlighted the need for different layers of control—from setting stable inputs like allergies and goals, to turning features on and off, to re ning recommendations in the moment—and preferred explanations that tied back to their health objectives rather than the algorithm itself.

Beyond Preference: The Role of Priorities and Existing Practices in Decision-making

First, our study demonstrated that individuals' nutritional choices are often driven by particular priorities. For some, cost of meals was the main consideration when choosing what and where to eat, while others focused on convenience, including required time and effort, and yet others primarily focused on taste and chose foods they like and enjoy eating. The study showed how these priorities become foundations for individuals eating habits and routines. For some of these priorities all participants had a clear directionality in their preference, for example all participants preferred to minimize cost or maximize convenience — though, how much these factors drove the nal meal selection varied across individuals. However other priorities, such as taste, varied and participants had taste preferences that ranged across the entire spectrum. All these different factors impact how individuals go about making their nutritional choices and must be accounted for in the design of truly human-centered recommender systems.

These ndings are consistent with previous research in HCI that highlighted the importance of values in individuals' choices and actions and of accounting for these values in the design of human-centered computing technologies. Speci cally, prior work examined value-sensitive design, and focused on eliciting the values and needs of users and on building technologies rooted in these values and attuned to users' needs. Similar efforts exist in the space of technologies for health and wellness, where recent work has explored the importance of building value-sensitive tools to support health, and explored building value-aware recommender systems [23, 116].

In contrast to these broader efforts, nutrition recommendation systems have primarily focused

on individuals preferences for certain foods. One might argue that inferences made with individuals' past choices re ect both preferences and values. However, in this approach it is easy to convolute these different factors that drive choices and tease out distinction between them, for example, between taste preferences and broader values. Furthermore, it may be challenging to untangle values and preferences on one hand and situational constraints on another. As such, the design space for value-sensitive recommender systems is relatively unexplored. Prior research suggested techniques for eliciting individuals' values in the context of health through collaborative re ection with healthcare providers [117]. However, outside of the healthcare settings, this more explicit approach to value elicitation could increase user burden and reduce their willingness to engage with technologies, as was demonstrated by research in privacy that required users to explicitly state their privacy preference [2]. Further research is needed to identify easy ways to incorporate individuals' priorities into the design of nutritional recommender systems without increasing user burden.

The Role of Situational Settings: Internal and External Context Shape Decision-making

In addition to pointing out the importance of internal values, our study also showed some variability in the degree to which individuals accounted for these values in their daily choices. Some of this variability was driven by internal factors and participants often talked about situations when they felt “in the mood” for foods they generally try to avoid. Participants noted that the situational setting impacted what they perceived the function of the recommendation to be: inspiration in the open setting, and instruction in the constrained setting. Consequently, they preferred seeing similarity-based meal recommendations in constrained settings as these recommendations closely aligned with their existing meal plans and were considered more practically actionable. All these ndings highlight the importance of considering context in nutritional choices and tuning to which dimensions of context are relevant for informing and shaping future action.

These ndings resonate with previous research. First, past work explored the role of internal context in individuals' decisions. For example, research in context-aware recommendations

and human-centered recommendation systems has shown that momentary needs sometimes take precedent over established preferences and priorities [118, 112, 119]. Furthermore, prior work emphasized the importance of incorporating these momentary shifts in priorities and intention (a person's state-of-mind) alongside their more established preferences [86, 120]. Past research in other application domains such as travel have explored this idea as well. Viswanathan et al. developed “Mindsets” a point-of-interest recommender where users could select mindset and explore suggestions through queries related to their mindset (e.g., “I'm hungry”, “Surprise Me”, “What's free”) [121]. In a health context, Rohani et al. described the development of MUBS, a recommendation system to support behavioral activation in metal health. Through this system users received recommendations across a range of dif culty levels to lend users the exibility to select an activity they felt they could do based on their general preferences and momentary needs [65]. Second, past research explored the role of external context in individuals' choices. For example, situation-aware recommender systems integrate multiple aspects of context to understand decision making and discreet situations at the intersection of these factors (e.g., a weekend in New York City with friends) [86, 36]. Past studies have hinted at the role of constraints in determining differences between situational settings and explored how priorities may be dynamic based on the situation. For example, Viswanathan et al. explored this in the context of three travel situations: an on-the-spot setting (highly constrained), a re ning plans situation (moderately constrained) and moments of boredom (where users were least constrained and focused on exploration). Similar to our study, these authors cataloged how individuals priorities shifted depending on their situational environment (for example they were less willing to consider activities at a distance in on the spot settings, but more willing to do so when re ning plans) [121]. Knowing how situational context shifts priorities and impacts what recommendations are considered more useful has important implications for the design of future systems.

The ndings from these studies have implications for future work. First, future nutrition recommendation system could take steps to account for individuals' internal context. For example, such systems could incorporate the notion of common mindsets people nd themselves in when eating

(e.g., “munching on the go”, “good for groups”, “comforting meals”) and use these mindsets to generate appropriate recommendations. Furthermore, future recommender systems could identify new opportunities to incorporate external context and constraints into the design of recommendation engines. While collaborative ltering and content-based recommendation algorithms continue to be dominating recommendation-generating engines, our study showed the plausibility of using other computational mechanisms, such as measures of meal similarity, as an alternative approach to selecting situationally appropriate recommendations. Finally, another approach to accounting for the impact of context is to vary prescriptiveness and speci city of recommendations. Past work on health recommendations within the HCI community have communicated suggestions with varying levels of prescriptiveness from identifying a general trend [57], to suggesting a category for action [64], to highly speci c suggestions about a behavior to do [65]. Recommendations that are less speci c (such as “add more vegetables” or “add more lean protein to your meal”) might enable users to nd situationally appropriate way to act on them. However, in the context of nutrition, such less speci c recommendation might require high nutritional literacy (for example, knowing what a lean protein is and what foods count as lean proteins), which may present a barrier to individuals from low resource communities with limited access to health education. Further research is needed to nd appropriate and user-friendly ways to account for situational context in the design of future nutritional recommender systems.

Enabling Control: Mechanisms to Support User Autonomy

Beyond this our study highlighted the importance of including mechanisms for user control when building truly human-centered systems. Throughout the participatory design session users were particularly interested in features that allowed them to control their interaction experience based on the setting (this was the case across all contexts). Users talked about features like lters for location, equipment availability, and preparation time. They also wanted to be able to turn on and off additional features, such as explanations or recipes when they were not situationally required to reduce cognitive burden. Participants also noted how having control of features and

being able to dynamically add or remove could empower them with an ability to tailor systems to better support their personal needs. For example, participants had different experiences with the feature of nutrition tracking: while some participants found tracking helpful, others found such features triggering (past studies have shown this as well [122]. Participants who found tracking triggering voiced wanting the ability to opt out of this feature, while still using other app features.

These ndings are consistent with prior literature that highlights the importance of autonomy and control in shaping individuals' experiences and choices, particularly in the context of health and health behaviors. Bandura's Social Cognitive Theory includes autonomy as one of the principle human needs that in uence individuals' motivation to engage in different behaviors [120]. The notion of autonomy is also prevalent in literature on chronic disease self-management and patient empowerment [123, 81, 124]. It is also one of the prevalent themes in the emerging body of literature on Human-Centered AI that increasingly highlights the importance of users' control and autonomy in their interaction with AI-powered interactive systems [125, 51, 37, 39].

Our ndings suggest that future nutritional recommendations systems should align with these prior efforts and strive to give users appropriate level of control that could enhance their sense of autonomy. For example, early recommender systems allowed users to explore a broad space of recommendations and dynamically apply different lters to narrow the space down, which gave users control over the selection of nal recommendations [126]. However, more recent RecSys mostly automated this process and obscured it from users, presenting them with only the nal few and sometimes even just one nal recommendation. Future work could explore how to combine computational mechanisms for ltering recommendations (such as collaborating ltering) with user-driven ltering to enable negotiation and promote user control.

Additionally, users highlighted that their needs and skills evolve over time and raised concern about outgrowing particular features (such as explanations) as they gained knowledge. This challenge has been echoed in prior HCI work and has been highlighted as one reason for lapsing use of informatics technologies to support self-management [25]. A exible design approach could lend users greater autonomy to turn on or off particular features such as explanations or meal logging.

Users could also customize the format of elements like how recommendations are presented. This might allow multiple participants use the same system to meet diverse preferences and needs and support participants as their own needs for decision support evolve. This may be another opportunity for a exible design approach to allow users to dynamically adjust features to their needs over time. However more research is needed to understand how to build exibility into such tools in a way that does not put the burden back on the user but at the same time lends them some control to better situate a tool in the context of their lives.

Context-aware computing can be a powerful tool to tailor interventions to people's values and situational setting to provide actionable support [88, 87, 15]. At the same time, it is important to recognize the limits of computation and invite users to provide input. These ndings highlight the importance of design systems that provide clear and intentional opportunities for user autonomy and control. Empowering users through providing mechanisms for control may support interaction experiences that are in the end more human-centered.

4.4 Key Takeaways

Aim 2 characterized user needs and assessed the feasibility of similarity-based recommendations in everyday decision settings.

Findings from RQ 2.1 showed that meal choices were shaped not only by individual priorities such as cost, taste, and preparation time, but also by the situational context in which decisions occurred. In open settings, participants balanced multiple priorities with support from existing nutrition technologies, whereas in constrained settings they made on-the-spot compromises and rarely used technology.

RQ 2.2 demonstrated that similarity-based meal recommendations were particularly valued in constrained settings, as they aligned closely with users' existing meal plans and were considered more actionable. However, the speci c dimensions shaping perceptions of similarity (e.g., meal form, cuisine, ingredients, or nutritional value) varied in ways that were not always predictable.

Finally, in RQ 2.3, Participants emphasized the importance of autonomy and control in meal

recommender systems. They wanted exibility at the system level (e.g., turning features on or off, updating dietary restrictions), the algorithm level (e.g., adjusting similarity criteria and dimensions), and the moment-to-moment level (e.g., ltering results by cost, reviews, or substitutions). These preferences highlight that users value tools that adapt to their evolving needs and give them agency over how recommendations are generated and presented, rather than passively receiving xed suggestions.

Implications for Human-Centered Nutrition Technologies: This study underscores that building human-centered AI for nutrition means going beyond static personalization to support decisions as they actually unfold. Systems need to be aware of shifting contexts, responsive to both long-term goals and immediate constraints, and designed to give users meaningful control at different levels. By centering user priorities and offering actionable, context-aware recommendations that can be re ned in the moment, AI systems can feel less like a black box and more like a collaborative tool that works with people in their everyday meal decisions.

Chapter 5: Aim 3 — Explore the use of large language models (LLMs) to elicit users on in-the-moment priorities and tailor meal recommendations

Prior to this aim, the dissertation examined how people reason about meal similarity and how those perceptions shape meal decision-making, establishing the foundation for similarity-based meal recommendation in-the-moment.

Aim 1 established the technical feasibility of computationally assessing the similarity between meals in a way aligned with human judgment. Four key dimensions emerged as most in uential in shaping human perceptions of meal similarity: meal form, cuisine, macronutrients, and key ingredients. It also showed that meals judged to be similar were more likely to be considered substitutable, supporting the rationale for similarity-based recommendations as a useful approach for meal decision support.

Aim 2 demonstrated the utility of similarity-based recommendations for supporting in-the-moment meal decisions, particularly in constrained settings. Participants identi ed the same meal similarity dimensions, but noted that their importance shifts based on the setting. For example, cuisine was the most salient when dining at a restaurant, whereas form mattered more in on-the-go contexts. They also expressed a desire for more direct and ne-grained control over which dimensions of similarity to prioritize during each interaction.

This work highlighted a critical user need for actionable similarity-based recommendations: the ability support recommendation around many different aspects of similarity and enable users dynamically specify which dimensions are a priority during decision-making. A promising approach to enabling such in-the-moment recommendation is by leveraging LLMs. For example, one might expect that contemporary LLMs are able to engage individuals in a conversation to elicit their nutritional goals and in-the-moment preferences and generate meal recommendations

that align with both.

This approach, however, requires reliance on generative capabilities of LLMs and is sensitive to their common challenges, including hallucinations, deviation from prompts, and loss of context overtime. Healthy eating and nutrition are also sensitive domains, and prior work has shown that unchecked generative recommendations can inadvertently propagate harm or reinforce disordered eating behaviors [127].

For this reason, I made an intentional decision to limit the system to retrieval-based recommendations. This provided a safer evaluation context. To support this retrieval-based approach, I drew on the 3K meal dataset used in previous aims. This dataset includes meals captured and consumed by individuals in the real world and was labeled and curated by expert dietitians on alignment with health goals, thus avoiding pitfalls of generative AI.

However, leveraging this dataset as a recommendation catalog requires three components. First, meals in the recommendation catalog must be clearly mapped to how they align with the four similarity dimensions in order to enable the selection of recommendation aligned with a specific dimension. Second, because different contexts call for different considerations, there is a need for mechanisms that elicit and update a user's priorities around these dimensions in real time, to guide which recommendations are selected. Finally, these elements must be integrated into a system that suggests similarity-based recommendations aligned with users' momentary priorities, ensuring that suggestions remain relevant and actionable in the setting at hand.

Here, I explore an approach that leverages the generative and interactive capabilities of LLMs to support these tasks. As such, Aim 3 of the thesis explored LLM-based methods for supporting dynamic, in-the-moment meal recommendation. This aim focuses on the following research questions, explored through three separate studies.

- '& $_{3.1}$: How effectively can LLMs generate accurate and appropriate tags from free-text meal descriptions for four dimensions of meal similarity?
- '& $_{3.2}$: How can LLMs be used to elicit in-the-moment meal preferences to support recommendation?

- '& $_{3.3}$: What are users' experiences and perceptions of interacting with an LLM-based chatbot for in-the-moment meal recommendation?

The approach proposed here aims to provide similarity-based recommendations for meals that do not meet an individual's speci ed nutritional goal. This requires a mechanism for assessing alignment between meals and nutritional goals. As the development of such a mechanism is beyond the scope of this dissertation, here, I rely on an approach developed by Hu et al [128]. Because the goal assessment step is handled separately, the studies in this aim take place in a controlled lab setting. This allows us to focus speci cally on how similarity-based recommendations can be presented and re ned after a meal has been identi ed as not meeting the goal.

Study 1 (RQ 3.1) assessed the use of LLMs to enrich the existing meal catalog of free-text meal descriptions with additional tags for each of the four meal similarity dimensions. This ensures meals in the recommendation catalog are clearly mapped to how they align with the four similarity dimensions, ensuring it is easier to lter and re ne recommendations along these dimensions.

Study 2 (RQ 3.2) explored how LLM-based dialog in combination with these tags can enable real-time, user-driven meal recommendation re nement along the different dimensions. I propose a "This or That" conversational dialog to elicit user preferences in an intuitive, low-effort way that also gives users dynamic control of recommendations.

Finally, Study 3 (RQ 3.3) evaluated users' attitudes toward this approach through a lab-based assessment of the “This or That” conversational ow in different meal decision-making contexts. Users interacted with two prototype chatbots, one that sought to elicit users needs through a "This or That" dialog and one that did not. They evaluated the chatbots on perceived utility and the feasibility of conversational, dimension-based re nement in practice.

5.1 Study 1 – Using LLMs to Generate Meaningful and Human-Aligned Tags for Meal Dimensions

This study validated the use of LLMs to generate tags for free-text meal descriptions aligned with the four meal similarity dimensions identi ed in Aim 1 and Aim 2. Tags were generated using

a variety of prompting approaches and validated by human crowdworkers. The best performing approach was used to tag the 3k meal dataset.

'& $_{3 \bullet 1}$: How effectively can LLMs generate accurate and appropriate tags from free-text meal descriptions for four dimensions of meal similarity?

5.1.1 Related Work

Prior work on LLMs for tagging and labeling highlights their potential as scalable alternatives to manual annotation. Traditional methods rely on experts or crowdworkers, but these are time-consuming, costly, and prone to inconsistency at scale [129]. By contrast, LLMs have been used for tasks such as named entity recognition, document classi cation, and qualitative coding, including in health contexts [130, 131, 106, 132]. Increasingly, researchers have explored human-in-the-loop pipelines where LLMs generate preliminary labels that humans then verify or re ne [133, 134, 135, 136]. These hybrid approaches reduce burden while maintaining contextual accuracy.

A central factor shaping the quality of LLM-based tagging is prompting strategy [133, 137, 138]. There are several commonly used prompting strategies for generating LLM output. Zero-shot prompts are the simplest and provide only a task directive, relying entirely on the model's existing knowledge. Few-shot prompts build on this by including a few labeled examples, giving the model a pattern to generalize from. Other approaches provide greater context for how the LLM should go about conducting the task at hand. For example, instruction-based prompts include detailed de nitions, constraints, or contextual guidance that standardize how the task should be carried out. In contrast, chain-of-thought prompts explicitly walk the model encourage the model to reason step-by-step through the task before producing an answer, improving consistency for tasks requiring multi-step judgment.

Studies show that performance of these prompting approaches varies with task complexity, and even small differences in examples or structure can affect outputs [139]. Recent work also highlights that hybrid approaches (such as combining instruction with few-shot examples or layering chain-of-thought reasoning onto example-based prompts) often yield stronger and more consistent

outcomes and that iterative, user-driven re nement often outperforms automated prompt optimization [138, 130].

At the same time, scholars caution that tagging frameworks re ect deeper assumptions about categories and meaning. For example, Zhang et al. show how prompt structures encode speci c views of pluralism and groundedness, raising concerns in culturally sensitive domains like food, where rigid or homogenized labels risk erasing diversity [140, 141, 142].

Taken together, this literature suggests that LLMs can be powerful tools for metadata generation, but success depends on prompt design and human veri cation. Additionally such automated tagging is usually completed using a set of structured tags. This study builds on prior work by applying LLM-based tagging to the domain of nutrition self-management, enriching a large corpus of unstructured meal descriptions with structured, human-centered metadata aligned to user-de ned dimensions of similarity.

5.1.2 Methods

For this aim I explored different model and prompt combinations to generate appropriate tags for each meal dimension. These tags were validated by human raters. Then the best performing approach was used to tag the full 3k meal dataset.

Validating LLM Tagging Approach: Model & Prompt Evaluation

Candidate LLMs: I explored the following commercial LLMs to tag a corpus of meal descriptions across the identi ed meal dimensions: form, cuisine, macronutrients, and key ingredients. Prior work (described above) has shown that these models performed well on free-text tagging and annotation tasks.

1. ChatGPT 4o (OpenAI): ChatGPT is a LLM developed by OpenAI that has demonstrated strong performance on classi cation and text generation tasks [143]. It was accessed via OpenAI's HIPAA-compliant API through Columbia University, using the gpt-4o-mini model with a temperature of 0.7 and standard chat-based prompting.

2. LLAMA (Meta): LLaMA is an open-weight, transformer-based LLM developed by Meta and trained on large-scale text corpora, with demonstrated effectiveness for reasoning and classi cation [144, 145]. It was accessed through the Groq API, using the llama3-8b-8192 version with a temperature of 0.7 and message-based prompting.

3. Gemini (Google): Gemini is a LLM developed by Google DeepMind that leverage large-scale data resources for ef cient and accurate reasoning [146]. It was accessed via the Gemini API, using the gemini-1.5- ash model with a temperature of 0.7

Candidate Prompt Evaluation & Selection: Four candidate prompts were explored to determine the best approach to eliciting dimension-speci c tags. Examples of candidate prompts are noted in Table 5.1.

Evaluation of Candidate Models & Prompts: Tags were generated for a test set of meal descriptions using each model–prompt combination. These tags were subsequently reviewed by crowdworkers to assess their validity. The evaluation was conducted on the gold-standard meal-pair dataset developed in RQ 1.1. This dataset included 50 meal pairs (59 unique meals) with human assessments of the relevant meal attributes.

Crowdworker Validation of LLM Tags: A crowdworking study was conducted to evaluate the accuracy of the LLM-generated tags for all meals in the test dataset. Two crowdworkers reviewed each assigned tag and assess the appropriateness of the tag by responding to the following questions: (1) Does the tag t the meal described? (Yes/No/Not Sure) and (2) Does the tag t the meal dimension? (Yes/No/Not Sure)

Validator Recruitment: Crowdworkers were recruited through Amazon Mechanical Turk to validate the LLM-generated meal tags. Eligibility criteria required uency in English and “Master” quali cation status, which indicates consistently high-quality performance on prior tasks. Each tag was independently assessed by two crowdworkers.

Data Analysis: To assess the validity of the approach, percent accuracy was calculated for tags at both the meal-description level (question 1) and the meal-dimension level (question 2). A tag was considered valid only if it was approved by both raters. Tags that failed validation went

LLM Role	Prompt
Zero-Shot	Task: Read the following meal description and generate up to 3 relevant tags that describe the meal form. Use concise, descriptive tags only, and avoid generic words like "food" or "meal", "entree" or "dish".
Zero-Shot (with dimension de - nition)	Task: Read the following meal description and generate up to 3 relevant tags that describe the meal form. Meal form refers to the physical structure or compositional format of the meal — how it is assembled, served, or eaten. Examples include "bowl", "wrap", "salad", "sandwich", "stir-fry", or " atbread", but do not limit responses to these. Focus on how the meal is constructed or presented — such as being layered, handheld, scooped, rolled, or portioned. Use concise, descriptive tags only, and avoid generic words like "food" or "meal", "entree" or "dish".
Two-Shot	Task: Read the following meal description and generate up to 3 relevant tags that describe the meal form. Here are some examples: • “spaghetti and meatballs” –> Tag(s): pasta, noodles, saucy • “chicken stir-fry with brown rice” –> Tag(s): stir-fry, rice bowl Use concise, descriptive tags only, and avoid generic words like "food" or "meal", "entree" or "dish".
Two-Shot (with dimension de - nition)	Task: Read the following meal description and generate up to 3 relevant tags that describe the meal form. Meal form refers to the physical structure or compositional format of the meal — how it is assembled, served, or eaten. Examples include "bowl", "wrap", "salad", "sandwich", "stir-fry", or " atbread", but do not limit responses to these. Focus on how the meal is constructed or presented — such as being layered, handheld, scooped, rolled, or portioned. Here are some examples: • “spaghetti and meatballs” –> Tag(s): pasta, noodles, saucy • “chicken stir-fry with brown rice” –> Tag(s): stir-fry, rice bowl Use concise, descriptive tags only, and avoid generic words like "food" or "meal", "entree" or "dish".

Table 5.1: Candidate Prompt Structure for LLM Tagging Task

through a qualitative error analysis to identify possible patterns in mislabeling (e.g., vague tags, semantic drift, irrelevant keywords).

Dimension	De nition
Meal Form	Meal form is the physical form or structure of the meal (for example, terms like "bowl", "wrap", "salad", or "sandwich", but do not limit yourself to these). Focus on how the meal is assembled, served, or eaten.
Meal Cuisine	Meal cuisine refers to the regional or cultural culinary tradition that the meal belongs to or is inspired by. Examples include "Italian", "Mexican", "Indian", "Mediterranean", or "Japanese", but do not limit responses to these. Focus on the overall style, ingredient combinations, and preparation patterns typically associated with a cuisine. If the meal blends multiple cuisines, include each one. Avoid tagging speci c ingredients or meal forms.
Macronutrient Composition	Macronutrient composition describes the dominant macronutrient pro le of the meal — whether it is primarily composed of proteins, carbohydrates, or fats, or is balanced. Valid tags include terms like "high-protein", "low-carb", "high-fat", "balanced", or "carb-heavy", but you may also infer combinations (e.g., "protein and carb rich"). Base your tags on the major components described in the meal. Avoid micronutrients or dietary labels like "vegan" or "keto" unless they relate directly to macronutrient balance.
Key Ingredients and Flavors	Key ingredients or avors are the primary food components or avor elements that de ne the character of the meal. These may include main proteins, vegetables, spices, sauces, or distinctive taste notes. Examples include "chicken", "tomato", "basil", "soy sauce", "garlic", or "spicy", but do not limit responses to these. Focus on recognizable ingredients or dominant avor descriptors (e.g., "umami", "tangy", "creamy"). Avoid describing the form or cultural origin of the meal.

Table 5.2: De nitions for four main dimensions of meal similarity used to guide labeling and instruction-based prompting

Generation of Dimension-Speci c Tags of Meal Description Dataset

After identifying the optimal model and prompt combination, this approach was used to generate tags for the full 3k meal dataset (described in Aim 1 with details in the appendix) to produce an enriched dataset with tags for meal form, cuisine, macronutrients, and key ingredients. This dataset served as the meal library for the conversational recommendation system developed in Aims 3.2 and 3.3.

5.1.3 Results

Across all prompt–model conditions, inter-rater reliability among crowdworkers was consistently high (above 85%), indicating strong agreement on the appropriateness of tags across both

evaluation questions. This provides con dence that the observed differences in performance across models and prompts re ect true variation in tagging quality.

Performance varied by both model and prompting strategy as is shown in Table 5.3. All models showed high levels of accuracy in producing valid tags (with accuracies ranging from 73% and 100%), though LLAMA generally outperformed GPT ad Gemini when evaluated on whether the generated tags matched the intended dimension, while GPT tended to perform slightly better than LLAMA and Gemini on whether tags captured speci c meals.

Few-shot prompting combined with the inclusion of a dimension de nition produced the most consistently strong results across both models. For instance, the LLAMA Few-Shot + De nition condition achieved the highest overall accuracy across dimensions, with performance at or above 96% for form, cuisine, macronutrients, and key ingredients when evaluating dimension-matching (Supplementary Materials, Table D.1) and achieving perfect performance (100%) on macronutrient and key ingredient tags when evaluating tag–meal alignment (Supplementary Materials, Table D.2). Accuracy varied slightly by dimension, as described below:

Model - Prompting Approach	Form	Cuisine	Macronutrient	Key Ingredient
GPT - Zero Shot	95%	86%	76%	97%
GPT - Zero Shot + Def	93%	88%	78%	92%
GPT - Few Shot	97%	86%	92%	95%
GPT - Few Shot + Def	95%	88%	92%	98%
LLAMA - Zero Shot	100%	98%	93%	86%
LLAMA - Zero Shot + Def	98%	98%	93%	92%
LLAMA - Few Shot	95%	98%	97%	72%
LLAMA - Few Shot + Def	98%	98%	97%	97%
Gemini - Zero Shot	86%	69%	69%	95%
Gemini - Zero Shot + Def	92%	93%	71%	93%
Gemini - Few Shot	71%	90%	83%	93%
Gemini - Few Shot + Def	88%	93%	75%	90%

Table 5.3: Performance of each prompt x model for generating tags accurate to the speci c dimension and meal of focus. The best performing approach for each dimension is noted in bold, with the second best approach in italics.

- Meal Form: All models performed exceptionally well, with multiple conditions achieving near-perfect accuracy. GPT Few-Shot + De nition and LLAMA Zero-Shot achieved 100% accuracy on this dimension.
- Cuisine: LLAMA conditions, particularly Few-Shot and Few-Shot + De nition, reached 100% accuracy for cuisine tags, outperforming GPT variants.
- Macronutrients: This dimension proved more challenging across models. GPT Zero-Shot variants dropped below 80% accuracy, while LLAMA Few-Shot + De nition achieved the highest performance (96.61% for dimension matching and 100% for meal matching).
- Key Ingredients: GPT Few-Shot and Few-Shot + De nition produced the strongest results (98.31%), though LLAMA Few-Shot + De nition was comparable (96.61%).

Optimal Approach: Overall, LLAMA Few-Shot + De nition emerged as the optimal model–prompting combination. This condition consistently produced the highest accuracy across both evaluation criteria, achieving near-ceiling performance for cuisine and macronutrient tagging while maintaining strong results for meal form and key ingredients. GPT Few-Shot + De nition also performed well, particularly for key ingredient tagging, though it underperformed relative to LLAMA on cuisine and macronutrient dimensions.

Based on these ndings, the LLAMA Few-Shot + De nition approach was selected for tagging the full 3,000-meal dataset in order to generate the enriched meal library used in subsequent aims. Figure 5.1 summarizes the most common tags in each dimension for the meals in the 3k Meal Dataset.

5.1.4 Discussion

This study explored the feasibility of using LLMs to tag meal descriptions across multiple dimensions relevant to nutrition self-management, including form, cuisine, macronutrients, and key ingredients. Our ndings demonstrate that LLMs can produce structured annotations that align closely with human-generated labels in certain domains, particularly for broad categories