UNIVERSITI TEKNOLOGI MARA

DEVELOPMENT, VALIDATION, AND EFFECTIVENESS OF A PERSONALIZED AI PROMPT GENERATOR (NEXGEN-CHATGPT) FOR WEIGHT MANAGEMENT IN ADULTS WITH OVERWEIGHT AND OBESITY

AZWA SURAYA BINTI MOHD DAN

MSc

September 2026

UNIVERSITI TEKNOLOGI MARA

DEVELOPMENT, VALIDATION, AND EFFECTIVENESS OF A PERSONALIZED AI PROMPT GENERATOR (NEXGEN-CHATGPT) FOR WEIGHT MANAGEMENT IN ADULTS WITH OVERWEIGHT AND OBESITY

AZWA SURAYA BINTI MOHD DAN

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science
(Sport Science and Recreation)

Faculty of Sports Science and Recreation

September 2026

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a Panel of Examiners has met on 20 May 2026 to conduct the final examination of Azwa Suraya Binti Mohd Dan on her Masters of Science thesis entitled "Development, Validation, and Effectiveness of a Personalized AI Prompt Generator (NExGEN-ChatGPT) for Weight Management in Adults with Overweight and Obesity" in accordance with Universiti Teknologi MARA Act 1976 (Akta 173). The Panel of Examiners recommends that the student be awarded the relevant degree. The Panel of Examiners was as follows:

Rozita Abdul Latif, PhD
Professor
Faculty of Sports Science and Recreation
Universiti Teknologi MARA
(Chairperson)

Maisarah Shari, PhD
Senior Lecturer
Faculty of Sports Science and Recreation
Universiti Teknologi MARA
(Internal Examiner)

Miharaini Md Ghani, PhD
Senior Lecturer
School of Communication
Universiti Sains Malaysia
(External Examiner)

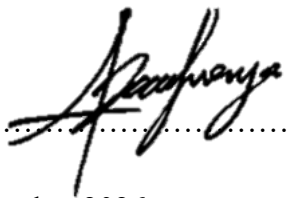
PROF DR HJH ZURAEDA BINTI IBRAHIM
Dean
Institute of Postgraduates Studies
Universiti Teknologi MARA

Date: 13 September 2026

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Postgraduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Azwa Suraya Binti Mohd Dan
Student ID. No.	:	2023653718
Programme	:	Master of Science (Sport Science and Recreation) – (Research) – SR750
Faculty	:	Sports Science and Recreation
Thesis Title	:	Development, Validation, and Effectiveness of a Personalized AI Prompt Generator (NExGEN-ChatGPT) for Weight Management in Adults with Overweight and Obesity
Signature of Student	:	
Date	:	September 2026

ABSTRACT

Artificial intelligence (AI) chatbots have shown promise for lifestyle-based weight management, yet no validated mechanism exists for generating personalized, obesity-specific prompts for large language models, and evidence comparing personalized versus generic AI-driven approaches remains limited. This thesis comprised two sequential studies to develop, validate, and evaluate the effectiveness of the NExGEN-ChatGPT personalized prompt generator for weight management in adults with overweight and obesity. In Study 1, a Fuzzy Delphi Method (FDM) was employed whereby 21 multidisciplinary experts evaluated 111 criteria across 6 constructs and 12 elements using a bilingual 7-point Likert scale questionnaire. All 111 criteria met the three FDM acceptance thresholds (expert agreement ≥ 75%, threshold value $d \leq 0.20$, defuzzified α-cut ≥ 0.50), confirming comprehensive expert consensus for the framework. The highest-ranked criteria included initial and ongoing physical activity assessment (0.859, 100% consensus) and clear weight change goal definition (0.806, 100% consensus), while instrument reliability was confirmed with Cronbach's α ranging from 0.714 to 0.891 across constructs. In Study 2, the validated NExGEN framework was deployed in a two-arm parallel randomized controlled trial whereby 160 adults (BMI 27.5–34.9 kg/m²) were randomized to either the NExGEN personalized prompt generator integrated with ChatGPT (NEX, n = 81) or a structured manual ChatGPT control program (CON, n = 79) over a 12-week intervention with 12-week follow-up. Compared to CON, the NEX group achieved significantly greater weight loss of 6.6 kg (7.5%) at 12 weeks and 5.5 kg (6.2%) at 24 weeks, with between-group differences of 3.2 kg ($p = 0.016$) and 3.5 kg ($p = 0.018$), respectively. Fat mass decreased significantly in NEX while fat-free mass remained stable across all time points ($p > 0.05$), confirming weight loss was primarily from adipose tissue. Dietary energy density and fat intake were significantly lower in NEX compared to CON at both 12 and 24 weeks ($p < 0.05$), while objectively measured physical activity via accelerometry was significantly higher in NEX at both time points ($p < 0.05$). Cardiometabolic markers including total cholesterol and LDL cholesterol showed significant time effects within NEX ($p < 0.05$), though no significant between-group differences were observed, likely reflecting the metabolically healthy baseline profile of participants. Taken together, the present thesis establishes the NExGEN-ChatGPT framework as the first expert-validated, personalized prompt engineering system for AI-driven obesity management and demonstrates that personalized prompt generation produces clinically meaningful weight loss and favorable behavioral modifications sustained beyond the active intervention period.

ACKNOWLEDGEMENT

All praise is due to *Allah*, the *Most Gracious* and *Most Merciful*, whose guidance carried me through every uncertain moment of this journey. There were nights I questioned whether I could finish this, and mornings where the path forward appeared only because He made it visible. Every breakthrough, every resolved setback, every ounce of clarity belongs to Him. *Alhamdulillah*. While I bear full responsibility for the content of this work, its completion would not have been possible without the intellectual contributions, emotional support, and honest (sometimes painfully honest) feedback from those who walked this journey alongside me.

To my supervisor, *Adam Linoby*, or as I have called him throughout this journey, "*Sir*," and sometime, "*Boss*." If anyone deserves a co-authorship credit on my sanity, it is him. I have spent more evenings in his office than I care to admit, planning experiments when most people were planning dinner, debugging statistics when the campus was empty, and rewriting sections I was convinced were already perfect (they were not). He never once told me to go home. *Sir* has been my only supervisor for this thesis, and looking back, I would not have wanted it any other way. He wore every hat this research demanded: methodologist when the Fuzzy Delphi numbers needed precision, clinician when the RCT logistics threatened to unravel, technologist when NExGEN needed architectural decisions, and motivator when I needed someone to remind me why any of this mattered. His patience is not the quiet, passive kind. What I value most is that he never made me feel like just a student. He treated my ideas as worth debating, my mistakes as worth understanding, and my ambitions as worth protecting. The late evening in his office were not just supervision. They were where I learned how research actually works: messy, iterative, and deeply human. Sir, this thesis carries your fingerprints on every page, and I carry your influence far beyond it. Thank you for believing in *Wawa* before *Wawa* believed in herself.

To my parents, who watched their daughter disappear into this thesis and never stopped believing she would come out the other side with something to show for it. Every prayer you whispered when I was too exhausted to whisper my own carried me further than you know. To my (hopefully) future husband, who spent countless hours behind the wheel driving me from Melaka to Seremban for every data collection and supervisory session. Thank you for choosing to show up, every single time, before you were even obligated to.

I extend my sincere gratitude to *Pragraph Digital Resources* for sponsoring the ChatGPT Business accounts powered for all study participants. Finally, to the 160 participants who trusted this research with their time, their data, and, in many ways, their health. You showed up for baseline assessments, wore accelerometers for seven days, logged dietary records, completed 24 weeks of commitment, and did so voluntarily. This thesis exists because you chose to be part of it. I acknowledge your essential role in this work with deep humility and respect.

TABLE OF CONTENTS

		Page
CONFIRMATION BY PANEL OF EXAMINERS		**ii**
AUTHOR'S DECLARATION		**iii**
ABSTRACT		**iv**
ACKNOWLEDGEMENT		**v**
TABLE OF CONTENTS		**vi**
LIST OF TABLES		**xi**
LIST OF FIGURES		**xiii**
LIST OF SYMBOLS		**xvi**
LIST OF ABBREVIATIONS		**xviii**
CHAPTER 1 INTRODUCTION		**1**
1.1	Research Background	1
1.2	Problem Statement	4
1.3	Research Question	6
1.4	Research Objectives	6
1.5	Research Hypothesis	7
1.6	Operational Definition	8
1.7	Limitation	11
1.8	Delimitation	14
1.9	Assumptions	16
1.10	Significance of Study	17
CHAPTER 2 LITERATURE REVIEW		**19**
2.1	Introduction	19
2.2	National and Global Obesity Crisis	19
2.3	Role of Nutritional and Exercise Interventions in Obesity Management	21
2.4	Digital Health Technologies in Addressing Obesity	26

2.5 Overview of Artificial Intelligence 28
2.6 Artificial Intelligence in Personalized Weight Management 30
2.7 Artificial Intelligence Chatbots as a Weight Management Tool 32
2.8 The Potential of ChatGPT in Personalized Obesity Treatment 34
2.8.1 Meal and Exercise Planner 35
2.8.2 Dietary and Physical Activity Advice and Guide 35
2.8.3 Nutrition and Fitness Material 37
2.9 Engineering for Health-Specific AI Applications 37
2.9.1 Prompt Engineering for Nutrition and Fitness Applications 37
2.9.2 Prompt Engineering in Clinical and Health Contexts 38
2.9.3 Personalized Prompt Systems for Health Interventions 39
2.10 Quality Assessment for AI-Generated Health Recommendations 40
2.10.1 Existing Quality Assessment Instruments 40
2.10.2 Limitations of Current AI-Driven Health Tools 42
2.10.3 Domain-Specific Criteria for AI-Generated Recommendations 43
2.11 Expert Consensus Validation Methods in Health Tools 44
2.11.1 Overview of Delphi Methods in Health Research 44
2.11.2 Fuzzy Delphi Method: Principles and Applications 45
2.11.3 Expert Consensus Methods in AI Health Tool Validation 46
2.12 Outcome Assessment in Weight Management Research 48
2.12.1 Body Composition Assessment 48
2.12.2 Behavioral Outcome Measurement 49
2.12.3 Cardiometabolic Risk Markers in Obesity 50
2.13 Previous Research on the AI Chatbots for Exercise and Nutrition 52
2.14 Summary 60

CHAPTER 3 RESEARCH METHODOLOGY 62
3.1 Introduction 62
3.2 Validation of a Personalized AI Prompt Generator (NExGEN-ChatGPT) for Obesity Management Using Fuzzy Delphi Method (Study 1) 63
3.2.1 Research Design 63
3.2.2 Overview of NExGEN Workflow 64
3.2.3 NExGEN-ChatGPT System Design and Process Flow 67

3.2.4 Refinement Process of NExGEN-ChatGPT Fitness and Nutrition Prompts 80
3.3 Personalized AI Prompt Generator and ChatGPT for Weight Loss: Randomized Controlled Trial in Adults with Overweight and Obesity (Study 2) 87
3.3.1 Research Design 87
3.3.2 Participant Recruitment 88
3.3.3 Sample Size Determination 90
3.3.4 Pilot Study 91
3.3.5 Randomization and Blinding 92
3.3.6 Research Procedure 93
3.3.7 Measurement 98
3.3.8 Statistical Analysis 100
3.3.9 Methodological 101

CHAPTER 4 RESULTS 103
4.1 Introduction 103
4.2 Results of Study 1: Validation of a Personalised AI Prompt Generator (NExGEN-ChatGPT) for Obesity Management Using Fuzzy Delphi Method 103
4.2.1 Baseline Descriptive Characteristics 104
4.2.2 Construct 1: Goal Settings 106
4.2.3 Construct 2: Demographics and Physical Characteristics 112
4.2.4 Construct 3: Health and Medical History 118
4.2.5 Construct 4: Lifestyle and Behaviour 122
4.2.6 Construct 5: Motivation and Social Support 129
4.2.7 Construct 6: Environmental Factors 131
4.2.8 Summary of Fuzzy Delphi Method Findings 133
4.3 Study 2: Personalised AI Prompt Generator and ChatGPT for Weight Loss: Randomised Controlled Trial in Adults with Overweight and Obesity 135
4.3.1 Baseline Descriptive Characteristics 135
4.3.2 Statistical Assumptions 139

4.3.3 Changes in Body Weight 146
4.3.4 Changes in Waist Circumference 147
4.3.5 Changes in Body Composition 149
4.3.6 Changes in Dietary Intake 153
4.3.7 Changes in Physical Activity 159
4.3.8 Changes in Metabolic Parameters 163
4.3.9 Changes in Blood Pressure 167
4.3.10 Summary of Study Randomised Controlled Trial Findings 171

CHAPTER 5 DISCUSSION AND CONCLUSION 173
5.1 Introduction 173
5.2 Expert Consensus of the NExGEN-ChatGPT Framework (Study 1) 174
5.2.1 Overall Expert Consensus and Frameworks Acceptance 174
5.2.2 Physical Activity Assessment, Goal Setting, and Stress Monitoring 175
5.2.3 Medication Interactions and Physiological Monitoring 176
5.2.4 Technology Dependence and Sleep-Exercise Interactions 177
5.2.5 NExGEN Versus Existing Health Technology Assessment Instruments 178
5.3 Effectiveness of the NExGEN-ChatGPT Intervention (Study 2) 179
5.3.1 Body Weight Reduction and Clinical Significance 179
5.3.2 Fat Mass Loss with Fat Free Mass Preservation 181
5.3.3 Energy Density as the Primary Mechanism 182
5.3.4 Physical Activity Measures 184
5.3.5 Cardiometabolic Markers 185
5.4 Personalized Versus Generic AI-Driven Weight Management 187
5.4.1 The Role of Input Quality and Prompt Personalization 187
5.4.2 User Engagement and System Utilization 188
5.5 Expert Validation to Clinical Effectiveness 189
5.6 Implications of the Study 190
5.6.1 Practical and Clinical Implications 190
5.6.2 Policy Implications 191
5.7 Contribution to Knowledge 192

5.8 Limitations of the Study 193
5.8.1 Study 1 Limitations 193
5.8.2 Study 2 Limitation 194
5.9 Recommendation for Future Research 196
5.10 Conclusion 197

REFERENCES **199**

APPENDICES **235**

AUTHOR'S PROFILE **305**

LIST OF TABLES

Tables	Title	Page
Table 2.1	Body Mass Index-based Weight Classification for Adults (>18 years old), Adopted from (Ministry of Health Malaysia, 2023)	20
Table 2.2	Overview of Randomized Controlled Trials on AI-based Chatbots for Recommendations in Physical Activity and Nutrition	58
Table 2.3	Overview of Non-randomized Studies on AI-based Chatbots for Recommendations in Physical Activity and Nutrition	59
Table 3.1	Likert Scale to Triangular Fuzzy Number Conversion	84
Table 4.1	Demographic Characteristics of the Expert Panel	105
Table 4.2	Fuzzy Delphi Method Analysis for 'Desired Weight Management Outcomes' Element	108
Table 4.3	Fuzzy Delphi Method Analysis for 'Target Weight Management Outcomes' Element	110
Table 4.4	Fuzzy Delphi Method Analysis for 'Timeline for Achieving Goals' Element	112
Table 4.5	Fuzzy Delphi Method Analysis for 'Age' Element	114
Table 4.6	Fuzzy Delphi Method Analysis for 'Body Composition' Element	116
Table 4.7	Fuzzy Delphi Method Analysis for 'Physiological Measurements' Element	118
Table 4.8	Fuzzy Delphi Method Analysis for 'Medical Conditions' Element	120
Table 4.9	Fuzzy Delphi Method Analysis for 'Medications' Element	122
Table 4.10	Fuzzy Delphi Method Analysis for 'Sleep Pattern' Element	124
Table 4.11	Fuzzy Delphi Method Analysis for 'Stress Level' Element	125
Table 4.12	Fuzzy Delphi Method Analysis for 'Physical Activity Habits' Element	127
Table 4.13	Fuzzy Delphi Method Analysis for 'Dietary Habits' Element	129

Table 4.14	Fuzzy Delphi Method Analysis for 'Motivations for Weight Management' Element	131
Table 4.15	Fuzzy Delphi Method Analysis for 'Environmental Factors' Element	133
Table 4.16	Summary of Fuzzy Delphi Method Findings Across All Constructs and Elements	134
Table 4.17	Baseline Demographic and Clinical Characteristics of Study Participants	138
Table 4.18	Metabolic Parameters Measured at Baseline and Across 12 Weeks of NEX Intervention and 3-Month Follow-Up in Intervention (NEX) and Control (CON) Groups (Intention-to-Treat Analysis)	167
Table 4.19	Blood Pressure Parameters Measured at Baseline and Across 12 Weeks of NEX Intervention and 3-Month Follow-Up in Intervention (NEX) and Control (CON) Groups (Intention-to-Treat Analysis)	171

LIST OF FIGURES

Figures	Title	Page
Figure 2.1	Conceptual Framework for Lifestyle Interventions for Management of Dietary- and Physical Inactivity-related Health Issues.	25
Figure 2.2	Multifaceted Applications of Artificial Intelligence in the Healthcare Domain.	30
Figure 3.1	Architecture Framework of NExGEN-ChatGPT System Integration	66
Figure 3.2	Starting interface of the NExGEN system	67
Figure 3.3	Goal Setting Interface of the NExGEN System	69
Figure 3.4	Snippet from Demographic and Physical Characteristics Interface of NExGEN	70
Figure 3.5	Snippet from Physiological Measurements Interface of NExGEN	70
Figure 3.6	Snippet from Women-Specific Survey Interface of NExGEN	71
Figure 3.7	Snippet from Medical History Section Interface of NExGEN	72
Figure 3.8	Snippet from Sleep Pattern Section Interface of NExGEN	72
Figure 3.9	Snippet from Stress Assessment Section Interface of NExGEN	73
Figure 3.10	Snippet from Physical Activity Section Interface of NExGEN	74
Figure 3.11	Snippet from Exercise Preferences Section Interface of NExGEN	75
Figure 3.12	Snippet from Dietary Intake Section Interface of NExGEN	75
Figure 3.13	Comprehensive Health Profile Prompt of NExGEN	76
Figure 3.14	Personalized Health Assessment and Recommendations Prompt of NExGEN	77
Figure 3.15	Plan Duration Selection Interface of NExGEN	77
Figure 3.16	Personalized Exercise Plan Prompt of NExGEN	78
Figure 3.17	Personalized Nutrition Plan Prompt of NExGEN	78
Figure 3.18	Personalized Groceries Planning Prompt of NExGEN	79
Figure 3.19	Personalized Meal Preparation Guide Prompt of NExGEN	79

Figure 3.20	Participant Recruitment Flowchart	90
Figure 3.21	Central and Noncentral Distributions from G*Power Output	91
Figure 3.22	Intervention Procedures for the NExGEN-ChatGPT Weight Management Trial.	97
Figure 4.1	CONSORT Flow Diagram of Participant Recruitment, Allocation, Follow-up, and Analysis in Study 2	136
Figure 4.2	Change (Δ) in Body Weight Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$); [a] Difference from CON ($P < 0.05$)	147
Figure 4.3	Change (Δ) in Waist Circumference Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$)	149
Figure 4.4	Change (Δ) in Fat Mass Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$)	151
Figure 4.5	Change (Δ) in Fat-Free Mass Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis).	152
Figure 4.6	Change (Δ) in Body Weight Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (○) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$); [a] Difference from CON ($P < 0.05$)	159
Figure 4.7	Change (Δ) in Physical Activity Level Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (○) (Group Means ± SE) (Intention-	

to-Treat Analysis). *Difference from Baseline (P < 0.05); [a] Difference from CON (P < 0.05) 162

LIST OF SYMBOLS

Symbols

α	Level of Significance / Cronbach's Alpha Reliability Coefficient / Alpha-Cut Value
α-cut	Defuzzified Score Derived from Fuzzy Membership Function
A_{max}	Maximum Defuzzified Value (Centroid Formula)
d	Threshold Value (Inter-Expert Vertex Distance)
η^2	Partial Eta-Squared (Effect Size)
m	Number of Multiple Imputations
m_1, m_2, m_3	Lower, Middle, And Upper Values of a Triangular Fuzzy Number
$\tilde{m}$, $\tilde{n}$	Fuzzy Numbers in Vertex Distance Calculation
n	Sample Size
P	Probability Value (P-Value)
r	Correlation Coefficient
SD (±)	Standard Deviation
≥	Greater Than or Equal To
≤	Less Than or Equal To
%	Percentage
cm	Centimetre
g	Gram
kcal	Kilocalorie

kg	Kilogramme
kg/m²	Kilogramme Per Metre Squared (Body Mass Index Unit)
mg/dL	Milligrammes Per Decilitre
min	Minute
mm	Millimetre
mmHg	Millimetres of Mercury (Blood Pressure Unit)
mmol/L	Millimoles Per Litre

LIST OF ABBREVIATIONS

Abbreviations

AI	Artificial Intelligence
ANOVA	Analysis of Variance
BIA	Bioelectrical Impedance Analysis
BMI	Body Mass Index
ChatGPT	Chat Generative Pre-trained Transformer
CHART	Chatbot Assessment Reporting Tool
CI	Confidence Interval
CON	Control Group
CVI	Content Validity Index
FDM	Fuzzy Delphi Method
GDPR	General Data Protection Regulation
HbA1c	Glycated Haemoglobin
HDL	High-Density Lipoprotein Cholesterol
IPAQ-LV	International Physical Activity Questionnaire – Long Version
IPAQ-SF	International Physical Activity Questionnaire – Short Form
ITT	Intention-to-Treat
LDL	Low-Density Lipoprotein Cholesterol
LLM	Large Language Model
MAP	Mean Arterial Pressure

MARS	Mobile Application Rating Scale
mBCA	Medical Body Composition Analyser
MET	Metabolic Equivalent of Task
MHO	Metabolically Healthy Obesity
MMRM	Mixed Model for Repeated Measures
MyFCD	Malaysian Food Composition Database
NEX	NExGEN Intervention Group
NExGEN	Nutrition and Exercise Prompt Generator
ORCHA	Organisation for the Review of Care and Health Applications
PP	Per-Protocol
PSQI	Pittsburgh Sleep Quality Index
QAMAI	Quality Assessment of Mobile Artificial Intelligence in Medicine
RCT	Randomised Controlled Trial
RRApp	Robust Randomisation Application
SDT	Self-Determination Theory
SPSS	Statistical Package for the Social Sciences
TFN	Triangular Fuzzy Number
UMIN	University Hospital Medical Information Network
WHO	World Health Organization

CHAPTER 1
INTRODUCTION

1.1 Research Background

Obesity has emerged as one of the most pressing global public health challenges of the twenty-first century. The most comprehensive population-level analysis to date, pooling data from 3,663 studies across 222 million participants in 200 countries, confirmed that over one billion people worldwide were living with obesity by 2022 (Phelps et al., 2024). This figure represented a fourfold increase in obesity rates among children and adolescents and a near threefold increase among adult males compared to 1990 levels (Phelps et al., 2024). The World Health Organization has estimated that in 2021, high body mass index (BMI) was responsible for approximately 3.7 million deaths from noncommunicable diseases including cardiovascular disease, type 2 diabetes, and several cancers (World Health Organization, 2025). If current trajectories persist without meaningful policy intervention, the World Obesity Federation projects that over half of the global adult population will be classified as overweight or obese by 2035 (Lobstein et al., 2023).

The economic consequences of unchecked obesity are substantial. Annual global costs attributable to overweight and obesity are projected to reach 4.32 trillion United States dollars by 2035, equivalent to approximately three percent of global gross domestic product (Lobstein et al., 2023). These costs encompass direct healthcare expenditure for treating obesity-related conditions and indirect losses through reduced workforce productivity, absenteeism, and premature mortality. Within Southeast Asia, Malaysia bears a disproportionate burden. The National Health and Morbidity Survey 2023 reported that 54.4 percent of Malaysian adults were overweight or obese, with obesity alone accounting for 21.8 percent (National Institutes of Health (NIH), 2023). This rate positions Malaysia among the highest in the Southeast Asian region (Tham et al., 2023). The convergence of rapid urbanization, nutrition transition toward energy-dense processed foods, and increasingly sedentary lifestyles has accelerated this trajectory in Malaysia, making the development of scalable, accessible, and effective weight management approaches an urgent national priority.

Conventional weight management interventions, typically involving face-to-face dietary counselling and structured exercise programmes, achieve clinically meaningful weight reductions of five to ten percent of baseline body weight over six to twelve months (Morgan-Bathke et al., 2023). These approaches, however, face significant barriers to long-term effectiveness. Systematic evidence demonstrates that weight regain begins as early as 36 weeks after intervention completion, with some individuals returning to baseline weight within one year (Machado et al., 2022). A meta-analysis of behavioural weight management programmes estimated that weight regain occurs at a rate of 0.12 to 0.32 kilograms per year relative to control conditions, with a between-group difference persisting for approximately five years before dissipating (Hartmann-Boyce et al., 2023). Traditional programmes are further constrained by labour-intensive delivery requirements, limited clinician availability, high per-patient costs, and insufficient capacity within primary care systems to provide the sustained, individualised support that obesity management demands (H. S. J. Chew, 2022).

Digital health technologies have shown promise in addressing the scalability limitations of conventional approaches. A systematic review and meta-analysis of eHealth interventions for weight management reported mean weight loss differences ranging from negative 1.07 to negative 3.10 kilograms compared to control conditions, with tailored interventions producing greater effects than generic ones (Sorgente et al., 2017). Digital self-monitoring interventions have demonstrated a pooled mean difference of approximately negative 2.87 kilograms in favour of the intervention group (Kankanhalli et al., 2019). These findings confirm the feasibility of technology-mediated weight loss yet also reveal a persistent gap between digital intervention outcomes and the effects achieved through intensive, personalised face-to-face programmes. The central challenge remains achieving deep personalisation at scale, a goal that early-generation digital tools, relying on rule-based algorithms and static content libraries, have not been able to fulfil. This limitation has prompted researchers to explore the capabilities of artificial intelligence (AI), and specifically large language models, as a mechanism for delivering dynamically personalised lifestyle guidance.

The rapid advancement of conversational AI has introduced new possibilities for health behaviour change. A systematic review of AI chatbot interventions targeting physical activity, diet, and weight management found that the majority of studies reported improvements in at least one lifestyle outcome, though definitive conclusions regarding overall efficacy could not yet be drawn due to heterogeneous study designs

and small sample sizes (Oh et al., 2021). A subsequent meta-analysis confirmed that chatbot interventions significantly improve physical activity and fruit and vegetable consumption, with AI-powered chatbots outperforming rule-based systems on several outcomes (Singh et al., 2023). Early evidence from AI-guided weight management programmes employing individualised data has shown favourable results, with one trial reporting approximately twelve percent weight loss in the AI-personalised group compared to seven percent in physician-guided controls (Michaelides et al., 2016). These findings collectively indicate that AI chatbot-based interventions hold considerable potential, yet critical questions about the optimal degree of personalisation, the quality of prompt engineering that drives large language model outputs, and the durability of effects remain unanswered.

Several interconnected research gaps underpin the need for further investigation. First, no validated mechanism previously existed for generating obesity-specific lifestyle planning prompts for large language model chatbots. Existing AI health tools have relied on generic conversational frameworks without structured, domain-validated prompt engineering that accounts for the clinical complexity of obesity management. Second, current quality assessment instruments for health technologies, such as the Quality Assessment of Mobile AI and the Mobile App Rating Scale, evaluate static application features rather than the dynamic, prompt-dependent outputs produced by large language models. These instruments fail to capture how input quality and prompt specificity determine the relevance, safety, and personalisation depth of AI-generated recommendations. Third, domain-specific quality criteria for AI-generated dietary and exercise recommendations tailored to populations with obesity are absent from the literature. No prior framework has incorporated obesity-specific safety considerations including medication interaction avoidance, comorbidity management, physiological monitoring thresholds, and culturally appropriate dietary guidance within a single validated instrument. Fourth, evidence comparing personalised versus generic AI-driven weight management remains limited, particularly for systems that integrate both exercise and dietary strategies through a personalised prompt generator combined with a large language model platform.

In response to these gaps, this thesis reports on the development, validation, and effectiveness evaluation of the NExGEN-ChatGPT personalised prompt generator system for weight management in adults with overweight and obesity. The research programme comprised two sequential studies. Study 1 employed the Fuzzy Delphi

Method (FDM) to evaluate and refine the NExGEN-ChatGPT framework through multidisciplinary expert consensus, validating six constructs, twelve elements, and 111 criteria for personalised AI-driven obesity management. Study 2 investigated the effectiveness of the validated NExGEN personalised exercise and dietary prompt generator integrated with Chat Generative Pre-trained Transformer (ChatGPT), compared to a structured manual ChatGPT control programme, through a 24-week randomised controlled trial measuring changes in body weight, body composition, dietary intake patterns, objectively measured physical activity, and cardiometabolic markers. Together, these two studies address the identified research gaps by establishing the first expert-validated prompt engineering instrument for obesity-specific AI interventions and providing the first clinical trial evidence for personalised versus generic prompt-driven weight management.

1.2 Problem Statement

As established in the preceding section, obesity prevalence in Malaysia has reached 54.4%, among the highest in Southeast Asia (National Institutes of Health (NIH), 2023; Tham et al., 2023), with escalating global economic and health consequences (World Health Organization, 2025). These statistics require urgent and scalable intervention strategies.

Despite the availability of conventional and digital weight management approaches, a critical deficiency persists: neither consistently delivers the depth of personalisation required for effective, sustained behaviour change. Standard clinical programmes are labour-intensive and lack scalability (Buratta et al., 2016; Kupila et al., 2023), while existing digital tools rely on standardised content unable to accommodate individual variability in health profiles, preferences, and behavioural contexts (Abrahams & Matusheski, 2020; Brug et al., 2003).

Large language model chatbots such as ChatGPT have demonstrated potential for personalised health behaviour support (Aggarwal et al., 2023; Oh et al., 2021; Singh et al., 2023), with AI-guided programmes producing superior outcomes over standard care (Graham et al., 2022; Mathioudakis et al., 2025). Despite these advances, critical gaps persist. No validated mechanism currently exists for generating obesity-specific lifestyle planning prompts for large language model chatbots. Existing AI health tools rely on generic conversational frameworks without structured, domain-validated

prompt engineering that accounts for the clinical complexity of obesity management (Chen et al., 2025; Schulhoff et al., 2025; White et al., 2023; Zaghir et al., 2024). Current quality assessment instruments for health technologies, such as the Quality Assessment of Medical AI scale and the Mobile App Rating Scale (Stoyanov et al., 2015; Vaira et al., 2024), evaluate static application features rather than the dynamic, prompt-dependent outputs generated by large language models. These frameworks fail to account for how input quality and prompt engineering determine the relevance, safety, and personalization depth of AI-generated health recommendations.

Domain-specific quality criteria for AI-generated dietary and exercise recommendations tailored to populations with overweight and obesity remain absent (Bayram & Arslan, 2025; Havers et al., 2025; Kaya et al., 2025). Existing evaluation frameworks do not incorporate obesity-specific considerations such as medication interaction avoidance (Al-Ashwal et al., 2023; Ong et al., 2025), physiological monitoring thresholds, or culturally appropriate dietary guidance (Telle-Hjellset et al., 2013). The multidimensional elements of obesity management, from physiological, psychological, behavioral, and environmental factors, requires important validation criteria that no prior instrument has established. Evidence comparing personalized versus generic AI-driven approaches for weight management is also limited. Despite growing interest in conversational AI chatbots for health coaching (Chew, 2022), no prior trial has combined a personalized prompt generator with an AI chatbot platform integrating both exercise and dietary strategies. Furthermore, most AI-based weight management interventions report only body weight changes without comprehensive evaluation of body composition, behavioral modifications, and cardiometabolic health markers. Few AI-based interventions incorporate validated bioelectrical impedance body composition analysis (Merrigan et al., 2022) or objectively measured physical activity assessment (Choi et al., 2012).

This thesis addresses these gaps through the development, validation, and effectiveness evaluation of the NExGEN-ChatGPT personalized prompt generator system for weight management in adults with overweight and obesity. The first study validates the NExGEN-ChatGPT framework through multidisciplinary expert consensus using the Fuzzy Delphi Method, encompassing 6 constructs, 12 elements, and 111 criteria for personalized AI-driven obesity management (Dan et al., 2025). The second study investigates the effectiveness of NExGEN integrated with ChatGPT compared to a structured manual ChatGPT control program through a randomized

controlled trial, measuring changes in body weight, body composition, dietary intake patterns, objectively measured physical activity, and cardiometabolic markers over a 24-week period. The integration of expert-validated prompt engineering with clinical effectiveness testing represents a systematic approach to establishing evidence-based AI-driven weight management. The expected contribution is a validated, scalable AI intervention framework that bridges the gap between generic digital health tools and the personalized, comprehensive approach required for effective obesity management.

1.3 Research Question

The following research questions guide the investigations in this thesis:

a) To what extent does the NExGEN-ChatGPT personalized prompt generator framework achieve expert consensus when evaluated using the Fuzzy Delphi Method? **(Study 1)**.
b) Does the NExGEN personalised prompt generator integrated with ChatGPT produce superior weight management outcomes compared to structured manual ChatGPT guidance in adults with overweight and obesity **(Study 2)**.

1.4 Research Objectives

This research develops, validates, and evaluates the NExGEN-ChatGPT personalized AI prompt generator system for weight management in adults with overweight and obesity. Across two phases of studies, Study 1 addressed the development and expert validation of the NExGEN-ChatGPT framework, while Study 2 evaluated its clinical effectiveness through a randomized controlled trial.

c) To develop a comprehensive personalized AI prompt generator framework (NExGEN-ChatGPT) for generating individualized exercise and dietary prompts through a chatbot platform for adults with overweight and obesity **(Study 1)**.
d) To evaluate and refine the quality of the NExGEN-ChatGPT framework by multidisciplinary expert consensus using the Fuzzy Delphi Method **(Study 1)**.
e) To determine the effectiveness of the validated NExGEN with ChatGPT

(NExGEN group) compared to structured manual ChatGPT guidance (control group) on body weight over a 12-week intervention period followed by a 3-month follow-up in overweight and obese adults **(Study 2)**.

f) To compare changes in waist circumference between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

g) To compare changes in body composition, specifically fat mass and fat-free mass, between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

h) To assess changes in dietary intake patterns between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

i) To evaluate changes in objectively measured and self-reported physical activity levels between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

j) To assess changes in metabolic parameters, including glycated haemoglobin (HbA1c), total cholesterol, LDL cholesterol, High-Density Lipoprotein (HDL) Cholesterol, and triglycerides, between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

k) To assess changes in blood pressure parameters, including systolic blood pressure, diastolic blood pressure, and mean arterial pressure, between the NExGEN group and the control group at post-intervention and 3-month follow-up in overweight and obese adults **(Study 2)**.

1.5 Research Hypothesis

Given that the experimental design applies only to Study 2, the list below contains the research hypotheses only for this phase of the study. The following hypotheses were formulated:

a) **H_a (Body Weight – Primary Outcome; Study 2)**: Participants receiving the NExGEN personalised prompt generator integrated with ChatGPT will achieve significantly greater reductions in body weight compared to participants

receiving structured manual ChatGPT guidance at both 12-week post-intervention and 3-month follow-up assessments.

b) **H_a (Waist Circumference; Study 2)**: Participants in the NExGEN group will demonstrate significantly greater reductions in waist circumference compared to the control group at both assessment timepoints.

c) **H_a** (Body Composition; Study 2): Participants in the NExGEN group will demonstrate greater reductions in fat mass with preservation of fat-free mass, compared to the control group at both assessment timepoints.

d) **H_a (Dietary Intake; Study 2)**: Participants in the NExGEN group will exhibit significantly greater improvements in dietary intake patterns, including reductions in dietary energy density, energy intake, and fat intake, with maintenance of protein intake and improvements in fibre and carbohydrate intake, compared to the control group at both assessment timepoints.

e) **H_a (Physical Activity; Study 2)**: Participants in the NExGEN group will demonstrate significantly greater increases in both objectively measured (accelerometer) and self-reported (IPAQ-LV) physical activity levels compared to the control group at both assessment timepoints.

f) **H_a (Metabolic Parameters; Study 2)**: Participants in the NExGEN group will show significantly greater improvements in metabolic parameters, including glycated haemoglobin (HbA1c), total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides, compared to the control group at both assessment timepoints.

g) **H_a (Blood Pressure; Study 2)**: Participants in the NExGEN group will show significantly greater improvements in blood pressure parameters, including systolic blood pressure, diastolic blood pressure, and mean arterial pressure, compared to the control group at both assessment timepoints.

1.6 Operational Definition

The operationalisation of each term within the study methodology is specified as follows:

a) **NExGEN-ChatGPT Personalized Prompt Generator:** The NExGEN-ChatGPT Personalized Prompt Generator is a validated, web-based system that

captures various individual-level variables across six constructs and twelve elements. These variables then translate into six sequential structured prompts for delivery to a large language model chatbot (ChatGPT). The six prompts address, respectively: a comprehensive personalised profile, behavioural recommendations, a weekly exercise plan, a weekly dietary plan, a grocery shopping list, and a meal preparation guide. This research operationalises the NExGEN system as the object of expert validation in Study 1 and the independent variable (intervention tool) in Study 2.

b) **Large Language Model (LLM):** A large language model is a deep neural network, typically built on the transformer architecture, containing billions of parameters pre-trained on large amount of text to linguistic patterns, and perform diverse natural language processing tasks. In this thesis, the term refers specifically to OpenAI's ChatGPT (ChatGPT-o3), which served as the generative engine that processed NExGEN- prompts to deliver tailored weight management recommendations.

c) **Prompt Engineering:** Prompt engineering is the systematic process of designing and refining natural language inputs to guide generative AI model outputs toward relevant, accurate, and contextually appropriate responses, without modifying core model parameters. In this research, prompt engineering constitutes the core operational mechanism of the NExGEN system: the 111-variable assessment instrument captures individual data that the system translates into structured prompts optimised for generating personalised exercise, dietary, and behavioural recommendations from ChatGPT.

d) **Fuzzy Delphi Method (FDM):** The Fuzzy Delphi Method is an extension of the traditional Delphi technique that integrates fuzzy set theory to aggregate expert opinions while accommodating the inherent vagueness of human judgement. The method converts Likert-scale responses into triangular fuzzy numbers and applies defuzzification procedures to quantify consensus. In Study 1, multidisciplinary experts evaluated the NExGEN framework's 111 criteria against three acceptance thresholds (see *Methodology* chapter for full details). Items meeting all three thresholds were retained.

e) **Expert Consensus:** Expert consensus, as operationalised in this thesis, denotes formal agreement among a purposively selected panel of 21 multidisciplinary specialists, each with a minimum of ten years of professional experience.

Consensus was defined as the simultaneous satisfaction of three Fuzzy Delphi Method thresholds.

f) **Overweight and Obesity:** The World Health Organization defines overweight as a BMI of 25.0 kg/m² or above and obesity as a BMI of 30.0 kg/m² or above, characterised as chronic conditions of excessive fat accumulation posing health risks. The WHO Expert Consultation recommended additional action points for Asian populations, with 23.0 kg/m² indicating increased risk and 27.5 kg/m² indicating high risk. This thesis adopted the Asian-adjusted range of 27.5 to 34.9 kg/m² as the inclusion criterion for Study 2.

g) **Body Mass Index (BMI):** Body mass index is a standardised anthropometric measure calculated by dividing body weight in kilograms by the square of height in metres (kg/m²), widely used for population-level weight status classification. In this research, BMI was measured at baseline, 12 weeks, and 24 weeks using the Seca mBCA 515 system under standardised fasting conditions and served as the screening variable for participant eligibility (27.5–34.9 kg/m²).

h) **Weight Management:** Weight management, as defined in this thesis, encompasses integrated evidence-based dietary, physical activity, and behavioural strategies aimed at achieving and sustaining clinically meaningful body weight reductions. This research adopted a target of 5% to 10% baseline weight reduction over 12 weeks, a threshold sufficient to confer clinically significant cardiometabolic improvements. Both groups received programming aligned with WHO physical activity guidelines (150 minutes moderate-intensity aerobic activity plus two resistance training sessions per week).

i) **Body Composition:** Body composition refers to the proportional distribution of distinct tissue compartments within the human body, principally fat mass, fat-free mass, and total body water. This thesis operationalised body composition as a secondary outcome measured using the Seca mBCA 515 medical bioelectrical impedance analysis system under standardised conditions.

j) **Dietary Energy Density:** Dietary energy density is the amount of energy per unit weight of food consumed, expressed as kcal/g, determined primarily by water content (0 kcal/g) and fat content (9 kcal/g). In this thesis, dietary energy density was calculated from seven-day dietary records analysed using Nutritionist Pro™ v7.9 with referenced to Malaysian Food Composition Database (MyFCD).

k) **Cardiometabolic Markers:** Cardiometabolic markers are measurable biochemical and physiological indicators of cardiovascular and metabolic health status that predict future risk for cardiovascular disease and type 2 diabetes, commonly altered in the presence of excess adiposity. In this thesis, specific markers assessed at baseline, 12 weeks, and 24 weeks comprised: glycated haemoglobin (HbA1c) and fasting glucose; blood cholesterols, and blood pressure.

l) **Bioelectrical Impedance Analysis:** Bioelectrical impedance analysis is a non-invasive method for estimating body composition by measuring the impedance of body tissues to a small alternating electrical current, based on the principle that fat-free tissues with high water content conduct current more readily than adipose tissue. This thesis employed the Seca mBCA 515. All measurements followed a standardised protocol.

m) **Objectively Measured Physical Activity:** Objectively measured physical activity refers to the quantification of human movement using wearable sensor technology, as distinct from self-reported measures susceptible to recall bias and social desirability effects. This thesis employed ActiGraph GT3X+ triaxial accelerometers, requiring a minimum of five valid days. Self-reported physical activity was assessed in parallel using IPAQ-LV.

n) **Personalization (*in AI-Driven Health Interventions*):** Personalization in AI-driven health interventions refers to the systematic tailoring of intervention content and recommendations to individual user characteristics through algorithmic processing of individual-level data, as contrasted with generic, one-size-fits-all digital health content. The NExGEN system operationalised personalisation by synthesising 111 assessment variables. The control group received identical content domains through structured manual prompts without individual adaptation.

1.7 Limitation

This section outlines the methodological and practical constraints that may have influenced the findings, discussed to support accurate interpretation and guide future research.

a) **Geographic and cultural concentration of the expert panel:** All 21 experts were based in Malaysia. Despite diverse professional backgrounds spanning nutrition, exercise science, medicine, psychology, and AI, the absence of international representation may limit the cross-cultural transferability of the validated criteria. Obesity management practices and health system structures vary across regions, and consensus achieved within a Malaysian context may not generalize to other settings.
b) **Theoretical validation without end-user testing:** Study 1 relied solely on expert professional judgment, with no end-user trials or usability assessments conducted during this phase. Expert consensus on appropriateness does not guarantee practical effectiveness. Although Study 2 addressed this gap through a randomized controlled trial, the validation phase remains limited to theoretical endorsement.
c) **Single artificial intelligence model dependency**: The framework was validated exclusively against ChatGPT-o3 model. Because large language models differ in training data, response patterns, and safety mechanisms, the validated criteria may perform differently on alternative models such as Gemini, Llama, or Claude. No cross-platform comparison was conducted, limiting claims of universal applicability.
d) **Absence of socioeconomic and ethnic diversity assessment**: The validation did not assess whether the 111 criteria adequately address populations across different socioeconomic strata or ethnic backgrounds. Strategies suited to urban, digitally literate populations may not translate to communities with limited technology access or distinct cultural attitudes toward weight management.
e) **Single geographic setting**: The trial was conducted exclusively in Greater Kuala Lumpur (Klang Valley), Malaysia, with assessments at the Exercise Science Laboratory, Universiti Teknologi MARA. Participants were urban adults with smartphone access and digital literacy. This setting excludes rural populations, communities with limited digital infrastructure, and individuals outside the Malaysian context, leaving generalizability to other settings untested.
f) **Incomplete participant blinding**: Complete participant blinding was not feasible given the visibly different intervention modalities. NExGEN participants interacted with a personalized system referencing their individual

characteristics, which may have created expectation bias and heightened motivation relative to the control group. Assessor and analyst blinding mitigated measurement bias but could not eliminate the influence of participant expectations.

g) **Self-reported dietary data**: Dietary intake was assessed via seven-day records analyzed through Nutritionist Pro™ v7.9 (Axxya Systems) with referenced to Malaysian Food Composition Database (MyFCD) (Tay et al., 2023). Despite requiring five valid recording days including one weekend day, dietary self-report remains susceptible to underreporting and social desirability bias. This standard approach was applied consistently across groups but introduces uncertainty in dietary intake precision.

h) **External engagement-enhancing conditions**: Two study-wide features may have elevated participant motivation and engagement beyond what routine clinical practice would yield. First, both groups received weekly WhatsApp reminders from the research team, introducing a human contact element that may have independently influenced adherence. While contact was equivalent across groups, it complicates attribution of within-group changes solely to the ChatGPT interventions. Second, all participants received complimentary six-month ChatGPT Business subscriptions (valued at approximately RM830; sponsored by Pragraph Digital Resources, SA0614953-A), which may have further enhanced engagement compared to a real-world scenario in which participants fund their own access.

i) **Suboptimal intervention utilization**: Variable engagement with allocated ChatGPT systems may have attenuated intervention effectiveness. Usage analysis indicated higher NExGEN group engagement, but individual differences in technology comfort, time availability, and motivation produced heterogeneous engagement patterns across both groups.

j) **Limited duration for cardiometabolic detection**: The 24-week duration may have been insufficient to capture full cardiometabolic benefits. Participants presented with metabolically healthy profiles at baseline, and cardiometabolic improvements in such populations typically require longer observation or greater weight loss. Fewer than half of NExGEN participants surpassed the five percent weight loss threshold considered necessary for clinically meaningful cardiometabolic effects.

k) **Measurement Constraints of dependent variables**: The primary dependent variable, body weight, and the secondary dependent variable, waist circumference, were measured under standardised fasting conditions; however, inherent day-to-day biological variability in hydration status and gastrointestinal content introduces measurement fluctuation that standardised protocols cannot fully eliminate. Body composition outcomes (fat mass and fat-free mass) were assessed using the Seca mBCA 515 bioelectrical impedance analysis system, which, while validated for clinical use, carries greater measurement error than the reference standard of dual-energy X-ray absorptiometry. These constraints should be considered when interpreting magnitude of change across dependent variables.

l) **Dropout attrition**: Thirty-three participants dropped out during the trial. Both per-protocol and intention-to-treat analyses (with multiple imputation under missing-at-random) were conducted. The dropout rate fell within the anticipated 20 percent buffer but still reduces the effective per-protocol sample and may limit effect estimate precision.

1.8 Delimitation

This section defines the boundaries within which this thesis was conducted. Delimitations are intentional and justified choices made by the researcher that restrict the scope of the study to ensure methodological rigour and feasibility. Each delimitation is justified below.

a) **Body mass index range based on Asian classification**: Eligibility was restricted to adults with BMI 27.5–34.9 kg/m², based on the Asian BMI classification where obesity-related risks occur at lower thresholds than in Western populations. The lower cut-off (27.5 kg/m²) corresponds to Asian obesity Class I, while the upper boundary excluded morbidly obese individuals requiring medical supervision beyond the scope of a lifestyle-based AI intervention.

b) **Exclusion of specific population groups**: Excluded groups included morbidly obese individuals (BMI ≥ 35 kg/m²) requiring closer medical monitoring, underweight and normal weight individuals outside the target population,

pregnant or lactating women whose nutritional requirements differ from weight loss protocols, individuals with severe psychiatric conditions, participants using weight-loss medications to avoid pharmacological confounding, and minors under 18 due to distinct pediatric considerations.

c) **Geographic focus on Greater Kuala Lumpur (Klang Valley), Malaysia,**: The trial was situated in the city of Greater Kuala Lumpur (Klang Valley), Malaysia, to increase recruitment likelihood of participant digital literacy and connectivity, maximize recruitment feasibility, and enable standardized clinical assessments at a single laboratory facility.

d) **ChatGPT as the sole large language model platform**: ChatGPT was selected as the exclusive AI platform based on popularity, its conversational capabilities, multi-turn dialogue capacity, and commercial accessibility. The objective was to evaluate personalized versus generic prompt engineering within a single controlled environment, not to benchmark across platforms.

e) **Intervention duration and follow-up period**: The 12-week active intervention followed by 12-week maintenance was selected based on established conventions in behavioral weight management research. The 24-week total duration balanced outcome detection against participant burden. Long-term maintenance beyond six months was outside the scope of this thesis.

f) **Adult age range of 19 to 59 years**: A pilot study conducted prior to the main study indicated that restricting eligibility to the middle-aged adult category (45–59 years), as classified by the World Health Organization (WHO), would not yield the required sample size. Consequently, the age range was expanded to include all adult categories from 19 to 59 years, excluding older adults (60 years and above), in accordance with the WHO age classification. This expanded range excluded paediatric populations and reduced confounding from age-related comorbidities prevalent among older adults Azmi et al. (2009), while capturing the working-age adult population most likely to engage with technology-mediated weight management.

g) **Malaysian expert panel for framework validation**: The panel was composed of Malaysian-based experts to ensure relevance to the local healthcare context and cultural practices where the NExGEN framework would first be deployed. This prioritized contextual validity for the initial setting, with cross-cultural validation expected in subsequent research.

1.9 Assumptions

This thesis operates on a set of assumptions accepted as true within the scope of the investigation. The assumptions are organized across four domains as follows:

a) **Honest and accurate participant self-reporting**: The NExGEN personalization depends on accurate responses to the 111-item assessment covering demographics, medical history, lifestyle, motivation, and environmental factors. Inaccurate self-reporting due to social desirability, poor recall, or misunderstanding would compromise prompt quality. However, response accuracy was not independently verified.

b) **Adherent use of allocated ChatGPT interventions**: Participants were assumed to use their assigned ChatGPT prompts as instructed throughout the 12-week period. In a free-living digital intervention, full protocol adherence could not be controlled, despite weekly WhatsApp reminders.

c) **Consistency and accuracy of ChatGPT-generated responses**. This research assumed ChatGPT (Model o3) produced clinically appropriate responses to NExGEN-generated prompts (as evidenced from Study 1). The structured, expert-validated prompts were designed to constrain responses, and the fixed model version (ChatGPT-o3) reduced mid-study variability.

d) **Accuracy of self-reported dietary records**: Seven-day dietary records assumed accurate and complete food reporting. Standard quality controls were applied (five valid days including one weekend day, validated nutrient analysis software), but the assumption that records reflect true consumption could not be verified through biomarker or other objective methods due to the limited resource.

e) **Accelerometer wear compliance as a valid indicator of physical activity**: Physical activity measurement relied on ActiGraph GT3X+ accelerometers worn for seven days (minimum five valid days, ten or more hours wear time). The research assumed consistent wear during waking hours, though accelerometers underestimate non-ambulatory activities such as swimming, cycling, and resistance exercises.

f) **Professional judgment of Fuzzy Delphi Method experts**: The Fuzzy Delphi validation assumed all 21 experts (each with over a decade of experience)

provided ratings based on authentic professional judgment. Response anonymity minimized conformity pressure and fuzzy logic accommodated variation, but the cognitive rigor applied by each panelist could not be independently verified.

1.10 Significance of Study

This research advances theoretical understanding of the relationship between prompt input quality and AI-generated health output quality, a dimension that has remained largely unexamined in digital health literature. By establishing that structured, domain-validated prompt engineering is a necessary precursor to clinically safe AI-generated obesity recommendations, the research contributes to emerging theory on how large language models can be responsibly integrated into evidence-based health behaviour change frameworks. The NExGEN framework extends personalisation theory in digital health by operationalising 111 individual-level variables as determinants of AI output specificity, moving the field beyond generic content delivery toward adaptive, data-driven intervention design.

This research develops and validates the first expert-consensus prompt engineering framework for AI-driven obesity management, comprising 111 criteria across six constructs and twelve elements. The application of the Fuzzy Delphi Method to this validation provides a more nuanced consensus mechanism than traditional approaches, accommodating inherent uncertainty in expert judgment through triangular fuzzy number analysis. Critically, the validated framework shifts the evaluation paradigm from static application feature assessment to dynamic prompt-output quality assessment, offering researchers, clinicians, and AI developers an evidence-based benchmark applicable to any domain where AI-generated health content requires quality assurance. The sequential progression from expert validation to randomised controlled effectiveness testing further establishes a replicable development-validation-effectiveness pipeline for AI health intervention research, a model that most deployed AI health tools have not followed.

The scalability and cultural relevance of this research carry direct implications for health policy. The NExGEN system delivers personalised dietary plans incorporating locally available Malaysian ingredients, culturally appropriate cooking methods, and region-specific grocery strategies, addressing the recognised

underrepresentation of Southeast Asian populations in digital health intervention research. As Malaysia's ratio of dietitians and exercise professionals to its population with obesity remains insufficient for intensive lifestyle therapy at scale, this fully automated system offers a cost-effective adjunct to conventional management strategies aligned with national health priorities. The demonstrated effectiveness provides evidence that AI-driven personalised interventions could be integrated into primary care and community health frameworks, particularly in resource-constrained settings where access to specialised obesity services remains limited. This methodological template is transferable to other health domains including diabetes management, cardiovascular rehabilitation, and mental health where personalised AI-driven interventions require systematic development and evaluation.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

The aim of the chapter is to identify the primary challenges and difficulties inherent in implementing artificial intelligence (AI) technologies for managing obesity. It focuses specifically on the challenges in designing AI-driven dietary and exercise plans for obese individuals. The first part reviews the literature on the global and national obesity epidemic and the need for a significant rethink of the current approach, including innovative solutions for weight management strategies. Subsequently, this chapter examines existing scholarly works concerning the efficacy of digital health technologies in obesity management alongside the emerging potential of AI-driven solutions in the field of obesity and weight control. Finally, this chapter reviews existing research on the efficacy of AI-driven chatbots in diet and exercise management for weight control.

2.2 National and Global Obesity Crisis

Obesity, defined as a disease marked by excessive fat accumulation that poses significant health risks, is a global challenge with escalating prevalence. Notably, Malaysia consistently records among the highest obesity prevalence in Southeast Asia, with the National Health and Morbidity Survey 2023 reporting that 54.4% of Malaysian adults were classified as overweight or obese, compared to 44.5% in 2011, representing a near 10% increase over twelve years (National Institutes of Health (NIH), 2023). This escalating trend is accentuated by projections indicating that obesity prevalence will reach 41% by 2035 (Lobstein et al., 2023). This alarming trend is particularly critical in the Asian context, where obesity-related health risks manifest at lower BMI levels compared to non-Asian populations. Consequently, Asian-specific BMI classifications adopt lower cut-off points, a practice reflected in the latest Malaysian Clinical Practice Guidelines for the Management of Obesity – Second Edition 2023 (Table 2.1), which delineate obesity into Obese class 1 (BMI 27.5–32.4), Obese class 2 (BMI 32.5–37.4) and Obese class 3 (BMI ≥37.5) (Table 2.1) (Ministry of Health Malaysia, 2023)

However, concerns remain regarding the optimal metrics for measuring obesity, where several studies have questioned the use of body mass index (BMI) as a standalone measure, arguing that it fails to distinguish between muscle and fat mass and does not adequately account for fat distribution. The gravity of this issue extends beyond regional health concerns, as obesity significantly contributes to the global burden of non-communicable diseases like cardiovascular diseases, cancer, and diabetes mellitus, which account for over 70% of early deaths worldwide. As a major risk factor for non-communicable diseases, obesity is linked to a reduction in life expectancy by an estimated 5–20 years, contingent on the severity and accompanying comorbidities. This confluence of factors underscores the urgent need to address the obesity epidemic, particularly in contexts like Malaysia, where rapid escalation poses profound public health challenges.

Table 2.1
Body Mass Index-based Weight Classification for Adults (>18 years old), Adopted from (Ministry of Health Malaysia, 2023)

Classification	**BMI (kg/m^2)**	**Risk of comorbidities**
Underweight	<18.5	Low, but with increased risk of other clinical problems
Normal	18.5 – 22.9	Optimal
Pre-obese (Overweight)	23.0 – 27.4	Increased
Obese I	27.5 – 32.4	High
Obese II	32.5 – 37.4	Very high
Obese III	≥37.5	Extremely High

The etiology of obesity, a multifaceted health crisis, remains a topic of considerable debate. At the heart of current health guidelines is the notion that obesity arises from an energy imbalance, where calorie intake surpasses expenditure. This epidemic is significantly propelled by dietary changes, notably the increased consumption of highly palatable, energy-dense foods. Beyond diet, a myriad of social, economic, and environmental factors intersects, influencing individuals' capacity to maintain energy balance. Illustratively, a 13-year longitudinal study involving 3,000 youths revealed that high fast-food consumption correlates with increased weight gain

and waist size, alongside escalated risks of metabolic syndrome and elevated triglycerides (Duffey et al., 2007). This complex interplay of factors extends to familial and psychological domains, where both genetic predisposition and lifestyle choices like poor diet or inadequate exercise contribute to obesity risk. Notably, children with one or both obese parents face a significantly higher likelihood of adult obesity, a risk that escalates exponentially with two obese parents (Lake et al., 1997). Further emphasizing the familial influence, a study of 260 children highlighted a strong association between family histories of cardiometabolic diseases and the severity of childhood obesity. This intricate web of dietary, genetic, and socio-economic elements underscores the urgency and complexity in addressing obesity (Corica et al., 2018).

In the realm of obesity management, the absence of specific pharmacological treatments presents a significant challenge. This gap underscores the pivotal role of lifestyle modification, which stands as the cornerstone of combating obesity. Such modification encompasses a holistic strategy, including a balanced combination of diet, physical activity, and behavior therapy (Mohseni et al., 2023). This multifaceted approach advocates for individuals with obesity to target a minimum of 10% body weight loss. The adoption of portion-controlled diets has been instrumental in securing significant short-term weight loss (Nicklas et al., 2012). However, the journey towards long-term weight control hinges on sustained high levels of physical activity and continuous interaction with practitioners. Notably, lifestyle modifications have been linked to dramatic reductions in body weight, subsequently cascading into a substantial decrease in cardiovascular risks. Furthermore, interventions designed to foster behavioral changes, such as health promotion, nutrition education, and incentives for healthy living, are crucial. These interventions, alongside measures that tackle the root causes of obesity, are anticipated to exert a profound influence on mitigating the obesity crisis.

2.3 Role of Nutritional and Exercise Interventions in Obesity Management

In addressing the complex challenge of overweight and obesity, it is noteworthy that, aside from a minority of cases attributable to pharmacological, metabolic, or genetic etiologies, standard weight loss interventions predominantly prescribe diet and exercise plans. These efforts, fundamentally centered on diet and physical activity, incorporate a variety of dietary approaches, including caloric restriction, the

Mediterranean diet, inclusive dairy diets, and alternate-day fasting, complemented by an exercise regimen that may include high or low-intensity activities, resistance training, aerobic exercise, and endurance training (Widmer et al., 2015). However, a significant obstacle in these interventions is the lack of adherence to these restrictive lifestyle plans, often stemming from deficiencies in motivation and self-control (Buratta et al., 2016). To address these challenges, health coaching has emerged as a beneficial tool, enhancing the initiation and sustainability of weight loss efforts. Despite its benefits, such programs are criticized for their labor intensity and resource inefficiency. Alternative strategies like brief counseling techniques, including motivational interviewing, have shown promise in improving lifestyle behaviors, yet they display no significant superiority in effectiveness compared to other methods, such as health coaching (Markert et al., 2021). Intriguingly, these findings remain consistent across various age groups and modes of delivery, highlighting that while current interventions are effective, they are markedly hindered by their resource-intensive nature.

An unhealthy lifestyle, including low levels of physical activity and increased intake of calorie-rich foods, is a key factor in accelerating obesity (Faccioli et al., 2023). Implementation of appropriate nutrition and physical activity in individuals who are overweight and/or obese may improve the health of patients, especially those with adverse health consequences due to increased fat mass and metabolic consequences (Babak et al., 2019; Bharti et al., 2021). Nutrition education and/or appropriate physical activity are important components in the management of individuals with obesity (Bharti et al., 2021). Calorie restriction is essential to inducing clinically meaningful weight loss. The Obesity Guidelines recommend the consumption of a diet designed to achieve a deficit of 500–750 kcal·day^{-1}, with a resulting mean loss of 0.5–0.75 kg per week (Kim et al., 2023). Accordingly, women often are prescribed 1200–1500 kcal·day^{-1} and men 1500–1800 kcal·day^{-1}, with higher targets within the ranges for individuals with higher body weights (or physical activity levels) (Kline et al., 2018). Another option is to prescribe 1200–1500 kcal·day^{-1} for individuals who weigh <113 kg and 1500–1800 kcal·day^{-1} for those >113kg (Hajduk et al., 2001).

Reducing portion sizes, as well as excess sugar and fat, is a convenient way to achieve desired calorie deficits (Parry et al., 2018). Studies have also shown that interventions based on nutritional education can lead to weight loss, modifications in eating habits, and reduction of fat intake in the adult population (López-Hernández et al., 2020). In the context of personalized nutrition, a randomized trial assigned adults

with obesity to a 10-week dietary intervention based on different hypocaloric diets, demonstrating the potential of personalized gene-based dietary recommendations for weight management (Holzapfel & Drabsch, 2019).

Lifestyle modification programs typically prescribe 150–180 minutes per week of moderately vigorous aerobic activity, such as brisk walking or cycling (Wadden et al., 2012). Regular aerobic activity is associated with numerous benefits, including improvements in physical (e.g., lipids, blood pressure) and mental health (e.g., anxiety, depression)(Powell et al., 2019). Physical activity also is associated with improved fitness, which may attenuate the risk of mortality associated with obesity(Barry et al., 2014). Persons who report a lack of time to exercise are encouraged to engage in multiple brief bouts (e.g., 10 minutes) of activity throughout the day and to increase their lifestyle activity (e.g., by using stairs rather than elevators) (Jakicic & Rogers, 2014).

A new generation of standing desks, including those with treadmills, potentially provides a convenient means of increasing daily energy expenditure. Decreasing screen time and other sedentary behaviors is another important consideration for potentially increasing physical activity (and weight loss) (Otten et al., 2009). The addition of twice-weekly strength training to an activity program may protect against the loss of muscle that accompanies aging (Powell et al., 2019). Physical activity alone produces minimal short-term weight loss if not combined with calorie restriction (Jakicic & Davis, 2011), for example, found that groups of non-dieting individuals who were instructed to engage in either 150 or 300 minutes of weekly activity both lost less than 2% of initial weight at 6 months. Thus, in the short term, participants who use lifestyle modification are encouraged to exercise primarily for its important health benefits.

Nutrition knowledge refers to the comprehension of fundamental principles, concepts, and guidelines related to dietary choices and their impact on health. Individuals with adequate nutrition knowledge demonstrate the capacity to critically assess nutrition information through their understanding of dietary principles, nutritional guidelines, and appropriate portion sizes (Guttersrud et al., 2014). According to a systematic review by Spronk et al. (2014), nutrition knowledge and dietary intake have a weak but significant positive association. This finding suggests that individuals who possess a greater understanding of nutrition tend to make more informed and healthier dietary choices. A subsequent inquiry conducted by Yahia et al. (2016) indicated that nutrition knowledge appears to exert a more pronounced impact on

promoting healthier dietary habits among individuals with lower educational attainment despite the fact that these individuals possess comparatively lower levels of nutrition knowledge than their counterparts. From a public health standpoint, it highlights the importance of nutrition knowledge in shaping our dietary behaviors Worsley (2002), as well as access to credible nutrition knowledge (Jenkins et al., 2020).

Figure 2.1 illustrates a conceptual framework for lifestyle interventions targeting the prevention and management of main diet- and physical inactivity-related non-communicable diseases such as obesity, diabetes, and cardiovascular disease. Central to this framework is the implementation of lifestyle interventions that encompass a healthy diet and increased physical activity. These interventions are depicted as the primary modifiable determinants poised to impact the management of these conditions. A healthy diet, characterized by a balanced intake of fruits, vegetables, lean proteins, and whole grains, is essential for the prevention and management of obesity, diabetes, and cardiovascular disease (Mabry-Hernandez & Ojeda, 2019). Concurrently, regular physical activity, which includes both aerobic and strength training exercises, plays a critical role in maintaining optimal body weight and improving metabolic functions.

Accordingly, these interventions can lead to measurable improvements in key health outcomes such as improvement in BMI, glycated hemoglobin (HbA1c), waist-hip ratio, fasting blood sugar, lipid profiles, and blood pressure, alongside behavioral outcomes, including weight management, physical activity levels, and nutrition habits and behaviors (Silva et al., 2014). The anticipated result of implementing these lifestyle interventions is an improvement in these parameters, which are critical indicators of disease progression and risk reduction in individuals with diet-related non-communicable diseases. Hence, the comprehensive approach of NExGEN-ChatGPT aims to reduce the burden related to obesity through a multifaceted weight management strategy that includes diet, exercise, and behavioral modification, underpinning the thesis that individualized and integrated lifestyle interventions are pivotal in improving disease-specific health outcomes.

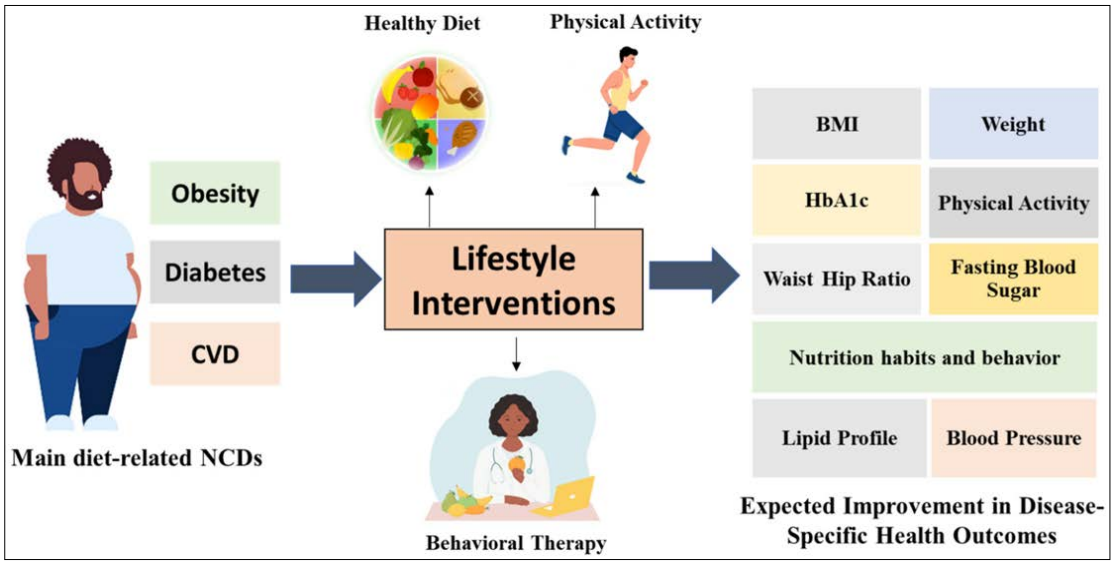


Figure 2.1 Conceptual Framework for Lifestyle Interventions for Management of Dietary- and Physical Inactivity-related Health Issues.

Despite numerous initiatives and the allocation of significant resources towards this goal, obtaining adequate nutritional knowledge remains a formidable challenge. A key obstacle is the lack of formal education, which hinders individuals' ability to make informed choices about healthier eating options. Even healthcare professionals such as general practitioners Vrkatić et al. (2022) and physicians Aggarwal et al. (2018) have been observed to exhibit deficiencies in nutrition education, knowledge, training, and aptitude. This challenge underscores the crucial function of dietitians and nutrition experts in providing specialized advice and assistance. Nevertheless, people globally often rely on sources like guidance from family and friends, books, television, and digital platforms for their nutritional information (Kołłajtis A & Żamojcin, 2016; Quaidoo et al., 2018). Many healthcare systems and communities likewise struggle to offer adequate resources and funding for dietetic services (Vilar et al., 2021). This deficiency further aggravates the existing barriers (e.g., cost of nutrition services and limited access to registered dietitians) in acquiring tailored information. information (Garcia et al., 2020; Stewart B. et al., 2016). Addressing this accessibility disparity is crucial to ensure equitable access to quality dietary guidance and promote overall health.

According to (Garcia & Garcia, 2023), worldwide initiatives to introduce dietary education programs aimed at enhancing nutritional knowledge remain insufficient due to the widespread nature of nutrition-related issues. Some persistent issues include the rising rates of obesity, increasing prevalence of diet-related chronic diseases such as

diabetes and heart disease, inadequate consumption of essential nutrients, and limited awareness of balanced and healthy eating habits among populations. This issue is of considerable importance because the deficiency in nutritional knowledge and education carries significant consequences for public health outcomes and the general well-being of the population (Mikkelsen et al., 2016). Bhawra et al. (2023) have emphasized the necessity for innovative approaches to address this problem. Salinari et al. (2023) share a similar perspective, advocating for the potential of artificial intelligence technologies to accelerate the pursuit of improved health and well-being. They highlight the importance of AI in providing accurate, personalized dietary recommendations and in furthering the development of predictive and preventive guidelines for enhanced health promotion and disease management. In a recent letter to the editor, Arslan (2023) conveyed optimism regarding ChatGPT's potential to provide personalized recommendations in the field of nutrition. Nonetheless, there remains a scarcity of literature on the application of ChatGPT within this area.

2.4 Digital Health Technologies in Addressing Obesity

The rapid expansion of technological advancements has played a pivotal role in bolstering digital transformation within the healthcare industry (Iyanna et al., 2022). These recent advancements have introduced a variety of novel health information technologies that empower healthcare professionals to augment patient care and optimize operational processes. The scholarly literature recognizes the significant impact of digital technologies, including telemedicine services (Tavares et al., 2023), blockchain technology Haleem et al. (2021), deep-learning-based medical image analysis (Maaliw et al., 2022), three-dimensional bioprinting Barua et al. (2023) digital game-based rehabilitation (Çaliş et al., 2023) and conversational chatbots (Solanki et al., 2023). By incorporating health-based technologies into daily operations, healthcare professionals can utilize digital tools and platforms to improve clinical decision-making, simplify administrative duties, and optimize the allocation of resources. Crucially, the transformative effect of health-centered technologies goes beyond individual patient interactions. It also supports the gathering and analysis of data, which can facilitate evidence-based practice and research (Sheikh et al., 2021). Overall, the integration of digital technologies in healthcare holds tremendous potential for revolutionizing the industry and driving improved outcomes for patients.

Nutritional and exercise interventions have emerged as a cornerstone in the realm of weight management, especially for individuals struggling with overweight or obesity. This strategy is crucial, as it not only promotes weight loss but also improves body composition. In fighting against obesity, nutritional management assumes a critical role, primarily by concentrating on modifiable dietary risk factors and the significance of nutritional components in intervention research. Indeed, utilizing digital tools for dietary self-monitoring has been demonstrated to be an effective approach for weight reduction. For instance, a study by Han & Rhee (2021) have shown the efficacy of mobile health applications in attaining a 5% reduction in body weight among obese and overweight adults. When these tools are paired with self-regulation strategies like goal setting and feedback, they produce significant outcomes.

Digital technologies are revolutionizing the methods and execution of personalized nutrition and physical activity planning. This involves customizing dietary and exercise recommendations and interventions to meet individual requirements, taking into account factors such as genetics, lifestyle, preferences, and health objectives. Experts are increasingly acknowledging that nutrition and physical activity applications can be utilized for behavior enhancement, dietary intake evaluation, and physical activity tracking (Abrahams & Matusheski, 2020; Hauptmann et al., 2022; Pantoja et al., 2017). These applications often include comprehensive nutritional databases, which enable them to offer recipe suggestions and personalized recommendations that are in line with individual goals, food preferences, and dietary restrictions.

The employment of online nutrition and physical activity monitoring services for weight management has attracted increased attention in recent years because of its ability to reach a broader audience and offer convenient access to nutrition counseling and support. Several studies have explored the effectiveness of online weight-loss interventions compared to traditional face-to-face interventions (Hegde et al., 2023; Rusali et al., 2018; K. Singh et al., 2021). Nonetheless, the provision of these online services is not a completely novel approach. The advent of digital dietary and exercise coaches has propelled the creation of various online applications, including Virtual Dietitian and Digital Trainer. These applications utilize a nutrition and exercise knowledge-based system that employs a forward chaining algorithm to generate personalized meal and exercise plans tailored to users' profiles, preferences, and restrictions (Garcia et al., 2021). These systems employ AI technologies to fast-track the achievement of enhanced health and well-being. They highlight the importance of

AI in providing accurate, customized dietary recommendations and in furthering the development of predictive and preventive guidelines for more effective health promotion. In a recent study by, Garcia (2023) has conveyed optimism regarding the potential of AI to provide personalized recommendations in the nutrition field. Despite the optimism reflected in recent scholarly works, there remains a scarcity of research concerning the use of AI-based personalized nutrition within this field.

2.5 Overview of Artificial Intelligence

Since the creation of the first computer, humans have concentrated on developing various approaches to decrease the computer size and increase its operational capacity. During the evolution of computer systems, researchers were interested in creating machines that think, work, and act like humans. This enthusiasm spurred the advancement of AI theory and led to the development of computer-based machines (e.g., *robots*) possessing intelligence that closely mimics human intelligence. It falls under a branch of computer science that investigates how computers can learn, think, and reason. According to John McCarthy, father of AI, AI is defined as "*the science and engineering of making intelligent machines, especially intelligent computer program*s".

As a discipline created in the 1950s, artificial intelligence is also defined as the "ability of a system to interpret and learn from exterior data correctly and to adopt the learning results to achieve specific objectives and solve problems through flexible adaptation" (Kaplan & Haenlein, 2019). Broadly defined, AI can be described as enabling computers to think and act both humanly and rationally, undertaking tasks that were previously performed better by humans and executing functions that would be deemed intelligent if carried out by a human being (Lombardo et al., 2022; Maknickiene, 2022; Wong et al., 2022; Yakan, 2022). As a member of the emerging technology, AI has been applied across a diverse range of fields for an extended period. Consequently, AI is increasingly acknowledged as a vital tool within the health sciences, with its applications extending from predicting disease outbreaks to medical imaging, from patient communication to behavioral modifications for weight management (Ronquillo et al., 2021).

The depicted schematic in Figure 2.2 provides a comprehensive view of the multifaceted applications of AI in the healthcare domain, emphasizing its potential to

revolutionize various aspects of patient care and medical administration. At the core of this framework is AI, serving as the central theme through which diverse facets of healthcare are integrated and optimized. In clinical environments, AI improves the analysis of medical imaging, including assessments of body composition, thus facilitating accurate diagnostics and interventions (Tang, 2020).

The AI further increases the efficiency of healthcare delivery by automating the scheduling of appointments, thereby reducing administrative burdens and decreasing wait times. AI significantly enhances telemedicine and remote patient monitoring by leveraging wearable technologies and mobile applications. This enables continuous care and real-time data analysis beyond the confines of traditional clinical (Haleem et al., 2021). Furthermore, the robotics sector, empowered by AI, brings new dimensions to patient care, ranging from surgery to rehabilitation, thereby enhancing precision and safety (O'Sullivan et al., 2019).

Precision medicine and predictive modeling, underpinned by AI, allow for treatments to be customized to the genetic and molecular profile of the individual, potentially increasing the efficacy and reducing the side effects of treatments. Moreover, AI is being utilized to personalize patient care through customized nutrition and physical activity plans based on individual health data, thus supporting behavior coaching and lifestyle modifications. The NExGEN-ChatGPT integration framework proposed in this study is designed to aid users in creating personalized fitness training and nutritional plans. This integrative approach to applying AI in weight management has the potential to enhance outcomes, optimize delivery, and lead the way toward more intelligent, efficient, and personalized dietary and physical exercise recommendations.



Figure 2.2 Multifaceted Applications of Artificial Intelligence in the Healthcare Domain.

2.6 Artificial Intelligence in Personalized Weight Management

Although AI initially emerged as a branch of computer science, it has now evolved into a concept discussed across nearly all disciplines. Application of AI in the field of nutrition and dietetics represents a relatively new and significant area of interest. Different apps related to nutrition are offered to the use of individuals. The importance of individual nutrition has also triggered the increase in AI apps (Ulker & Ayyildiz, 2021). It is believed that various applications, including those focused on food preferences and dietary intake, can play a crucial role in promoting health (Salinari et al., 2023). Similarly, AI has also emerged in the field of physical activity. As in most

other areas of society, there is a growing accumulation of data in diverse physical activities, and the application of AI for automated data analysis has emerged as a significant avenue, especially in addressing the global obesity crisis. Planning and analyzing physical activity data enhances the individual's capacity to refine weight management strategies effectively.

The use of technology-supported dietary and interventions has proven advantageous in aiding weight management among individuals. Before the widespread adoption of smartphones, earlier mHealth studies primarily depended on text messaging as a method to influence individuals' self-management behaviors (Horin et al., 2019). Evidence from previous studies on chronic diseases show that: medical institutions provided the tailored text messaging to support participants with self–management of diabetes Dobson et al. (2018); have been utilized to aid individuals with cardiovascular disease in improving their lifestyle and behavior or in making positive lifestyle and behavioral changes (Piette et al., 2015). Mobile text messaging was employed to disseminate health promotion content, and this form of intervention proved effective in inducing weight loss among overweight and obese individuals.

In recent years, with the swift evolution of information and communication technologies, there has been a shift in the use of digital devices as intervention tools for weight management, moving away from mobile text messaging P. Lin et al. (2014), websites Beleigoli et al. (2019) to mobile apps (Ufholz & Werner, 2023). In early weight loss literature, previous studies focused on investigating the self-monitoring of diet through the use of paper diaries or records over extended periods. As smartphones have rapidly gained popularity, the trend nowadays moves the weight loss study toward using smartphones to investigate self-monitoring. Recently, there has been a significant advancement with the emergence of AI chatbots, and some studies have begun to explore the application of medical AI in the management of diabetes care Contreras & Vehi (2018) and challenges of chatbot application for meal suggestion and healthy lifestyle promotion (Schachner et al., 2020).

The application of AI in weight loss has received considerable attention for its capacity to efficiently and intuitively track diet, exercise, and energy balance. Nonetheless, its effectiveness in offering recommendations and behavioral nudges to improve weight loss outcomes remains less explored (H. Chew et al., 2021). Chatbots hold significant promise as a means of communication for delivering behavioral nudges over a sustained period through health coaching, thus complementing the role of human

healthcare professionals in monitoring and counseling for weight loss. They also provide personalized recommendation services conveniently through an individual's preferred device and social communication platform.. These functionalities have been demonstrated to enhance usability, user acceptance, engagement, and the potential for weight loss success due to their convenience and immediacy (Gericke et al., 2021). This effectiveness is contingent on the ability to establish a human-like rapport with users, which represents one of the most significant challenges in chatbot development. Furthermore, there is limited knowledge regarding the chatbots that have been developed to tackle health issues necessitating multiple long-term behavior changes, such as those associated with overweight and obesity.

2.7 Artificial Intelligence Chatbots as a Weight Management Tool

Recent technological advancements have facilitated the development of computerized chatbots, also known as conversational agents, which can emulate the role of human health coaches. While terms like chatbots, conversational AI, intelligent chatbots, and conversational agents are frequently used interchangeably, chatbots can be categorized into those that incorporate AI and those that do not. In this thesis, chatbots are defined as computer software capable of engaging in conversation with individuals. AI is described as the emulation of human intelligence by machines to execute tasks typically associated with human capabilities, such as decision-making and problem-solving, primarily through the use of machine learning techniques.

Traditional rule-based chatbots, which do not incorporate AI, are limited in their capacity to recognize a finite set of client intents, relying solely on the interpretation of specific keywords in user utterances (Car et al., 2020). This limitation restricts the extent to which these chatbots can mimic human conversation, thereby also limiting the number of meaningful conversational exchanges necessary to establish a motivational human-chatbot rapport. In contrast, AI-powered chatbots are capable of machine learning, allowing them to comprehend human intents and sentiments. This capability enables them to engage in conversations with human-like nuances, enhancing the quality of human-chatbot interactions (Folstad et al., 2018). Recently, natural language processing has evolved from traditional recurrent neural network models, which process short texts for summarization, translation, and abstraction, to pretrained transformer

models. These advanced models analyze lengthy texts comprehensively, enabling more sophisticated tasks of understanding and contextualization (Tian, 2020).

As chatbots emerge as a popular digital communication medium, they offer vast potential for administering personalized behavior change programs for disease prevention and health promotion at scale. The benefits of AI chatbot programs primarily stem from their computational capability to create and provide personalized interventions (Conroy et al., 2019; Fukuoka et al., 2018; Hardeman et al., 2019). These interventions can surmount various constraints inherent in traditional, non-personalized approaches by being tailored to individual characteristics and behavioral trajectories. They are capable of progressively adjusting intervention strategies according to contextual conditions and the individual's cognitive and emotional states over time. In other words, chatbot technologies have the capacity to "*understand*" individuals through natural human conversations, persuade them towards change, and foster enduring, supportive relationships that help maintain healthy behaviors

AI's role in weight loss has been extensively researched for its proficiency in tracking diet, exercise, and energy balance both efficiently and intuitively. However, its effectiveness in offering impactful recommendations and behavioral nudges to boost weight loss success remains less explored (H. Chew et al., 2021). Chatbots hold significant promise as a means of communication for delivering behavioral nudges over a prolonged period through health coaching, thus augmenting the role of human healthcare professionals in monitoring and counseling for weight loss. Furthermore, chatbots offer continuous, real-time monitoring, instant counseling, and customized recommendation services around the clock. These capabilities have been demonstrated to enhance usability, user acceptance, engagement, and the likelihood of weight loss success due to their convenience and immediacy. However, this effectiveness depends on the capability to establish a human-like connection with users, representing one of the biggest challenges in chatbot development.

AI chatbots have been utilized to offer nutrition or food diary advice to users (Celis-Morales et al., 2015). Developing a chatbot for personalized nutrition recommendations necessitates producing responses that are coherent and suitable for the context. This complex task demands an extensive volume of training data, which is challenging and costly to acquire, particularly when the data must be annotated or assessed by healthcare experts. Such a limitation could restrict the chatbot's capacity to discern users' intentions and offer personalized, practical, and actionable nutrition

advice (Fadhil et al., 2019). However, the launch of 'ChatGPT' in late 2022, (as explored in the subsequent section of this literature review), has revealed the potential application of large language models across different healthcare domains, such as weight management. It offers advice and suggestions that are akin to human interaction to the target population.

2.8 The Potential of ChatGPT in Personalized Obesity Treatment

Launched in November 2022, ChatGPT is an AI-based large language model trained on massive text datasets in multiple languages with the ability to generate humanlike responses to text input. Developed by OpenAI (OpenAI, L.L.C., San Francisco, CA, USA), ChatGPT etymology is related to being a chatbot (a program able to understand and generate responses using a text-based interface) and is based on the generative pre-trained transformer (GPT) architecture (Brown et al., 2020a; Howard, 2019). The GPT architecture employs a neural network for processing natural language, thereby generating responses that are contextually aligned with the input text (Lund & Wang, 2023). The superiority of ChatGPT over its GPT-based predecessors is attributed to its capability to respond in multiple languages, producing refined and highly sophisticated responses through advanced modeling techniques (Tustumi et al., 2023). The advent of ChatGPT marks a significant progression in the realm of weight management. Nonetheless, the potential applications and limitations of ChatGPT as a tool for intelligent dietary and exercise planning in weight management remain largely unexplored.

Given the scant research on the use of ChatGPT in the nutrition and exercise field, its effectiveness as a tool within this domain is yet to be determined. This thesis suggests conducting a study to explore ChatGPT's potential in personalized obesity treatment. While ChatGPT may not match the expertise level of healthcare professionals, it has the potential to be a valuable resource for individuals without professional expertise in the field. Findings from prior studies have shown that ChatGPT surpassed other large language models in delivering accurate medical information. Despite not having specialized training, the findings indicated that ChatGPT scored at or near the passing threshold for all three exams it was tested on. These positive results suggest the potential for ChatGPT to also show promising

capabilities in the domain of nutrition and exercise planning. However, further assessment is needed to confirm this hypothesis.

2.8.1 Meal and Exercise Planner

Meal and exercise planning plays a crucial role in achieving successful outcomes in weight management. Crucially, tailoring meals and physical activities to fit individual preferences, dietary restrictions, and daily caloric needs is essential for enhancing adherence to a specific weight loss plan (Garcia et al., 2021). When ChatGPT is fed this variety of variables, it becomes better equipped to produce personalized meal and exercise plans that address individual requirements. Similar to other knowledge-based systems, ChatGPT can act as an effective recommender system, leveraging its sophisticated algorithms and natural language processing capabilities (Mustafa et al., 2022). This key feature aids users in diversifying their diets, exploring new healthy recipes, and optimizing nutrient intake.

Nutrition- and exercise-focused applications depend extensively on detailed databases that provide accurate information about the nutritional composition of various foods and the specifics of exercise programs (Hsu et al., 2019). One illustrative example of this is the 'Plan-Cook-Eat' application, which employed a mix of the Philippine Food Composition Table, My Food Data, FooDb, and the USDA Food Composition Databases (Mainardi et al., 2023). These nutrition databases are essential for analyzing dietary intake, offering personalized recommendations, and enabling nutritional tracking. As a result, users can search for specific foods and discover their macro- and micronutrient compositions (for example, the nutritional value of 100 grams of broccoli). Furthermore, a recent update to ChatGPT has introduced the capability to manually load a custom knowledge base, potentially mitigating these limitations (Casas et al., 2018). This feature enables developers or personal users to upload pertinent information related to their task, thereby addressing the previously mentioned limitation.

2.8.2 Dietary and Physical Activity Advice and Guide

The ability of a nutrition and exercise tool to provide advice and guidance is crucial in enabling individuals to participate in interactive discussions about their

nutrition and exercise-related questions (Keaver et al., 2023). Many current applications struggle to fully understand human language. ChatGPT, however, excels in this area due to its advanced capability for human-like conversation. It engages users through interfaces that can handle a variety of mediums, such as text, voice, speech, images, simulated movements, tactile feedback, and other data file formats. This allows users to ask questions about different food and exercise groups, learn about the importance of specific nutrients, or seek clarification on common dietary and exercise misconceptions (Sallam, 2023). This capability enhances nutrition and physical activity education by enabling users to participate in meaningful and informative conversations. However, it's crucial to recognize ongoing concerns about the accuracy of the advice provided, despite previous reports of positive outcomes. Guaranteeing the precision, reliability, and accuracy of health advice continues to be a significant challenge in applications driven by AI.

Nutrition and exercise applications often feature daily tips and reminders sent to users, aiming to educate and motivate them towards making healthier choices and enhancing their nutritional habits (Tonkin et al., 2017). These automated messages act as gentle nudges and ongoing sources of inspiration, aiding individuals in their quest to adopt better eating habits. The daily tips and reminders prove invaluable in various respects. First, they serve as timely reminders that keep nutrition top of mind for users. This continuous reinforcement motivates users to consciously consider their choices throughout the day, gradually guiding them towards healthier food options. Second, these messages frequently offer practical advice, actionable suggestions, and insights into healthier eating habits (Klassen et al., 2018). They equip users with the knowledge necessary to make educated dietary choices. Nonetheless, it's crucial to recognize that ChatGPT operates differently compared to a conventional nutrition and exercise tool (Natalucci et al., 2023). Unlike applications that can send direct notifications and reminders, ChatGPT functions as a conversational AI tool. Nonetheless, users can utilize its capabilities to obtain strategies and supplementary information about maintaining a healthy lifestyle.

2.8.3 Nutrition and Fitness Material

Interactive lessons and modules offering detailed insights into different facets of nutrition represent another common feature. Nutrition and exercise tools frequently possess the necessary technical infrastructure and innovative capabilities to actively engage users (Baretta et al., 2019). They extend beyond simple information dissemination by integrating interactive components that promote user involvement. This could encompass quizzes, assignments, and self-assessments, all designed to enhance the learning experience (Brug et al., 2003). As users advance through these modules, they acquire not just knowledge but also practical skills pertinent to nutrition. Conversely, ChatGPT presents a unique method of learning and disseminating information. Although it may not directly interface with educational materials on nutrition and physical activity, it acts as a readily available learning platform in its own right. Users can engage with ChatGPT by posing questions, seeking clarification, and requesting information related to nutrition or any other subject matter (Garcia, 2023). Its strength lies in its ability to provide quick and responsive answers to user queries. For instance, users can ask questions such as "*What are some good dietary sources of omega-3 fatty acids?*" and receive immediate responses. This interaction fosters a dynamic and individualized learning experience.

2.9 Engineering for Health-Specific AI Applications

2.9.1 Prompt Engineering for Nutrition and Fitness Applications

Prompt engineering has developed from an informal practice into a structured research discipline focused on the systematic design of inputs that guide large language model behavior (Chen et al., 2025). Zero-shot prompting enables models to respond to task descriptions without prior examples, relying on pre-trained knowledge, while few-shot prompting provides exemplars that improve performance on complex tasks (Brown et al., 2020b). Chain-of-thought prompting embeds step-by-step reasoning within prompts, yielding measurable gains on multi-step benchmarks (Wei et al., 2023). System-level strategies, including role prompting, assign the model a clinical or professional identity to improve domain-specific output alignment (Chen et al., 2025).

Empirical research confirms that prompt design significantly influences large language model output quality in nutrition and fitness contexts. Azimi et al. (2025) evaluated multiple large language models on 1,050 registered dietitian examination questions, finding that chain-of-thought with self-consistency prompting achieved the highest accuracy, with performance varying considerably by domain and model. In fitness applications, Havers et al. (2025) found that detailed prompts incorporating participant characteristics produced resistance training plans with moderate expert ratings, though deficiencies in periodization persisted. Bayram & Arslan (2025) demonstrated that AI-generated dietary plans for individuals with obesity contained insufficient calcium and potassium, indicating that even structured prompts require dietitian oversight to ensure nutritional adequacy.

The rapid proliferation of prompting techniques has driven the development of comprehensive classification frameworks for structured application in health domains. Schulhoff et al. (2025) assembled a taxonomy of 58 text-based prompting techniques through a systematic review of 1,565 papers, organizing them into categories including few-shot prompting, thought generation, decomposition, ensembling, and self-criticism. White et al. (2023) extended this by presenting prompt patterns analogous to software design patterns, covering output customization and error identification to enable domain-agnostic knowledge transfer including healthcare. Chen et al. (2025) further surveyed advanced techniques such as plan-and-solve reasoning and automated prompt optimization, consolidating a foundation for engineering clinically appropriate prompts in nutrition and fitness.

2.9.2 Prompt Engineering in Clinical and Health Contexts

The application of prompt engineering to clinical contexts has evolved from exploratory experimentation into a recognized strategy for improving large language model accuracy and safety in healthcare. Singhal et al. (2023) demonstrated that instruction prompt tuning of a large language model achieved 67.6% accuracy on United States Medical Licensing Examination-style questions, surpassing prior performance by over 17 percentage points. Sivarajkumar et al. (2024) further showed that heuristic prompts, which encode domain-specific rules within the prompt structure, achieved 0.96 accuracy on clinical word sense disambiguation tasks, establishing

prompt architecture as a critical determinant of output quality independent of the underlying model.

Structured prompt frameworks represent an advance beyond ad hoc prompt construction toward validated architectures for generating health recommendations. Boit & Patil (2025) proposed a layered framework for mental health chatbots incorporating proactive risk detection, retrieval-augmented dialogue grounded in cognitive behavioral therapy, and post-generation ethical filtering. In nutritional AI applications, Kaya et al. (2025) found substantial variability across large language model platforms under standardized prompts, with inconsistent micronutrient adequacy and limited cultural sensitivity. Bayram & Arslan (2025) extended this by demonstrating that energy outputs deviated significantly from targets across models, establishing that minimal prompts produce variable nutritional outputs and require structured front-loading of clinical parameters.

Hallucinations, defined as fluent but factually incorrect large language model outputs, represent a consequential risk in health AI deployment (Anh-Hoang et al., 2025). Prompt-level mitigation strategies, including chain-of-thought decomposition, self-consistency prompting, and explicit output constraints, reduce confabulation but do not eliminate it, as the fundamental tension between generative fluency and factual grounding persists (Anh-Hoang et al., 2025). Evidence from the dietary domain reinforces this concern: Liao et al. (2024) found that ChatGPT produced adequate general nutritional knowledge but inconsistent accuracy in individualized scenarios, while Shi et al. (2023) demonstrated that iterative three-turn prompting improved both accuracy and specificity in nutrition-related large language model outputs compared to single-pass generation.

2.9.3 Personalized Prompt Systems for Health Interventions

The emergence of large language models in healthcare has catalyzed interest in systems that dynamically generate personalized prompts from user-specific health data rather than relying on static templates. The NExGEN system collects data across 111 assessment variables, including nutritional needs, fitness levels, anthropometric measurements, medical history, stress levels, and sleep patterns, and algorithmically converts those variables into structured ChatGPT prompts (Dan et al., 2025). A scoping review of 23 AI chatbot studies for weight loss found that most systems relied on pre-

defined dialogue flows or rule-based logic rather than dynamically generated prompts informed by comprehensive user assessments (Chew, 2022), establishing the distinct architectural advantage of multi-variable prompt generation.

Effective personalization requires systematic capture of diverse user variables and a mechanism that translates those variables into prompts improving large language model output quality. A systematic review of eHealth interventions for weight management found that personalized feedback from multiple data streams produced weight loss outcomes up to 4.3 kg greater than minimal interventions, attributing the effect to improved adherence grounded in user-specific data (Kupila et al., 2023). A benchmarking study evaluating 56,000 large language model responses across 25 biomarker-defined profiles found that retrieval-augmented generation exhibited limitations in contraindication checking and dosage personalization, demonstrating that prompt structure and specificity, rather than the mere presence of user data, determine clinical output quality (Jarchow et al., 2025).

The credibility of personalized prompt systems depends on the psychometric rigor of instruments used to capture individual variables. The NExGEN 111-item assessment was validated through the Fuzzy Delphi Method (FDM) with a 21-member multidisciplinary panel, achieving consensus across all criteria (Dan et al., 2025). Established instruments such as the International Physical Activity Questionnaire (IPAQ) Craig et al. (2003) and the Pittsburgh Sleep Quality Index Buysse et al. (1989) provide validated inputs for prompt personalization, though systematic integration into automated generation pipelines beyond NExGEN remains limited. A mixed-methods study subsequently developed a quality evaluation index for health conversational AI encompassing safety, accuracy, and personalization dimensions (Liao et al., 2026).

2.10 Quality Assessment for AI-Generated Health Recommendations

2.10.1 Existing Quality Assessment Instruments

The Mobile App Rating Scale (MARS) remains the most widely adopted quality assessment instrument for mobile health applications. Developed by Stoyanov et al. (2015), the MARS evaluates applications across four objective subscales covering engagement, functionality, aesthetics, and information quality, alongside one subjective quality subscale. Initial validation with 50 wellbeing applications demonstrated

excellent internal consistency (Cronbach's alpha = .90) and strong interrater reliability (Stoyanov et al., 2015). A German adaptation validated across 104 applications confirmed high convergent validity with the original instrument, with correlation coefficients between 0.93 and 0.98 (Messner et al., 2020). A simplified user version subsequently enabled end-user quality assessment without specialized training, retaining excellent internal consistency and good test-retest reliability (Stoyanov et al., 2016).

The Organisation for the Review of Care and Health Applications (ORCHA) represents a distinct organizational assessment approach. The ORCHA Baseline Review evaluates digital health interventions across five domains: data and privacy, clinical regulation and safety, evidence and outcomes, usability and accessibility, and ongoing compliance, using binary evidence-based scoring to maximize assessor consistency (Hyzy et al., 2023). ORCHA has completed over 30,000 assessments globally, requiring a minimum score of 65% across all areas for inclusion in curated health application libraries. In 2023, ORCHA partnered with the American College of Physicians and the American Telemedicine Association to develop the Digital Health Assessment Framework, comprising approximately 400 criteria tailored to United States healthcare regulations and best practices (American College of Physicians (ACP) et al., 2026).

The Quality Assessment of Medical Artificial Intelligence (QAMAI) instrument addresses a critical gap left by prior tools designed for conventional mobile health applications. Developed by Vaira et al. (2024), the QAMAI evaluates six dimensions of AI-generated health information: accuracy, clarity, relevance, completeness, usefulness, and information sourcing. Validation across 792 assessments by 27 reviewers from 25 international centers demonstrated high internal consistency (Cronbach's alpha = 0.837) and excellent interrater reliability (intraclass correlation coefficient = 0.983) (Vaira et al., 2024). Systematic evidence from a scoping review of digital health implementation further documents that fewer than 12% of WHO member states have conducted formal evaluations of their digital health activities, and only eight of 94 identified assessment frameworks have undergone validity testing (American College of Physicians (ACP) et al., 2026).

2.10.2 Limitations of Current AI-Driven Health Tools

Conventional health technology assessment frameworks were designed for static medical devices and predetermined application features, creating a structural mismatch when applied to AI systems that generate variable outputs with each interaction. A Delphi study involving 46 international experts found that the established EUnetHTA Core Model fails to capture topics relevant to AI-based health technologies, with model accuracy (97.78% consensus), data bias (91.30% consensus), and algorithmic transparency identified as critical additions to any adequate assessment framework (Di Bidino et al., 2024). A scoping review of 78 studies on AI-based medical device assessment similarly concluded that no existing health technology assessment domain adequately accounts for data quality dependencies, continuous learning characteristics, or evolving performance profiles that distinguish AI tools from static interventions (Farah et al., 2024).

The MARS, validated through pooled data from 15 international reviews with acceptable construct validity across engagement, functionality, aesthetics, and information quality dimensions (Terhorst et al., 2020), cannot account for stochastic variability in large language model outputs, where identical inputs may produce substantially different recommendations across sessions. Analysis of AI-generated diet plans documented temporal inconsistencies across repeated prompts, with significantly higher fat and saturated fat content relative to dietitian-prepared menus and frequent micronutrient inadequacy in calcium, potassium, and vitamin B1 (Ozlu Karahan & Kenger, 2025). Expert panel evaluations of AI dietary advice further found appropriateness rates ranging from only 55.5% to 73.3% across clinical conditions, with accuracy declining substantially when overlapping comorbidities required integrated management (Ponzo et al., 2024).

The most consequential gap in current evaluation practice is the absence of criteria spanning the full input-process-output chain of large language model health applications. Tam et al. (2024) reviewed 142 healthcare large language model evaluation studies and identified systematic deficiencies in reliability and generalizability, proposing the QUEST framework addressing quality of information, reasoning, expression, safety, and trust. A scalable health language model evaluation framework subsequently proposed template-driven criteria based on personalization data elements including blood biomarkers and wearable device data (Mallinar et al.,

2026). Neither framework evaluates the prompt engineering methodology that generates outputs, nor whether personalization mechanisms produce clinically appropriate individualized recommendations across sustained interactions, representing a critical evaluation vacuum for personalized AI health systems.

2.10.3 Domain-Specific Criteria for AI-Generated Recommendations

Obesity management involves complex pharmacological considerations that AI health systems must address through structured comorbidity screening. The Obesity Medicine Association identifies medication interaction screening as a foundational safety concern, noting that challenges of accuracy, reliability, and potential perpetuation of systemic biases in AI remain unresolved (Bays et al., 2023). A scoping review of generative AI and large language models in mitigating medication-related harm, covering 30 studies, found that while large language models show promise in drug-drug interaction identification, no studies tested these models prospectively and performance varied considerably across applications (Ong et al., 2025). Al-Ashwal et al. (2023) further demonstrated that ChatGPT models exhibited lower sensitivity, specificity, and accuracy compared to conventional pharmacological databases, establishing that standalone large language models cannot reliably perform medication safety screening without integration with authoritative drug interaction resources.

AI-generated dietary recommendations require evaluation across multiple dimensions that current instruments do not systematically capture. A systematic review of 23 empirical studies on ChatGPT's performance in personalised meal planning identified nutritional safety as a primary concern, noting that allergen inclusion in meal plans posed direct risks to individuals with declared food sensitivities (Guo et al., 2025). Niszczota & Rybicka (2023) confirmed this finding, demonstrating that ChatGPT produced diets containing allergens for hypothetical individuals with documented food allergies, alongside inaccuracies in portion sizes and caloric content. Adilmetova et al. (2025) further documented that AI dietary advice performance weakened markedly in underrepresented languages, with responses containing hallucinated food names, indicating that cultural and linguistic appropriateness constitutes a distinct evaluative dimension requiring formal criterion development.

Expert-derived evaluation criteria for AI-generated exercise prescriptions have been formalized through coding schemes benchmarked against established clinical

guidelines. Zaleski et al. (2024) developed a ten-category coding scheme and found that AI-generated exercise recommendations were only 41.2% comprehensive, with the majority of inaccuracy attributable to deficient exercise preparticipation health screening. Dergaa et al. (2024) found that AI-generated prescriptions prioritized excessive safety over therapeutic effectiveness, lacking precision in addressing individual health conditions and goals. Lee et al. (2025) extend these evaluation concerns to multidimensional validation, arguing that AI systems for obesity management must integrate physical activity, dietary behaviour, psychological factors, and environmental determinants, as no existing evaluation instrument integrates all domains for AI-generated weight management content.

2.11 Expert Consensus Validation Methods in Health Tools

2.11.1 Overview of Delphi Methods in Health Research

The Delphi technique is one of the most widely deployed consensus methods in health sciences, providing a structured process for aggregating expert judgment where empirical evidence is incomplete, conflicting, or ethically difficult to obtain (Nasa et al., 2021). A scoping review of 287 consensus Delphi studies in health sciences published between 2018 and 2021 confirmed the technique's growing adoption across clinical guideline formulation, curriculum development, terminology standardization, and research prioritization (Schifano & Niederberger, 2025). In health technology development specifically, Delphi methods have been applied to validate evaluation frameworks for digital health tools; the EVALAPPS project used a modified Delphi process to reach stakeholder consensus on 114 evaluation criteria for weight management mobile applications, ultimately prioritizing 53 criteria across security, privacy, usability, and clinical effectiveness domains (Robles et al., 2020).

Three dominant Delphi variants have emerged in health sciences research: the classical Delphi, the modified Delphi, and the e-Delphi. The classical Delphi is defined by anonymous iteration across multiple rounds, controlled feedback between rounds, statistical aggregation of group responses, and open-ended first-round questions that allow items to emerge from expert input (Nasa et al., 2021). The modified Delphi departs from this by replacing the open-ended round with a structured questionnaire generated from a prior literature review, with 43% of health sciences Delphi studies in

the Schifano & Niederberger (2025) review employing this variant. The e-Delphi adapts the process to web-based platforms and has been combined with content validity procedures including the Content Validity Index (CVI), Content Validity Ratio, and kappa statistics to strengthen psychometric rigor (Barrett & Heale, 2020).

Methodological rigor in Delphi studies depends critically on transparent expert panel selection, defensible panel size, and a priori consensus threshold specification. Systematic reviews recommend a minimum of 10 to 18 panelists per stakeholder group to ensure adequate diversity while maintaining feasible logistics, with a median panel size of 17 members reported across healthcare quality indicator studies (Boulkedid et al., 2011). Percentage agreement is used by 81% of health sciences Delphi studies to define consensus, yet threshold values range widely across publications, with 75% agreement representing the median standard (Diamond et al., 2014; Schifano & Niederberger, 2025). The ACCORD reporting guideline, developed through a three-round Delphi process with 72 panelists, established 35 reporting items to address the long-standing problem of inconsistent documentation that undermines confidence in consensus-derived recommendations (Gattrell et al., 2024).

2.11.2 Fuzzy Delphi Method: Principles and Applications

The Fuzzy Delphi Method (FDM) integrates Zadeh's fuzzy set theory with the classical Delphi technique to address the inherent imprecision of human judgment that crisp Likert scales fail to capture (Roldán López et al., 2021). Each expert's judgment is represented as a triangular fuzzy number (TFN) defined by three parameters: a minimum value, a most plausible value, and a maximum value, expressed as the tuple (m1, m2, m3) (Roldán López et al., 2021). FDM analysis proceeds by converting individual expert responses to TFNs, aggregating these into a group fuzzy number, and calculating the threshold value d using the vertex method to measure Euclidean distance between each expert's evaluation and the group average (Manakandan et al., 2017). An item achieves consensus when d does not exceed 0.2, expert agreement meets or exceeds 75%, and the defuzzified alpha-cut score is at or above 0.5 (Sulaiman et al., 2020).

FDM offers several methodological advantages over the classical Delphi approach in health instrument validation. By representing each expert response as a range rather than a single point, FDM preserves the degree of certainty alongside the

direction of agreement, overcoming the measurement compression that crisp ordinal scales impose on complex professional judgments (Roldán López et al., 2021). Consensus can be achieved in a single survey administration because fuzzy aggregation and threshold calculations inherently quantify agreement without requiring re-polling across multiple rounds, reducing time, costs, and the social desirability pressures that can distort opinions when earlier-round feedback is shared with panelists (Mohd Yusoff et al., 2024). The three acceptance criteria, d value, percentage agreement, and defuzzified score, function as a conjunctive rule requiring simultaneous satisfaction, providing a more stringent and defensible standard than percentage-agreement-only approaches (G. Lin et al., 2023; Manakandan et al., 2017).

FDM has been adopted across instrument validation, clinical framework prioritization, and curriculum development in health sciences, demonstrating consistent application of the three-criterion acceptance framework. Sulaiman et al. (2020) validated an occupational zoonotic disease questionnaire with 14 experts across 32 items, achieving consensus agreement ranging from 78.8% to 91.9% with threshold d values between 0.07 and 0.14. Selvaraju & Nik (2024) applied FDM to determine new norm elements for emergency departments, evaluating 35 items with a panel of 15 emergency physicians and retaining 25 items meeting all three acceptance criteria. In curriculum development, G. Lin et al. (2023) employed FDM with 62 dental materials science experts, accepting 33 of 36 topics based on average fuzzy numbers ranging from 0.526 to 0.784. These applications reflect a consistent pattern of beginning with a literature-derived item pool, submitting it to a purposively selected expert panel of 10 to 20 members, and applying the triple threshold to produce an empirically defensible, expert-endorsed instrument (Sulaiman et al., 2020).

2.11.3 Expert Consensus Methods in AI Health Tool Validation

The Delphi technique has emerged as the predominant expert consensus methodology for validating AI-driven health interventions, particularly where empirical evidence for implementation decisions remains insufficient (Nasa et al., 2021). A two-round Delphi survey involving 46 experts across 14 countries established a health technology assessment framework for AI-based health technologies, identifying that standard assessment domains require supplementary evaluation of data quality, algorithmic transparency, and technology-specific clinical safety (Di Bidino et al.,

2024). In parallel, a Delphi study on AI-based clinical decision support systems for antibiotic prescribing convened 36 experts and reached consensus on 34 implementation factors, ranking perceived clinical added value as the most critical determinant of successful adoption (Tokgöz et al., 2025).

Expert consensus methodology has been applied specifically to AI-driven clinical decision support for chronic disease management, with multidisciplinary panel composition consistently identified as essential. The European Diabetes Forum employed nominal group technique and Delphi methodology to formulate 14 consensus recommendations addressing the development, regulation, and deployment of AI-driven decision support for diabetes care, explicitly addressing the regulatory challenges posed by adaptive AI systems that can change performance characteristics after deployment (Bajramagic et al., 2026). A validated framework for responsible AI in healthcare autonomous systems was subsequently developed through expert interviews across clinical, technical, ethical, and regulatory domains, yielding consistently high scores for relevance, clarity, and usability across ten framework dimensions (Alelyani, 2025). The Chatbot Assessment Reporting Tool (CHART) statement employed an international modified Delphi process involving 531 stakeholders, followed by synchronous panel meetings with 48 stakeholders, producing a 12-item checklist addressing model identifiers, prompt engineering, and performance evaluation (Huo et al., 2025).

Despite the rapid proliferation of AI health technologies, a pronounced validation gap persists for generative AI applications in chronic disease management. A systematic review of guidelines and standard frameworks for AI in medicine found 11 distinct reporting guidelines operating in parallel without standardized coordination, creating an evaluative landscape that novel applications such as LLM-based weight management tools must navigate without a unified framework (Shiferaw et al., 2025). A systematic review of LLM applications in obesity management identified 33 eligible studies spanning personalized nutrition, educational interventions, and surgical decision-making, yet none employed expert consensus methods to validate the prompts or frameworks guiding LLM behavior (Suenghataiphorn et al., 2025). A scoping review of prompt engineering paradigms in medical applications similarly found that most studies presented results from a single prompt choice without expert-validated selection (Zaghir et al., 2024). This thesis addresses this gap by applying the FDM with a multidisciplinary expert panel spanning nutrition, exercise science, medicine,

psychology, and AI to validate the NExGEN-ChatGPT personalized prompt generator, representing the first application of expert consensus methodology to validate a prompt engineering framework for an LLM-based obesity intervention.

2.12 Outcome Assessment in Weight Management Research

2.12.1 Body Composition Assessment

Dual-energy X-ray absorptiometry (DXA) remains the clinical reference standard for body composition assessment in obesity research, providing non-invasive quantification of fat mass, fat-free mass, and bone mineral density (Lemos & Gallagher, 2017). Bioelectrical impedance analysis (BIA) represents a practical, cost-effective alternative that estimates body composition by measuring resistance and reactance of electrical current through body tissues. A retrospective study of 3,655 measurements found that for individuals with BMI between 18.5 and 40 kg/m2, BIA systematically overestimated fat-free mass by 3.38 to 8.28 kg and underestimated fat mass by 2.51 to 5.67 kg compared with DXA; however, this systematic bias remained relatively stable within overweight and grade I to II obesity categories (Achamrah et al., 2018). Multifrequency BIA devices demonstrate good to excellent agreement with DXA at the population level, with intraclass correlation coefficients ranging from 0.77 to 0.93 for body fat percentage (Merrigan et al., 2022).

Waist circumference is widely recognized as a practical, inexpensive surrogate marker for visceral adipose tissue, the metabolically active abdominal fat depot most strongly associated with cardiometabolic disease risk (Ross et al., 2020). An International Atherosclerosis Society and International Chair on Cardiometabolic Risk consensus statement recommended that waist circumference be treated as a vital sign in clinical practice, arguing that BMI alone is insufficient to assess cardiometabolic risk (Ross et al., 2020). Even modest weight loss produces disproportionately large reductions in visceral fat: a 5.8% weight reduction in men produced an 11.2% reduction in visceral adipose tissue, and a 4.1% reduction in women yielded an 11.1% decrease in visceral fat as measured by computed tomography (Pinho et al., 2018). Pooled MRI-assessed data from two lifestyle weight loss trials further confirmed that participants achieved a mean visceral adipose tissue reduction of 22.5% over 18 months, with

changes independently associated with improvements in metabolic syndrome markers, blood pressure, and inflammatory biomarkers (Klein et al., 2025).

Preserving fat-free mass during intentional weight loss is a clinical imperative with direct implications for metabolic rate, physical function, and long-term weight maintenance. Fat-free mass is the primary determinant of resting energy expenditure; its reduction during weight loss directly lowers metabolic rate and creates a physiological predisposition toward weight regain (Heymsfield et al., 2014). The degree of caloric restriction is positively correlated with the proportion of weight lost as fat-free mass, while resistance exercise and adequate protein intake attenuate this lean tissue depletion (Cava et al., 2017; Chaston et al., 2007). A systematic review and meta-analysis demonstrated that protein intake exceeding approximately 1.05 g/kg/day during weight loss improved muscle mass retention in adults with overweight and obesity (Kokura et al., 2024). Sarcopenic obesity, affecting approximately 11% of individuals aged 60 years and above, underscores the need for weight management strategies that explicitly protect lean tissue (Donini et al., 2022).

2.12.2 Behavioral Outcome Measurement

Accurate physical activity measurement is fundamental to evaluating weight management interventions. The ActiGraph accelerometer has emerged as the most widely used research-grade device for capturing tri-axial acceleration data, and has been adopted by major surveillance studies including the United States National Health and Nutrition Examination Survey and the United Kingdom Biobank (Troiano et al., 2008). Self-report instruments such as the International Physical Activity Questionnaire Short Form (IPAQ-SF) demonstrate substantially lower criterion validity against accelerometry, with correlations ranging from 0.09 to 0.39 in a systematic review of 23 validation studies, none reaching the accepted minimum validity threshold of 0.50 (Lee et al., 2011). The discrepancy is amplified in populations with overweight and obesity: Hicks et al. (2024) found that IPAQ-derived energy expenditure exceeded accelerometer-measured values by more than two-fold in obese participants, who simultaneously underestimated sedentary time by approximately 23 hours per week.

Dietary assessment in weight management trials presents distinct methodological challenges, as all self-report methods are subject to substantial random and systematic measurement error (Thompson & Subar, 2017). Food records, 24-hour

dietary recalls, and food frequency questionnaires each carry distinct strengths and limitations. When validated against recovery biomarkers, food records explained 7.8% of biomarker-measured energy variation in postmenopausal women, compared with 3.8% for food frequency questionnaires and 2.8% for 24-hour recalls (Prentice et al., 2011). All three traditional methods underestimate energy intake in populations with overweight and obesity, typically by 15% to 30% when compared with recovery biomarkers, with approximately 30% of participants classified as under-reporters (Poslusna et al., 2009). Dietary energy density, calculated as kilocalories per gram of food consumed excluding beverages, has gained prominence as a clinically meaningful secondary outcome; lower dietary energy density diets are consistently associated with lower body weight and reduced obesity risk (Pepe et al., 2023).

Social desirability bias constitutes a persistent threat to the internal validity of behavioral weight management trials. Hebert et al. (1995) demonstrated that social desirability bias produced a downward bias in dietary energy reporting equivalent to approximately 450 kilocalories across the interquartile range of the Marlowe-Crowne Social Desirability Scale, with the bias approximately twice as large in women as in men. A multi-ethnic Asian cohort study (n = 2,995) confirmed that higher social desirability scores were associated with systematic overreporting of physical activity and healthy dietary habits (Teh et al., 2023). Kirkpatrick et al. (2019) recommended that validation studies should always include assessment of social desirability bias and explore its role as a potential confounder, particularly in populations with overweight and obesity where reporting biases are most pronounced. Valid wear-time criteria for accelerometry further require specification of non-wear algorithms, minimum daily wear-time thresholds, and minimum valid-day requirements, with the Choi algorithm and a 10-hour per valid day criterion representing current methodological consensus (Choi et al., 2012).

2.12.3 Cardiometabolic Risk Markers in Obesity

The relationship between weight loss magnitude and cardiometabolic improvement follows a graded, dose-dependent pattern. Current clinical guidelines endorse a minimum 5% weight loss as the threshold for clinically meaningful metabolic benefit, associated with improvements in blood pressure, triglycerides, fasting glucose, and reduced incidence of type 2 diabetes (Gilbert et al., 2025). A randomized controlled

trial evaluating 40 adults with obesity confirmed that a 5% weight loss significantly reduced intra-abdominal adipose tissue volume, intrahepatic triglyceride content, systolic blood pressure, and plasma triglyceride concentrations, while simultaneously improving multi-organ insulin sensitivity and beta-cell function; progressive loss to 11% and 16% produced additional stepwise improvements in skeletal muscle insulin sensitivity and adipose tissue gene expression (Magkos et al., 2016). The Look AHEAD trial further confirmed that each successively higher weight loss category was associated with progressively greater improvements across six cardiovascular risk factors, with those achieving 5 to 10% loss demonstrating significantly increased odds of a 5 mmHg systolic blood pressure reduction (OR: 1.56; 95% CI: 1.27 to 1.91) and a 40 mg/dL triglyceride decrease (OR: 2.20; 95% CI: 1.71 to 2.83) (Wing et al., 2011).

The metabolic heterogeneity of obesity carries critical implications for how weight loss interventions are measured and interpreted. Approximately one-third of individuals with obesity present as metabolically healthy obesity (MHO), typically defined by the absence of hypertension, dyslipidaemia, and insulin resistance despite elevated BMI (Blüher, 2020). However, longitudinal evidence consistently demonstrates that MHO is a transient state: a meta-analysis estimated that approximately 49% of MHO individuals develop at least one metabolic abnormality within 10 years, with a pooled relative risk of incident metabolic abnormalities reaching 1.80 (95% CI: 1.53 to 2.11) compared to metabolically healthy normal-weight individuals (Lin et al., 2017). Metabolomics profiling has identified elevated branched-chain amino acids, aromatic amino acids, and specific acylcarnitines in metabolically unhealthy obesity compared to MHO, suggesting that circulating signatures may partially explain differential cardiometabolic risk (Telle-Hansen et al., 2020). When metabolically healthy obesity individuals undergo weight loss interventions, they show less absolute improvement in cardiometabolic markers than metabolically unhealthy counterparts due to ceiling effects at near-normal baseline values (Lin et al., 2017).

The temporal dynamics of cardiometabolic improvement following lifestyle-induced weight loss are non-uniform across biomarkers. Blood pressure responds earliest, with reductions emerging within four to eight weeks of caloric restriction and closely paralleling the trajectory of weight change; individuals with elevated baseline values experienced approximately twice the blood pressure reduction per kilogram of weight lost compared to normotensive individuals (Morris et al., 2021). Glycaemic markers exhibit a more complex pattern: fasting glucose responds within weeks through

acute caloric-deficit effects on hepatic glucose output, while HbA1c inherently lags by its erythrocyte turnover window of eight to twelve weeks, meaning that intervention effects on HbA1c cannot be reliably captured at timepoints shorter than three months (Magkos et al., 2016; Pidd et al., 2025). Combined dietary and physical activity interventions produce lipid improvements that exceed either intervention in isolation, with dietary change driving atherogenic lipoprotein reduction and exercise driving High-Density Lipoprotein (HDL) cholesterol elevation and triglyceride clearance; a meta-analysis of 80 randomized controlled trials with 4,804 participants confirmed that combined interventions reduced total cholesterol, triglycerides, and LDL cholesterol more effectively than exercise alone (Khalafi et al., 2023; Smart et al., 2025).

2.13 Previous Research on the AI Chatbots for Exercise and Nutrition

Table 2.2 provides details of the previous randomized controlled trials on AI chatbots for recommendations in physical activity and nutrition. Kramer et al. (2020) evaluated the effects of incentives, weekly planning, and daily self-monitoring prompts that were used as intervention components as part of The Assistant to Lift your Level of activity (Ally) chatbot app. The researcher conducted an 8-week optimisation trial with policyholders (n = 274) of a health insurance company in Switzerland. However, the study found that engagement with the chatbot for physical activity and nutrition planning interventions, as well as self-monitoring prompts was low. Further, 30% of participants stopped using the app over the course of the study. This was attributed to several factors, including prompts provided by the chatbot, which might not have been sufficiently personalized to the individual needs, preferences, and fitness levels of the participants. Generic advice or goals of the chatbot that do not match a user's personal context can reduce engagement and perceived value.

Kunzler et al. (2019) explored the factors affecting users' receptivity towards a chatbot app. This study was conducted with 189 participants, over a period of 6 weeks, where participants received interventions to improve their physical activity levels. The interventions were delivered by Ally chatbot. This investigation found that (1) several participant-specific characteristics (age, personality, and device type) show significant associations with the overall participant receptivity over the course of the study, and that (2) several contextual factors (day/time, phone battery, phone interaction, physical activity, and location), show significant associations with the participant receptivity.

The research speculates that personal desire to engage in an activity for its own sake, is often more sustainable for long-term behavior change. Accordingly, the chatbot must be designed to foster this intrinsic motivation adequately.

Piao et al. (2020) assessed the efficacy of the Healthy Lifestyle Coaching Chatbot intervention presented via a messenger app aimed at stair-climbing habit formation for office workers. In the study, participants were randomly assigned to the intervention group or the control group. The intervention group received cues and intrinsic and extrinsic rewards for the entire 12 weeks. However, the control group did not receive intrinsic rewards for the first 4 weeks and only received all rewards as in the intervention group from the fifth to twelfth week. The study concluded that the chatbot app can be used to provide improvements in relevant variables in the long term and can be useful in producing new healthy habits. However, the study acclaimed that the chatbot might have limitations in terms of not being able to contextually give encouragement and meaningful support to users to continue using the app.

Finally, the study by Carfora et al. (2019) tested a daily messaging intervention aimed at promoting the reduction of red and processed meat consumption. The study randomly allocated 180 young adults to three different message conditions. Participants in the informational condition read messages on the consequences of excessive red and processed meat consumption on one's health and the environment. Participants in the emotional condition read messages eliciting anticipated regret for the consequences of excessive red and processed meat consumption on one's health and the environment. Participants in the control condition read messages on the health and environmental consequences of sugar consumption. The chatbot system sent messages through a chatbot every morning for two weeks. Compared to the control condition, participants exposed to emotional messages reduced red and processed meat consumption at follow up, while this was not the case for participants exposed to informational messages. While the chatbot focus on the red and processed meat consumption, it does not include other aspect of dietary education and planning.

A critical synthesis of the randomized controlled trials reviewed reveals significant contradictions and collective limitations that directly motivate the present study. A fundamental conflict emerges between the findings of Piao et al. (2020), who demonstrated sustained habit formation over 12 weeks through chatbot-delivered cue-and-reward mechanisms, and Kramer et al. (2020), who reported a contrasting 30% dropout rate and low overall engagement in a comparable chatbot intervention of similar

duration. Kramer et al. (2020) explicitly attributed this failure to messages that were insufficiently personalized to individual needs, preferences, and fitness levels, a finding consistent with Kunzler et al. (2019), who identified participant-specific characteristics and contextual factors as significant determinants of chatbot receptivity. Together, these conflicting outcomes converge on a single explanatory variable: the inadequacy of generic, non-adaptive content delivery. Carfora et al. (2019) further demonstrated that emotionally framed messages outperformed informational ones in reducing processed meat consumption; however, this intervention addressed only a single dietary behavior and did not integrate exercise or broader nutritional planning. Critically, no reviewed RCT was conducted in Southeast Asia, and none incorporated culturally specific dietary patterns, limiting their applicability to Malaysian adults with overweight and obesity. The collective gap across these trials is the absence of a comprehensive, multi-variable, culturally adapted personalization mechanism. The present study addresses this gap directly through NExGEN-ChatGPT, which generates prompts from a validated 111-item assessment encompassing nutritional needs, physical fitness levels, medical history, cultural dietary preferences, and behavioral factors specific to the Malaysian population.

Table 2.3 provides details of the previous non-randomized studies on AI chatbots for recommendations in physical activity and nutrition. The study by Maher et al. (2020) evaluated the feasibility and preliminary efficacy of a 12-week physical activity and Mediterranean-style dietary intervention, MedLiPal, delivered via an AI virtual health coach. The study demonstrated excellent feasibility in terms of recruitment and retention (90% at 12 weeks) and reported significant improvements in participants' physical activity (+109.8 minutes·week^{-1}), Mediterranean diet scores (+5.7 points), body weight (-1.3 kg), and waist circumference (-2.1 cm), without significant blood pressure changes. These findings support further exploration of AI-driven interventions for health improvement at scale. However, Maher et al. (2020) concentrated on interventions based on the Mediterranean diet, whereas our study proposes a more tailored dietary approach that considers geographical location and cultural preferences.

Fadhil et al. (2019) outlines the development and assessment of CoachAI, a novel system that enhances health interventions by pairing with a conversational agent to aid caregivers' routine functions. Unlike tools that supplant human contact, CoachAI boosts engagement in health interventions via natural language dialogues handled by a

finite state machine automaton, providing physicians with valuable data on patient compliance. Evaluated through a month-long study on exercise, diet, and stress control, it proved the human-AI partnership's effectiveness in health coaching. The research also links higher user involvement to better adherence to prescribed activities, underscoring the significant role of conversational agents in health promotion and habit integration within daily communications, thereby enriching the socio-technical landscape of healthcare assistance. Future research should focus on refining and enhancing the activity recommender by using machine learning that incorporate dietary and emotional pattern.

Research by Stephens et al. (2019) assess the efficacy of Tess, a digital coaching chatbot, in supporting young patients with weight control and prediabetes. Deployed through SMS and Facebook Messenger, Tess delivers 24/7 personalized educational and supportive interactions. The pilot study with 23 adolescents (mean age = 15.20) demonstrated the chatbot's practicality, with a preference for texting to bolster treatment compliance and well-being. Adolescents demonstrated notable health improvements 81% of the time and exchanged 4,123 messages, leading to high perceived value (96% usefulness rating). These outcomes highlight the promise of AI in pediatric health support, suggesting scalability and potential to augment ongoing care. Future studies should focus on enhancing Tess's algorithms for deeper customization to individual needs and cultural contexts, and on evaluating its scalability for broader deployment across multiple platforms to maximize accessibility and impact in pediatric healthcare.

Casas et al. (2018) study the creation of a conversational agent crafted to assist individuals seeking to enhance their dietary practices. The agent employs a rapid and straightforward data collection technique, aimed at user convenience. It enables users to set personal dietary goals with daily monitoring, offering options to either decrease meat intake or increase fruit and vegetable consumption. In trials involving 36 individuals, success was modest, with 11% fully achieving their objectives. However, 65% of participants improved their dietary patterns. The agent demonstrates preliminary effectiveness in fostering improved food choices. Despite the low completion rate, the general trend towards better eating habits suggests the utility of the agent, warranting further development and customization to improve effectiveness and encourage wider usage for more substantial health benefits. Future studies should prioritize refining the conversational agent's user experience through feedback-driven

enhancements and broadening the food database to encompass a diverse spectrum of dietary preferences and cultural cuisines, ensuring accessibility and relevance.

Kocielnik et al. (2018) introduced the Reflection Companion, a mobile conversational tool designed to enhance interaction with fitness tracker data, thereby encouraging reflective thought. Current self-tracking tools often fail to prompt meaningful action from the data collected due to a lack of reflective design. To combat this, Reflection Companion delivers personalized daily conversations and visual data insights on users' smartphones, derived from a refined set of 25 dialogues based on 275 reflection-inducing questions sourced from user workshops. A fortnight trial with 33 Fitbit users showed the system's efficacy in fostering reflection, resulting in increased motivation, empowerment, and behavioral change. Notably, almost half continued its use post-study without incentives, indicating its utility. These findings illuminate the potential of dialog systems in promoting active engagement with personal health data. Future studies should explore the application of behavioral science to enhance dialogue design in fitness trackers, alongside the integration of advanced analytics for deeper interpretation of data, to drive more personalized insights and promote healthier outcomes.

Taken together, the non-randomized studies reviewed demonstrate preliminary evidence supporting AI conversational agents in dietary and physical activity promotion, yet reveal important inconsistencies and systemic limitations that the present study is designed to resolve. Maher et al. (2020) and Fadhil et al. (2019) both reported positive adherence outcomes; however, these studies employed non-randomized designs with small samples drawn predominantly from Western and European populations, substantially limiting generalizability to Southeast Asian contexts where dietary culture, food availability, and cultural attitudes toward weight management differ markedly. Directly conflicting adherence findings emerge from Casas et al. (2018) and Kocielnik et al. (2018): Casas et al. (2018) reported that only 11% of participants fully achieved their dietary goals, Fadhil et al. (2019) found that nearly half of participants continued using the Reflection Companion without incentives post-study, suggesting that reflective dialogue designs sustain engagement more effectively than explicit goal-setting mechanisms. Critically, none of the reviewed non-randomized studies incorporated geographically specific food databases or culturally adapted physical activity recommendations. Fadhil et al. (2019) and Stephens et al. (2019) addressed distinct populations like caregivers and adolescents, respectively

making direct comparison with the adult obesity management context of the present study methodologically inappropriate. A further gap common to all reviewed studies is their reliance on pre-defined dialogue flows or single-domain recommendations, without dynamic generation of personalized content from a comprehensive, validated multi-variable user assessment. The present study resolves this through NExGEN-ChatGPT's expert-validated prompt generation framework, representing an architectural departure from all reviewed prior systems and addressing the primary unmet need identified in this literature.

Table 2.2
Overview of Randomized Controlled Trials on AI-based Chatbots for Recommendations in Physical Activity and Nutrition

First Author (Year)	Intervention Duration Usage Frequency	Technological Features	Chatbot System Initiation	System Capacity
Kramer, J. (2020)	6 weeks Daily	The chatbot app was developed using the MobileCoach platform.	Daily and weekly messages delivered at a random time between 10 AM and 6 PM/Not reported	Setting personalized goals, using action-planning, coping planning, and self-monitoring prompts.
Kunzler, F. (2019)	6 weeks Daily	The chatbot app is based on the MobileCoach platform. Physical activity was measured by subscribing to CoreMotion Activity Manager on iOS and Google Activity Recognition API	Daily messages delivered 3 times a day (between 8 and 10 AM, 10-6 PM, or 8 PM) and weekly messages delivered at random times	Using persuasive prompts such as goal setting, selfmonitoring, goal achievement, and weekly planning.
Piao, M. (2020)	12 weeks Daily	The chatbot was developed using the Watson Conversational tool (IBM Corp) and was linked to the KakaoTalk Smart Chat application programming interface. It was deployed through KakaoTalk messenger app	Daily message delivered at participants' specified time (on demand).	Setting personalized goals, providing extrinsic (e.g., financial incentives) and intrinsic (e.g., pleasure, satisfaction) were used.
Carfora, V. (2019)	2 weeks Daily	The chatbot was deployed through Facebook Messenger.	Daily messages delivered at 7:30 AM	Informing participants about the health and environmental impact of excessive RPMC; Using persuasive message appeal.

Table 2.3
Overview of Non-randomized Studies on AI-based Chatbots for Recommendations in Physical Activity and Nutrition

First Author (Year)	Intervention Duration Usage Frequency	Technological Features	Chatbot System Initiation	System Capacity
Maher, C. (2020)	12 weeks Weekly	The chatbot was developed using the IBM Watson Virtual Assistant AI software and was hosted on the cloudbased instant messaging platform Slack. The program also used the MedLipal website and the Garmin Vivofit4 physical activity monitor.	Weekly messages (on demand)	Assisting participants to set a personal daily step count goal based on age-based normative + values 2000 steps. This daily step goal was revisited and edited at each weekly check-in.
Fadhil, A. (2019)	3 weeks Daily	The chatbot used the coaching portal combined with the dialogue engine which consists of rails state machine, user clustering model and fitbit wearable data. The chatbot was deployed through Telegram Messenger.	Messages delivered recurrently (e.g., daily, weekly, on weekends, or weekdays) or at a scheduled date and time.	Providing motivational messages consisted of positive reinforcement and a feedback on user's overall adherence for the week to increase adherence
Stephens, T. (2019)	10–12 weeks Not reported	The chatbot was deployed through multiple channels (i.e., SMS text message, Slack, WhatsApp or Facebook Messenger) and also was integrated with Google Home and Amazon Alexa.	Not reported demand.	Specific goals and targeted behaviors were entered to the system to offer individualized conversations. Delivered customized integrative support, psychoeducation, and interventions
Casas, J. (2018)	Seven days Daily	The chatbot was developed using the Chatfuel service, and deployed through Facebook Messenger.	Daily messages	Asking participants to choose their goals, monitoring food intake, answering questions, and giving recommendations
Kocielnik, R. (2018)	2 weeks Daily	The chatbot app was developed with Twillio API used to communicate with users via mobile phones through SMS/MMS. Fitbit API was queried for user activity data to generate activity graphs. LUIS API offered automated recognition of free-text user responses.	Daily messages	The mini dialogues were delivered with a graph showing user's physical activity metrics

2.14 Summary

This chapter synthesised the current evidence base across six interrelated domains: the obesity epidemic, conventional and digital weight management, AI chatbot interventions, prompt engineering for health AI, quality assessment instruments, expert consensus methods, and outcome measurement. Several critical observations emerge from this synthesis.

Across conventional and digital interventions, a consistent pattern is evident: effect sizes decrease as delivery intensity decreases. Face-to-face lifestyle programmes achieve 5 to 10% weight loss but cannot scale; eHealth tools produce modest weight reductions of approximately 1 to 3 kg but lose personalisation depth; early AI chatbots improve engagement but rely on rule-based logic that cannot adapt to individual clinical complexity. This progressive dilution of effect with scale is the central problem this thesis addresses

Studies reviewed on AI chatbot interventions (Kramer et al., 2020; Kunzler et al., 2019; Piao et al., 2020; Maher et al., 2020) consistently attribute attrition and reduced engagement to insufficient personalisation. However, none of these systems addressed the upstream problem: the absence of a validated mechanism for translating individual clinical data into structured AI prompts. The distinction is critical. Prior chatbots personalised the delivery of pre-defined content; the NExGEN-ChatGPT system personalises the content itself through a prompt engineering framework validated against domain-specific clinical criteria. No prior study reviewed has employed this architecture.

The quality assessment literature reveals a structural mismatch: instruments such as the MARS and QAMAI evaluate static application features or post-hoc output quality, but neither evaluates the prompt engineering methodology that determines output quality in the first place. The NExGEN validation addresses this gap by establishing upstream criteria validated through expert consensus, shifting quality assurance from output review to input design. This represents a fundamentally different evaluation paradigm from any prior instrument reviewed.

The expert consensus literature confirms that Fuzzy Delphi Method is appropriate for domains where clinical uncertainty and expert judgment are primary evidence sources, and that a 21-member multidisciplinary panel meets established size and diversity recommendations. However, no prior application of the FDM reviewed

here has targeted a prompt engineering framework for an AI-driven obesity intervention, making this thesis methodologically novel within that tradition.

Taken together, the literature identifies four convergent gaps: lack of validated prompt mechanisms, inadequacy of current assessment tools for dynamic AI outputs, absence of domain-specific criteria for AI-generated obesity recommendations, and limited comparative trial evidence for personalised versus generic AI prompt engineering. The research programme reported in this thesis directly addresses all four gaps through a sequential development, validation, and effectiveness design.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Weight management in adults with overweight and obesity requires integrated dietary and exercise interventions. Traditional one-to-one counseling remains clinically effective, producing 5–10% body weight reduction over 6–12 months, yet these approaches are resource-intensive and difficult to scale (Aggarwal et al., 2023; Wharton et al., 2020). Digital health tools and AI-powered chatbots, including ChatGPT, offer a more scalable alternative; however, their effectiveness depends critically on the quality and specificity of user-provided prompts (Oh et al., 2021; Ponzo et al., 2024). Documented risks of AI hallucination, inconsistent outputs, and inadequate clinical personalization further limit current applications in obesity management (Aljamaan et al., 2024; Omar et al., 2025). No validated mechanism currently exists for generating obesity-specific, user-targeted prompts for large language model chatbots (Iqbal et al., 2021).

This chapter describes the research methodology employed across the two sequential studies that constitute this thesis. Study 1 details the development and expert validation of the NExGEN-ChatGPT personalized prompt generator framework using the Fuzzy Delphi Method (FDM) with a multidisciplinary expert panel. Study 2 describes the design, participant recruitment, intervention procedures, outcome measurements, and statistical approach of the 24-week randomized controlled trial evaluating the effectiveness of the validated NExGEN-ChatGPT system against a structured manual ChatGPT control programme. Together, the two studies address the identified research gaps by establishing the first expert-validated prompt engineering instrument for obesity-specific AI interventions and generating clinical trial evidence for personalized versus generic prompt-driven weight management.

3.2 Validation of a Personalized AI Prompt Generator (NExGEN-ChatGPT) for Obesity Management Using Fuzzy Delphi Method (Study 1)

3.2.1 Research Design

This study adopted a structured quantitative design, employing the FDM to systematically refine and validate the NExGEN-ChatGPT prompt generation framework for delivering personalized fitness and nutritional recommendations. This study has been published online in the Journal of *Biology Methods and Protocols*, *Oxford University Press* (*Web of Science and Scopus-indexed, SJR Quartile 1, Impact Factor: 1.3)*. The FDM was specifically selected over alternative consensus methods on three grounds. First, compared to traditional Delphi, FDM achieves consensus in a single data collection round by mathematically quantifying the distribution of expert opinion through triangular fuzzy numbers, thereby eliminating the need for multiple iterative rounds and reducing expert fatigue and attrition (Murry & Hammons, 1995). Second, compared to nominal group technique and expert panel discussions, FDM produces objective, reproducible consensus metrics through a defuzzification process, preventing the social influence dynamics and dominant-voice bias that can distort group discussion outcomes (Chang et al., 2000). Third, compared to the analytic hierarchy process, FDM accommodates large numbers of assessment criteria without requiring pairwise comparisons, making it practically feasible for the 111-criteria framework evaluated in this study (Jakubczyk & Kamiński, 2017).

The specific suitability of FDM for this study is further justified by three contextual factors. First, the NExGEN framework spans five interdisciplinary domains in which experts hold differing degrees of certainty and apply domain-specific terminology to evaluate the same construct; fuzzy triangular numbers capture this variation rather than forcing false precision onto ordinal scale responses. Second, the complexity and volume of the validation task (111 criteria across six constructs) render multi-round iterative Delphi impractical within the study timeline. Third, FDM has been applied and validated in analogous health technology and e-health tool validation studies, establishing methodological precedent for its use in this context (Roldán López et al., 2021; Yin & Hanif, 2024). In this approach, experts' qualitative judgments, initially captured using a seven-point Likert scale, were translated into quantitative Triangular Fuzzy Numbers represented by minimum (m1), moderate (m2), and

maximum (m3) values. This preserved the nuanced and varying levels of agreement among experts while providing quantifiable consensus metrics for each criterion.

3.2.2 Overview of NExGEN Workflow

The Nutrition and Exercise Planning Artificial Intelligence Prompt Generator, or NExGEN, is a web-based platform specifically developed to generate personalized dietary and exercise planning prompts for integration with AI-driven chatbots. This tool aims to support individuals in attaining their health and weight management goals through tailored nutrition and exercise plans. NExGEN is designed to support individualized health and weight management goals by systematically incorporating diverse user-specific factors such as nutritional needs, fitness levels, anthropometric data, medical history, socioeconomic status, geographic location, stress levels, sleep patterns, and motivational aspects into its framework.

The development of the NExGEN system took into account several interpersonal variabilities relevant to nutrition and exercise planning, including nutrition, fitness status, anthropometrics, demographics, socioeconomics, medical history, physiological health, geographical location, sleep, stress, and motivational factors to help the user attain their personal health and weight management goals. Implemented as a responsive web application with intuitive user interface design, NExGEN produces highly tailored prompts designed explicitly for large language models such as ChatGPT, thereby facilitating precise AI-generated recommendations.

The completed prompt comprises personalized directive input intended to be used with large language model chatbots like ChatGPT. Using the AI decision-making capabilities of ChatGPT, the integration (NExGEN-ChatGPT) aims to facilitate prompt generation, enabling the efficient creation of detailed exercise routines and nutritional plans that promote user autonomy and sustained adherence to lifestyle changes. This integration aligns with Self-Determination Theory, where autonomy is supported through individualized choice architecture and recommendations that adapt to personal preferences, dietary habits, and environmental constraints (D. Ryan & Yockey, 2017). This synergy is expected to enhance user autonomy in maintaining personalized and continuous health strategies.

The NExGEN-ChatGPT framework employs end-to-end encryption and General Data Protection Regulation GDPR-equivalent compliance protocols. Data

minimization principles ensure only essential information is transmitted, with sensitive data remaining client-side until authorized. Users maintain complete control through explicit consent mechanisms, addressing ethical concerns and regulatory requirements for AI-driven healthcare applications. Figure 3.1 illustrates the architecture framework of the NExGEN-ChatGPT system integration.

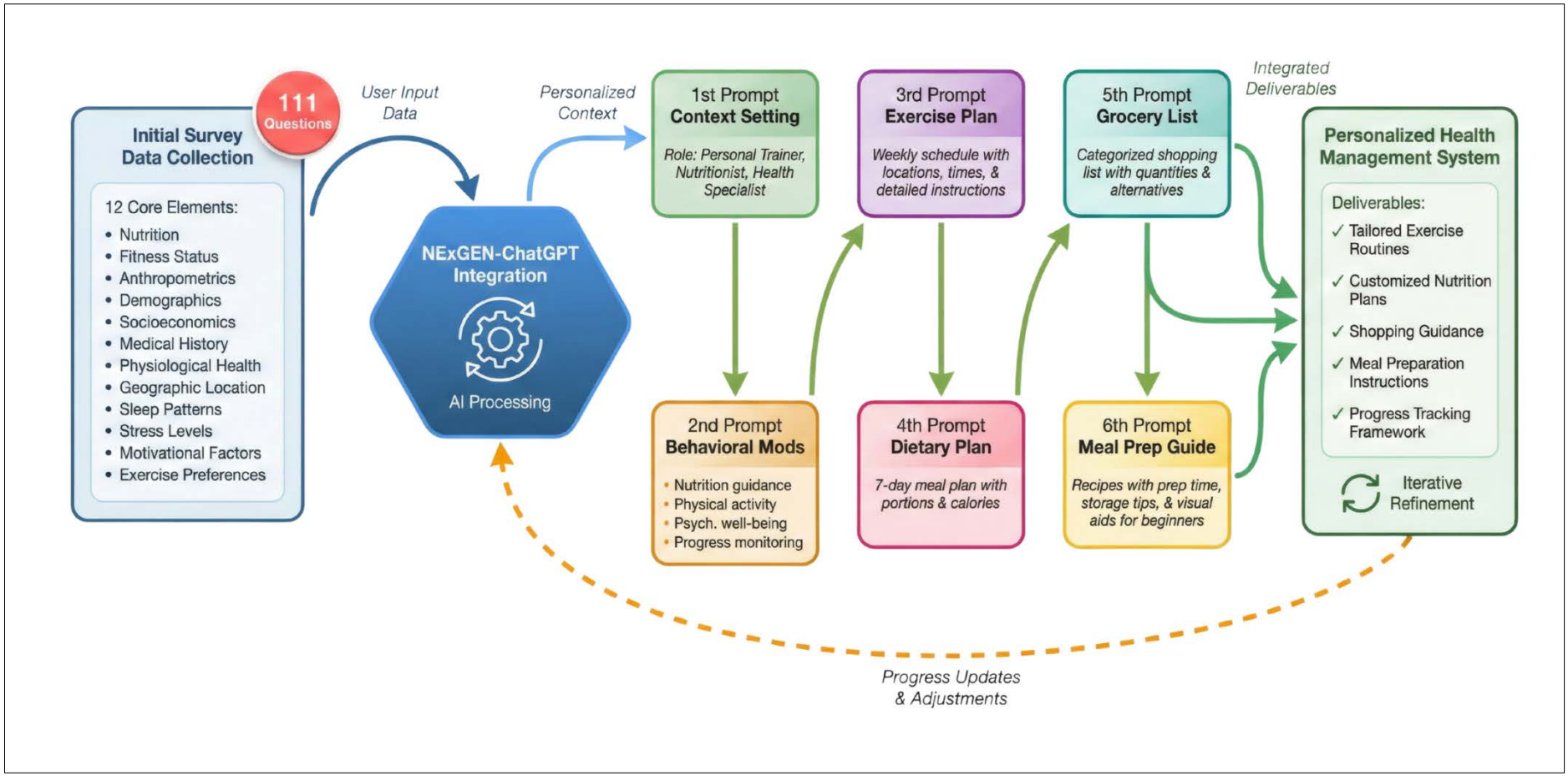


Figure 3.1 Architecture Framework of NExGEN-ChatGPT System Integration

3.2.3 NExGEN-ChatGPT System Design and Process Flow

NExGEN is a web-based application designed to generate personalized nutrition and exercise plans through structured user profiling and AI-powered prompt engineering. The starting interface of the NExGEN system is shown in Figure 3.2. The system operates as an intermediary between the user and a large language model of their choice, such as ChatGPT, Gemini or Claude. Rather than embedding a proprietary AI engine, NExGEN collects comprehensive health data through a guided questionnaire and transforms user responses into optimized prompts that can be submitted to any compatible AI chatbot. This architectural decision decouples the data collection layer from the plan generation layer, allowing the framework to remain model-agnostic.

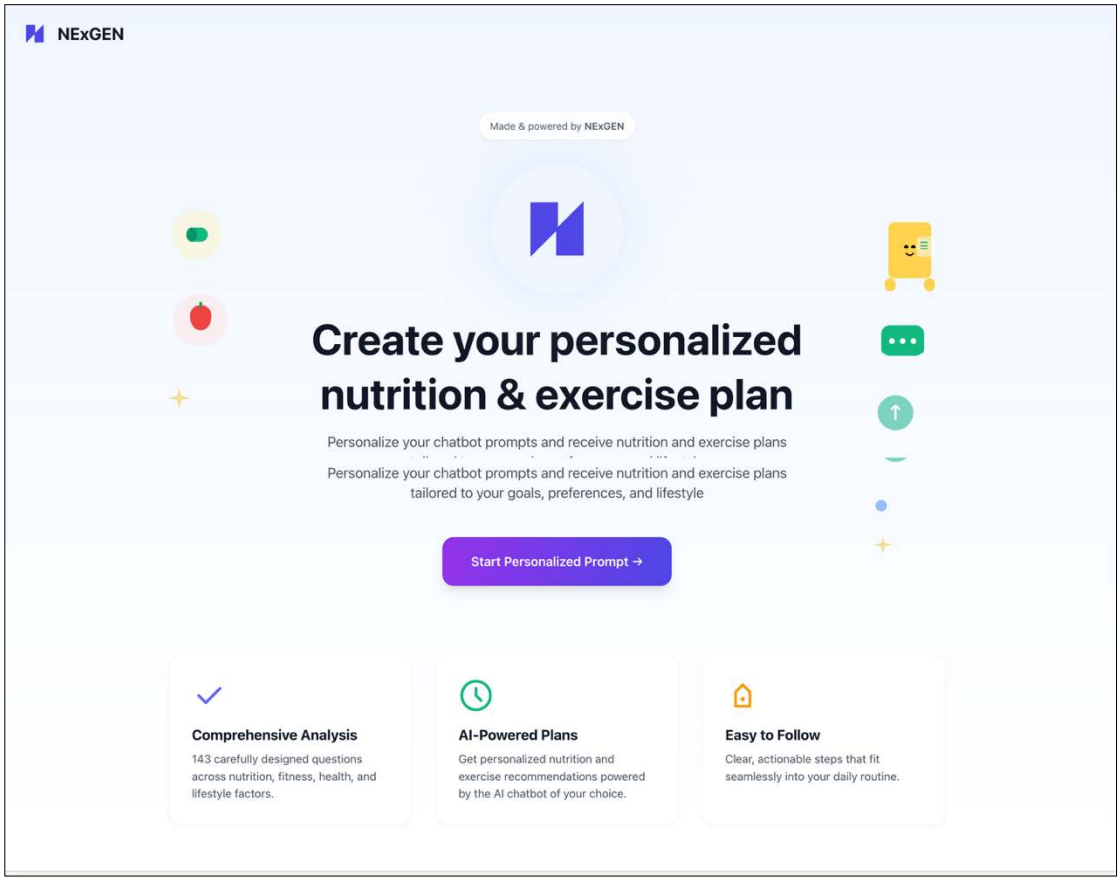


Figure 3.2 Starting interface of the NExGEN system

The data collection process is organized into ten sequential sections comprising 111 questions. These sections span goal setting, basic demographic and physical characteristics, physiological measurements, women-specific health considerations,

medical history, sleep patterns, stress assessment, physical activity levels, exercise preferences, and dietary intake. Each question is accompanied by a contextual hint that explains its relevance to the personalization process. All questions are optional, and users may proceed to the next section at any point. This design balances comprehensiveness with user flexibility, reducing abandonment risk while still enabling granular personalization when users provide detailed responses.

The goal setting section captures the user's desired weight management outcome, target weight change, timeline, start date, willingness to adjust timelines, and underlying motivation. The goal setting interface is presented in Figure 3.3. The demographic and physiological sections collect age, height, weight, gender, place of residence, occupation, income bracket, ethnicity, religious affiliation, and disability status, as illustrated in Figure 3.4. Physiological measurements extend to body fat percentage, lean body mass, waist and hip circumference, resting and maximum heart rate, VO2 max, blood pressure, fasting blood glucose, and cholesterol levels. The physiological measurements interface is shown in Figure 3.5. These data points allow the generated prompts to account for metabolic baselines and clinical risk factors.

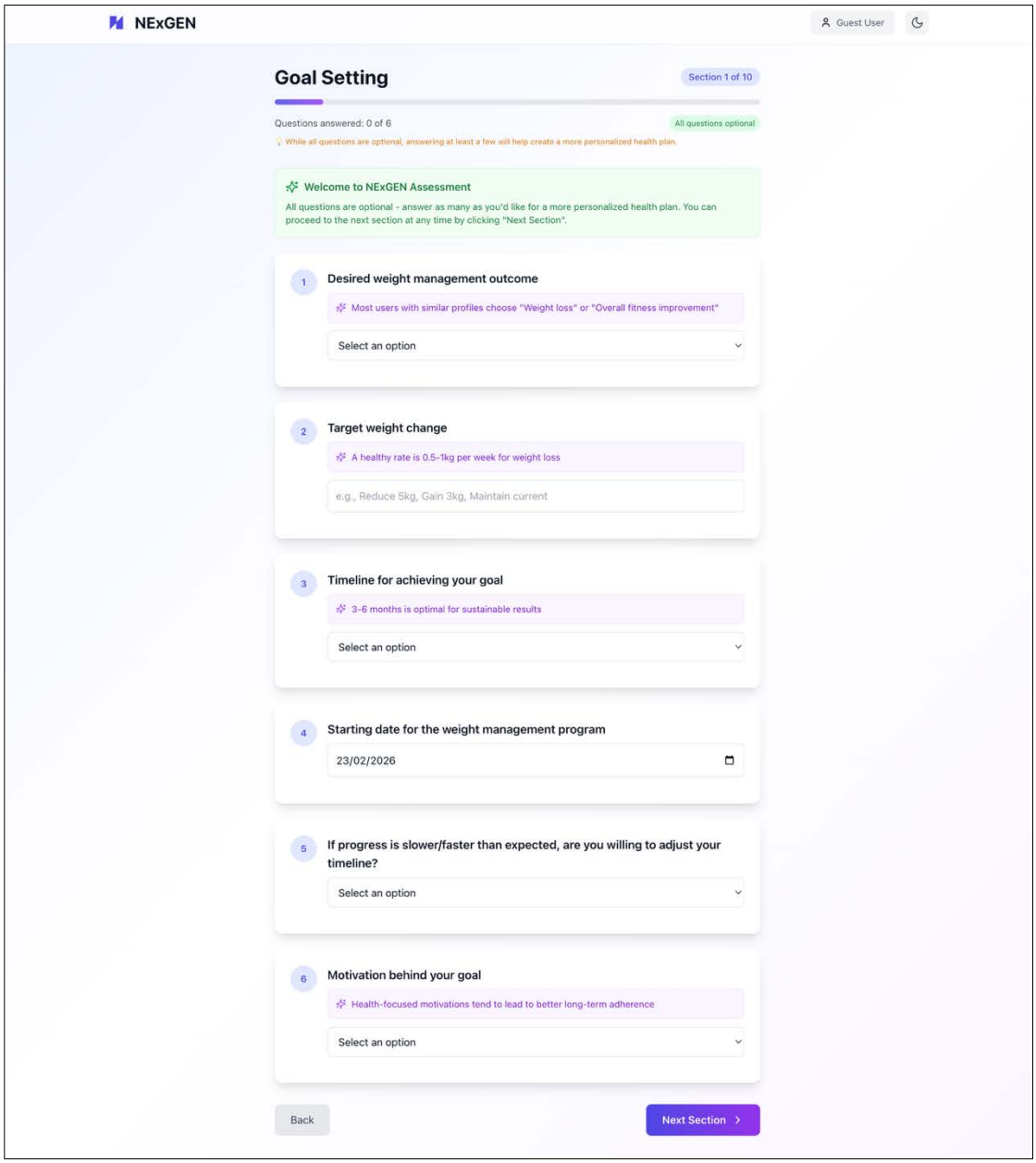


Figure 3.3 Goal Setting Interface of the NExGEN System

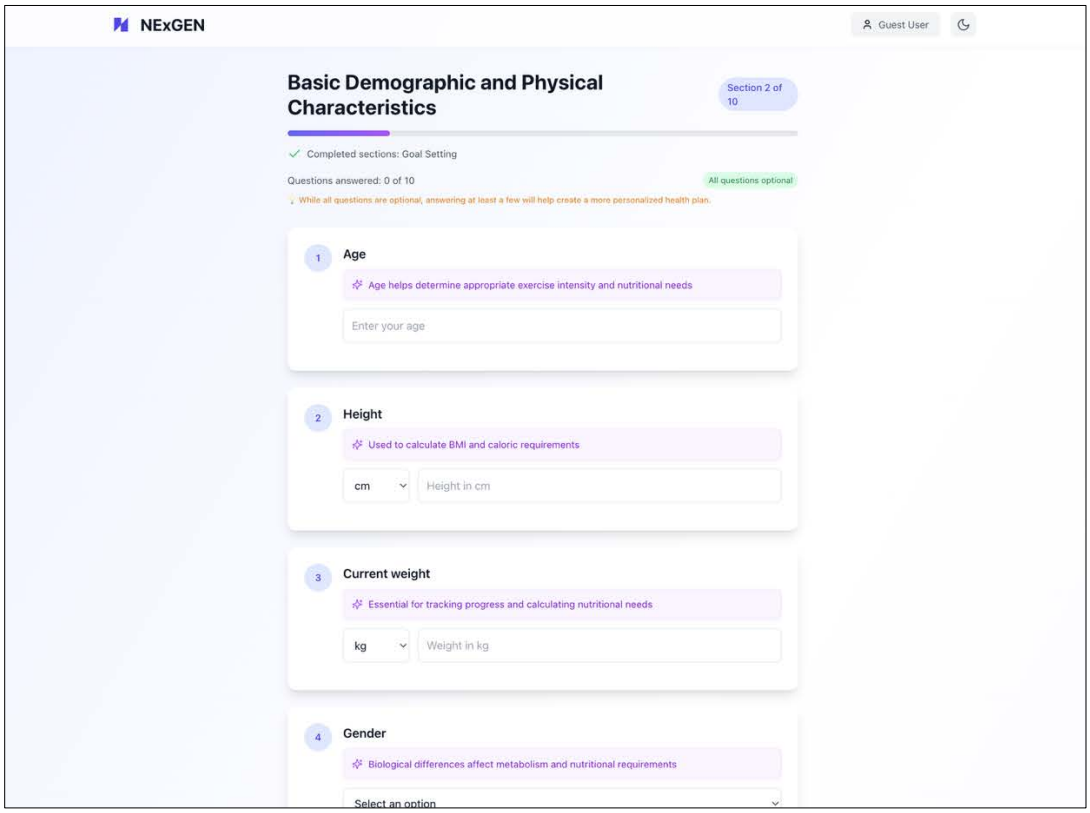


Figure 3.4 Snippet from Demographic and Physical Characteristics Interface of NExGEN

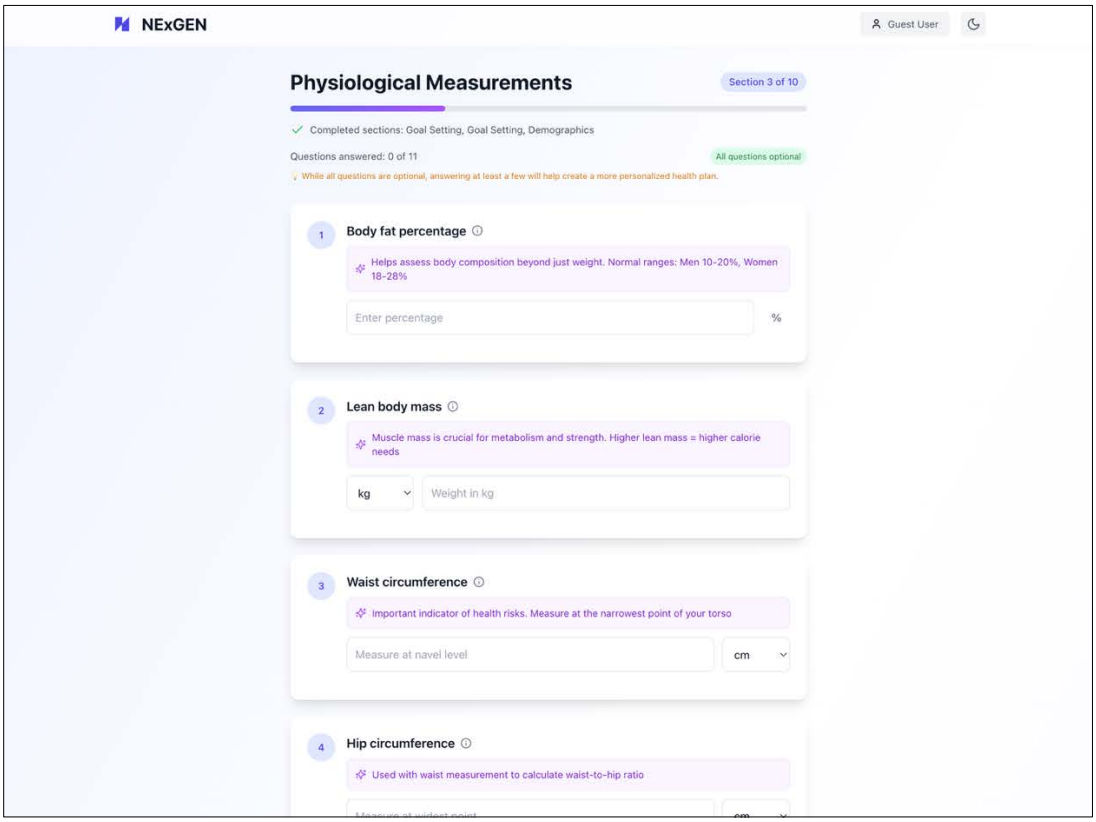


Figure 3.5 Snippet from Physiological Measurements Interface of NExGEN

For female users, a dedicated section addresses pregnancy status, breastfeeding, menstrual cycle regularity, menopause, hormonal contraceptive use, and postpartum considerations, as shown in Figure 3.6. The medical history section, presented in Figure 3.7, addresses diagnosed conditions, current medications, allergies, surgical history, hospitalizations, and specific concerns including chest pain during activity, respiratory issues, cardiovascular problems, family disease history, musculoskeletal limitations, and digestive intolerances. Sleep patterns are assessed through bedtime, sleep latency, nocturnal awakenings, wake time, subjective sleep quality, sleeping environment ratings, napping habits, and sleep aid usage. The sleep patterns interface is illustrated in Figure 3.8

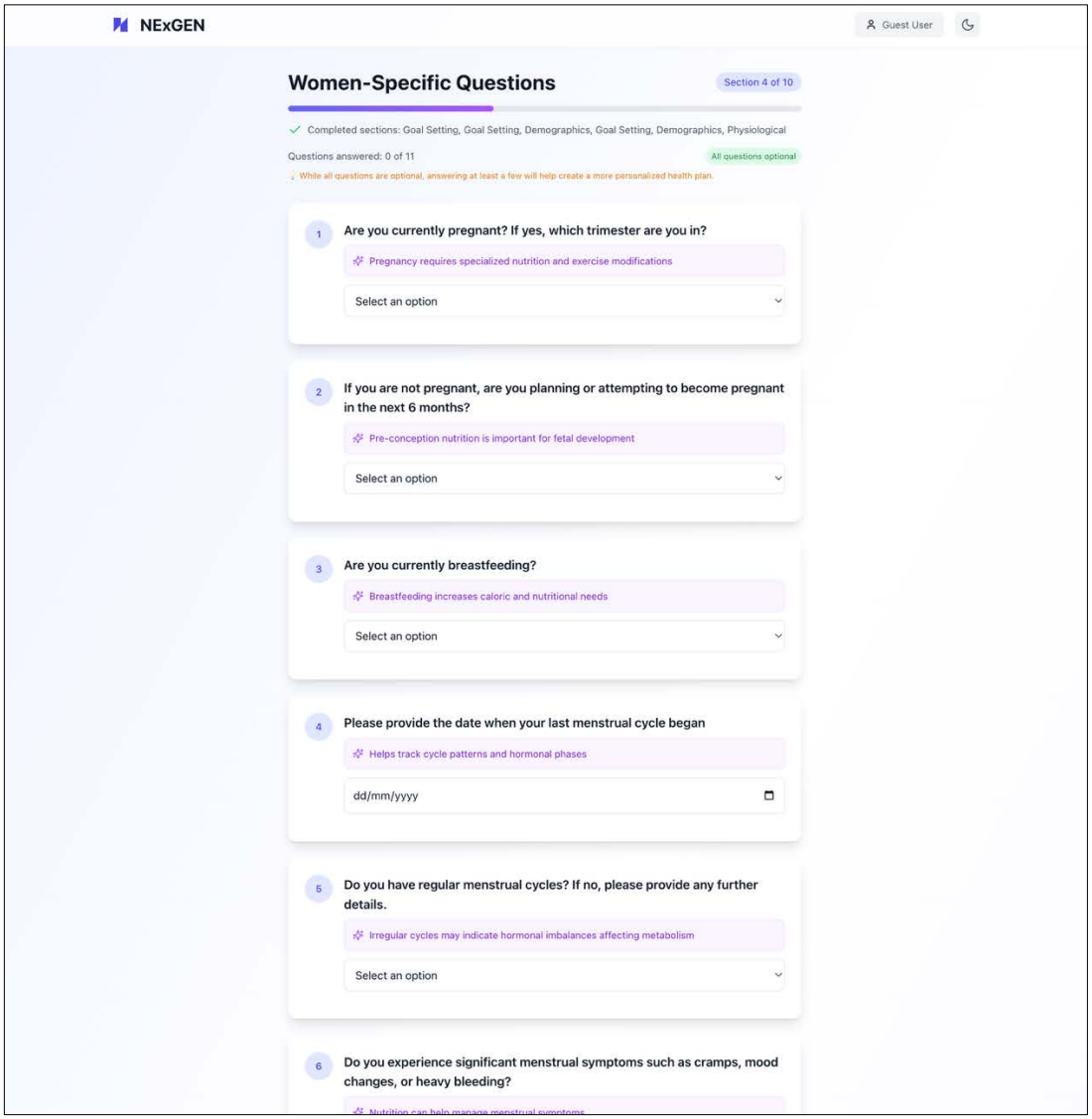


Figure 3.6 Snippet from Women-Specific Survey Interface of NExGEN

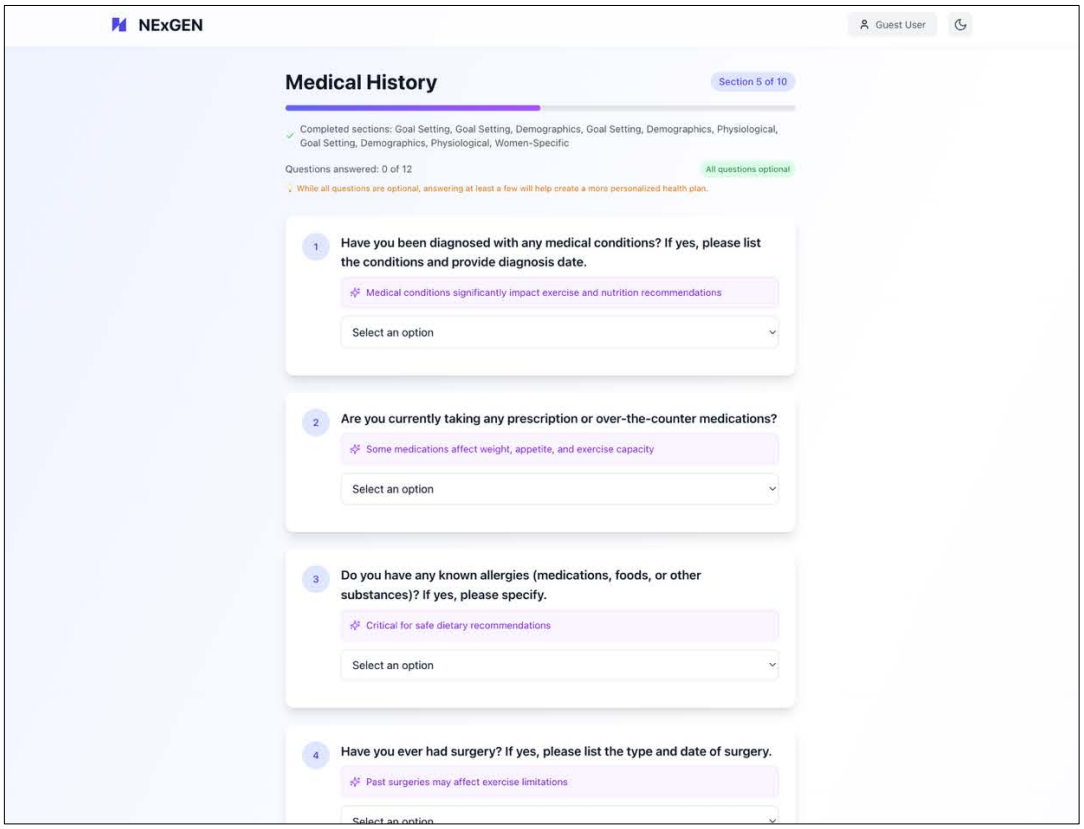


Figure 3.7 Snippet from Medical History Section Interface of NExGEN

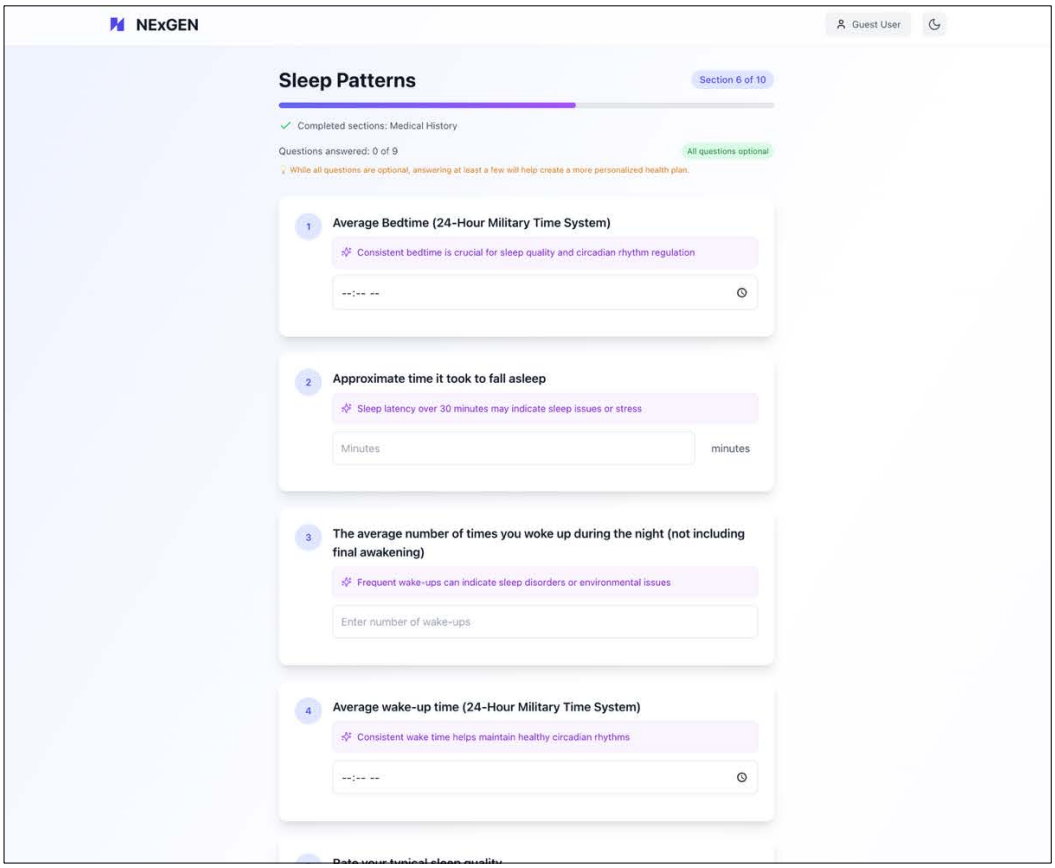


Figure 3.8 Snippet from Sleep Pattern Section Interface of NExGEN

Stress levels are measured using a validated scale with nine items rated from 0 (never) to 4 (very often), capturing perceived control, coping ability, and emotional reactivity over the preceding month. The stress assessment interface is shown in Figure 3.9. Physical activity assessment follows the International Physical Activity Questionnaire format, measuring vigorous activity days and duration, moderate activity days and duration, walking frequency and duration, and sedentary time, as presented in Figure 3.10

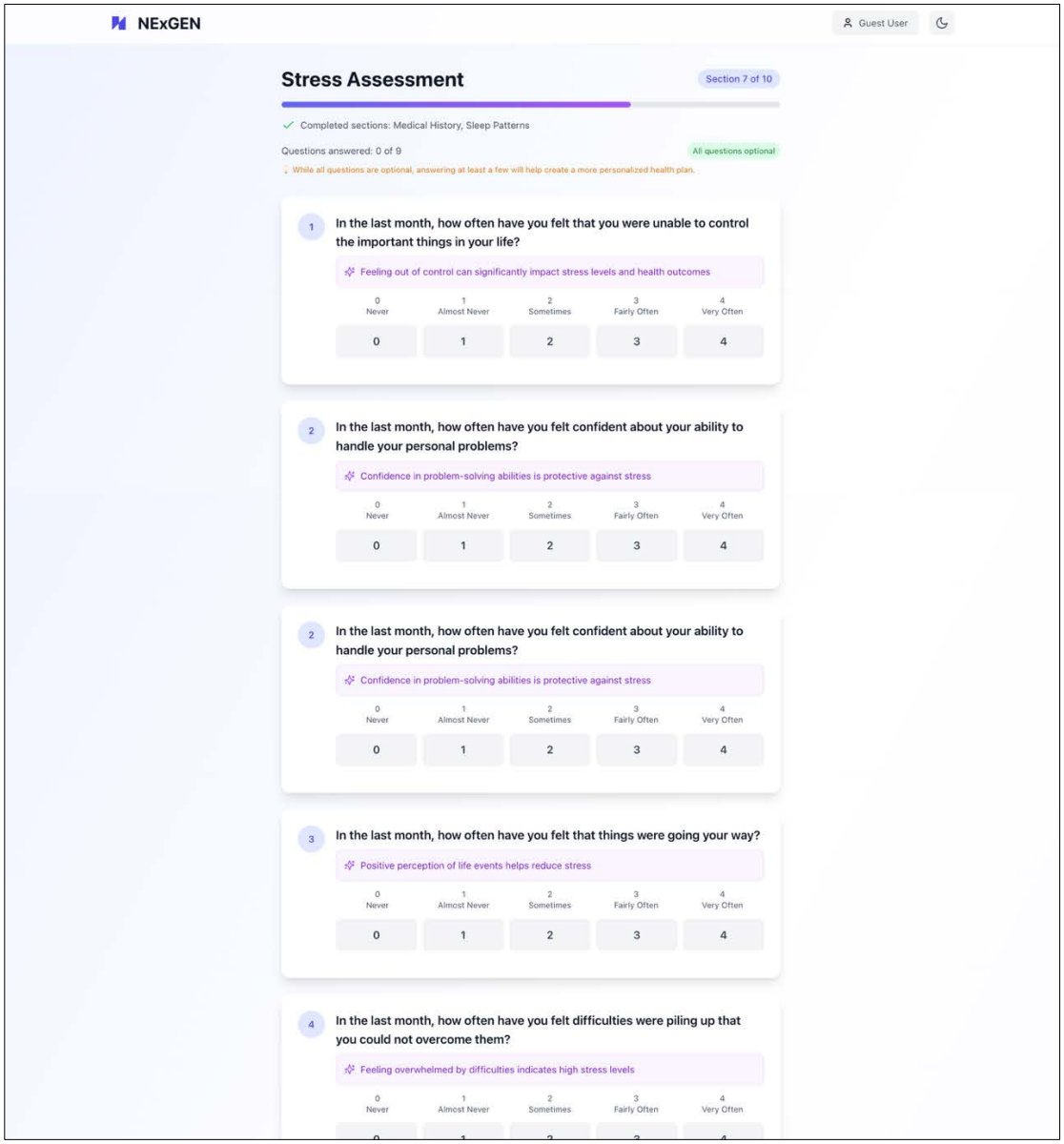


Figure 3.9 Snippet from Stress Assessment Section Interface of NExGEN

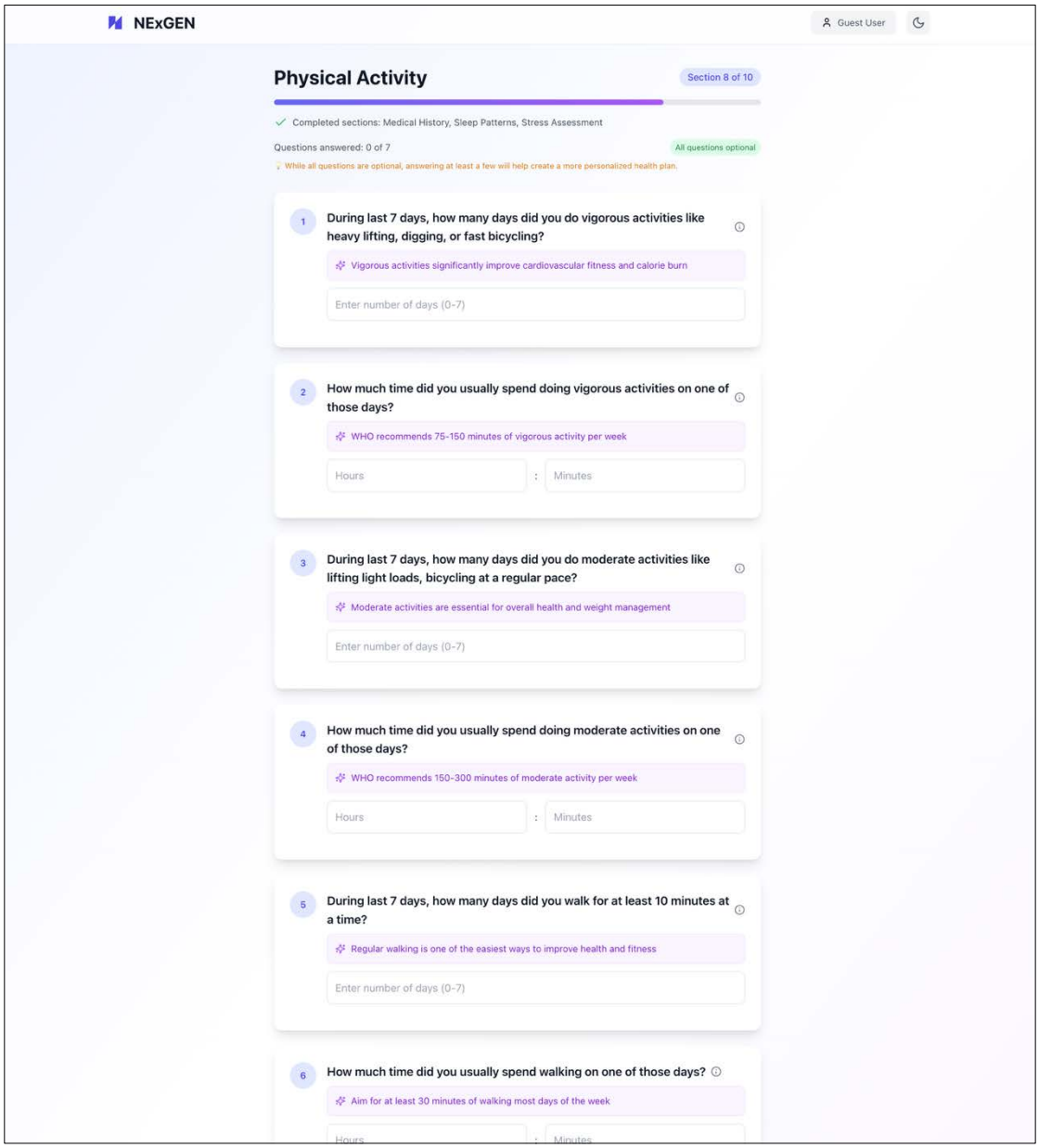


Figure 3.10 Snippet from Physical Activity Section Interface of NExGEN

Exercise preferences capture available days, preferred times, session duration, equipment access, gym availability, location suitability for outdoor exercise, venue preferences rated on a five-point scale, trainer preference, exercise motivations, enjoyed activities, disliked activities, injury history, and perceived barriers. The exercise preferences interface is illustrated in Figure 3.11. The dietary intake section, shown in Figure 3.12, addresses meal frequency, meal skipping, dietary restrictions, food preferences and aversions, specific diet adherence, and nutrition label reading habits, among other items.

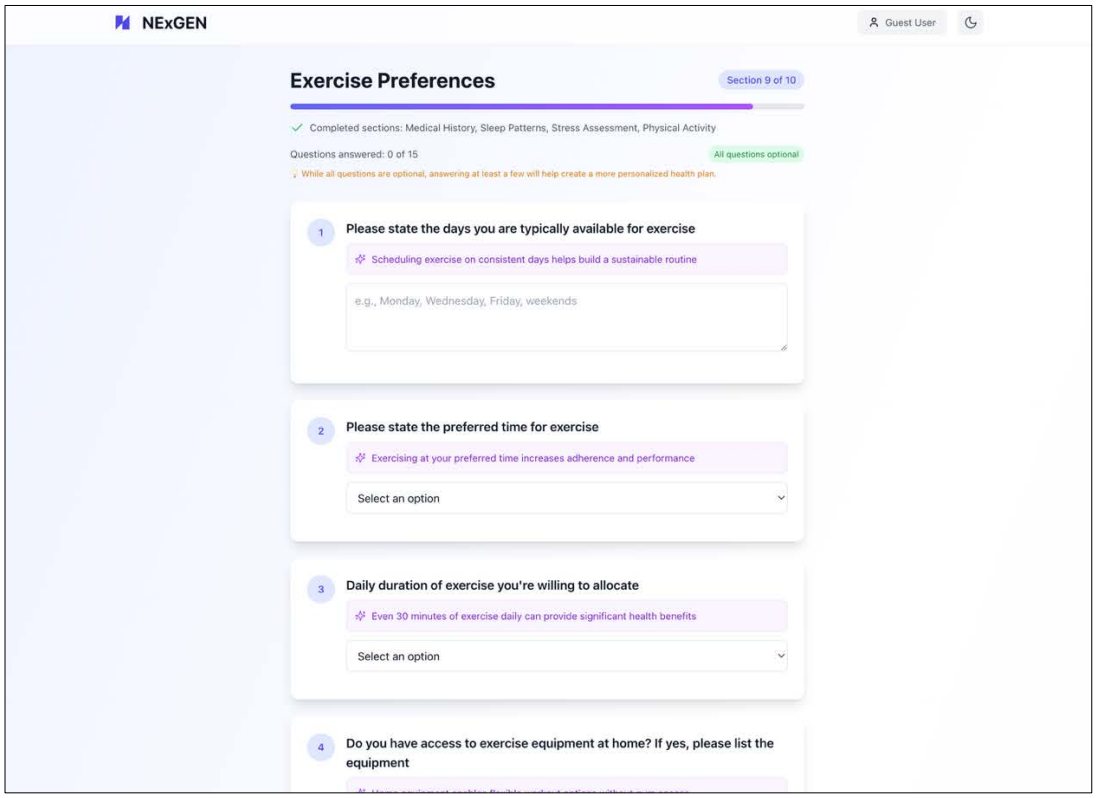


Figure 3.11 Snippet from Exercise Preferences Section Interface of NExGEN

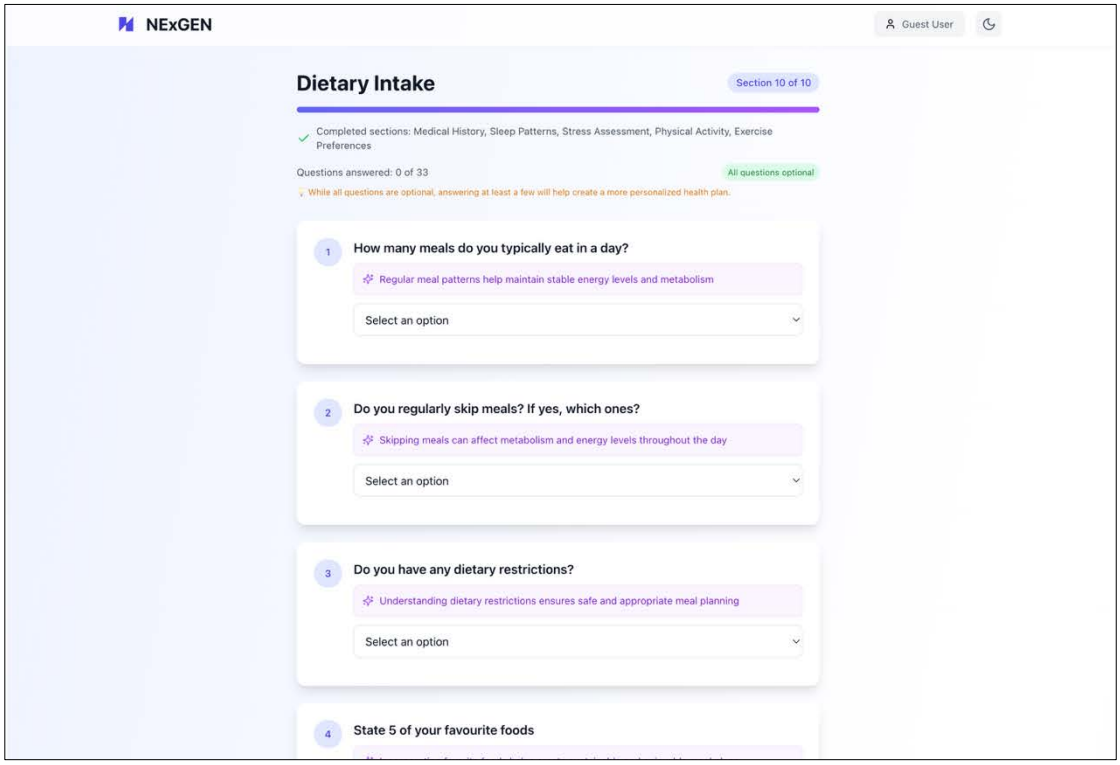


Figure 3.12 Snippet from Dietary Intake Section Interface of NExGEN

Upon completion of the questionnaire, NExGEN generates a structured output sequence of seven steps. Step 1 serves as the foundational prompt in the NExGEN output sequence. It compiles all user responses from the ten-section questionnaire into a single, structured health profile prompt. This prompt instructs the AI chatbot to assume the role of a qualified personal trainer, nutritionist, and health specialist. It then presents the user's data organized by section, covering goal setting, demographics, physiological measurements, medical history, sleep patterns, stress levels, physical activity, exercise preferences, and dietary intake. The user copies this prompt and pastes it into their preferred AI chatbot to establish the session context. All subsequent prompts reference the data introduced in this step. The comprehensive health profile prompt interface is shown in Figure 3.13

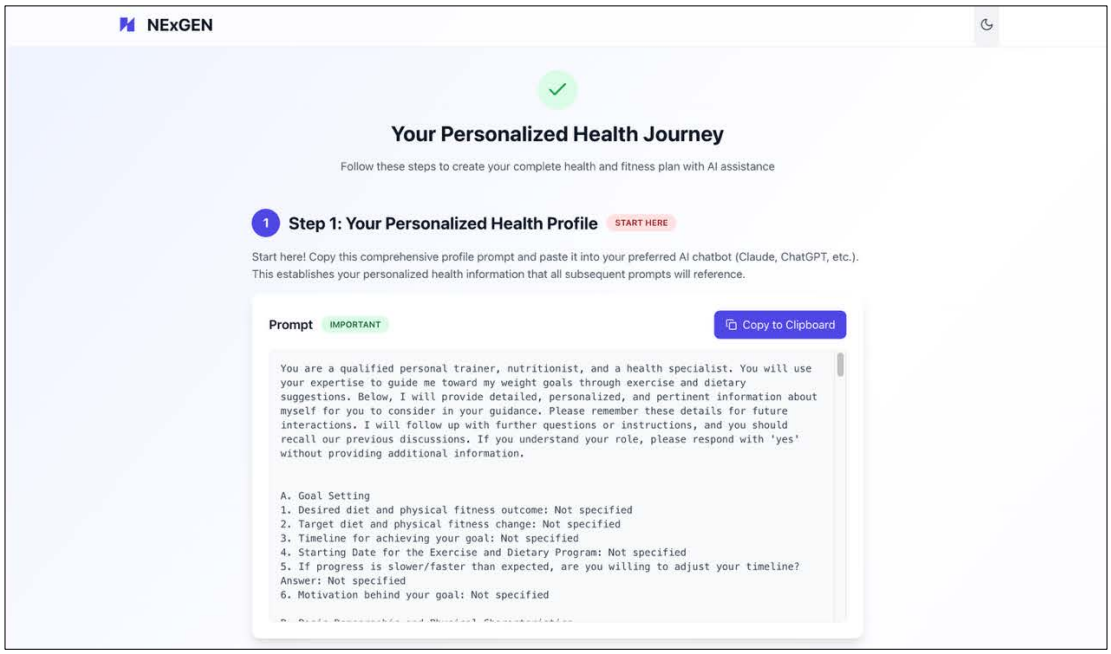


Figure 3.13 Comprehensive Health Profile Prompt of NExGEN

Step 2 is an optional prompt that requests the AI chatbot to evaluate the user's current health status based on the profile established in Step 1. The prompt instructs the AI to provide personalized observations and improvement suggestions grounded in the user's physical, physiological, and psychological data. It explicitly directs the AI to present advice in an accessible format and to avoid opinionated or biased recommendations. Users who prefer to proceed directly to plan generation may skip this step. The health assessment and recommendations prompt is presented in Figure 3.14.

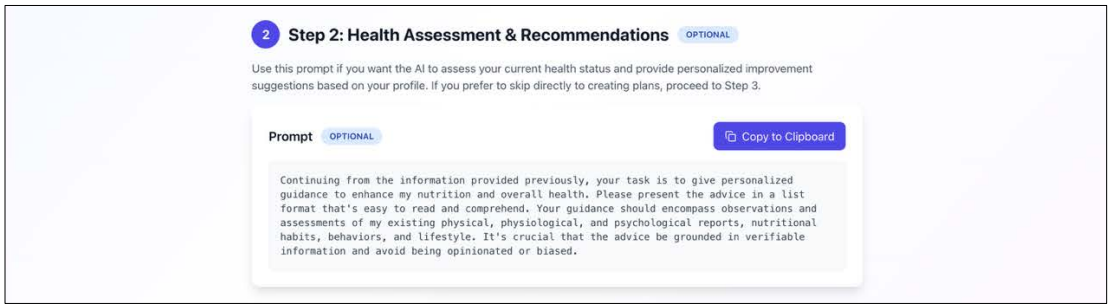


Figure 3.14 Personalized Health Assessment and Recommendations Prompt of NExGEN

Step 3 allows the user to select the duration of their personalized plans. Three options are available: one week, two weeks, or one month. This selection determines the scope and structure of the exercise and nutrition plan prompts generated in Steps 4 and 5. The plan duration selection interface is illustrated in Figure 3.15.

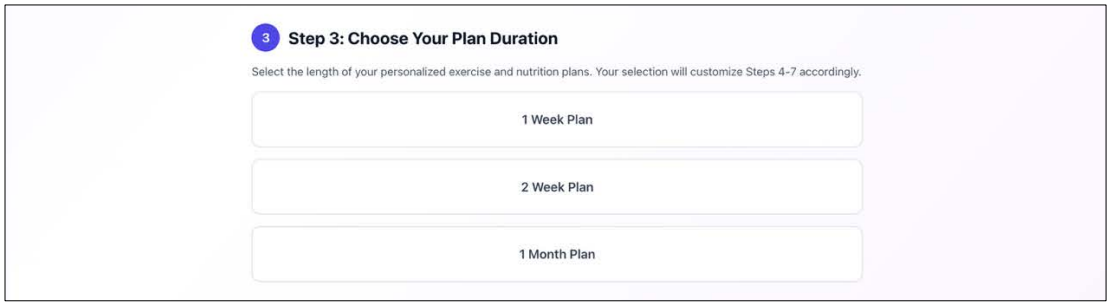


Figure 3.15 Plan Duration Selection Interface of NExGEN

Step 4 generates the exercise plan prompt. The prompt requests a comprehensive exercise plan tailored to the selected duration and instructs the AI to account for the user's demographic profile, physiological measurements, and exercise preferences. It specifies that the plan should be presented in table format and should include recommended locations, specific times, and detailed instructions for each exercise. The exercise plan prompt is shown in Figure 3.16.

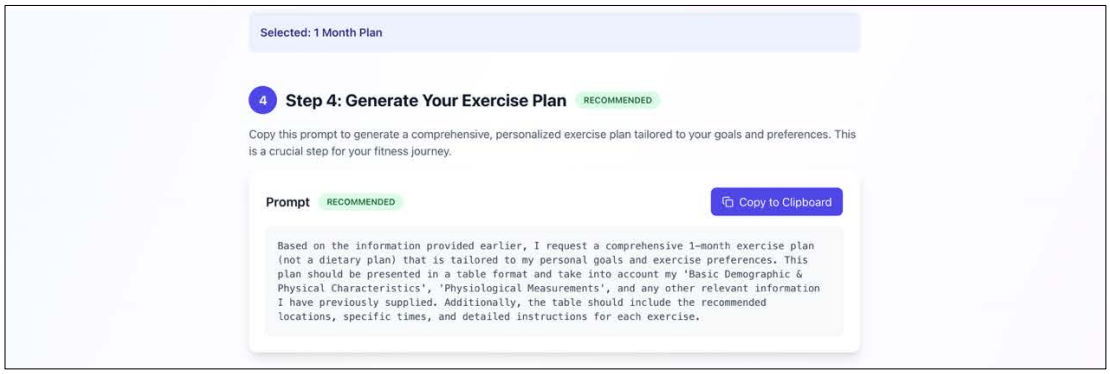


Figure 3.16 Personalized Exercise Plan Prompt of NExGEN

Step 5 generates the nutrition plan prompt. This prompt requests a detailed meal plan covering each day of the selected duration, customized according to the user's goals from the goal setting section and dietary intake data. It instructs the AI to reflect the user's dietary preferences, physical characteristics, and physiological measurements. The prompt further specifies that each meal suggestion should include recommended portion sizes, calorie estimates, and food choices accessible based on the user's place of residence. The nutrition plan prompt is presented in Figure 3.17.

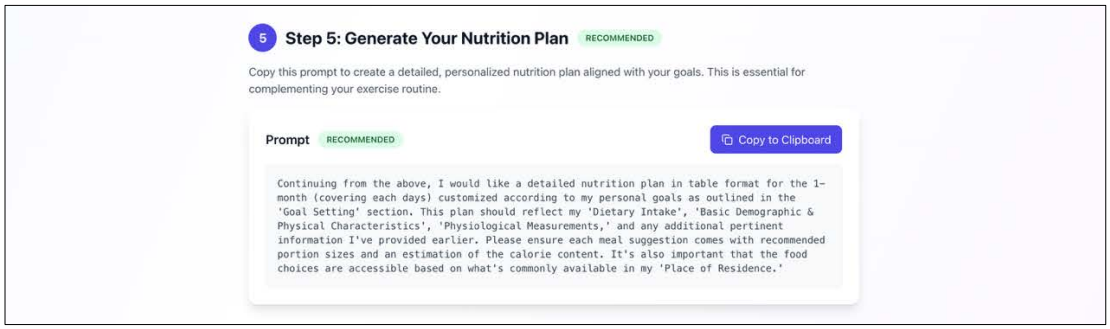


Figure 3.17 Personalized Nutrition Plan Prompt of NExGEN

Step 6 is an optional prompt that generates a grocery shopping list based on the nutrition plan produced in Step 5. The prompt instructs the AI to list all required ingredients with quantities matched to the plan's portion sizes and calorie estimates, covering the full plan duration. It requests that alternatives be provided for ingredients that may be difficult to source locally and that items be categorized by food type for ease of shopping. The groceries planning prompt is shown in Figure 3.18.

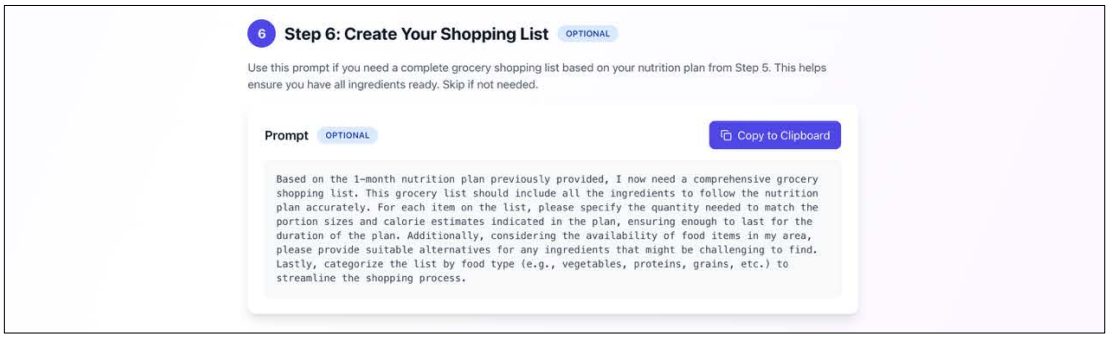


Figure 3.18 Personalized Groceries Planning Prompt of NExGEN

Step 7 is an optional prompt that requests a meal preparation guide based on the nutrition plan and grocery list from the preceding steps. The prompt instructs the AI to provide concise recipes with preparation and cooking times, portion size guidance, food storage and safety tips, and alternative ingredient options. It specifies that the guide should be organized by meal times or days and designed to be accessible for users with varying levels of cooking experience. The meal preparation guide prompt is illustrated in Figure 3.19.

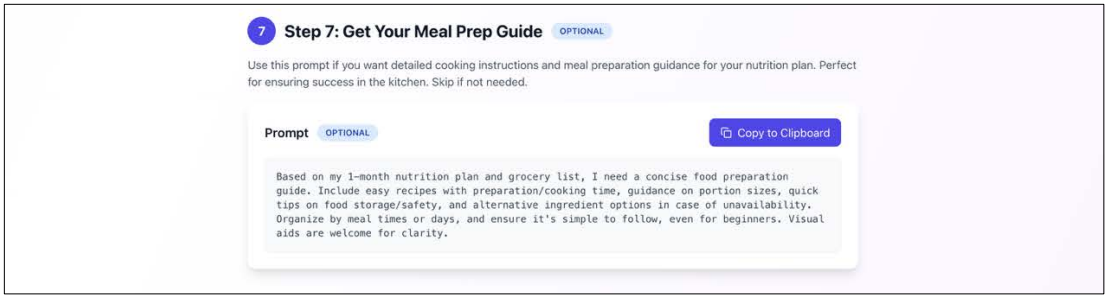


Figure 3.19 Personalized Meal Preparation Guide Prompt of NExGEN

This sequential prompt architecture enables NExGEN to leverage the contextual reasoning capabilities of large language models while maintaining a reproducible and standardized input structure. The consistency of prompt formatting across all users establishes the controlled conditions necessary for evaluating the quality of AI-generated health plans, which the following section addresses through the FDM.

3.2.4 Refinement Process of NExGEN-ChatGPT Fitness and Nutrition Prompts

The NExGEN web application serves as an electronic assessment tool designed to create personalized dietary and exercise prompts for AI chatbot integration. The platform supports individuals in achieving their health goals through customized plans, systematically considering factors like fitness level, demographics, medical history, sleep patterns, stress levels, and motivational factors. The prompts generated through NExGEN's web interface are designed for seamless integration with large language model chatbots like ChatGPT to produce detailed exercise routines and nutritional plans. This integration leverages ChatGPT's AI decision-making capabilities to efficiently create personalized health strategies, enhancing user autonomy and adherence.

3.2.4.1 Selection of Expert

This study employed purposive sampling to select a panel of 21 experts, each possessing specialized knowledge in one or more of the five core domains underpinning the NExGEN framework: nutrition, exercise science, medicine, psychology, and AI. Purposive sampling was selected over probability-based approaches because expert validation requires individuals with directly relevant domain knowledge; random sampling from the broader professional population would not guarantee the specialized expertise necessary to evaluate an interdisciplinary AI-based health tool (Skulmoski et al., 2007).

Domain representation was deliberately structured to ensure comprehensive evaluative coverage of all constructs within the NExGEN questionnaire. Registered dietitians and nutritionists were recruited to assess the dietary planning criteria, macronutrient recommendations, and food preference personalization embedded within the prompts. Exercise scientists and sports medicine practitioners assessed the physical activity programming elements, including exercise prescription principles, progressive overload, and safety considerations for overweight populations. Medical practitioners provided clinical oversight, evaluating the framework's appropriateness for individuals presenting with cardiometabolic risk factors. Psychologists and behavioral scientists assessed the motivational constructs, behavioral modification strategies, and

psychological support elements incorporated into the prompts. Researchers and practitioners with expertise in AI and machine learning evaluated the technical architecture, prompt engineering quality, and inherent limitations of large language model integration for health applications.

Inclusion criteria for the expert panel were: (1) a minimum of five years of active professional experience in the relevant domain at the time of recruitment; (2) current engagement in clinical practice, applied research, or technology development within their field; and (3) demonstrable expertise in at least one of the five core NExGEN domains. The five-year minimum experience threshold was adopted consistent with conventions in comparable FDM and Delphi validation studies in health technology representing the threshold at which practitioners are considered to possess sufficient practical judgment to provide valid evaluative input beyond entry-level competence.

The panel size of 21 is consistent with the guidelines provided by Jones & Twiss (1978), who recommend an expert panel of 10 to 50 for Delphi-type studies. This size is sufficiently large to capture divergent professional perspectives and limit the influence of individual outliers on consensus calculations, while remaining small enough to sustain active engagement and achieve high response rates (Hsu & Sandford, 2007). The multidisciplinary composition ensures that the validated NExGEN framework reflects collective expert judgment across all operational domains of the system, strengthening the content validity and real-world applicability of the finalized prompt structure.

3.2.4.2 Questionnaire Development

The development of the NExGEN framework starts by identifying constructs and variables relevant to exercise and dietary planning for weight management. The development entails a comprehensive review across the disciplines of nutrition, exercise, medicine, and psychology to identify the essential factors relevant to individualized exercise and dietary planning. The derivation of these constructs integrates theoretical findings from published studies and the objectives of the NexGen project. Potential constructs encompass user interaction with fitness and nutrition programs, tailored behavioral modification recommendations, and engagement's psychological effects. In developing the questionnaire, these constructs are dissected into quantifiable variables for examination via targeted questions. This step considers

the operationalization of each variable in a manner that aligns with the study's goals and is clear to the participating experts.

Following the identification of key constructs and variables, the next step is to draft the questionnaire. The questionnaire used a combination of Likert scales (that quantify attitudes and perceptions) and open-ended questions (that allow for detailed, qualitative insights). Each question is crafted to elicit information on specific variables, focusing on aligning the questions with the Study 2 goals. For instance, in the 'Goal Settings' section of NExGEN, the experts were given the following instructions and questions:

> *This section focuses on understanding and evaluating the personalization capabilities of the NExGEN–ChatGPT system in facilitating weight management goals. Your expertise is sought to assess the system's approach to tailoring fitness and nutrition plans based on user-specific weight management outcomes, target weight loss, timelines, and program start dates.*
>
> *Question 1 (Desired Weight Management Outcome Evaluation): In your opinion, how effectively does the NExGEN–ChatGPT system identify and incorporate users' desired weight management outcomes (e.g., weight loss, gain muscle) into personalized plans?*
>
> *Response Options: (1) Strongly Disagree; (2) Disagree; (3) Somewhat Disagree; (4) Neutral; (5) Somewhat Agree; (6) Agree; (7) Strongly Agree.*
>
> *Follow-Up (Open-Ended): Please provide suggestions for improving the system's ability to tailor plans based on desired weight management outcomes.*
>
> *Question 2 (Target Weight Loss Specification): Evaluate the system's mechanism for setting target weight loss goals (e.g., 6 kg, 5% of current weight, 20 lbs). How adequately does it align with best practices in health and fitness?*
>
> *Response Options: (1) Strongly Disagree; (2) Disagree; (3) Somewhat Disagree; (4) Neutral; (5) Somewhat Agree; (6) Agree; (7) Strongly Agree.*
>
> *Follow-Up (Open-Ended): Can you suggest any enhancements to improve the specificity and health alignment of weight loss targets within the system?*
>
> *Question 3 (Timeline for Achieving Weight Loss Goals): Assess the system's*

functionality for defining timelines for weight loss goals (e.g., 12 weeks, 3 months).

Response Options: (1) Strongly Disagree; (2) Disagree; (3) Somewhat Disagree; (4) Neutral; (5) Somewhat Agree; (6) Agree; (7) Strongly Agree.

Follow-Up (Open-Ended): Please provide insights on optimizing timeline.

Question 4 ('Program Start Date Setting): How effectively does the system facilitate the setting of a starting date for the weight management program?

Response Options: (1) Strongly Disagree; (2) Disagree; (3) Somewhat Disagree; (4) Neutral; (5) Somewhat Agree; (6) Agree; (7) Strongly Agree.

Follow-Up (Open-Ended): Suggestions for enhancing user engagement through the program start date feature are welcome."

The drafting stage of this questionnaire is both iterative and collaborative, which involved consultations with experts and the interdisciplinary team to refine questions, validate the questionnaire's content, and ensure it comprehensively covers the intended constructs.

3.2.4.3 Data Collection

In this study, the data collection process involves an open-ended questionnaire, distributed to 21 experts to evaluate the efficiency of the NExGEN Prompt Generation system. To accommodate the varying availabilities of these experts, the questionnaire was disseminated using both face-to-face interactions and online methods. For the latter, the questionnaire is formatted in Portable Document Format (PDF) and shared with the experts. A critical preliminary step is to contact each expert to secure their consent for participation. Following the acquisition of consent, the questionnaire was provided in both hardcopy format for face-to-face meetings and in a softcopy PDF format for those unable to meet in person. The softcopy was distributed via email and WhatsApp. This allows experts to independently complete it at their convenience.

3.2.4.4 Data Analysis

Fuzzy Delphi Scale: This method was employed to refine the NExGEN initial survey and pre-defined prompts, involving the conversion of all linguistic variables into Triangular Fuzzy Numbers. Each Triangular Fuzzy Number is represented by three parameters: m1, m2, and m3, where m1 denotes the minimum value, m2 is the moderate value, and m3 is the maximum value. This conversion is pivotal in establishing a Fuzzy scale from the data, which was developed using a 7-point Likert scale. This scale encapsulates seven levels of agreement: (1) Strongly Disagree; (2) Disagree; (3) Somewhat Disagree; (4) Neutral; (5) Somewhat Agree; (6) Agree; (7) Strongly Agree. Experts were required to express their level of agreement, thereby translating the Likert scale responses into Fuzzy numbers. The Fuzzy scale is its odd-numbered level structure, with the understanding that a higher Fuzzy ranking implies more precise data. The study involved an analysis of all Likert scale data using Microsoft Excel, converting these data into the form of Triangular Fuzzy Numbers based on a seven-point Fuzzy system. This methodological approach ensures an accurate interpretation of linguistic variables, significantly enhancing the reliability and depth of the study's findings.

This study adopted a structured quantitative design, employing the FDM to systematically refine and validate the NExGENChatGPT prompt generation framework for delivering personalized fitness and nutritional recommendations. The FDM was specifically selected due to its unique combination of iterative consensus-building typical of traditional Delphi enhanced by fuzzy set theory (Yin & Hanif, 2024), which accommodates subjective expert opinions and uncertainties inherent in interdisciplinary fields such as nutrition, fitness, psychology, medicine, and AI. In this approach, experts' qualitative judgments, initially captured using Likert scales, were translated into quantitative Triangular Fuzzy Numbers represented by minimum (*m1*), moderate (*m2*), and maximum (*m3*) values (Table 3.1). This allowed for capturing the nuanced and varying levels of agreement among experts effectively.

Table 3.1
Likert Scale to Triangular Fuzzy Number Conversion

Likert Value	Linguistic Term	Fuzzy Number (m_1, m_2, m_3)
1	Strongly Disagree	(0.0, 0.0, 0.1)

2	Disagree	(0.0, 0.1, 0.3)
3	Somewhat Disagree	(0.1, 0.3, 0.5)
4	Neutral	(0.3, 0.5, 0.7)
5	Somewhat Agree	(0.5, 0.7, 0.9)
6	Agree	(0.7, 0.9, 1.0)
7	Strongly Agree	(0.9, 1.0, 1.0)

Data Analytics: The data analysis in Study 1 primarily focuses on the use of Triangular Fuzzy Numbers to determine the value of the Threshold (d), which is important for achieving a level of expert agreement, as emphasized by (Thomaidis et al., 2006). In line with (Cheng & Lin, 2002), for a criterion to be recognized as having expert consensus, the Threshold (d) value must not surpass 0.2. The vertex method was utilized to calculate the distance between the statistic for reporting ranked responses averages, enabling precise measurement of consensus among the experts. The distances for each Fuzzy number m = (m1, m2, m3) and n = (n1, n2, n3) calculated using the formula:

$$d(\tilde{m},\tilde{n}) = \sqrt{\frac{1}{3}\left[(m_1 - n_1)^2 + (m_2 - n_2)^2 + (m_3 - n_3)^2\right]}$$

An important aspect of this analysis is determining the percentage of expert agreement. It is necessary for group consensus to exceed a threshold of 75.0% for each criterion, as suggested by (Chu & Hwang, 2008) and (Murry & Hammons, 1995). If consensus falls below 75.0%, a second round of evaluation is typically required to achieve at least 75.0% consensus, as per (Chang et al., 2011). In such instances, the results from the fuzzy Delphi analysis was sent back to the panelists for reconsideration, a process to be repeated until a satisfactory degree of consensus is reached, as indicated by (Zhang & Lam, 2019).

Defuzzification Process: Defuzzification was employed to determine the scores and rankings for each item, with a primary focus on ascertaining the Fuzzy score Amax (A) value. This process is pivotal, as the Fuzzy score is set to be equal to or greater than

0.5, aligning with Bodjanova (2006), who states that a Fuzzy score above the median α-cut value of 0.5 indicates acceptance by expert consensus. The significance of the Fuzzy score (A) is important in establishing the ranking and priority of the elements under scrutiny. To calculate the Fuzzy score value (A), the formula Amax = 1/4 x (a1 + 2am + a2) was applied. The α-cut value, serving as a critical benchmark, is computed as the median of '0' and '1', resulting in α-cut = (0 + 1)/2 = 0.5. If the calculated Fuzzy score (A) is found to be below this α-cut threshold, the criteria was deemed unacceptable, as this indicates a lack of expert agreement. The formula for the defuzzification process is as follows:

i. Amax = 1/3 * (a1 + am + a2)
ii. Amax = 1/4 * (a1 + 2am + a2)
iii. Amax = 1/6 * (a1 + 4am + a2)

3.3 Personalized AI Prompt Generator and ChatGPT for Weight Loss: Randomized Controlled Trial in Adults with Overweight and Obesity (Study 2)

3.3.1 Research Design

Study 2 employed a two-arm parallel randomized controlled clinical trial comprising a 12-week active intervention phase followed by a 12-week maintenance and follow-up phase. This study has been accepted for 2026 publication in the *Journal of Technology in Behavioral Science*, *Springer Nature* (*Scopus-indexed, SJR Quartile 1, H-index: 24)*. Ethical approval was obtained from the Universiti Teknologi MARA Institutional Review Board (Ethics REC/05/2024 (PG/FB/30)), and the study conformed to the principles of the Declaration of Helsinki (World Medical Association, 2013). The trial was registered at the UMIN Trials Registry (registration number UMIN000053570). Study 2 was designed to investigate the effectiveness of the NExGEN personalized prompt generator integrated with ChatGPT compared to a structured manual ChatGPT control program for weight management in adults with overweight and obesity. Specifically, Study 2 addressed two objectives:

a) To investigate the effectiveness of the NExGEN personalized prompt generator integrated with ChatGPT compared to a structured manual ChatGPT control program, as measured by changes in body weight (primary outcome) over a 24-week period (12-week intervention followed by 12-week follow-up) in adults with overweight and obesity (BMI 27.5 to 34.9 kg/m²).
b) To evaluate the comparative effects of the NExGEN-ChatGPT personalized system versus the structured manual ChatGPT control on secondary outcomes, including body composition (fat mass, fat-free mass, waist circumference), dietary intake patterns (energy density, macronutrient distribution), objectively measured physical activity, and cardiometabolic markers (HbA1c, lipid profile, blood pressure)

To achieve these objectives, the research employed quantitative data analysis of body composition changes, dietary intake patterns, objectively measured physical activity levels, and cardiometabolic markers from participants across both intervention

and control groups. The study further aimed to examine the association between the personalized AI-driven prompt engineering approach and changes in health outcomes over the 24-week study period. This involved a systematic comparison of the extent to which participants following personalized versus generic AI-generated guidance achieved differential weight management outcomes. Through this investigation, the research intended to establish whether validated, domain-specific prompt engineering produces superior clinical outcomes relative to standardized AI-mediated health guidance, thereby advancing the evidence base for personalized AI interventions in obesity management.

The study was designed to evaluate the effectiveness of a personalized prompt generator (NExGEN) integrated with the commercially available AI chatbot ChatGPT for delivering tailored, text-based interventions through a novel approach to weight management. Existing AI chatbot interventions for health behavior change rely predominantly on generic conversational frameworks without structured, domain-validated prompt engineering that accounts for the clinical complexity of obesity management (Oh et al., 2021). The NExGEN system addresses this limitation by systematically capturing 111 user-specific variables across six constructs and 12 elements, validated through expert consensus using the FDM in Study 1.

The system was intended to serve as an adjunct to conventional weight management approaches, extending support beyond the clinical setting through on-demand, AI-mediated access *via* the ChatGPT-o3 model through ChatGPT Business accounts (sponsored by *Pragraph Digital Resources, SA0614953-A*). For this study, the system was customized through a six-prompt sequential process integrating intensive nutrition and exercise planning with behavioral modification strategies recommended for obesity management Jensen et al. (2014), paired with motivational and educational support approaches suited to sustaining lifestyle change. The intervention was delivered across three structured phases: Exploration (weeks 1 to 3), Consolidation (weeks 4 to 8), and Reinforcement (weeks 9 to 12).

3.3.2 Participant Recruitment

Study participants residing in Greater Kuala Lumpur (Klang Valley), Malaysia, who expressed interest through online and offline recruitment channels were enrolled in this randomized controlled trial between May 2024 and August 2025. Clinical

assessments took place at the Exercise Science Laboratory, Universiti Teknologi MARA. Participants were recruited from the Greater Kuala Lumpur (Klang Valley), Malaysia,area through targeted recruitment strategies. Prior to participant recruitment, institutional research ethics approval was obtained and the research protocol was registered. A preliminary feasibility study was then conducted with 12 participants (n = 6 NExGEN group, n = 6 control group) to assess the practicality of the study procedures and identify potential issues before full-scale enrollment. Findings from the feasibility study informed refinements to the recruitment process and intervention delivery prior to the main study.

Following the feasibility study, potential participants were screened against predefined eligibility criteria. Individuals were eligible for the study if, at the time of recruitment, they (1) were aged 19 to 59 years, (2) had a BMI of 27.5 to 34.9 kg/m² based on the WHO Asia-Pacific classification for overweight and obesity Rush (2004), (3) were men or women, with postmenopausal women eligible only if at least 12 months had elapsed since their last menstrual period Harlow et al. (2012), (4) were not pregnant and not currently taking weight loss medication, (5) had no history of chronic disease requiring medical nutrition therapy, and (6) possessed the ability to use a smartphone, computer, or tablet. Potential participants were excluded if they did not meet the inclusion criteria or declined to participate. Eligible participants were enrolled into one of two groups, the NExGEN intervention group (NEX) or the control group (CON), with a target sample size of 70 participants per group. The participant recruitment process is visualized in Figure 3.20 Participant Recruitment Flowchart.

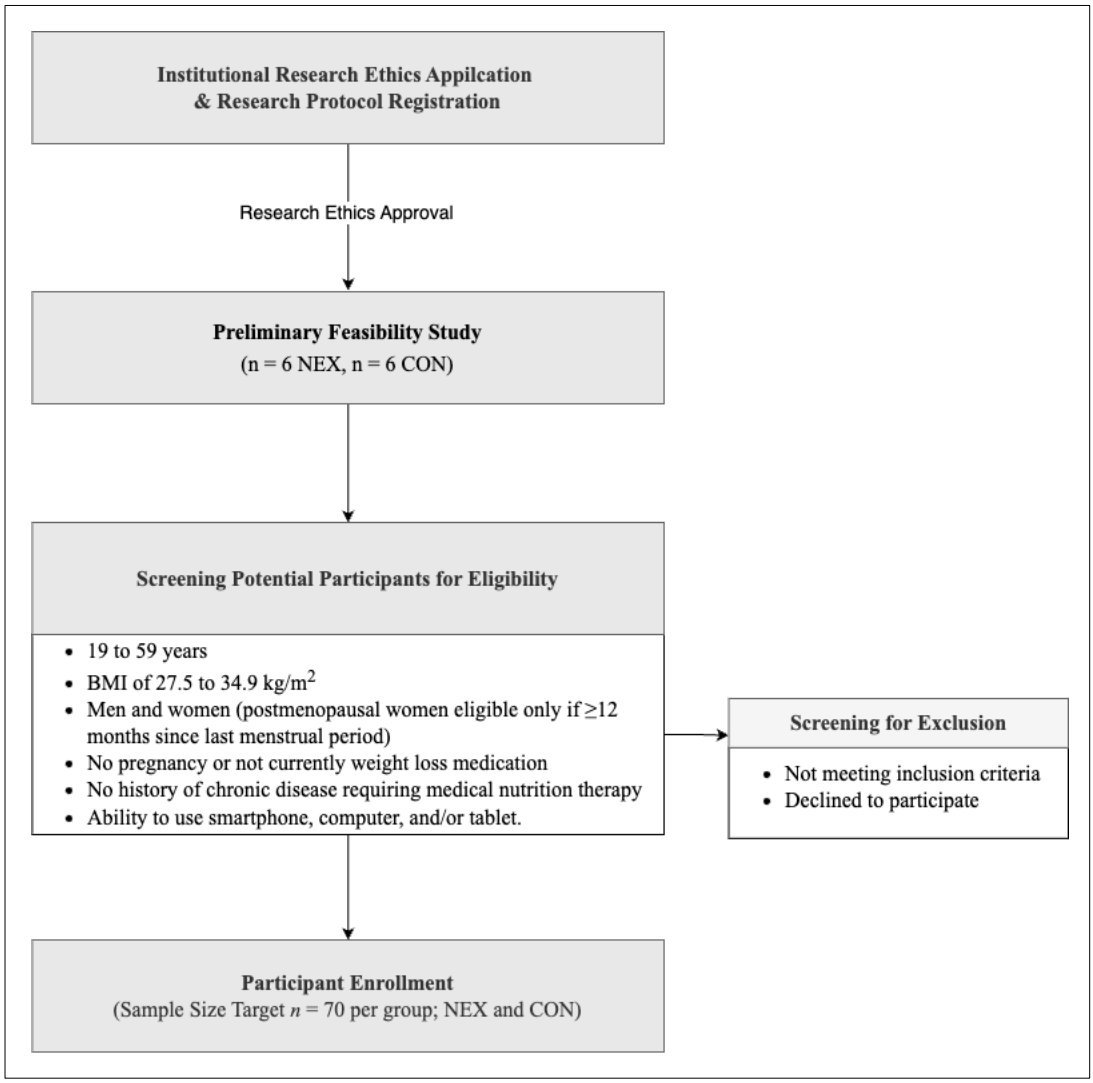


Figure 3.20 Participant Recruitment Flowchart

3.3.3 Sample Size Determination

A review of the literature identified limited evidence comparing personalized versus generic AI-driven approaches for weight management, with no prior trial combining a validated prompt generation system with an AI chatbot platform integrating both exercise and dietary strategies (Oh et al., 2021). The expected effect size was therefore derived from a comparable app-based lifestyle intervention trial. Gemesi et al. (2024) conducted a randomized controlled trial evaluating a digital multimodal weight loss program in 168 adults with obesity. Their intervention group achieved a mean weight loss of 3.2 ± 3.2 kg compared to 0.4 ± 2.6 kg in the control group after 12 weeks, yielding a between-group difference of approximately 3.0 kg. Given that the NExGEN intervention relies on AI-generated plans rather than a

validated commercial app with established behavior change mechanisms, a conservative pooled standard deviation of 6.3 kg was adopted to account for the novelty of the approach, longer follow-up period, and anticipated population variability among Malaysian adults.

Power analysis was conducted using GPower software (version 3.1.9.7, Heinrich-Heine-Universität Düsseldorf) (Faul et al., 2007). The calculation was based on an independent samples t-test comparing two groups, using a two-tailed test with a Cohen's d effect size of 0.476 (derived from a mean difference of 3.0 kg and pooled SD of 6.3 kg), a significance level (α) of 0.05, statistical power (1-β) of 0.80, and an allocation ratio of 1:1. The analysis indicated a required sample size of 142 participants (71 per group) to achieve an actual power of 0.8042. To accommodate a potential 15% dropout rate based on attrition rates observed in similar digital health intervention trials, a targeted enrollment of 168 participants. The GPower output for this calculation is presented in Figure 3.21

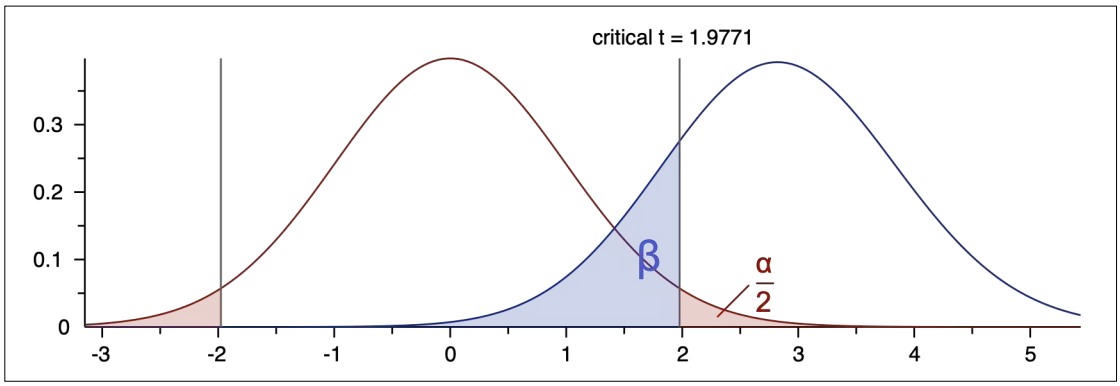


Figure 3.21 Central and Noncentral Distributions from G*Power Output

3.3.4 Pilot Study

Prior to the main trial, a preliminary feasibility study was conducted during May 2024 to evaluate the operational readiness of the trial protocol. Twelve participants (*n* = 6 NEX, *n* = 6 CON) were recruited separately from the main trial sample through identical eligibility criteria and screening procedures. Consistent with recommendations for external pilot studies preceding definitive randomized controlled trials (Thabane et al., 2010), the pilot assessed five operational domains: recruitment feasibility, screening protocol efficiency, intervention delivery logistics, outcome measurement procedures,

and participant burden. Participants completed the full study protocol, including randomization, baseline assessments, intervention delivery, and endpoint measurement. Pilot participants were excluded from the main analysis to prevent contamination of the definitive trial data (Thabane et al., 2010). No protocol modifications were required following the pilot, confirming that all trial procedures functioned as intended and could proceed without alteration for the main enrolment phase.

3.3.5 Randomization and Blinding

Participants were randomly assigned to one of two groups, the NEX intervention group or the CON control group, following completion of all baseline assessments. Block randomization with computer-generated random allocation sequences was performed using the Robust Randomization App (RRApp) (Tu & Benn, 2017). Variable block sizes of 4, 6, and 8 were employed to minimize the predictability of allocation sequences while maintaining approximately equal group sizes throughout the enrollment period (Efird, 2010). Allocation concealment was maintained until enrollment completion, with automated randomization occurring only after each participant had completed all baseline measurements. This sequence ensured that neither the enrolling researcher nor the participant could anticipate group assignment prior to commitment to the study, consistent with CONSORT recommendations for minimizing selection bias (Schulz et al., 2010).

Participant blinding was not feasible due to the inherently different intervention modalities, as the NEX group interacted with a personalized prompt-driven system while the CON group received standardized manual prompts. This open-label design is consistent with behavioral intervention trials in which the nature of the intervention precludes participant masking (Boutron et al., 2008). To mitigate potential assessment bias, research personnel conducting outcome assessments and data analysts remained blinded to group allocation throughout data collection and initial analysis phases. Outcome assessors were not involved in intervention delivery and had no access to randomization records.

3.3.6 Research Procedure

3.3.6.1 Intervention Group

Participants in the NEX group received personalized prompts generated by the NExGEN system, which synthesized data from 111 assessment variables spanning demographics, physiological measurements, medical history, lifestyle behaviors, and motivational factors. Participants logged into designated ChatGPT Business accounts (sponsored by *Pragraph Digital Resources, SA0614953-A*), initiated sessions with the ChatGPT-o3 model, and entered the NExGEN-generated prompts. Six sequential prompts delivered structured personalized weight management support through dynamic prompt-response cycles with ChatGPT (Oh et al., 2021). The system provided personalized feedback, adaptive recommendations, and ongoing dialogue tailored to each participant's profile. Participants performed daily adherence logging of dietary intake and physical exercise within the chat dialogue. The 12-week active intervention was structured across three progressive phases: Exploration (weeks 1 to 3), Consolidation (weeks 4 to 8), and Reinforcement (weeks 9 to 12). Personalized grocery lists and culturally appropriate recipes were adapted for local Malaysian ingredients to enhance dietary compliance.

Participants in the CON group received six structured manual prompts for ChatGPT covering equivalent content domains, including dietary guidance, exercise programming, and behavioral strategies. These prompts mirrored the categorical structure of the NEX intervention but delivered generic, non-individualized recommendations. The content was static and non-adaptive, consistent across all CON participants regardless of personal characteristics, preferences, or progress feedback. The critical methodological distinction between groups was data utilization, not data collection. Both groups provided identical baseline information through the NExGEN assessment. However, the NEX group's data was algorithmically synthesized across all 111 variables to generate tailored outputs, while the CON group received uniform prompts irrespective of personal profile. This design isolated the effect of personalized AI-driven prompt generation from the effect of ChatGPT access alone, controlling for platform exposure and content domain coverage (Collins et al., 2007).

Standardized parameters applied to both groups. Weight loss targets were set at 5 to 10% of baseline body weight over 12 weeks, equivalent to 0.5 to 1.0 kg per week,

consistent with clinical guidelines for obesity management (Jensen et al., 2014). Physical activity goals comprised 150 minutes of moderate-intensity aerobic exercise plus two resistance training sessions per week, aligned with World Health Organization recommendations (Bull et al., 2020). Dietary principles emphasized whole food consumption, reduced ultra-processed food intake, and appropriate portion sizes. All participants received a complimentary 6-month ChatGPT Business subscription (sponsored by *Pragraph Digital Resources*) to ensure equitable platform access throughout the intervention and follow-up phases.

3.3.6.2 Study Timeline

Prior to the study's commencement, the researchers engaged with eligible participants, provided a detailed overview of the study protocol, and obtained written informed consent from each willing participant. Following this, participants were asked to visit the Exercise Science Laboratory, Universiti Teknologi MARA, where they gained access to the NExGEN system and a ChatGPT Business account (OpenAI, San Francisco, USA), encompassing access to the ChatGPT-o3 model. A structured orientation session was dedicated to comprehensively instructing each participant on the use of the NExGEN-ChatGPT system during the intervention phase. Specifically, participants completed the initial NExGEN survey encompassing 111 assessment questions across 12 elements that captured personalized baseline data spanning goal settings, demographics, physiological measurements, medical history, lifestyle and behavior patterns, motivation, and environmental factors. Following this, participants transferred the six sequential prompts from NExGEN to their personal ChatGPT account and generated personalized weight management plan information. Participants were also briefed on a secure method for sharing their ChatGPT chat dialogue records with the study investigators.

Each participant received a validated waist-mounted ActiGraph GT3X+ accelerometer (ActiGraph LLC, Pensacola, USA) to measure and record physical activity and movement (McMinn et al., 2013). Participants were instructed to wear the accelerometer for seven consecutive days at each assessment timepoint, with a minimum of five valid days including one weekend day. Before departing from the laboratory, participants completed baseline assessments. Body weight and body composition were measured using the validated Seca mBCA 515 bioelectrical

impedance analysis system (Seca GmbH & Co. KG, Hamburg, Germany) following a standardized protocol: minimal clothing, bare feet, 12-hour overnight fast, empty bladder, private rooms, and same-sex assessors. Secondary endpoint assessments included waist circumference measured using a Seca 201 measuring tape, height recorded using a Seca 274 stadiometer (Seca GmbH & Co. KG, Hamburg, Germany), 7-day dietary records for analysis via Nutritionist Pro™ v7.9 (Axxya Systems) with reference to the Malaysian Food Composition Database (MyFCD) Tay et al. (2023), physical activity via the International Physical Activity Questionnaire-Long Form (Craig et al., 2022), and cardiometabolic markers via point-of-care analyzers.

Following the initial laboratory visit, participants were given a seven-day window to independently complete the NExGEN-ChatGPT system setup and familiarize themselves with the platform. This approach ensured that participants completed the initial system usage without investigator assistance, confirming independent operability. Prior to the start of the study intervention, participants shared the chat dialogue records of the initial ChatGPT interactions with the investigators through a secure, encrypted messaging system. A 12-week active weight management intervention then commenced, during which participants in both groups were expected to adhere to the exercise and dietary regimens delivered through their respective ChatGPT-based systems. Participants were encouraged to engage frequently with the generated dialogue in ChatGPT during the intervention period, where they could offer feedback and pose pertinent questions. All interactions for each participant were maintained within a single chat dialogue throughout the intervention duration. At the conclusion of each week, participants utilized the feedback and revision feature within NExGEN to request adjustments to their exercise and dietary plans from ChatGPT. These requests were based on self-assessed weight changes at home and feedback on the plan from the preceding week. Weekly records of ChatGPT interactions were shared with the investigators through a secure, encrypted messaging platform. Following the 12-week active phase, a 12-week maintenance and follow-up phase commenced. During this period, participants continued to have access to the ChatGPT platform and could sustain their established dietary and exercise routines without additional structured prompts. No new prompts or investigator-initiated adjustments were introduced during the maintenance phase.

At the conclusion of week 12 and again at week 24, participants attended laboratory visits to complete assessments identical to the baseline protocol, including

body weight, body composition, waist circumference, height, dietary records, physical activity measures, and cardiometabolic markers. Participants did not receive any honorarium but were provided with a complimentary 6-month ChatGPT Business subscription (approximately ~RM830; sponsored by *Pragraph Digital Resources, SA0614953-A*). Figure 3.22 depicts the intervention procedures for the NExGEN-ChatGPT weight management trial.

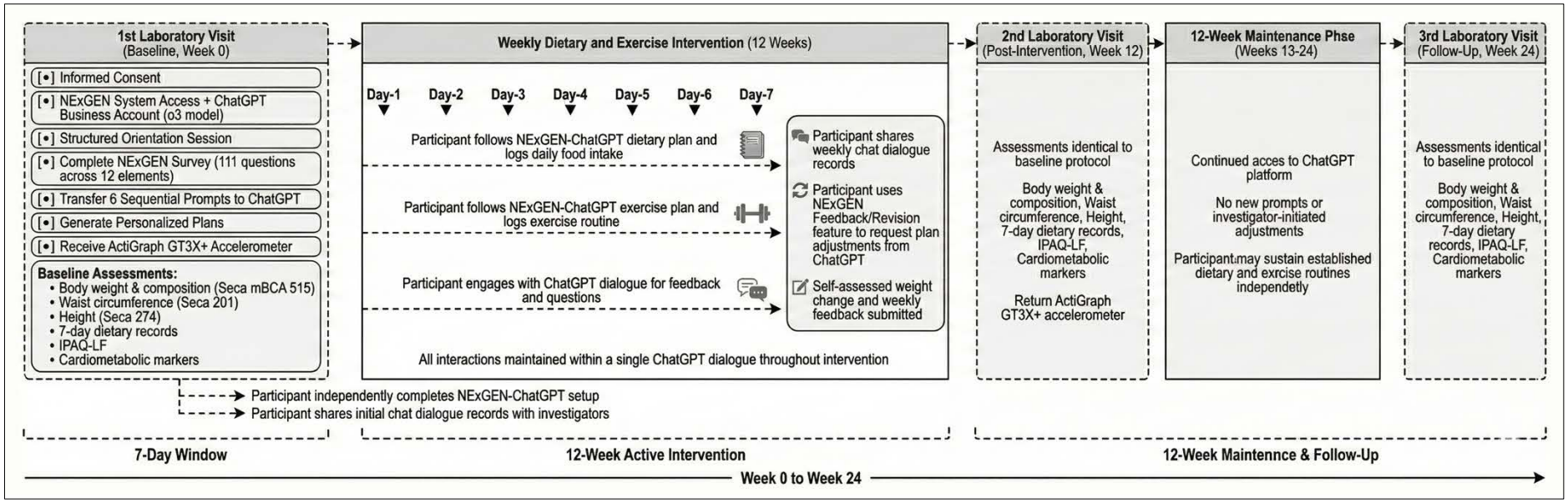


Figure 3.22 Intervention Procedures for the NExGEN-ChatGPT Weight Management Trial.

3.3.7 Measurement

3.3.7.1 Body Composition Assessment

Body composition variables, including body weight, fat mass, fat-free mass, body fat percentage, and BMI, were measured at baseline, post-intervention, and follow-up using validated equipment. Body weight and composition were assessed using the Seca mBCA 515 bioelectrical impedance analysis system (Seca GmbH & Co. KG, Hamburg, Germany), which employs 8-point segmental impedance across 19 frequencies (Bosy-Westphal et al., 2017). Participants attended assessments barefoot and in minimal clothing following a 12-hour overnight fast with an empty bladder, in private rooms with same-sex assessors. Height was recorded using a Seca 274 stadiometer (Seca GmbH & Co. KG, Hamburg, Germany). Waist circumference was measured at the midpoint between the pelvic crest and lower rib using a Seca 201 measuring tape to the nearest 1 mm. The coefficient of variation for all equipment was determined by repeated triplicate assessments.

3.3.7.2 Dietary Intake

Dietary intake was assessed using 7-day dietary records at each assessment timepoint: baseline (week 0), post-intervention (week 12), and follow-up (week 24). Participants recorded all foods and beverages consumed using a combination of household measures (e.g., cups, tablespoons, pieces), standard portion size photographs provided at orientation, and gram weights where available, with reference to the Malaysian Food Composition Database (MyFCD) portion guides. Completed records were reviewed by the research team for completeness and plausibility before data entry. A minimum of five valid recording days, including one weekend day, was required for each assessment period to account for day-to-day variability in eating patterns (Basiotis et al., 1987). All dietary data were analyzed using Nutritionist Pro™ v7.9 (Axxya Systems), with referenced to Malaysian Food Composition Database (MyFCD) (Tay et al., 2023). Analysis of dietary records yielded total energy intake (expressed in kilocalories per day; kcal/day), macronutrient distribution (carbohydrates, fats, and proteins reported in absolute amounts in grams per day and as a percentage of total energy), dietary energy density (expressed in kilocalories per gram; kcal/g, excluding

beverages), micronutrient adequacy (reported in absolute daily units compared against Malaysian Recommended Nutrient Intakes), and food group patterns (reported as number of servings and estimated portion sizes in grams per day). Dietary energy density was calculated by dividing total food energy by total food weight, excluding all beverages, consistent with established methods for assessing diet quality in weight management interventions (Ledikwe et al., 2006).

3.3.7.3 Physical Activity

Physical activity was objectively measured using ActiGraph GT3X+ accelerometers (ActiGraph LLC, Pensacola, USA) worn at the right hip for seven consecutive days at each assessment timepoint (McMinn et al., 2013). Raw triaxial acceleration data were sampled at 30 Hz and processed using 60-second epochs via ActiLife software (version 6.13.4, ActiGraph LLC, Pensacola, USA). The primary accelerometer metrics derived included total daily steps and vertical axis counts per minute (CPM). Physical activity intensity was classified using Freedson adult cut-points (Freedson et al., 1998) as follows: sedentary behavior (< 100 CPM), light-intensity physical activity (100–2,019 CPM), moderate-intensity physical activity (2,020–5,998 CPM), and vigorous-intensity physical activity (≥ 5,999 CPM). Total moderate-to-vigorous physical activity (MVPA) in minutes per day was derived by summing moderate and vigorous intensity time. Non-wear periods were defined as 60 or more consecutive minutes of zero activity counts, permitting up to two minutes of non-zero interruption (Troiano et al., 2008). A minimum of five valid days with 10 or more hours of wear time, including one weekend day, was required for data inclusion. Self-reported physical activity was additionally assessed using the International Physical Activity Questionnaire-Long Form at each timepoint (Craig et al., 2003). Metabolic equivalent values were derived from the 2024 Adult Compendium of Physical Activities (Herrmann et al., 2024)

3.3.7.4 Cardiometabolic Markers

Cardiometabolic markers were assessed at baseline, post-intervention, and follow-up. HbA1c and lipid profile, including total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides, were

measured using the Cobas b 101 point-of-care system (Roche Diagnostics, Basel, Switzerland). This device employs reflectance photometry on whole blood capillary samples and has demonstrated acceptable agreement with laboratory-based enzymatic methods (Li et al., 2023). Fasting glucose was measured using the HemoCue HB 201 analyzer (HemoCue AB, Ängelholm, Sweden). All blood-based assessments were conducted following a 12-hour overnight fast. Blood pressure (systolic and diastolic) was measured using a validated DINAMAP PRO 420 oscillometric device (GE Healthcare, Chicago, Illinois, USA). Triplicate measurements were taken after five minutes of seated rest, and the mean of the final two readings was recorded, consistent with established clinical measurement protocols (Muntner et al., 2019).

3.3.8 Statistical Analysis

Descriptive statistics were employed to characterize the baseline characteristics of participants in both groups and to evaluate outcome measures across the three assessment timepoints. To validate the data's normality assumption, the Shapiro-Wilk test was applied. Non-normally distributed data were addressed through appropriate transformations or non-parametric alternatives. The primary analysis employed linear mixed models for repeated measures (MMRM) with treatment group, time, and group-by-time interaction as fixed effects (Prakash et al., 2008). MMRM accommodates incomplete data under a missing-at-random assumption without requiring explicit imputation, making it particularly suitable for longitudinal clinical trials with participant dropout (Fitzmaurice et al., 2011). Both intention-to-treat (ITT) and per-protocol (PP) analyses were performed. The ITT analysis included all randomized participants and employed multiple imputation (m = 50 imputations) under a missing-at-random assumption. Imputation models incorporated baseline characteristics, treatment group, and all available follow-up measurements. The PP analysis excluded participants with major protocol violations or missing primary outcome data. ITT results were primarily reported, with PP results presented for sensitivity verification. Effect sizes and 95% confidence intervals were reported to delineate the probable range of true values within the sample. IBM SPSS Statistics (version 30, IBM Corp., Armonk, New York, USA) and GraphPad Prism software (version 10.5, GraphPad Software, Boston, Massachusetts, USA) were used for all data analyses, with statistical significance accepted at $p < 0.05$

3.3.9 Methodological

Two methodological limitations inherent to this study design require explicit acknowledgment: the reliance on self-report data for key outcome measures, and the potential for AI response inconsistency arising from the generative nature of large language models.

Self-report data limitations. Dietary intake and behavioral adherence data in this study were partially collected through self-report instruments, including 7-day dietary records and the International Physical Activity Questionnaire-Long Form (IPAQ-LF). Self-report methods are subject to well-documented systematic biases. Social desirability bias may lead participants to under-report energy intake or over-report physical activity to align with perceived investigator expectations (Dhurandhar et al., 2015). Retrospective recall bias is an additional concern, particularly for the IPAQ-LF, where accuracy of recalled activity duration and frequency diminishes over longer reference periods (Prince et al., 2008). Furthermore, the act of prospective dietary recording may itself alter eating behavior through heightened self-awareness, a phenomenon known as dietary reactivity, meaning that recorded days may not fully represent habitual intake patterns (Nisbett & Wilson, 1977). To mitigate these limitations, objective accelerometry was used to provide an independent measure of physical activity, and point-of-care biochemical markers (HbA1c, lipid profile, fasting glucose) provided objective corroboration of cardiometabolic outcomes. These objective measures partially triangulate self-reported behavioral data; however, dietary intake assessment remained dependent on self-report, and this constitutes a limitation in the precision of energy intake quantification across the study period.

AI response inconsistency. The NExGEN-ChatGPT framework delivers its intervention through a large language model, which introduces a source of output variability not present in conventional manual or scripted interventions. Large language models generate responses through stochastic token sampling processes governed by temperature and top-p parameters, meaning that two participants submitting identical prompts may receive substantively different dietary and exercise plans (Aljamaan et al., 2024). This variability limits the standardization of intervention content compared to fixed-protocol behavioral programs. Additionally, large language models are susceptible to hallucination, in which the model generates clinically plausible but factually incorrect information presented with apparent confidence (Omar et al., 2025).

To reduce these risks, the study implemented three safeguards: (1) prompts used in the trial were derived from the expert-validated NExGEN framework from Study 1, constraining the solution space of AI-generated outputs to evidence-based parameters; (2) the ChatGPT-o3 reasoning model was selected based on its demonstrated improvements in factual accuracy and logical consistency relative to standard generative models; and (3) weekly investigator review of participant chat logs was conducted throughout the 12-week active phase to identify and address any clinically inappropriate or potentially harmful recommendations before participants acted on them. These measures reduced but cannot entirely eliminate AI output variability, and findings should be interpreted with the understanding that the intervention was not fully standardized across all participants in the manner of a scripted behavioral protocol.

CHAPTER 4
RESULTS

4.1 Introduction

This chapter presents the findings of this research, organised according to the studies conducted. The results are reported in two main sections corresponding to the two studies undertaken: Study 1, which addresses the validation of the Personalised AI Prompt Generator (NExGEN-ChatGPT) for obesity management using the Fuzzy Delphi Method (FDM); and Study 2, which examines the effectiveness of the NExGEN-ChatGPT personalised AI prompt generator, validated through the FDM in Study 1, on weight loss and associated health outcomes in adults with overweight and obesity. Each section follows a systematic reporting structure, beginning with baseline descriptive characteristics, followed by statistical assumptions, and subsequently presenting the findings for each construct and element examined.

4.2 Results of Study 1: Validation of a Personalised AI Prompt Generator (NExGEN-ChatGPT) for Obesity Management Using Fuzzy Delphi Method

Study 1 employed the Fuzzy Delphi Method to establish expert consensus on the constructs, elements, and criteria necessary for the development of the NExGEN-ChatGPT personalised AI prompt generator for obesity management. Part of Study 1 results has been published online in the Journal of *Biology Methods and Protocols*, *Oxford University Press* (*Web of Science* and *Scopus-indexed*, *SJR Quartile 1*, *Impact Factor: 1.3*).The findings are presented according to six main constructs: (1) Goal Settings, (2) Demographics and Physical Characteristics, (3) Health and Medical History, (4) Lifestyle and Behaviour, (5) Motivation and Social Support, and (6) Environmental Factors. Overall, all the criteria accepted for each construct and element met the requirements and conditions to pass the expert consensus of >75%, the threshold value (d) < 0.2, and the α-cut value exceeding 0.5.

4.2.1 Baseline Descriptive Characteristics

4.2.1.1 Expert Panel Demographics

A total of 21 experts were recruited for the FDM panel. The panel comprised 11 male and 10 female experts, with 52.4% and 47.6% representation, respectively. The ages of the experts ranged from 36 to 58 years, with a mean age of 47.3 ± 7.6 years, with the youngest expert being 36 years old and the oldest being 58 years old. Table 4.1

Demographic Characteristics of the Expert Panel presents the professional background of the expert panel. The majority of the experts possessed doctoral-level qualifications (12, 57.14%), followed by master's degree holders (8, 38.10%) and one bachelor's degree holder (1, 4.76%). In terms of area of expertise, 5 experts (23.81%) specialised in accredited nutrition, 5 (23.81%) in exercise science, 5 (23.81%) in medical practice, 4 (19.05%) in accredited psychology, and 2 (9.52%) in AI research. The mean years of professional experience was 12.8 years (SD = 2.8), with a minimum of 10 years and a maximum of 19 years.

Table 4.1
Demographic Characteristics of the Expert Panel

Expert Category	Position	Level of Education	Years of Experience
Accredited Nutritionist	Dietitian at a hospital in the Kuala Lumpur, Malaysia.	Master	19 Years
	Nutritionist at a wellness center in Selangor, Malaysia.	Doctorate	12 Years
	Dietitian at a clinic in Penang, Malaysia.	Doctorate	18 Years
	Dietitian at a health center in Ipoh, Malaysia.	Master	14 Years
	Dietitian at a healthcare facility in Petaling Jaya, Malaysia.	Doctorate	12 Years
Exercise Specialist	Fitness Director at a commercial fitness center in Kuala Lumpur, Malaysia.	Master	12 Years
	Exercise Physiologist at Malaysia Sports Institute.	Bachelor	11 Years
	Sports Therapist at a sports center in Selangor, Malaysia.	Master	10 Years
	Wellness Coach at a fitness studio in Johor, Malaysia.	Master	11 Years
	Exercise and Rehabilitation Specialist at a rehabilitation clinic in Johor, Malaysia.	Doctorate	13 Years
Medical Personnel	General Practitioner, at a public hospital in Serdang, Malaysia.	Doctorate	18 Years
	General Practitioner at a private hospital in Johor, Malaysia.	Doctorate	15 Years
	General Practitioner at a health center in Kuala Lumpur, Malaysia.	Doctorate	12 Years
	General Practitioner at a public hospital in Penang, Malaysia.	Doctorate	11 Years
	General Practitioner at a health center in Malacca, Malaysia.	Doctorate	16 Years
Accredited Psychologist	Educational Psychologist at a university in Penang, Malaysia.	Master	10 Years
	Research Psychologist at a university in Johor, Malaysia.	Doctorate	12 Years
	Clinical Psychologist at a health clinic in Selangor, Malaysia.	Master	11 Years
	Organizational Psychologist in Selangor, Malaysia.	Doctorate	10 Years
AI Researcher	AI Lab Director at a public university in Kuala Lumpur, Malaysia.	Doctorate	10 Years
	Machine-learning Specialist at an AI-based company in Penang, Malaysia.	Master	12 Years

Note. The names and full institutional affiliations of the expert panelists are not disclosed to preserve anonymity in accordance with the ethical requirements approved by the Universiti Teknologi MARA Institutional Review Board (Ethics REC/05/2024 (PG/FB/30)).

4.2.2 Construct 1: Goal Settings

Construct 1 focuses on goal setting and comprises three elements and eighteen criteria. The elements specify the desired outcome for weight management, the targeted weight change, and the timeline for achieving these goals. The findings from the FDM analysis for each element are presented in the following subsections. Each table lists only the criteria that met the acceptance threshold as described in Section 4.1.2.1.

4.2.2.1 Element: Desired Weight Management Outcomes

Table 4.2 presents the collective expert judgment on each criterion's importance and applicability for the element of 'Desired Weight Management Outcomes'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions of expert consensus >75%, threshold value (d) < 0.2, and α-cut > 0.5.

The most significant criterion identified was the thorough initial and ongoing assessment of physical activity, which achieved unanimous expert agreement (100%), a threshold value of 0.127, and the highest fuzzy score of 0.859. This finding underscores that consistent assessment was regarded as essential for programme effectiveness. Ranking second, the criterion of utilising specific metrics to measure improvements received a 90.48% consensus and a fuzzy score of 0.821, reinforcing the necessity for measurable outcomes in the weight management programme.

The third and fourth-ranked criteria demonstrated nearly identical fuzzy scores (0.765 vs. 0.762), indicating that the experts viewed goal alignment and tracking systems as complementary priorities requiring concurrent implementation rather than hierarchical distinction. Specifically, aligning physical activity habits with overall health and weight management goals ranked third (90.5% consensus, fuzzy score 0.765), emphasising alignment with broader health objectives. Effective tracking systems and participant feedback ranked fourth (95.24% consensus, fuzzy score 0.762), highlighting the role of supportive infrastructure in programme delivery.

Similarly, the fifth and sixth criteria demonstrated close scoring (0.667 vs. 0.665), where diverse physical activities marginally outranked technology tools despite a lower consensus percentage. This narrow difference suggests that expert opinion was divided on activity variety versus technological support, indicating both approaches may be equally valid depending on individual user preferences and programme

contexts. Incorporating diverse physical activities catering to individual interests showed moderate consensus (81%, fuzzy score 0.667), whilst the use of technology tools such as applications and wearable devices received a lower but notable acceptance (76.19%, fuzzy score 0.665).

These findings reveal a clear expert prioritisation hierarchy emphasising assessment rigour and measurable outcomes as foundational elements for effective obesity management. The unanimous consensus (100%) on thorough physical activity assessment reflects expert recognition that baseline and ongoing monitoring provides essential data for personalising interventions and detecting progress plateaus requiring strategy adjustment. Notably, the close fuzzy scores between goal alignment (0.765) and tracking systems (0.762) suggest these elements function synergistically rather than independently, as effective tracking mechanisms enable goal alignment by providing real-time feedback on progress toward health objectives. The relatively lower consensus on technology tools (76.19%, fuzzy score 0.665) compared to activity diversity (81%, fuzzy score 0.667) indicates expert ambivalence regarding technological dependence, possibly reflecting concerns about accessibility barriers, user technological literacy, or the risk of over-reliance on devices rather than intrinsic motivation.

Table 4.2
Fuzzy Delphi Method Analysis for 'Desired Weight Management Outcomes' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
How thoroughly are initial and ongoing assessments of physical activity levels integrated into the program?	0.127	100.00%	0.859	1	Accepted
How effectively are physical activity habits integrated with overall health and weight management goals?	0.160	90.50%	0.765	3	Accepted
Does the program include a wide range of physical activities to cater to different interests and capabilities?	0.151	81.00%	0.667	5	Accepted
Are there effective systems in place for tracking physical activity and providing feedback to participants?	0.136	95.24%	0.762	4	Accepted
How is technology (e.g., apps, wearable devices) used to support and enhance engagement in physical activities?	0.174	76.19%	0.665	6	Accepted
Are specific metrics used to measure improvements or changes in physical activity habits?	0.179	90.48%	0.821	2	Accepted

4.2.2.2 Element: Target Weight Management Outcomes

Table 4.3 presents the prioritised criteria for the element of 'Target Weight Management Outcomes' derived through the FDM, emphasising their importance and feasibility based on expert consensus percentages and fuzzy scores. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion was clearly and precisely defining target weight change goals, which achieved unanimous expert consensus (100%) and a fuzzy score of 0.806, highlighting clarity and specificity as essential for effective weight management. The alignment of weight change goals with established health standards, such as Body Mass Index (BMI) and body fat percentage, ranked second, receiving

90.48% consensus and a fuzzy score of 0.786, underscoring the need for evidence-based objectives. Goals tailored realistically to individual health and fitness status ranked third (76.2% consensus, fuzzy score 0.724), reinforcing the value of practicality in goal setting. Personalisation based on age, gender, and lifestyle ranked fourth (76.19% consensus, fuzzy score 0.656), further supporting individualised goal-setting approaches.

The safety of weight change goals ranked fifth (80.95% consensus, fuzzy score 0.640), whilst realistic timelines, despite a high consensus (90.5%), had a comparatively lower fuzzy score (0.576), suggesting that timeline practicality was perceived as less foundational compared to other criteria. The unanimous expert consensus (100%, fuzzy score 0.806) on clearly defining target weight change goals underscores the fundamental importance of specificity in obesity interventions, consistent with SMART goal-setting principles. This finding suggests that vague aspirations are insufficient for AI-driven interventions, which require quantifiable targets to generate appropriate recommendations.

The high ranking of alignment with health standards (90.48%, fuzzy score 0.786) reflects expert concern that AI systems might accommodate potentially unsafe user preferences without medical contextualisation. Interestingly, the relatively lower fuzzy score for realistic timelines (0.576) despite high consensus (90.5%) suggests that whilst experts agree timelines matter, they may be less critical than goal clarity and safety parameters. This could indicate that timeline flexibility is viewed as more individually variable and thus less appropriate for rigid standardisation compared to absolute safety thresholds.

Table 4.3
Fuzzy Delphi Method Analysis for 'Target Weight Management Outcomes' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are the target weight change goals clearly and precisely defined?	0.152	100.00%	0.806	1	Accepted
Are the target weight change goals realistic and feasible given the user's current health and fitness status?	0.187	76.20%	0.724	3	Accepted
Are the weight management outcomes realistic and achievable within the specified timelines?	0.180	90.50%	0.576	6	Accepted
Do the target weight change goals align with recognized health and medical standards (e.g., BMI, body fat percentage)?	0.196	90.48%	0.786	2	Accepted
Are the target weight change goals personalized based on individual characteristics such as age, gender, health conditions, and lifestyle?	0.188	76.19%	0.656	4	Accepted
Are the target weight change goals set in a manner that ensures the health and safety of the individual (e.g., avoiding rapid weight loss/gain that could be harmful)?	0.180	80.95%	0.640	5	Accepted

4.2.2.3 Element: Timeline for Achieving Goals

Table 4.4 presents the criteria ranked through the FDM concerning the element of 'Timeline for Achieving Goals', based on expert consensus and fuzzy scores, emphasising their relative importance and practical relevance. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The top-ranked criterion, assessing whether timelines are realistic and achievable considering individual fitness levels and lifestyles, attained an 85.7% consensus and the highest fuzzy score of 0.725, indicating the importance of aligning timelines with personal capabilities. The second-ranked criterion, structuring timelines to sustain motivation and engagement, received identical consensus (85.71%) but a lower fuzzy score (0.595), underscoring the role of motivation throughout the goal

process. Flexibility in adjusting goals and timeframes ranked third (80.95% consensus, fuzzy score 0.562), reinforcing the need for adaptable planning.

Fourth, incorporating incremental milestones for monitoring progress and adjustments scored 85.7% consensus and a fuzzy score of 0.557, highlighting the importance of structured checkpoints. Personalisation of timelines based on age, gender, health, and lifestyle ranked fifth (85.71% consensus, fuzzy score 0.530). Lastly, prioritising health and safety by avoiding rapid changes ranked sixth (81% consensus, fuzzy score 0.529), stressing cautious timeline planning. These rankings indicate that experts prioritised the feasibility and motivational aspects of timeline setting over rigid demographic-based personalisation, reflecting the understanding that adherence to weight management programmes is strongly influenced by the perceived achievability of set timelines.

Table 4.4
Fuzzy Delphi Method Analysis for 'Timeline for Achieving Goals' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Is the proposed timeline realistic and achievable based on the user's current fitness level and lifestyle?	0.165	85.70%	0.725	1	Accepted
Does the timeline ensure that health and safety are prioritized, avoiding overly rapid or extreme changes?	0.175	81.00%	0.529	6	Accepted
Are there clear incremental milestones within the timeline to monitor progress and make adjustments?	0.187	85.70%	0.557	4	Accepted
Does the timeline allow for flexibility to adjust goals and timeframes based on progress and unforeseen challenges?	0.171	80.95%	0.562	3	Accepted
Is the timeline structured in a way that maintains motivation and engagement over the duration?	0.198	85.71%	0.595	2	Accepted
Is the timeline personalized according to individual factors such as age, gender, health status, and lifestyle?	0.174	85.71%	0.530	5	Accepted

4.2.3 Construct 2: Demographics and Physical Characteristics

Construct 2 addresses demographic and physical characteristics and comprises three elements and eighteen criteria. The elements encompass age considerations, body composition (height, weight, and BMI), and physiological measurements (body-fat percentage and lean body mass). The findings from the FDM analysis for each element are presented in the following subsections.

4.2.3.1 Element: Age

Table 4.5 presents the ranked criteria concerning the element of 'Age' in nutrition and exercise recommendations using the FDM. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The top criterion, assessing whether strategies are safe across age groups given health risks, had the highest expert consensus at 95.24% and a fuzzy score of 0.681, emphasising safety and risk mitigation as paramount considerations in age-related recommendations. The second-ranked criterion addressed the clarity and appropriateness of recommendations for diverse age groups (95.2% consensus, fuzzy score 0.673), highlighting the necessity of clear, adaptable guidelines. Adaptability to individual differences ranked third (80.95% consensus, fuzzy score 0.648), underscoring personalised recommendations that account for within-group variation.

Specific nutritional and exercise requirements for each age group ranked fourth (90.5% consensus, fuzzy score 0.603). Consideration of developmental and physiological variations was fifth (85.7% consensus, fuzzy score 0.576), whilst meeting specific nutritional needs ranked sixth (85.71% consensus, fuzzy score 0.511), establishing foundational importance. The high consensus on safety (95.24%) across age groups reflects the expert panel's awareness that age-related physiological differences significantly influence both exercise tolerance and nutritional requirements, necessitating careful calibration of AI-generated recommendations to prevent adverse health outcomes.

Table 4.5
Fuzzy Delphi Method Analysis for 'Age' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are the nutrition and exercise recommendations appropriate for different age groups (e.g., children, adolescents, adults, elderly)?	0.181	95.20%	0.673	2	Accepted
Do the strategies consider the developmental and physiological differences across age groups?	0.198	85.70%	0.576	5	Accepted
Are the specific nutritional and exercise needs of each age group adequately addressed?	0.179	90.50%	0.603	4	Accepted
Are the proposed strategies safe for all age groups, considering potential age-related health risks?	0.189	95.24%	0.681	1	Accepted
Can the recommendations be easily adapted to fit the unique requirements of individuals within various age groups?	0.179	80.95%	0.648	3	Accepted
Do the dietary guidelines meet the specific nutritional requirements for different ages?	0.138	85.71%	0.511	6	Accepted

4.2.3.2 Element: Body Composition

Table 4.6 presents the prioritised criteria for nutrition and exercise recommendations based on the element of 'Body Composition' using the FDM. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The top-ranked criterion, addressing whether recommendations align with specific BMI categories, achieved unanimous expert consensus (100%) and the highest fuzzy score of 0.681, highlighting the importance of BMI-targeted strategies. The second criterion, assessing the safety of strategies across different height and weight profiles, also received full consensus (100%) and a fuzzy score of 0.605, emphasising safety considerations. Personalisation according to individual height and weight ranked third (95.2% consensus, fuzzy score 0.576), underscoring tailored approaches.

Realistic and achievable weight management goals placed fourth (85.71% consensus, fuzzy score 0.557), ensuring practicality. Addressing nutritional needs across body sizes was fifth (90.48% consensus, fuzzy score 0.517), emphasising dietary inclusivity. Lastly, suitability of exercise intensity and type ranked sixth (76.19% consensus, fuzzy score 0.511), indicating lower relative emphasis. The unanimous consensus (100%) on BMI-targeted recommendations and safety considerations reflects the expert panel's recognition that body composition parameters serve as critical inputs for personalising AI-generated nutrition and exercise recommendations.

Table 4.6
Fuzzy Delphi Method Analysis for 'Body Composition' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are the nutrition and exercise recommendations tailored to different BMI categories (e.g., underweight, normal weight, overweight, obese)?	0.186	100.00%	0.681	1	Accepted
Are the recommendations personalized based on individual height and weight measurements?	0.162	95.20%	0.576	3	Accepted
Are the proposed strategies safe for individuals with different height and weight profiles, considering potential health risks?	0.152	100.00%	0.605	2	Accepted
Are the specific nutritional needs of individuals with different body sizes adequately addressed?	0.164	90.48%	0.517	5	Accepted
Is the intensity and type of exercise recommended appropriate for different height and weight categories?	0.193	76.19%	0.511	6	Accepted
Are weight management goals realistic and achievable for individuals with different height and weight profiles?	0.187	85.71%	0.557	4	Accepted

4.2.3.3 Element: Physiological Measurements

Table 4.7 presents the prioritised criteria related to the element of 'Physiological Measurements'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion was evaluating whether physiological measurements ensure participant safety and health monitoring during nutrition and exercise programmes. This criterion achieved an expert consensus of 85.7% and the top fuzzy score of 0.659, highlighting participant safety as the primary concern. Ranking second, the criterion on personalised strategies based on individual physiological data also received 85.7% consensus and a fuzzy score of 0.630, emphasising tailored

interventions. Employing modern technologies to enhance physiological data collection, analysis, and feedback ranked third (90.48% consensus, fuzzy score 0.613), underscoring technology's role in data accuracy.

Ensuring reliable and accurate measurement tools ranked fourth (95.2% consensus, fuzzy score 0.576). Adjusting fitness and nutrition plans according to changing physiological data was fifth (80.95% consensus, fuzzy score 0.573), with tracking and reporting systems ranking sixth (80.95% consensus, fuzzy score 0.511), reflecting their supporting roles in the overall framework. The prioritisation of safety monitoring (0.659) over personalisation (0.630) and technology (0.613) suggests that the expert panel considered the protective function of physiological measurements as more fundamental than their personalisation or technological enhancement capabilities.

Table 4.7
Fuzzy Delphi Method Analysis for 'Physiological Measurements' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Do the recommendations personalize strategies based on individual physiological measurements?	0.180	85.70%	0.630	2	Accepted
Are the methods and tools used for measuring physiological metrics accurate and reliable?	0.162	95.20%	0.576	4	Accepted
Are the physiological measurements used to ensure the safety and monitor the health of participants during exercise and nutrition programs?	0.145	85.70%	0.659	1	Accepted
Can fitness and nutrition plans be adjusted based on changes in physiological measurements over time?	0.195	80.95%	0.573	5	Accepted
Are there effective systems in place for tracking and reporting physiological measurements over time?	0.166	80.95%	0.511	6	Accepted
Are modern technologies employed to enhance the collection, analysis, and feedback of physiological measurements?	0.176	90.48%	0.613	3	Accepted

4.2.4 Construct 3: Health and Medical History

Construct 3 addresses health and medical history and comprises two elements and twelve criteria. The elements encompass medical conditions and medications. The findings from the FDM analysis for each element are presented in the following subsections.

4.2.4.1 Element: Medical Conditions

Table 4.8 presents the analysis results related to the element of 'Medical Conditions', emphasising tailored and inclusive strategies for fitness and nutrition management. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The criterion advocating comprehensive inclusion of diverse medical conditions received 81% expert consensus and the highest fuzzy score of 0.741, emphasising broad applicability of the framework across various health conditions. Ranked second, customising exercise and diet plans for specific medical conditions obtained unanimous consensus (100%) and a fuzzy score of 0.683, highlighting personalised adaptation as a critical component. Accessibility and feasibility for individuals with medical limitations ranked third (85.71% consensus, fuzzy score 0.630), reinforcing inclusivity in programme design.

Fourth-ranked was the presence of feedback mechanisms to assess programme effectiveness in managing medical conditions (90.48% consensus, fuzzy score 0.613), emphasising continuous improvement. Safety considerations for participants with varying medical conditions ranked fifth (85.7% consensus, fuzzy score 0.594), underlining risk reduction. Lastly, monitoring the impact of fitness and nutrition programmes on medical conditions ranked sixth (80.95% consensus, fuzzy score 0.530), underscoring adaptability based on changing health conditions. The high priority given to comprehensive inclusion (0.741) and personalised adaptation (0.683) indicates that the expert panel recognised the heterogeneity of medical conditions among individuals with obesity and the necessity for the AI system to accommodate this diversity.

Table 4.8
Fuzzy Delphi Method Analysis for 'Medical Conditions' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are a wide range of known medical conditions considered in the fitness and nutrition recommendations (e.g., diabetes, heart disease, arthritis)?	0.181	81.00%	0.741	1	Accepted
How effectively are the exercise and diet plans customized to accommodate specific medical conditions?	0.161	100.00%	0.683	2	Accepted
Are the recommendations safe for individuals with different medical conditions, minimizing health risks?	0.194	85.70%	0.594	5	Accepted
Are there mechanisms in place to monitor the impact of the fitness and diet plans on the medical conditions and adjust them as necessary?	0.199	80.95%	0.530	6	Accepted
Are the programs accessible and feasible for individuals with varying degrees of medical limitations?	0.180	85.71%	0.630	3	Accepted
Is there a feedback system to gather input from participants on the effectiveness of the programs in managing their medical conditions?	0.176	90.48%	0.613	4	Accepted

4.2.4.2 Element: Medications

Table 4.9 presents the prioritised criteria related to the element of 'Medications', emphasising medication considerations in fitness and nutrition planning. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion, ensuring exercise and nutrition recommendations avoid negative interactions with prescribed medications, achieved 95.2% consensus and the highest fuzzy score of 0.698, underscoring the significance of safety in medication-aware programme design. Second-ranked was effective customisation of health plans

based on medication effects (81% consensus, fuzzy score 0.667), highlighting personalised approaches. Adaptability of plans to accommodate medication regimen changes ranked third (76.2% consensus, fuzzy score 0.663), emphasising programme flexibility.

Fourth-ranked was responsiveness regarding necessary medication adjustments due to fitness and nutrition changes (95.24% consensus, fuzzy score 0.644). Ethical and legal considerations for medication-related privacy ranked fifth (85.71% consensus, fuzzy score 0.622), reinforcing participant trust and compliance with data protection requirements. Lastly, integrating medical records and medication histories into planning ranked sixth (80.95% consensus, fuzzy score 0.529), indicating foundational yet secondary importance. The results emphasise that safety, personalisation, adaptability, and ethical compliance are the primary considerations in health interventions that account for medication use.

These findings reveal a clear expert prioritisation hierarchy emphasising assessment rigour and measurable outcomes as foundational elements for effective obesity management. The unanimous consensus (100%) on thorough physical activity assessment reflects expert recognition that baseline and ongoing monitoring provides essential data for personalising interventions and detecting progress plateaus requiring strategy adjustment. Notably, the close fuzzy scores between goal alignment (0.765) and tracking systems (0.762) suggest these elements function synergistically rather than independently, as effective tracking mechanisms enable goal alignment by providing real-time feedback on progress toward health objectives. The relatively lower consensus on technology tools (76.19%, fuzzy score 0.665) compared to activity diversity (81%, fuzzy score 0.667) indicates expert ambivalence regarding technological dependence, possibly reflecting concerns about accessibility barriers, user technological literacy, or the risk of over-reliance on devices rather than intrinsic motivation.

Table 4.9
Fuzzy Delphi Method Analysis for 'Medications' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are the exercise and nutrition recommendations designed to avoid negative interactions with commonly prescribed medications?	0.181	95.20%	0.698	1	Accepted
How effectively are health and fitness plans customized to accommodate the effects of current medications on individuals?	0.151	81.00%	0.667	2	Accepted
Can the fitness and nutrition plans be easily adjusted based on changes in medication regimes?	0.190	76.20%	0.663	3	Accepted
Are medical records that include medication histories integrated into the fitness and nutrition planning process?	0.175	80.95%	0.529	6	Accepted
Are considerations for medication adjustments included if fitness and nutritional changes necessitate such adjustments?	0.140	95.24%	0.644	4	Accepted
Are ethical and legal considerations regarding medication information and privacy maintained throughout the program?	0.183	85.71%	0.622	5	Accepted

4.2.5 Construct 4: Lifestyle and Behaviour

Construct 4 addresses lifestyle and behavioural factors and comprises four elements and twenty-four criteria. The elements encompass sleep patterns, stress levels, physical activity habits, and dietary habits. The findings from the FDM analysis for each element are presented in the following subsections.

4.2.5.1 Element: Sleep Pattern

Table 4.10 presents the findings for the element of 'Sleep Pattern'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion, adjusting programmes for participants with diagnosed sleep disorders or irregular sleep patterns, achieved 90.48% consensus and a fuzzy score of 0.673, indicating strong agreement on addressing individual sleep-related challenges. Addressing the impact of sleep quality and duration on fitness and health outcomes ranked second (76.2% consensus, fuzzy score 0.665). This criterion represents a borderline case requiring sensitivity analysis, given its proximity to the 75% threshold. Nevertheless, the finding underscores sleep's critical role in recovery, weight management, and sustaining energy levels.

Educating participants on sleep's connection to health ranked third (80.95% consensus, fuzzy score 0.638). Integrating sleep patterns into personalised plans ranked fourth (81% consensus, fuzzy score 0.637), emphasising tailored approaches. Customised recommendations based on individual sleep quality ranked fifth (85.7% consensus, fuzzy score 0.602), followed by monitoring systems for tracking sleep (85.71% consensus, fuzzy score 0.583), highlighting data-driven approaches to improving health outcomes.

The expert prioritisation pattern reveals a sophisticated understanding of sleep's multifaceted role in obesity management. The highest ranking for addressing diagnosed sleep disorders (90.48%, fuzzy score 0.673) acknowledges that conditions such as sleep apnoea are both obesity consequences and weight loss barriers, creating a bidirectional relationship requiring explicit intervention. The criterion examining sleep's impact on fitness and health outcomes (76.2%, fuzzy score 0.665) represents a borderline case near the 75% consensus threshold, suggesting some expert uncertainty about whether AI-generated recommendations can adequately address the complex psychophysiological mechanisms linking sleep deprivation to weight gain, including hormonal disruption, increased appetite, and reduced exercise capacity.

Table 4.10
Fuzzy Delphi Method Analysis for 'Sleep Pattern' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
How well are sleep patterns considered and integrated into personalized fitness and nutrition plans?	0.198	81.00%	0.637	4	Accepted
Are the effects of sleep quality and duration on health and fitness outcomes sufficiently addressed in the program?	0.174	76.20%	0.665	2	Accepted
How effectively are recommendations customized based on individual sleep quality and patterns?	0.191	85.70%	0.602	5	Accepted
Are participants educated about the importance of sleep and its impact on weight management and overall health?	0.189	80.95%	0.638	3	Accepted
Are there effective systems in place for monitoring and tracking changes in sleep patterns as part of the fitness and nutrition program?	0.184	85.71%	0.583	6	Accepted
Are adjustments made for participants with diagnosed sleep disorders or irregular sleep patterns?	0.186	90.48%	0.673	1	Accepted

4.2.5.2 Element: Stress Level

Table 4.11 presents the prioritised criteria for the element of 'Stress Level'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The top-ranked criterion, considering stress levels in relation to health outcomes such as weight management, achieved the highest expert consensus (95.2%) and fuzzy score of 0.784, highlighting stress as a significant factor affecting physical and mental health. Ranked second, conducting initial and ongoing stress assessments to tailor programme recommendations obtained 81% consensus and a fuzzy score of 0.738,

emphasising personalised evaluation. Integrating stress management into fitness and nutrition programmes ranked third (76.2% consensus, fuzzy score 0.663).

Ensuring the safety and effectiveness of stress management methods ranked fourth (80.95% consensus, fuzzy score 0.621), followed by promoting relaxation techniques such as meditation and yoga (80.95% consensus, fuzzy score 0.538). Customising stress management based on individual triggers ranked sixth (80.95% consensus, fuzzy score 0.529), supporting a holistic, personalised approach to wellness. The high consensus and fuzzy score for stress consideration (95.2%, 0.784) reflects the expert panel's recognition of the well-established relationship between chronic stress, cortisol elevation, and weight gain, particularly visceral adiposity.

Table 4.11
Fuzzy Delphi Method Analysis for 'Stress Level' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are initial and ongoing assessments of stress levels included in the program to tailor recommendations?	0.182	81.00%	0.738	2	Accepted
How are stress levels considered in relation to their impact on health outcomes, including weight management?	0.161	95.20%	0.784	1	Accepted
How effectively are stress management strategies integrated into fitness and nutrition plans?	0.190	76.20%	0.663	3	Accepted
Are stress management techniques customized based on individual stress triggers and responses?	0.175	80.95%	0.529	6	Accepted
Are the recommended stress management techniques safe and proven effective?	0.198	80.95%	0.621	4	Accepted
Are relaxation techniques (e.g., meditation, deep breathing, yoga) incorporated into the program?	0.188	80.95%	0.538	5	Accepted

4.2.5.3 Element: Physical Activity Habits

Table 4.12 presents the prioritised criteria for integrating the element of 'Physical Activity Habits'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion, effectively aligning physical activity habits with overall health and weight management goals, achieved the top fuzzy score (0.781) and expert consensus (85.7%), emphasising alignment with broader health objectives. The use of technology, such as applications and wearable devices, ranked second (85.71% consensus, fuzzy score 0.760), highlighting its role in enhancing participant engagement. Consistent initial and ongoing assessments of physical activity levels ranked third (95.2% consensus, fuzzy score 0.689), underscoring the importance of regular monitoring.

Specific metrics to measure physical activity improvements ranked fourth (90.48% consensus, fuzzy score 0.630), followed by tracking and feedback systems (80.95% consensus, fuzzy score 0.519). Finally, incorporating diverse physical activities to accommodate varied interests ranked sixth (76.2% consensus, fuzzy score 0.510), stressing inclusivity and sustained engagement. Notably, the high ranking of technology use (0.760) in this element contrasts with the relatively lower ranking of technology tools in the 'Desired Weight Management Outcomes' element (0.665), suggesting that experts view technology as more valuable for tracking physical activity specifically than for general weight management outcomes.

Table 4.12
Fuzzy Delphi Method Analysis for 'Physical Activity Habits' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
How thoroughly are initial and ongoing assessments of physical activity levels integrated into the program?	0.197	95.20%	0.689	3	Accepted
How effectively are physical activity habits integrated with overall health and weight management goals?	0.190	85.70%	0.781	1	Accepted
Does the program include a wide range of physical activities to cater to different interests and capabilities?	0.166	76.20%	0.510	6	Accepted
Are there effective systems in place for tracking physical activity and providing feedback to participants?	0.158	80.95%	0.519	5	Accepted
How is technology (e.g., apps, wearable devices) used to support and enhance engagement in physical activities?	0.181	85.71%	0.760	2	Accepted
Are specific metrics used to measure improvements or changes in physical activity habits?	0.170	90.48%	0.630	4	Accepted

4.2.5.4 Element: Dietary Habits

Table 4.13 presents the prioritised criteria for the element of 'Dietary Habits'. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

The highest-ranked criterion, consideration of environmental impacts and sustainability in dietary recommendations, achieved the top fuzzy score (0.768) and expert consensus (90.48%), highlighting alignment with ecological sustainability. Ranked second, aligning dietary habits with individual fitness and weight management goals scored a fuzzy score of 0.700 with a consensus of 76.2%, emphasising personalised dietary integration. Thorough assessments of dietary habits during

programme initiation and progression ranked third (100% consensus, fuzzy score 0.689), underscoring the importance of continuous monitoring.

Fourth was the consideration of dietary habits' psychological effects (95.24% consensus, fuzzy score 0.684), followed by ensuring dietary recommendations' safety based on medical history (81% consensus, fuzzy score 0.648). Lastly, technology use, such as meal-tracking applications, ranked sixth (85.71% consensus, fuzzy score 0.511), enhancing adherence and participant engagement. The top ranking of environmental sustainability in dietary recommendations (0.768) is a notable finding, potentially reflecting growing expert awareness of the intersection between planetary health and individual dietary behaviour, and the recognition that sustainable dietary patterns often align with health-promoting food choices.

Table 4.13
Fuzzy Delphi Method Analysis for 'Dietary Habits' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
How thoroughly are initial and ongoing assessments of dietary habits integrated into the program?	0.195	100.00%	0.689	3	Accepted
How effectively are dietary habits aligned with and supportive of individual fitness and weight management goals?	0.154	76.20%	0.700	2	Accepted
Are the recommended diets safe and suitable for individuals based on their medical history and current health status?	0.179	81.00%	0.648	5	Accepted
How is technology (e.g., meal tracking apps, online nutrition counseling) used to support and enhance adherence to dietary recommendations?	0.138	85.71%	0.511	6	Accepted
Are the effects of dietary habits on psychological well-being considered and addressed?	0.139	95.24%	0.684	4	Accepted
Are environmental impacts and sustainability considered in the dietary recommendations?	0.156	90.48%	0.768	1	Accepted

4.2.6 Construct 5: Motivation and Social Support

Construct 5 addresses motivational factors and social support and comprises one element and six criteria. The findings from the FDM analysis are presented in the following subsection.

4.2.6.1 Element: Motivations for Weight Management

Table 4.14 presents the analysis of the element of 'Motivations for Weight Management', highlighting expert consensus on tailored motivational strategies. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

Regular feedback and positive reinforcement were ranked highest (fuzzy score 0.700, 76.19% consensus), emphasising continuous feedback for sustained motivation. Alignment of strategies with individual health goals ranked second (fuzzy score 0.698, 100% consensus), reinforcing goal-oriented personalisation. Third was tailoring strategies to personal backgrounds (fuzzy score 0.665, 76.2% consensus), followed by the clarity of motivational relevance to specific goals (fuzzy score 0.640, 85.7% consensus), integration into daily routines (fuzzy score 0.592, 85.71% consensus), and adaptability based on changing needs (fuzzy score 0.529, 80.95% consensus).

These findings support structured, individualised, and adaptable motivational frameworks within the NExGEN-ChatGPT system. The top ranking of regular feedback and positive reinforcement (0.700) despite having the lowest consensus percentage (76.19%) among the top criteria is noteworthy, as it suggests that the experts who endorsed this criterion did so with strong conviction. This pattern aligns with behavioural psychology principles indicating that consistent positive reinforcement is among the most effective strategies for sustaining behaviour change in weight management programmes.

Table 4.14
Fuzzy Delphi Method Analysis for 'Motivations for Weight Management' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
Are the motivational strategies clear and directly relevant to the individual's weight management goals?	0.171	85.70%	0.640	4	Accepted
How well are motivational strategies aligned with personal health and fitness objectives?	0.180	100.00%	0.698	2	Accepted
How personalized are the motivational strategies based on individual needs and backgrounds?	0.174	76.20%	0.665	3	Accepted
Are regular feedback and positive reinforcement used as part of the motivational strategies?	0.154	76.19%	0.700	1	Accepted
How well are motivational strategies integrated into participants' daily routines and lifestyles?	0.189	85.71%	0.592	5	Accepted
Are motivational strategies flexible and adjustable based on feedback and changing needs?	0.175	80.95%	0.529	6	Accepted

4.2.7 Construct 6: Environmental Factors

Construct 6 addresses environmental factors and comprises one element and six criteria. The findings from the FDM analysis are presented in the following subsection.

4.2.7.1 Element: Environmental Factors

Table 4.15 presents the analysis of grocery shopping habits within the element of 'Environmental Factors', highlighting their importance in supporting nutritional strategies. A total of six criteria were evaluated, and all six met the acceptance threshold conditions.

Personalised shopping strategies aligned with individual dietary preferences ranked highest (fuzzy score 0.768, 90.5% consensus). Feedback mechanisms evaluating shopping habit changes ranked second (fuzzy score 0.668, 95.2% consensus), emphasising regular assessment. Alignment of grocery habits with programme nutritional goals ranked third (fuzzy score 0.640, 81% consensus), followed by integration with meal planning (fuzzy score 0.529, 81% consensus), educational programmes promoting healthy choices (fuzzy score 0.517, 90.5% consensus), and the influence of shopping behaviour on dietary outcomes (fuzzy score 0.510, 76.2% consensus).

The results indicate that personalised, supported, and education-oriented grocery strategies enhance dietary success. The high priority placed on personalised shopping strategies (0.768) reflects expert recognition that the food environment and purchasing behaviour represent critical upstream determinants of dietary quality. By integrating grocery shopping guidance into the AI prompt generator, the NExGEN-ChatGPT framework addresses a practical barrier that is often overlooked in traditional dietary counselling.

Table 4.15
Fuzzy Delphi Method Analysis for 'Environmental Factors' Element

Item	Condition of Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Threshold Value, d	Percentage of Experts Group Consensus (%)	Fuzzy Scores (A)	Position	Expert Consensus
How well do grocery shopping habits support the nutritional goals set in the program?	0.180	81.00%	0.640	3	Accepted
How significantly do grocery shopping habits influence the dietary choices of participants?	0.166	76.20%	0.510	6	Accepted
Does the program provide adequate education on how to make healthy choices while grocery shopping?	0.164	90.50%	0.517	5	Accepted
Are grocery shopping strategies personalized based on individual dietary needs and preferences?	0.156	90.48%	0.768	1	Accepted
How effectively are grocery shopping habits integrated with meal planning within the program?	0.175	80.95%	0.529	4	Accepted
Are there feedback mechanisms in place to assess the effectiveness of changes in grocery shopping habits?	0.128	95.24%	0.668	2	Accepted

4.2.8 Summary of Fuzzy Delphi Method Findings

Table 4.16 presents the overall summary of the FDM findings across all six constructs, twelve elements, and their associated criteria. Overall, all criteria evaluated across the six constructs met the three acceptance threshold conditions: expert consensus exceeding 75%, threshold value (d) less than 0.2, and α-cut value exceeding 0.5. This indicates that the expert panel reached sufficient agreement on the importance and applicability of all proposed criteria for inclusion in the NExGEN-ChatGPT personalised AI prompt generator for obesity management.

The construct with the highest overall fuzzy scores was Construct 1 (Goal Settings), with the criterion of thorough initial and ongoing assessment of physical activity achieving the highest individual fuzzy score of 0.859 and unanimous expert

consensus. The construct with the lowest overall fuzzy scores was Construct 6 (Environmental Factors), with the criterion concerning the influence of shopping behaviour on dietary outcomes recording the lowest individual fuzzy score of 0.510, though this still exceeded the minimum acceptance threshold of 0.5.

Taken together, the FDM findings provide robust validation of the proposed constructs, elements, and criteria for the NExGEN-ChatGPT framework. The accepted criteria form the foundation for the personalised AI prompt generator, ensuring that the system's recommendations are grounded in expert consensus and address the multifaceted nature of obesity management, encompassing goal setting, demographic and physical considerations, health and medical history, lifestyle and behavioural factors, motivational support, and environmental influences.

Table 4.16
Summary of Fuzzy Delphi Method Findings Across All Constructs and Elements

No.	Construct	Elements (n)	Accepted Criteria (n)	Fuzzy Score Range	Expert Consensus Range (%)
1	Weight Management Outcomes	3	18	0.529 – 0.859	76.19 – 100.00
2	Individual Characteristics	3	18	0.511 – 0.681	76.19 – 100.00
3	Health Status	2	12	0.529 – 0.741	76.20 – 100.00
4	Lifestyle Factors	2	12	0.529 – 0.784	76.20 – 95.20
5	Behavioral Habits	2	12	0.510 – 0.781	76.20 – 100.00
6	Psycho-Environmental Factors	2	12	0.510 – 0.768	76.19 – 100.00
	Total	**14**	**84**	**0.510 – 0.859**	**76.19 – 100.00**

Note. Across the six constructs, a total of 84 criteria were validated, with fuzzy scores ranging from 0.510 to 0.859 and expert consensus percentages ranging from 76.19% to 100.00%. All criteria met the acceptance threshold ($d \leq 0.2$, consensus $\geq 75\%$, fuzzy score ≥ 0.5).

4.3 Study 2: Personalised AI Prompt Generator and ChatGPT for Weight Loss: Randomised Controlled Trial in Adults with Overweight and Obesity

A Randomized Controlled Trial (RCT) design to evaluate the effectiveness of the NExGEN-ChatGPT personalised AI prompt generator, validated through the Fuzzy Delphi Method in Study 1, on weight loss and associated health outcomes in adults with overweight and obesity. Part of Study 2 results has been accepted for publication in 2026 in the *Journal of Technology in Behavioral Science*, *Springer Nature* (*Scopus-indexed, SJR Quartile 1, H-index: 24).* The findings are presented systematically, beginning with participant recruitment and baseline descriptive characteristics, followed by statistical assumptions, and subsequently reporting the outcome measures: weight and body composition changes, dietary and physical activity changes, metabolic parameters, and blood pressure responses. Results are reported for both within-group (changes from baseline) and between-group (NEX vs. CON) comparisons across three time points: baseline, 12-week (post-intervention), and 3-month follow-up.

4.3.1 Baseline Descriptive Characteristics

4.3.1.1 Participant Recruitment and Flow

Recruitment of participants residing in Greater Kuala Lumpur (Klang Valley), Malaysia, was conducted from May 2024 to August 2025 through both online and offline media, including news items, flyers, and social media platforms. During this extended recruitment period, a total of 243 potential participants formally expressed their interest in the study. Following screening procedures comprising telephone interviews and preliminary examinations, 160 subjects met the inclusion criteria and were enrolled in the study. Participants were sequentially booked for baseline assessments and subsequently randomised into the NEX intervention group or the CON control group.

Throughout the study duration, there were 36 dropouts across both groups. In the per-protocol analysis, three additional participants were excluded: two from the NEX group (one due to incomplete baseline data identified during data cleaning, and one due to unreliable self-reported data detected through quality assurance procedures) and one from the CON group (who missed the standardised clinical assessment due to

relocation outside the study area). All excluded participants were retained in the intention-to-treat (ITT) analysis per original group assignment. The study participant flow, detailing participant progression through enrolment, allocation, follow-up, and analysis, is shown in Figure 4.1.

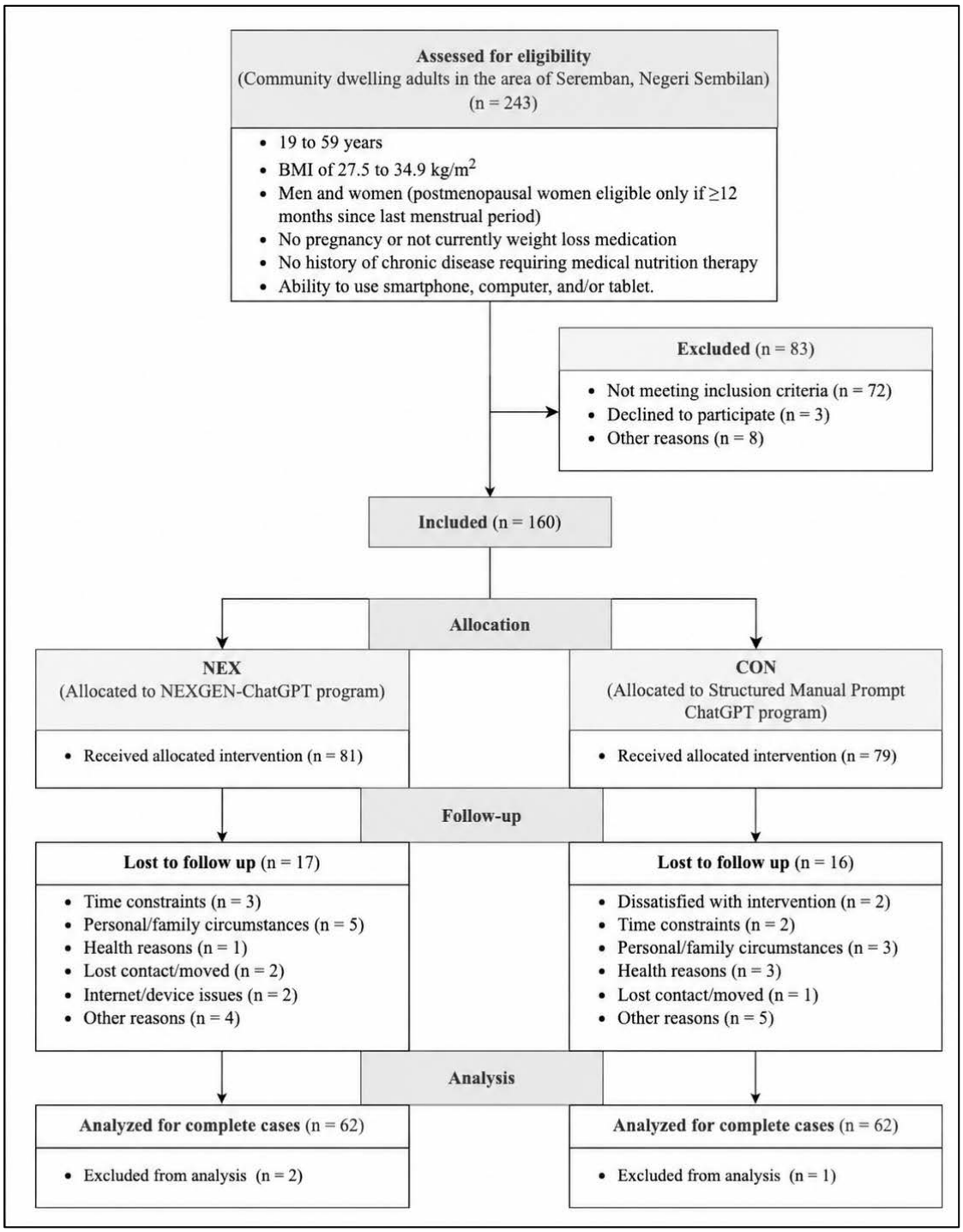


Figure 4.1 CONSORT Flow Diagram of Participant Recruitment, Allocation, Follow-up, and Analysis in Study 2

4.3.1.2 Anthropometric and Demographic Characteristics

Table 4.17 presents the baseline demographic and clinical characteristics of the 160 participants randomised into the exercise (NEX, n = 81) and control (CON, n = 79) groups. The mean age of participants was 35.90 ± 9.39 years, with a predominantly female sample (67.5%). The ethnic composition reflected the Malaysian population, comprising Malay (48.8%), Chinese (28.7%), Indian (21.2%), and other ethnicities (1.2%). The majority of participants held an undergraduate degree (59.4%), were employed full-time (61.3%), and were single (64.4%). At baseline, participants had a mean body weight of 88.01 ± 10.24 kg, a mean height of 1.66 ± 0.10 m, and a mean BMI of 31.20 ± 2.58 kg/m², classifying the sample as obese (Class I). A large proportion of participants (81.2%) reported being physically inactive, exercising fewer than five times per week. The overall dropout rate was 20.63% (n = 33), distributed similarly between the NEX (20.99%) and CON (20.25%) groups.

The distribution of demographic and anthropometric variables appeared comparable between the two groups. Both groups demonstrated similar age profiles (NEX: 36.04 ± 9.48 years; CON: 35.75 ± 9.34 years), sex distribution, ethnic composition, and education levels. Baseline clinical measures were also closely matched, with mean BMI values of 31.28 ± 2.60 kg/m² and 31.12 ± 2.57 kg/m² for the NEX and CON groups, respectively. Physical activity patterns were nearly identical, with 81.5% of NEX and 81.0% of CON participants classified as inactive. These balanced baseline characteristics suggest that the randomisation process was effective in producing equivalent groups, thereby reducing the likelihood that observed between-group differences in outcomes could be attributed to pre-existing participant difference.

Table 4.17
Baseline Demographic and Clinical Characteristics of Study Participants

Variable	All participants (N=160)	NEX (n=81)	CON (n=79)
Age (yrs)[a]	35.90 ± 9.39	36.04 ± 9.48	35.75 ± 9.34
Sex[b]			
Male	52 (32.5)	25 (30.9)	27 (34.2)
Female	108 (67.5)	56 (69.1)	52 (65.8)
Ethnicity[b]			
Malay	78 (48.8)	40 (49.4)	38 (48.1)
Chinese	46 (28.7)	24 (29.6)	22 (27.8)
Indian	34 (21.2)	16 (19.8)	18 (22.8)
Others	2 (1.2)	1 (1.2)	1 (1.3)
Education Level[b]			
Primary or secondary school	7 (4.4)	4 (4.9)	3 (3.8)
STPM/pre-university/diploma	23 (14.4)	10 (12.3)	13 (16.5)
Undergraduate (eg., Degree)	95 (59.4)	49 (60.5)	46 (58.2)
Postgraduate (eg., Master, PhD)	35 (21.9)	18 (22.2)	17 (21.5)
Work status[b]			
Employed–full-time	98 (61.3)	49 (60.5)	49 (62.0)
Employed–part-time	4 (2.5)	2 (2.5)	2 (2.5)
Self-employed	16 (10.0)	8 (9.9)	8 (10.1)
Retired	3 (1.9)	1 (1.2)	2 (2.5)
Not employed (eg., student)	39 (24.4)	21 (25.9)	18 (22.8)
Marital status[b]			
Single	103 (64.4)	54 (66.7)	49 (62.0)
Married	57 (35.6)	27 (33.3)	30 (38.0)
Body weight (kg)[a]	88.01 ± 10.24	88.41 ± 10.54	87.60 ± 9.97
Body height (m)[a]	1.66 ± 0.10	1.66 ± 0.10	1.65 ± 0.10
BMI (kg/m^2)[a]	31.20 ± 2.58	31.28 ± 2.60	31.12 ± 2.57

Variable	All participants (N=160)	NEX (n=81)	CON (n=79)
Physical activity frequency (per wk)[b]			
Active (≥5 times)	30 (18.8)	15 (18.5)	15 (19.0)
Inactive (<5 times)	130 (81.2)	66 (81.5)	64 (81.0)
Dropouts[b]	33 (20.63)	17 (20.99)	16 (20.25)

[a] values are presented as the mean ± SD. [b] values are presented as the n (%).

4.3.1.3 Baseline Dietary Intake and Physical Activity

Baseline dietary intake and physical activity levels were assessed prior to any intervention to confirm comparability between the two groups. At baseline, the NEX group had a mean energy density of 4.84 ± 0.25 kcal/g, energy intake of 1833.2 ± 659.7 kcal/day, and fat intake of 80.9 ± 33.0 g/day. The CON group showed comparable values of 4.84 ± 0.26 kcal/g, 1830.1 ± 656.7 kcal/day, and 80.5 ± 31.9 g/day, respectively. No significant differences in baseline dietary intake parameters were detected between groups (all $P > 0.05$), confirming dietary equivalence at the start of the intervention.

4.3.2 Statistical Assumptions

4.3.2.1 Normal Distribution

In order to test for normal distribution, data were explored and the significance value of the Shapiro-Wilk normality test (W) statistic was reported. Data were considered normally distributed if the P-value statistics were not significant ($P > 0.05$) (Pallant, 2020). In addition to the Shapiro-Wilk test, data were considered within the limits of a normal distribution provided that the magnitude of the skewness and kurtosis statistics did not surpass ± 2.0 (Schmider et al., 2010). The quantile-to-quantile (Q-Q) plots were also examined as a graphical method to assess normal distribution, whereby data points lying on or near the straight diagonal line (i.e., 45°) with minimal deviations were considered normally distributed (Chambers et al., 2018).

The Shapiro-Wilk test, skewness and kurtosis statistics, and Q-Q plots were used to evaluate the normality of all primary and secondary outcome variables across the NEX (n = 81) and CON (n = 79–81) groups at baseline, 12-week, and follow-up time points. These variables encompassed weight and body composition measures (body weight, waist circumference, fat mass, fat-free mass), dietary intake parameters (energy density, energy intake, carbohydrate intake, fat intake, protein intake, fibre intake), physical activity indices (accelerometer-derived MET-min/wk and IPAQ-LV min/week), metabolic parameters (HbA1c, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides), and blood pressure measures (systolic BP, diastolic BP, mean arterial pressure). The full normality test results are presented in Appendix 25 (NEX condition) and Appendix 25 (CON condition). The complete Q-Q plots for all datasets are presented in Appendix 26.

Across the 120 group-by-time combinations assessed (20 variables × 3 time points × 2 groups), the Shapiro-Wilk test indicated statistically significant departures from normality ($p < 0.05$) in 97 instances. HDL cholesterol passed the Shapiro-Wilk test in all six group-by-time combinations. Energy density and fat-free mass each passed in five of six combinations. A small number of additional isolated non-significant results were observed: HbA1c at baseline in both the NEX ($W = 0.974$, $p = 0.094$) and CON ($W = 0.974$, $p = 0.103$) groups, total cholesterol in the CON group at baseline ($W = 0.973$, $p = 0.092$), LDL cholesterol in the CON group at follow-up ($W = 0.981$, $p = 0.288$), body weight in the CON group at 12-week ($W = 0.973$, $p = 0.088$), fat intake in the CON group at 12-week ($W = 0.969$, $p = 0.054$), and carbohydrate intake in the CON group at follow-up ($W = 0.984$, $p = 0.448$). All remaining group-by-time combinations, including all blood pressure measures, physical activity indices, and the majority of metabolic and dietary variables, produced significant W statistics.

Despite these significant Shapiro-Wilk results, several converging lines of evidence supported the use of parametric statistical tests for all variables in the present study. First, the skewness and kurtosis statistics for every variable across all 120 group-by-time combinations fell well within the ±2.0 acceptable range (Andy Field, 2013; George & Mallery, 2010). Skewness values ranged from −0.580 to 0.523, indicating negligible asymmetry in all distributions. Kurtosis values ranged from −1.701 to 0.975, with the predominant pattern reflecting mild platykurtosis (flatter distributions with lighter tails) rather than any problematic leptokurtic tendency. These values represent trivial departures from normality in practical terms.

Second, with sample sizes of 79 to 81 participants per group, the Central Limit Theorem provides a robust theoretical basis for parametric inference. When sample sizes exceed approximately 30 to 40 observations per group, the sampling distribution of the mean converges to normality regardless of the underlying population distribution (Andy Field, 2013; Lumley et al., 2002). The Shapiro-Wilk test is known to be overly sensitive at larger sample sizes, flagging statistically significant but practically negligible deviations that carry no meaningful consequence for the validity of parametric analyses (Ghasemi & Zahediasl, 2012; Mohd Razali & Bee Wah, 2011). Accordingly, reliance on formal significance tests alone to determine analytical approach has been cautioned against in favour of evaluating the practical magnitude of any departure (Altman & Bland, 1995).

Third, parametric tests including the mixed Analysis of Variance (ANOVA) and independent-samples t-test are well-established as robust to moderate violations of the normality assumption, particularly when group sizes are approximately equal, as in the present study (Blanca et al., 2017; Glass et al., 1972; Schmider et al., 2010). Simulation studies have demonstrated that Type I error rates and statistical power remain largely unaffected by non-normality of the magnitude observed here, especially with balanced designs and sample sizes exceeding 40 per group (Blanca et al., 2017)

Taken together, the combination of skewness and kurtosis values within acceptable limits (±2.0), adequate sample sizes invoking the Central Limit Theorem, the established robustness of parametric methods to mild distributional violations, and consistent Q-Q plot inspection confirmed that the assumption of approximate normality was satisfactorily met. Parametric statistical tests were therefore applied to all weight, body composition, dietary intake, physical activity, metabolic, and blood pressure variables.

4.3.2.2 Outliers

Prior to the main statistical analyses, all datasets were examined for the presence of outliers. Screening for outliers is a critical step in data preparation, as extreme or anomalous values have the potential to inflate error rates, distort measures of central tendency and variability, and compromise the validity of subsequent parametric analyses (Aguinis et al., 2013). If left undetected, outliers can exert a disproportionate

influence on parameter estimates, leading to misleading conclusions regarding the magnitude and direction of treatment effects (Leys et al., 2013).

Outlier detection in the present study was conducted using the Robust Regression and Outlier Removal (ROUT) method within GraphPad Prism software. The ROUT method was selected over alternative approaches, such as the Grubbs' test, because it is capable of identifying multiple outliers simultaneously without being susceptible to masking effects, whereby the presence of one outlier conceals the detection of another (Motulsky & Brown, 2006). The ROUT procedure operates in two stages. In the first stage, a robust nonlinear regression model is fitted to the data using a method that minimises the influence of potential outliers on the model parameters, thereby producing a fit that accurately represents the majority of the data. In the second stage, residuals from this robust fit are evaluated using an adaptation of the false discovery rate (FDR) approach to determine which data points lie sufficiently far from the model's predictions to be classified as outliers (Motulsky & Brown, 2006).

The false discovery rate coefficient, denoted as Q, was set at 1%, the most conservative threshold available within the software. At this setting, only definitive outliers, those with a less than 1% probability of being falsely identified, are flagged for removal. This conservative threshold was deliberately chosen to minimise the risk of erroneously excluding legitimate data points while still ensuring the detection of genuinely aberrant values. The ROUT analysis was applied to all primary and secondary outcome variables across both the NEX and CON experimental conditions at all time points (baseline, 12-week, and follow-up). The analysis identified no outliers in any variable within any experimental condition, and consequently, no data points were excluded from subsequent statistical analyses on this basis.

4.3.2.3 Homogeneity of Variance

Homogeneity of variance is an assumption whereby the spread of the scores from the mean is considered equal across groups. To determine whether equal variances existed, all datasets were analysed with Levene’s Test of Equality of Error Variances. If $P > 0.05$, the data does not violate the assumption of equal variances. However, if the P-value of Levene’s test is less than 0.05, the variances for the groups have violated the assumption. Levene’s Test of Equality of Error Variances was conducted for all primary outcome variables across the experimental groups at each measurement time point. The

results for weight and body composition parameters, dietary changes parameters, physical activity parameters, metabolic changes parameters, and blood pressure parameters in Appendix 27.

For weight and body composition, body weight and waist circumference satisfied the assumption of equal variances at all three time points (all $P > 0.05$). Fat mass violated the assumption at 12-week ($F = 2.72$, $P = 0.05$) and follow-up ($F = 3.35$, $P = 0.02$), although the baseline assessment was non-significant ($P = 0.12$). Fat-free mass violated the assumption at 12-week ($F = 3.77$, $P = 0.01$), while baseline ($P = 0.91$) and follow-up ($P = 0.51$) remained non-significant. For dietary changes, energy intake, carbohydrate intake, and fat intake satisfied the assumption at all time points (all $P > 0.05$). Protein intake violated the assumption at 12-week ($F = 2.86$, $P = 0.04$) and follow-up ($F = 3.06$, $P = 0.03$), while baseline was non-significant ($P = 0.18$). Fiber intake similarly violated the assumption at 12-week ($F = 3.50$, $P = 0.02$) and follow-up ($F = 3.81$, $P = 0.01$), with a non-significant baseline ($P = 0.55$). Energy density demonstrated significant violations at all three time points ($P < 0.001$ for all), representing the most consistent departure from the assumption among dietary parameters.

For physical activity, both accelerometer-measured physical activity and self-reported physical activity (IPAQ-LV) satisfied the assumption at baseline ($P = 0.38$ and $P = 0.07$, respectively). The accelerometer measure violated the assumption at 12-week ($F = 3.45$, $P = 0.02$) and follow-up ($F = 3.07$, $P = 0.03$). The IPAQ-LV measure violated the assumption at 12-week ($F = 14.20$, $P < 0.001$) but was non-significant at follow-up ($P = 0.36$). For metabolic changes, triglycerides satisfied the assumption of equal variances at all three time points (all $P > 0.05$). HbA1c violated the assumption at baseline ($F = 2.90$, $P = 0.04$) and follow-up ($F = 5.16$, $P < 0.01$), while the 12-week assessment was non-significant ($P = 0.83$). Total cholesterol violated the assumption at all three time points ($P = 0.02$, $P < 0.01$, and $P = 0.02$ for baseline, 12-week, and follow-up, respectively). LDL cholesterol violated the assumption at baseline ($F = 4.09$, $P < 0.01$) and follow-up ($F = 4.07$, $P < 0.01$), with a non-significant 12-week ($P = 0.61$). HDL cholesterol violated the assumption at 12-week ($F = 4.30$, $P < 0.01$) and follow-up ($F = 3.34$, $P = 0.02$), while baseline remained non-significant ($P = 0.08$). For blood pressure, diastolic blood pressure and mean arterial pressure satisfied the assumption of equal variances at all three time points (all $P > 0.05$). Systolic blood pressure also satisfied the assumption at baseline ($P = 0.74$) and 12-week ($P = 0.46$), but violated the assumption at follow-up ($F = 4.83$, $P < 0.01$). Overall, the blood pressure domain

demonstrated the most favourable results, with only one of nine tests yielding a significant violation.

Across all five domains, a total of 60 Levene's tests were conducted. Of these, 37 tests (61.7%) satisfied the assumption of equal variances, while 23 tests (38.3%) produced statistically significant results. Despite these violations, the use of parametric tests, specifically the mixed-design ANOVA with repeated measures, remains appropriate and defensible for all parameters investigated. Several lines of evidence support this position. First, ANOVA is widely recognised as robust to moderate violations of the homogeneity of variance assumption, particularly when group sizes are approximately equal (Glass et al., 1972; Tabachnick et al., 2019) . In the present study, the four experimental subgroups maintained approximately balanced sample sizes ($n \approx 40$ per subgroup), which substantially mitigates the impact of unequal variances on Type I error rates. Balanced designs ensure that the pooled error variance remains an unbiased estimate, even under heterogeneity.

Second, the variance ratios for the overwhelming majority of violated conditions remained below the critical 4:1 threshold recommended by(Howell, 2013), above which ANOVA results may become unreliable. The variance ratios for fat mass (maximum 2.11:1), fat-free mass (2.03:1), protein intake (1.92:1), fiber intake (3.08:1), accelerometer-measured physical activity (3.22:1), HbA1c (2.98:1), total cholesterol (2.95:1), HDL cholesterol (2.55:1), systolic blood pressure (2.52:1), and the majority of LDL cholesterol time points (2.63:1 at baseline) all fell within this acceptable range. Three conditions exceeded the 4:1 threshold and warrant specific consideration. Energy density exhibited variance ratios up to 6.81:1, driven by the overweight subgroups displaying greater variability than the obese subgroups. The absolute magnitudes of these variances were very small (0.02 to 0.13 $kcal/g^2$), meaning that the elevated ratios reflect trivial practical differences in data spread rather than substantive heterogeneity that would compromise the F-statistic (Keppel & Wickens, 2004). The IPAQ-LV at 12-week exhibited a variance ratio of 5.56:1, attributable to the control overweight subgroup demonstrating greater within-group variability at a single time point, with an otherwise compliant baseline and follow-up. LDL cholesterol at follow-up exhibited a variance ratio of 4.33:1, driven by the NEX overweight subgroup displaying a narrower distribution (SD = 12.51 mg/dL) compared to the other three subgroups (SD = 22.08 to 26.04 mg/dL). This ratio only marginally exceeded the threshold and is unlikely to

meaningfully distort the F-statistic given the balanced group sizes (Tabachnick & Fidell, 2019).

Third, Norman (2010) demonstrated that parametric tests such as ANOVA are robust to violations of underlying assumptions when sample sizes exceed 30 per group and the distributions are not severely skewed. The present study satisfies both conditions, with each subgroup containing approximately 40 participants and the data distributions demonstrating adequate normality as verified in the preceding section. Taken together, these considerations provide sufficient justification for the use of parametric analysis across all weight and body composition, dietary changes, physical activity, metabolic changes, and blood pressure parameters investigated in this study.

4.3.2.4 Sphericity

For repeated measures analyses involving three or more time points, Mauchly's test of sphericity was conducted to determine whether the variances of the differences between all combinations of related groups were equal. When the assumption of sphericity was violated ($p < 0.05$), the Geisser-Greenhouse correction was applied to adjust the degrees of freedom, providing a more conservative test of significance. This correction was applied to all repeated measures analyses where sphericity was violated.

Mauchly's test indicated that the assumption of sphericity was met for body weight ($\chi^2(2) = 5.40$, $p = .067$), waist circumference ($\chi^2(2) = 4.68$, $p = .097$), fat mass ($\chi^2(2) = 0.11$, $p = .948$), fat-free mass ($\chi^2(2) = 2.45$, $p = .294$), energy density ($\chi^2(2) = 0.63$, $p = .729$), carbohydrate intake ($\chi^2(2) = 2.98$, $p = .225$), protein intake ($\chi^2(2) = 2.03$, $p = .363$), fibre intake ($\chi^2(2) = 0.48$, $p = .786$), accelerometer-derived physical activity ($\chi^2(2) = 5.53$, $p = .063$), HbA1c ($\chi^2(2) = 2.76$, $p = .252$), total cholesterol ($\chi^2(2) = 1.43$, $p = .489$), LDL cholesterol ($\chi^2(2) = 2.05$, $p = .359$), HDL cholesterol ($\chi^2(2) = 1.54$, $p = .463$), triglycerides ($\chi^2(2) = 1.12$, $p = .573$), systolic blood pressure ($\chi^2(2) = 0.86$, $p = .650$), and mean arterial pressure ($\chi^2(2) = 3.32$, $p = .190$). Uncorrected degrees of freedom were therefore used in the repeated measures analyses for these variables.

The assumption of sphericity was violated for energy intake ($\chi^2(2) = 15.20$, $p < .001$, $\varepsilon = 0.92$), fat intake ($\chi^2(2) = 14.12$, $p = .001$, $\varepsilon = 0.92$), self-reported physical activity via IPAQ-LV ($\chi^2(2) = 106.37$, $p < .001$, $\varepsilon = 0.67$), and diastolic blood pressure ($\chi^2(2) = 6.01$, $p = .049$, $\varepsilon = 0.96$). The Geisser-Greenhouse correction was applied to

adjust the degrees of freedom for all repeated measures analyses involving these variables.

4.3.3 Changes in Body Weight

Group mean changes in body weight relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.2. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for body weight ($P < 0.001$; effect size, 0.058), indicating that the response across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that body weight was significantly reduced in the NEX group from baseline at 12-week ($P < 0.001$) and at the 3-month follow-up ($P < 0.001$), representing reductions of 6.6 kg (7.5%) and 5.5 kg (6.2%), respectively. Between-group comparisons demonstrated that the NEX group achieved significantly greater weight loss than the CON group at 12-week (mean difference −3.2 kg; $P = 0.016$) and at follow-up (mean difference −3.5 kg; $P = 0.018$). In contrast, the CON group experienced a significant reduction in body weight of 3.0% at 12-week ($P = 0.003$), but this was not maintained at follow-up (1.3%; $P = 0.100$).

At baseline, body weight was comparable between the NEX (88.41 ± 10.54 kg) and CON (87.60 ± 9.97 kg) groups. At 12-week, the NEX group weighed 81.76 ± 9.27 kg compared to 84.98 ± 7.77 kg in the CON group. At follow-up, values were 82.94 ± 9.50 kg and 86.46 ± 9.11 kg for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant time × intervention interaction ($F(2, 316) = 9.811$, $P < 0.001$, $\eta p^2 = 0.058$), a significant main effect of time ($F(2, 316) = 38.716$, $P < 0.001$, $\eta p^2 = 0.197$), and a non-significant main effect of group ($F(1, 158) = 2.155$, $P = 0.144$, $\eta p^2 = 0.013$). Within-group pairwise comparisons for the CON group revealed a significant reduction in body weight from baseline to 12-week (−2.62 kg, 3.0%; 95% CI −4.32 to −0.91; $P = 0.003$), but this reduction was not maintained at follow-up (−1.14 kg, 1.3%; 95% CI −2.51 to 0.22; $P = 0.100$).

Overall, body weight was significantly reduced by 7.5% at 12-week and 6.2% at the 3-month follow-up in the NEX group relative to baseline, with the NEX group achieving 3.2 kg and 3.5 kg greater weight loss than the CON group at 12-week and follow-up, respectively. The sustained weight reduction at follow-up indicates that the behavioural changes induced by the NExGEN-ChatGPT intervention were durable

beyond the active intervention period, although a modest weight regain of approximately 1.1 kg (from 6.6 kg to 5.5 kg loss) was observed between 12-week and follow-up. Therefore, the null hypothesis stating no significant difference in body weight following the NEX intervention compared to CON in adults with overweight and obesity is rejected.

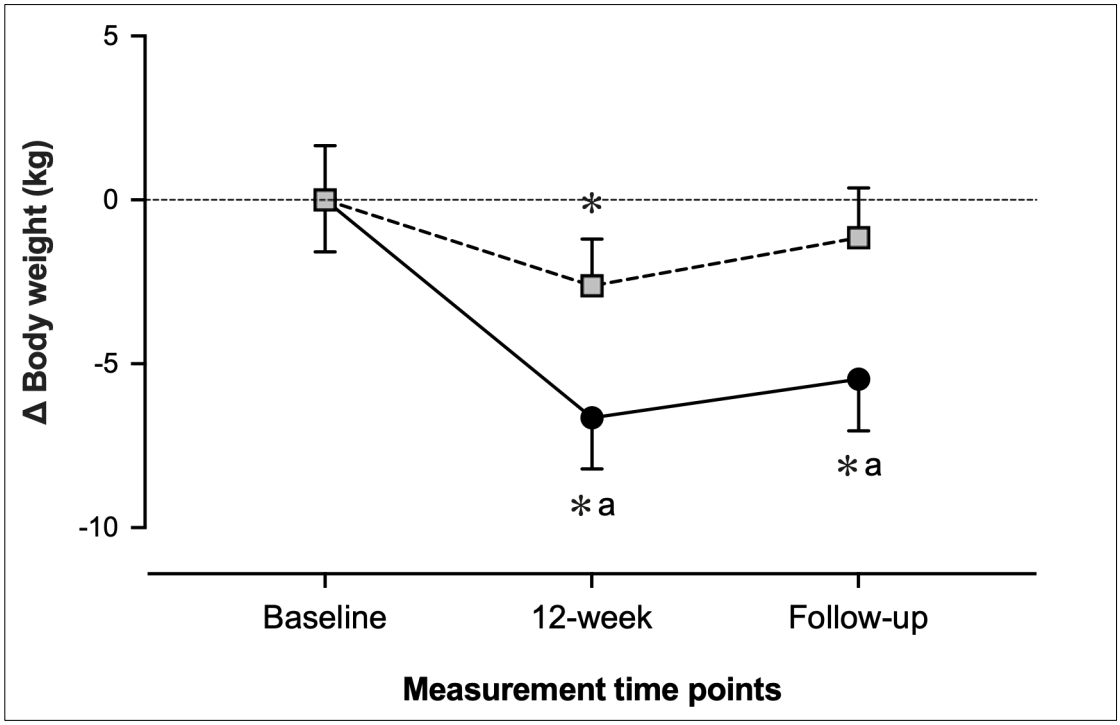


Figure 4.2 Change (Δ) in Body Weight Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$); [a] Difference from CON ($P < 0.05$)

4.3.4 Changes in Waist Circumference

Change in waist circumference relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.3 A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for waist circumference ($P < 0.001$; effect size, 0.13), indicating that the pattern of change across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that waist circumference was significantly reduced in the NEX group from baseline at 12-week and at the 3-month follow-up (both $P < 0.001$), representing reductions of 5.4 cm (5.3%) and 5.0 cm (4.9%), respectively. In contrast, the CON group experienced a reduction in waist circumference of 2.1% at 12-week before returning near baseline values at follow-up.

At baseline, waist circumference was comparable between the NEX (101.12 ± 8.50 cm) and CON (99.68 ± 8.73 cm) groups. At 12-week, the NEX group measured 95.66 ± 7.26 cm compared to 97.54 ± 7.02 cm in the CON group. At follow-up, values were 96.08 ± 8.67 cm and 98.39 ± 9.15 cm for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant time × intervention interaction ($F(2, 316) = 9.101$, $P < 0.001$, $\eta p^2 = 0.13$), a significant main effect of time ($F(2, 316) = 35.828$, $P < 0.001$, $\eta p^2 = 0.183$), and a non-significant main effect of group ($F(1, 158) = 0.612$, $P = 0.435$, $\eta p^2 = 0.004$). Within-group pairwise comparisons for the CON group revealed a significant reduction in waist circumference from baseline to 12-week (−2.14 cm, 2.1%; 95% CI −3.41 to −0.87; $P = 0.001$), but this reduction was not maintained at follow-up (−1.29 cm, 1.3%; 95% CI −2.65 to 0.06; $P = 0.061$).

Overall, waist circumference was significantly reduced by 5.3% at 12-week and 4.9% at the 3-month follow-up in the NEX group relative to baseline. The sustained reduction at follow-up, with only a modest attenuation from 5.4 cm to 5.0 cm loss, indicates that the abdominal adiposity reduction was largely maintained beyond the active intervention period. The minimal reductions in waist circumference observed in the CON group further highlights the divergent trajectories between the two groups. Therefore, the null hypothesis stating no significant difference in waist circumference following the NEX intervention compared to CON in adults with overweight and obesity is rejected.

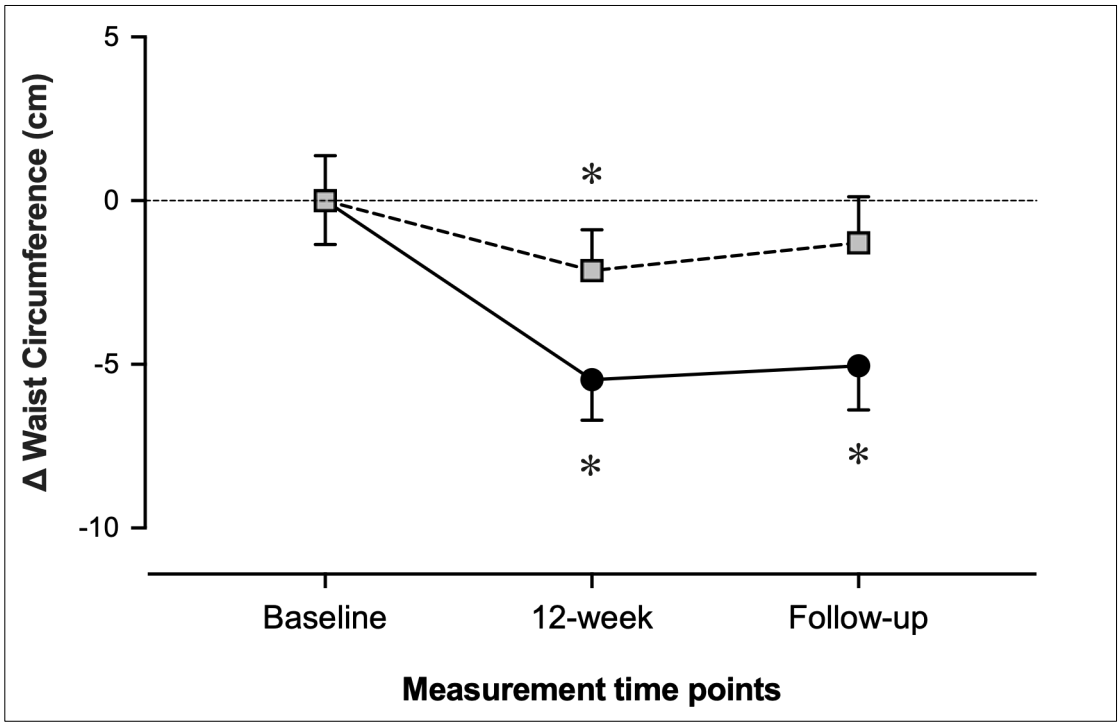


Figure 4.3 Change (Δ) in Waist Circumference Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▨) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$)

4.3.5 Changes in Body Composition

4.3.5.1 Fat Mass

Change in fat mass relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.4. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for fat mass ($P = 0.02$; effect size, 0.10), indicating that the pattern of fat mass change across time differed between the NEX and CON groups.

Follow-up tests revealed that fat mass was significantly reduced in the NEX group from baseline at 12-week and at the 3-month follow-up (both $P < 0.001$), representing decreases of 3.7 kg (10.5%) and 3.1 kg (8.8%), respectively. Between-group differences showed that the NEX group achieved 1.7 kg greater fat loss than the CON group at 12-week ($P = 0.082$) and 1.6 kg greater loss at follow-up ($P = 0.097$). Although these between-group differences did not reach conventional statistical significance at the $P < 0.05$ level, there was a notable trend toward significance,

suggesting a clinically meaningful fat mass reduction in the NEX group. In contrast, the CON group experienced reductions in fat mass of 5.7% at 12-week and 4.0% at follow-up.

At baseline, body fat mass was comparable between the NEX (35.2 ± 6.0 kg) and CON (35.1 ± 5.9 kg) groups. At 12-week, the NEX group measured 31.5 ± 5.5 kg compared to 33.1 ± 6.5 kg in the CON group. At follow-up, values were 32.1 ± 6.3 kg and 33.7 ± 5.5 kg for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant time × intervention interaction ($F(2, 316) = 3.98$, $P = 0.020$, $\eta p^2 = 0.11$), a significant main effect of time ($F(2, 316) = 37.32$, $P < 0.001$, $\eta p^2 = 0.191$), and a non-significant main effect of group ($F(1, 158) = 1.55$, $P = 0.215$, $\eta p^2 = 0.010$). The mean between-group difference (NEX minus CON) at baseline was 0.1 kg (95% CI: −1.8 to 2.0), at 12-week was −1.7 kg (95% CI: −3.5 to 0.2), and at follow-up was −1.6 kg (95% CI: −3.4 to 0.3).

Overall, fat mass was significantly reduced by 10.5% at 12-week and 8.8% at the 3-month follow-up in the NEX group relative to baseline, with a trend towards greater fat loss in the NEX group compared to the CON group. The greater magnitude of fat mass reduction in the NEX group compared to the CON group underscores the differential impact of the NExGEN-ChatGPT intervention on adipose tissue. Therefore, the null hypothesis stating no significant difference in fat mass following the NEX intervention compared to CON is rejected based on the significant time × intervention interaction ($P = 0.02$), with significant within-group reductions observed in the NEX group. However, pairwise between-group comparisons at individual timepoints did not reach conventional significance ($P = 0.082$ at 12-week; $P = 0.097$ at follow-up), indicating a trend toward greater fat mass reduction in the NEX group.

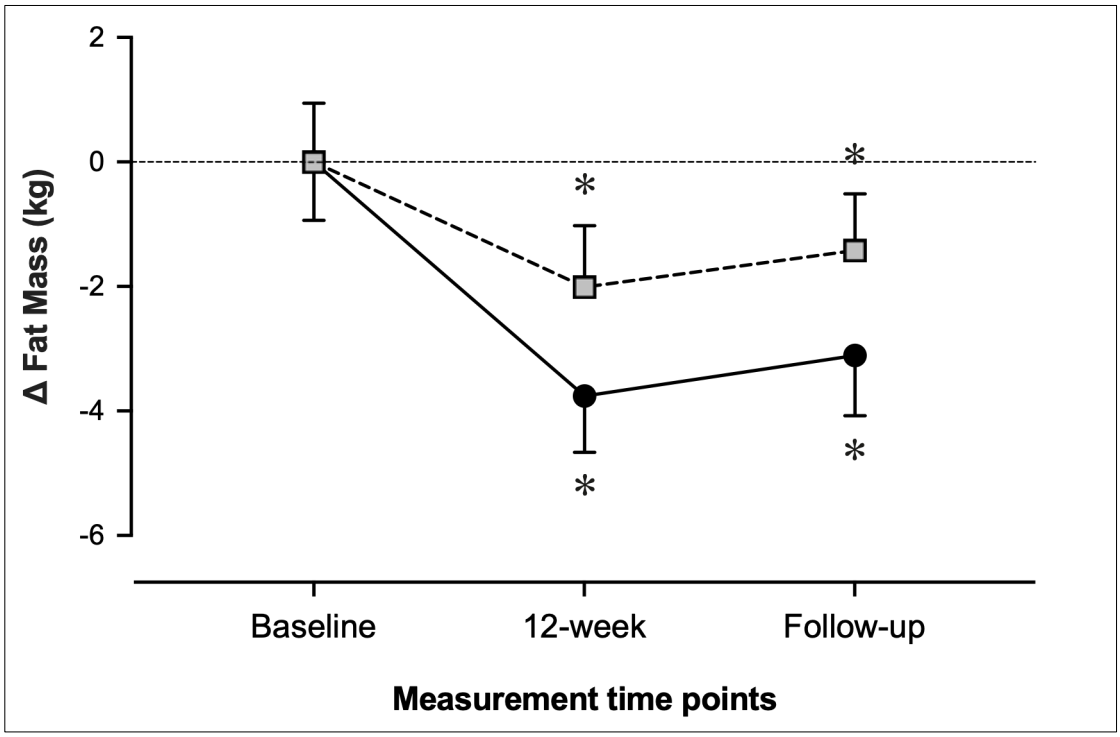


Figure 4.4 Change (Δ) in Fat Mass Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▧) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$)

4.3.5.2 Fat-Free Mass

Change in fat-free mass relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.5. A mixed-design ANOVA with repeated measures revealed no significant main effects or interaction effects for fat-free mass (all $P > 0.05$), indicating that fat-free mass remained stable across all time points in both groups.

At baseline, lean body mass was comparable between the NEX (53.2 ± 6.9 kg) and CON (52.5 ± 6.6 kg) groups. At 12-week, the NEX group measured 50.3 ± 6.4 kg compared to 51.9 ± 5.9 kg in the CON group. At follow-up, values were 50.8 ± 6.4 kg and 52.7 ± 6.8 kg for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant main effect of time ($F(2, 316) = 3.57$, $P = 0.092$, $\eta p^2 = 0.022$), while neither the main effect of group ($F(1, 158) = 1.78$, $P = 0.184$, $\eta p^2 = 0.011$) nor the time × intervention interaction ($F(2, 316) = 2.31$, $P = 0.101$, $\eta p^2 = 0.014$) reached statistical significance. The mean between-group difference (NEX minus CON)

at baseline was 0.7 kg (95% CI: −1.4 to 2.8), at 12-week was −1.6 kg (95% CI: −3.5 to 0.4), and at follow-up was −1.9 kg (95% CI: −4.0 to 0.1).

The preservation of fat-free mass throughout the study is a clinically important finding, as it indicates that the weight loss achieved through the NExGEN-ChatGPT intervention was primarily derived from adipose tissue rather than lean body mass (although the difference is not clinically significant). This pattern of body composition change is consistent with recommended weight loss strategies that combine dietary modification with physical activity to preserve muscle mass. The null hypothesis, stating no significant difference in fat-free mass following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

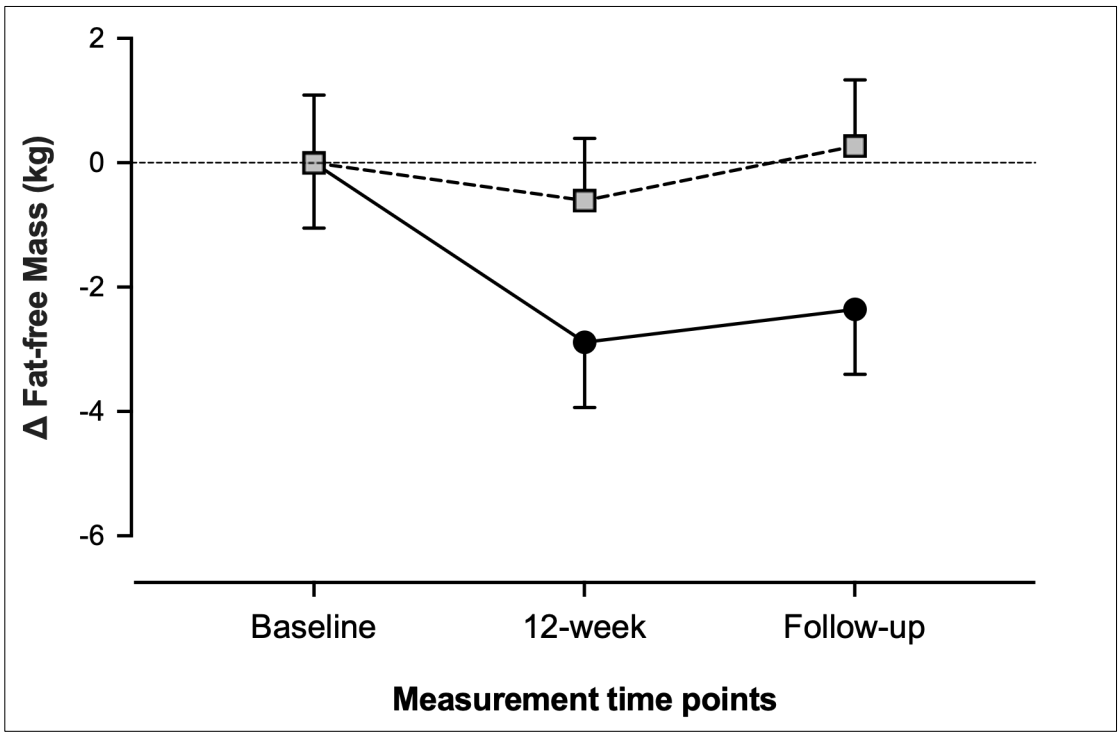


Figure 4.5 Change (Δ) in Fat-Free Mass Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (▧) (Group Means ± SE) (Intention-to-Treat Analysis).

4.3.6 Changes in Dietary Intake

4.3.6.1 Energy Density

Change in energy density relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.6A. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for energy density (P = 0.0014; effect size, 0.041), indicating that the dietary energy density response across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that energy density was significantly lower in the NEX group compared to baseline at 12-week (P < 0.001) and at follow-up (P = 0.0005). Between-group comparisons demonstrated that the NEX group had significantly lower energy density than the CON group at 12-week (95% CI −0.213 to −0.076; P < 0.001) and at follow-up (95% CI −0.203 to −0.008; P = 0.015). No significant changes in energy density were observed in the CON group throughout the study period.

At baseline, energy density was comparable between the NEX (4.84 ± 0.25 kcal/g) and CON (4.84 ± 0.26 kcal/g) groups. At 12-week, the NEX group reported lower energy density (4.64 ± 0.23 kcal/g) compared to the CON group (4.78 ± 0.21 kcal/g), and this difference persisted at follow-up (NEX: 4.71 ± 0.27 kcal/g; CON: 4.84 ± 0.19 kcal/g). The mixed-design ANOVA confirmed a significant time × intervention interaction ($F(2, 316) = 6.781$, $P = 0.001$, $\eta p^2 = 0.041$), along with significant main effects of time ($F(2, 316) = 20.198$, $P < 0.001$, $\eta p^2 = 0.113$) and group ($F(1, 158) = 11.029$, $P = 0.001$, $\eta p^2 = 0.065$).

Overall, the NEX intervention produced a significant and sustained reduction in dietary energy density, with between-group differences maintained at both 12-week and the 3-month follow-up. The reduction in energy density is a recognised dietary strategy for weight management, as lower energy density diets promote greater satiety per calorie consumed. Therefore, the null hypothesis stating no significant difference in dietary energy density following the NEX intervention compared to CON is rejected.

4.3.6.2 Energy Intake

Group mean changes in energy intake following the NEX and CON interventions are shown in Figure 4.6B. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for energy intake ($F(2, 316) = 5.612$, $P = 0.004$, $\eta p^2 = 0.034$), indicating that the dietary energy intake response across time differed significantly between the NEX and CON groups. A significant main effect of time was also observed ($F(2, 316) = 24.693$, $P < 0.001$, $\eta p^2 = 0.135$), with no significant main effect of group ($F(1, 158) = 0.704$, $P = 0.403$, $\eta p^2 = 0.004$).

Follow-up tests revealed that energy intake was significantly lower in the NEX group compared to baseline at both 12-week ($P < 0.001$) and follow-up ($P < 0.001$). In the CON group, energy intake was significantly reduced at 12-week ($P < 0.001$) but had returned to baseline levels at follow-up ($P = 0.429$). Between-group comparisons of change scores demonstrated that the NEX group achieved a significantly greater reduction in energy intake than the CON group at 12-week (−101.0 kcal/day, 95% CI −187.6 to −14.4; $P = 0.023$) and at follow-up (−133.6 kcal/day, 95% CI −221.9 to −45.3; $P = 0.003$).

At baseline, energy intake was comparable between the NEX (1833.19 ± 659.67 kcal/day) and CON (1830.15 ± 656.67 kcal/day) groups. At 12-week, the NEX group reported 1638.80 ± 571.16 kcal/day compared to 1736.77 ± 541.00 kcal/day in the CON group. At follow-up, values were 1676.43 ± 567.43 kcal/day and 1806.97 ± 504.86 kcal/day for the NEX and CON groups, respectively.

Overall, the NEX intervention produced a significant and sustained reduction in dietary energy intake, with the magnitude of between-group difference increasing from 12-week to follow-up. Notably, the CON group's initial reduction was not maintained at the 3-month follow-up, whereas the NEX group sustained its lower intake. Therefore, the null hypothesis stating no significant difference in energy intake following the NEX intervention compared to CON in adults with overweight and obesity is rejected.

4.3.6.3 Carbohydrate Intake

Change in carbohydrate intake relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure

4.6C. A mixed-design ANOVA with repeated measures revealed a significant time effect for carbohydrate intake (P = 0.0117), with reductions observed in both groups. However, no significant time × intervention interaction effect was detected. The mean difference between groups was −4.863 g/day (95% CI −24.89 to 15.17; P = 0.6322), indicating no differential effect of the NEX intervention on carbohydrate intake.

At baseline, carbohydrate intake was comparable between the NEX (203.26 ± 77.29 g/day) and CON (202.73 ± 80.90 g/day) groups. At 12-week, the NEX group reported 187.97 ± 65.52 g/day compared to 194.42 ± 66.86 g/day in the CON group. At follow-up, values were 191.56 ± 67.40 g/day and 200.24 ± 61.78 g/day for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant main effect of time ($F(2, 316) = 4.586$, $P = 0.011$, $\eta p^2 = 0.028$), a non-significant main effect of group ($F(1, 158) = 0.230$, $P = 0.632$, $\eta p^2 = 0.001$), and a non-significant time × intervention interaction ($F(2, 316) = 0.744$, $P = 0.476$, $\eta p^2 = 0.005$).

The absence of a significant time × intervention interaction for carbohydrate intake indicates that the NExGEN-ChatGPT intervention did not selectively modify carbohydrate consumption relative to the control condition. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on dietary outcomes, this finding suggests that the dietary modifications achieved by the NEX intervention were primarily driven by reductions in energy density and fat intake rather than carbohydrate restriction. Therefore, the null hypothesis stating no significant difference in carbohydrate intake following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

4.3.6.4 Fat Intake

Change in fat intake relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.6D. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for fat intake ($P < 0.001$; effect size, 0.066), indicating that the pattern of fat intake change across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that fat intake was significantly reduced in the NEX group from baseline at 12-week (95% CI 9.53 to 20.83; $P < 0.001$) and at follow-up (95% CI 5.57 to 18.07; $P < 0.001$). Between-group comparisons demonstrated that the NEX group had significantly lower fat intake than the CON group at 12-week (95% CI

−16.85 to −0.152; P = 0.046) and at follow-up (95% CI −19.30 to −2.10; P = 0.015). No significant changes in fat intake were observed in the CON group during the intervention period.

At baseline, fat intake was comparable between the NEX (80.88 ± 33.05 g/day) and CON (80.51 ± 31.90 g/day) groups. At 12-week, the NEX group reported 65.80 ± 28.62 g/day compared to 74.31 ± 24.73 g/day in the CON group. At follow-up, values were 69.05 ± 28.13 g/day and 79.76 ± 26.94 g/day for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant time × intervention interaction ($F(2, 316) = 11.120$, $P < 0.001$, $\eta p^2 = 0.066$), a significant main effect of time ($F(2, 316) = 37.460$, $P < 0.001$, $\eta p^2 = 0.192$), and a non-significant main effect of group ($F(1, 158) = 2.072$, $P = 0.152$, $\eta p^2 = 0.013$).

Overall, the NEX intervention produced a significant and sustained reduction in daily fat intake, with between-group differences maintained at both assessment time points. The wider confidence interval at follow-up (95% CI −19.30 to −2.10) compared to 12-week (95% CI −16.85 to −0.152) suggests increasing variability in dietary fat reduction at longer follow-up, though the effect remained statistically significant. Therefore, the null hypothesis stating no significant difference in fat intake following the NEX intervention compared to CON is rejected.

4.3.6.5 Protein Intake

Change in protein intake relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.6E. A mixed-design ANOVA with repeated measures revealed no significant main effects or interaction effects for protein intake (all $P > 0.05$), indicating that protein intake remained stable across all time points in both groups.

At baseline, protein intake was comparable between the NEX (73.07 ± 26.54 g/day) and CON (73.66 ± 24.69 g/day) groups. At 12-week, values were 73.70 ± 23.78 g/day and 72.62 ± 21.97 g/day for the NEX and CON groups, respectively. At follow-up, values were 72.18 ± 27.56 g/day and 72.07 ± 17.82 g/day, respectively. The mixed-design ANOVA confirmed non-significant effects for time ($F(2, 316) = 0.451$, $P = 0.637$, $\eta p^2 = 0.003$), group ($F(1, 158) = 0.003$, $P = 0.954$, $\eta p^2 < 0.001$), and time × intervention interaction ($F(2, 316) = 0.179$, $P = 0.836$, $\eta p^2 = 0.001$).

The stability of protein intake across both groups is a notable finding, as it suggests that the dietary modifications facilitated by the NExGEN-ChatGPT intervention were selective, targeting energy density and fat intake without compromising protein consumption. This selectivity is clinically important for maintaining lean body mass during weight loss, consistent with the preservation of fat-free mass reported earlier.

4.3.6.6 Fibre Intake

Change in fibre intake relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.6F. A mixed-design ANOVA with repeated measures revealed a significant time effect for fibre intake ($P < 0.001$; effect size, 0.128), with both groups showing improvements from baseline. The NEX group demonstrated significant increases from baseline at 12-week (95% CI −5.70 to −4.46; $P < 0.001$) and at follow-up (95% CI −4.27 to −6.06; $P < 0.001$). The CON group also showed significant improvements in fibre intake. However, the time × intervention interaction was not statistically significant ($P = 0.076$), though a trend approaching significance was observed.

At baseline, fibre intake was comparable between the NEX (20.84 ± 8.98 g/day) and CON (21.08 ± 8.54 g/day) groups. At 12-week, the NEX group reported 24.39 ± 9.97 g/day compared to 23.01 ± 10.16 g/day in the CON group. At follow-up, values were 22.57 ± 9.15 g/day and 21.19 ± 9.48 g/day for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant main effect of time ($F(2, 316) = 23.395$, $P < 0.001$, $\eta p^2 = 0.129$), a non-significant main effect of group ($F(1, 158) = 0.354$, $P = 0.553$, $\eta p^2 = 0.002$), and a non-significant time × intervention interaction ($F(2, 316) = 2.610$, $P = 0.075$, $\eta p^2 = 0.016$). Within-group pairwise comparisons for the CON group revealed a significant increase in fibre intake from baseline to 12-week (mean change 1.93 g/day, 95% CI 0.27 to 3.58; $P = 0.023$), but this improvement was not maintained at follow-up (mean change 0.11 g/day, 95% CI −1.50 to 1.71; $P = 0.895$).

The improvements in fibre intake observed in both groups, combined with the trend toward a significant interaction effect ($P = 0.076$), suggest that whilst the NEX intervention may have produced a slightly greater improvement in fibre intake, both conditions promoted increased fibre consumption. In relation to the study objective of

evaluating the effectiveness of NExGEN-ChatGPT on dietary quality outcomes, this finding indicates that the intervention did not produce a significant differential effect on fibre intake relative to the control condition. Therefore, the null hypothesis stating no significant difference in fibre intake following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

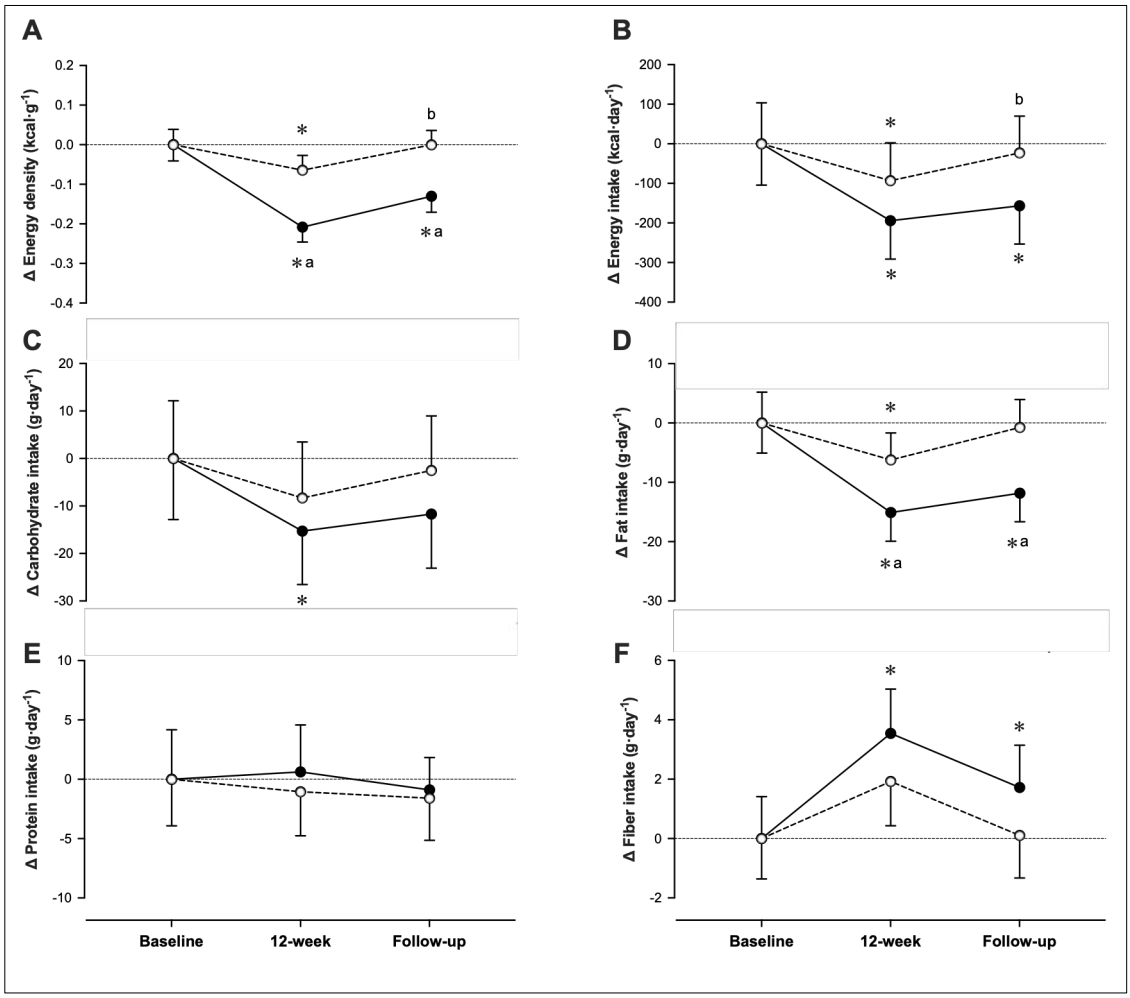


Figure 4.6 Change (Δ) in Body Weight Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (○) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$); [a] Difference from CON ($P < 0.05$)

4.3.7 Changes in Physical Activity

The physical activity changes to the NEX intervention across baseline, 12-week, and 3-month follow-up assessments are summarised in Figure 4.7. The following subsections present the findings for each of the different methods of measuring physical activity level individually.

4.3.7.1 Accelerometer-Measured Physical Activity

Change in accelerometer-measured physical activity relative to baseline during the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.7A. A mixed-design ANOVA with repeated measures revealed a significant

time × intervention interaction effect for accelerometer-measured physical activity (P = 0.0017; effect size, 0.040), indicating that the physical activity response across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that accelerometer-measured physical activity was significantly increased in the NEX group from baseline at 12-week (95% CI −367.4 to −214.9; P < 0.001) and at follow-up (95% CI −312.0 to −162.8; P < 0.001). Between-group comparisons demonstrated that the NEX group achieved significantly higher physical activity than the CON group at 12-week (95% CI 20.6 to 277.9; P = 0.023) and at follow-up (95% CI 25.63 to 360.7; P = 0.0241). The CON group showed modest increases in physical activity over time, but these changes were significantly smaller than those observed in the NEX group.

At baseline, accelerometer-measured physical activity was comparable between the NEX (1716.32 ± 454.94 MET-min/week) and CON (1707.82 ± 482.97 MET-min/week) groups. At 12-week, values were 2007.49 ± 340.68 MET-min/week and 1858.23 ± 470.50 MET-min/week for the NEX and CON groups, respectively. At follow-up, values were 1953.69 ± 472.71 MET-min/week and 1760.52 ± 591.66 MET-min/week for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed significant effects for time ($F(2, 316) = 35.546$, $P < 0.001$, $\eta p^2 = 0.184$), and time × intervention interaction ($F(2, 316) = 6.512$, $P = 0.0017$, $\eta p^2 = 0.040$), but not for group ($F(1, 158) = 2.934$, $P = 0.089$, $\eta p^2 = 0.018$). Within the CON group, physical activity increased significantly from baseline to 12-week (95% CI 79.0 to 221.8; P < 0.001), but this change was not maintained at follow-up (95% CI −38.6 to 144.0; P = 0.254).

Overall, accelerometer-measured physical activity was significantly enhanced by the NEX intervention, with sustained improvements maintained at the 3-month follow-up. The objective nature of accelerometer measurement strengthens the validity of these findings by eliminating self-report bias. Therefore, the null hypothesis stating no significant difference in objectively measured physical activity following the NEX intervention compared to CON is rejected.

4.3.7.2 Self-Reported Physical Activity (IPAQ-LV)

Change in self-reported physical activity measured using the International Physical Activity Questionnaire – Long Version (IPAQ-LV) relative to baseline during

the 12-week intervention period and 3-month follow-up with NEX and CON are shown in Figure 4.7B. A mixed-design ANOVA with repeated measures revealed a significant time × intervention interaction effect for self-reported physical activity ($P = 0.0021$; effect size, 0.038), indicating that the self-reported physical activity response across time differed significantly between the NEX and CON groups.

Follow-up tests revealed that self-reported physical activity was significantly increased in the NEX group from baseline at 12-week (95% CI −276.7 to −118.4; $P < 0.001$) and at follow-up (95% CI −264.1 to −116.1; $P < 0.001$). Both accelerometer and self-reported measures demonstrated sustained improvements at follow-up, indicating the durability of behavioural changes induced by the NEX intervention.

At baseline, self-reported physical activity was comparable between the NEX (939.16 ± 541.17 MET-min/week) and CON (944.54 ± 473.03 MET-min/week) groups. At 12-week, values were 1136.65 ± 247.22 MET-min/week and 1044.41 ± 451.33 MET-min/week for the NEX and CON groups, respectively. At follow-up, values were 1129.27 ± 271.55 MET-min/week and 1042.66 ± 314.76 MET-min/week for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed significant effects for time ($F(2, 316) = 60.199$, $P < 0.001$, $\eta p^2 = 0.276$) and time × intervention interaction ($F(2, 316) = 6.273$, $P = 0.0021$, $\eta p^2 = 0.038$), but not for group ($F(1, 158) = 0.916$, $P = 0.340$, $\eta p^2 = 0.006$). Between-group comparisons revealed no significant differences at 12-week (mean difference = 92.2 MET-min/week, 95% CI −22.1 to 206.6; $P = 0.110$) or follow-up (mean difference = 86.6 MET-min/week, 95% CI −5.3 to 178.5; $P = 0.064$), though the trend favoured the NEX group at both time points.

The concordance between accelerometer-measured and self-reported physical activity improvements provides triangulated evidence supporting the effectiveness of the NEX intervention in promoting sustained physical activity behaviour change. Therefore, the null hypothesis stating no significant difference in self-reported physical activity following the NEX intervention compared to CON is rejected.

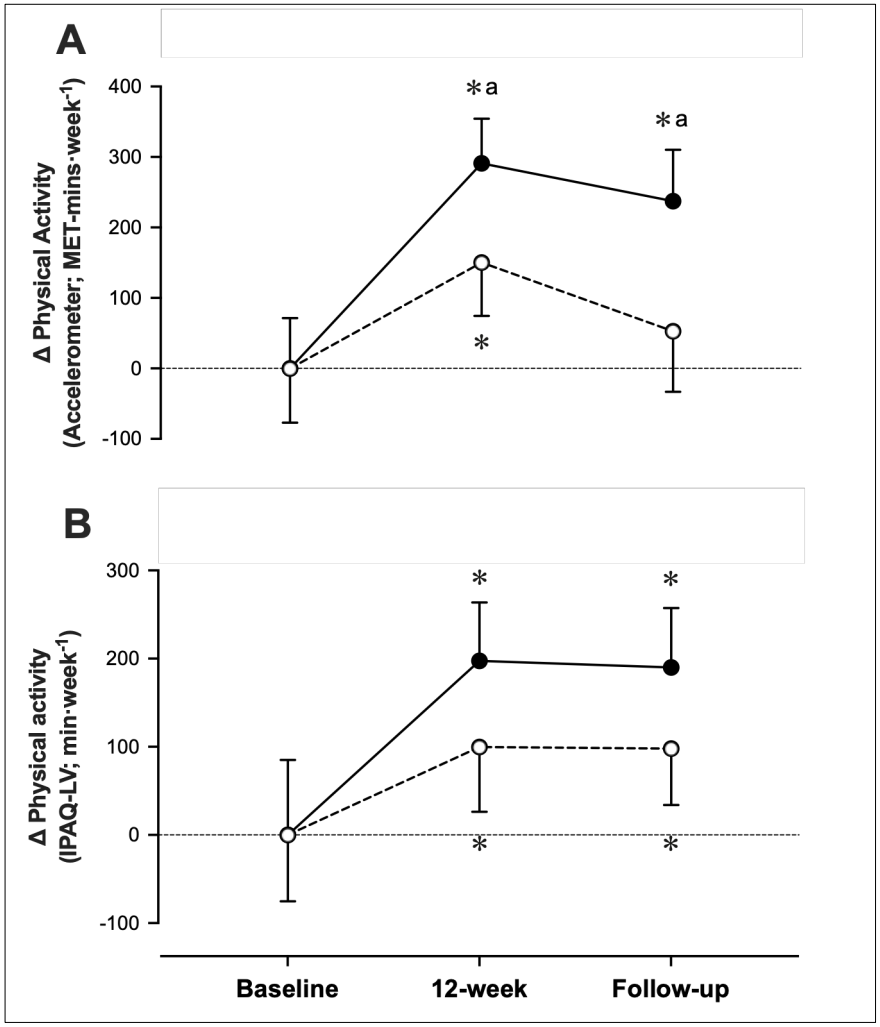


Figure 4.7 Change (Δ) in Physical Activity Level Relative to Baseline During the 12-Week Intervention Period and 3-Month Follow-Up with NEX (●) and CON (○) (Group Means ± SE) (Intention-to-Treat Analysis). *Difference from Baseline ($P < 0.05$); [a] Difference from CON ($P < 0.05$)

4.3.8 Changes in Metabolic Parameters

The metabolic parameters measured at baseline and across 12 weeks of intervention and 3-month follow-up in NEX and CON groups are summarised in Table 4.18. The following subsections present the findings for each metabolic parameter individually.

4.3.8.1 Glycated Haemoglobin (HbA1c)

Group mean changes in HbA1c levels following the NEX and CON interventions are presented in Table 4.18. A mixed-design ANOVA with repeated measures revealed no significant intervention effect for HbA1c ($P = 0.82$). Follow-up tests revealed no significant differences between the NEX and CON groups at any time point: baseline (95% CI −0.12 to 0.127; $P = 0.949$), 12-week (95% CI −0.153 to 0.089; $P = 0.607$), and follow-up (95% CI −0.115 to 0.099; $P = 0.879$). Similarly, neither group showed significant within-group changes from baseline values across the intervention and follow-up periods.

At baseline, HbA1c was comparable between the NEX (5.36 ± 0.35%) and CON (5.35 ± 0.43%) groups. At 12-week, values were 5.30 ± 0.37% and 5.34 ± 0.41% for the NEX and CON groups, respectively. At follow-up, values were 5.33 ± 0.31% and 5.34 ± 0.37% for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed non-significant effects for time ($F(2, 316) = 1.279$, $P = 0.280$, $\eta p^2 = 0.008$), group ($F(1, 158) = 0.051$, $P = 0.822$, $\eta p^2 < 0.001$), and time × intervention interaction ($F(2, 316) = 0.317$, $P = 0.728$, $\eta p^2 = 0.002$).

The absence of significant between-group differences in HbA1c across all time points indicates that the NExGEN-ChatGPT intervention did not produce a differential effect on glycaemic status relative to the control condition. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on metabolic health outcomes, this finding indicates that glycaemic control was not differentially altered by the intervention within this study period. Therefore, the null hypothesis stating no significant difference in HbA1c levels following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

4.3.8.2 Total Cholesterol

Group mean changes in total cholesterol following the NEX and CON interventions are presented in Table 4.18. A mixed-design ANOVA with repeated measures revealed a significant time effect for total cholesterol ($P < 0.0001$; Geisser-Greenhouse's epsilon = 0.99), with both groups showing reductions over time.

In the NEX group, total cholesterol was significantly reduced from baseline at 12-week (95% CI −21.48 to 7.419; $P < 0.0001$) and at follow-up (95% CI −20.44 to 6.813; $P < 0.0001$). The CON group showed a significant reduction only at follow-up (95% CI −17.94 to 0.380; $P = 0.039$). However, no significant between-group differences were observed at any time point.

At baseline, total cholesterol was comparable between the NEX (212.60 ± 38.02 mg/dL) and CON (210.28 ± 42.44 mg/dL) groups. At 12-week, values were 198.16 ± 40.74 mg/dL and 201.43 ± 42.00 mg/dL for the NEX and CON groups, respectively. At follow-up, values were 198.98 ± 36.37 mg/dL and 201.12 ± 43.29 mg/dL for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant effect for time ($F(2, 316) = 15.291$, $P < 0.0001$, $\eta p^2 = 0.088$), but non-significant effects for group ($F(1, 158) = 0.032$, $P = 0.859$, $\eta p^2 < 0.001$) and time × intervention interaction ($F(2, 316) = 0.751$, $P = 0.473$, $\eta p^2 = 0.005$).

Whilst significant within-group reductions were observed, particularly in the NEX group where reductions were evident at both 12-week and follow-up compared to only at follow-up in the CON group, the absence of significant between-group differences indicates that the NEX intervention did not produce a differential effect on total cholesterol relative to the CON condition. Therefore, the null hypothesis stating no significant between-group difference in total cholesterol following the NEX intervention compared to CON is accepted.

4.3.8.3 Low-Density Lipoprotein (LDL) Cholesterol

Group mean changes in LDL cholesterol following the NEX and CON interventions are presented in Table 4.18. A mixed-design ANOVA with repeated measures revealed a significant time effect for LDL cholesterol ($P = 0.001$), exhibiting patterns similar to total cholesterol.

In the NEX group, LDL cholesterol was significantly reduced from baseline at 12-week (95% CI −15.85 to 2.67; P = 0.0035) and at follow-up (95% CI −14.52 to 0.856; P = 0.023). The CON group changes remained non-significant at all time points. However, no significant between-group differences were observed.

At baseline, LDL cholesterol was comparable between the NEX (133.22 ± 34.58 mg/dL) and CON (137.72 ± 37.27 mg/dL) groups. At 12-week, values were 123.95 ± 32.48 mg/dL and 132.14 ± 37.09 mg/dL for the NEX and CON groups, respectively. At follow-up, values were 125.52 ± 32.09 mg/dL and 132.44 ± 39.92 mg/dL for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant effect for time ($F(2, 316) = 7.146$, $P = 0.001$, $\eta p^2 = 0.043$), but non-significant effects for group ($F(1, 158) = 1.667$, $P = 0.199$, $\eta p^2 = 0.010$) and time × intervention interaction ($F(2, 316) = 0.381$, $P = 0.683$, $\eta p^2 = 0.002$).

Although the NEX group demonstrated significant within-group reductions in LDL cholesterol at both 12-week and follow-up, no significant between-group difference was detected at any time point. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on metabolic health outcomes, this finding indicates that the intervention did not produce a superior LDL-lowering effect compared to the control condition. Therefore, the null hypothesis stating no significant between-group difference in LDL cholesterol following the NEX intervention compared to CON is accepted.

4.3.8.4 High-Density Lipoprotein (HDL) Cholesterol

Group mean changes in HDL cholesterol following the NEX and CON interventions are presented in Table 4.18. A mixed-design ANOVA with repeated measures revealed a significant between-group effect for HDL cholesterol (P = 0.014), with the NEX group displaying consistently higher HDL values throughout the study period.

At baseline, HDL cholesterol values were 56.07 ± 25.70 mg/dL and 48.82 ± 30.83 mg/dL for the NEX and CON groups, respectively. At 12-week, values were 52.88 ± 24.45 mg/dL and 46.41 ± 28.72 mg/dL for the NEX and CON groups, respectively. At follow-up, values were 51.51 ± 25.66 mg/dL and 45.87 ± 33.57 mg/dL for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant group effect ($F(1, 158) = 6.218$, $P = 0.014$, $\eta p^2 = 0.038$), but non-significant

effects for time ($F(2, 316) = 0.763$, $P = 0.467$, $\eta p^2 = 0.005$) and time × intervention interaction ($F(2, 316) = 0.032$, $P = 0.968$, $\eta p^2 < 0.001$). Neither group showed significant within-group changes from baseline values across the intervention and follow-up periods. However, no significant between-group differences were observed at any individual time point.

Therefore, the null hypothesis stating no significant between-group difference in HDL cholesterol following the NEX intervention compared to CON is accepted.

4.3.8.5 Triglycerides

Group mean changes in triglyceride levels following the NEX and CON interventions are presented in Table 4.18. A mixed-design ANOVA with repeated measures revealed no significant intervention effect for triglycerides ($P = 0.548$), with minimal between-group differences observed throughout the study period.

At baseline, triglyceride levels were comparable between the NEX (116.63 ± 56.26 mg/dL) and CON (118.71 ± 61.89 mg/dL) groups. At 12-week, values were 106.64 ± 48.04 mg/dL and 114.37 ± 48.58 mg/dL for the NEX and CON groups, respectively. At follow-up, values were 109.71 ± 56.61 mg/dL and 114.04 ± 50.84 mg/dL for the NEX and CON groups, respectively. The mixed-design ANOVA confirmed a significant effect for time ($F(2, 316) = 3.439$, $P = 0.033$, $\eta p^2 = 0.021$), but non-significant effects for group ($F(1, 158) = 0.362$, $P = 0.548$, $\eta p^2 = 0.002$) and time × intervention interaction ($F(2, 316) = 0.476$, $P = 0.622$, $\eta p^2 = 0.003$).

The finding indicates that the NExGEN-ChatGPT intervention did not produce a differential effect on triglyceride levels relative to the control condition over the 12-week period. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on metabolic health outcomes, this suggests that the significant weight loss achieved in the NEX group was not accompanied by a between-group differential effect on fasting triglycerides within this timeframe. Therefore, the null hypothesis stating no significant difference in triglyceride levels following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

Table 4.18
Metabolic Parameters Measured at Baseline and Across 12 Weeks of NEX Intervention and 3-Month Follow-Up in Intervention (NEX) and Control (CON) Groups (Intention-to-Treat Analysis)

Variable and group	Baseline (mean ± SD)	12 Week (mean ± SD)	Follow up (mean ± SD)
HbA1c (%)			
NEX	5.36 ± 0.35	5.30 ± 0.37	5.33 ± 0.31
CON	5.35 ± 0.43	5.34 ± 0.41	5.34 ± 0.37
Total Cholesterol (mg/dL)			
NEX	212.6 ± 38.0	198.1 ± 40.7*	199.0 ± 36.4*
CON	210.3 ± 42.4	201.4 ± 42.0	201.1 ± 43.3*
LDL Cholesterol (mg/dL)			
NEX	133.2 ± 34.6	123.9 ± 32.5*	125.5 ± 32.1*
CON	137.7 ± 37.3	132.1 ± 37.1	132.4 ± 39.9
HDL Cholesterol (mg/dL)			
NEX	56.1 ± 25.7	52.9 ± 24.5	51.5 ± 25.7
CON	48.8 ± 30.8	46.4 ± 28.7	45.9 ± 33.6
Triglycerides (mg/dL)			
NEX	116.6 ± 56.3	106.6 ± 48.0*	109.7 ± 56.6
CON	118.7 ± 61.9	114.4 ± 48.6	114.0 ± 50.8

Values are means ± SD. *Different from baseline, $P < 0.05$

4.3.9 Changes in Blood Pressure

The blood pressure parameters measured at baseline and across 12 weeks of intervention and 3-month follow-up in NEX and CON groups are summarised in Table 4.19. The following subsections present the findings for each blood pressure parameter individually.

4.3.9.1 Systolic Blood Pressure

Group mean changes in systolic blood pressure following the NEX and CON interventions are presented in Table 4.19. A mixed-design ANOVA with repeated measures revealed no significant intervention effect for systolic blood pressure (P =

0.402). Between-group differences remained non-significant at all time points: baseline (95% CI −4.317 to 3.19; P = 0.767), 12-week (95% CI −6.24 to 1.20; P = 0.183), and follow-up (95% CI −5.27 to 2.60; P = 0.503). Within-group analyses revealed no significant changes from baseline in either group across the study duration.

At baseline, systolic blood pressure was comparable between the NEX (127.95 ± 11.73 mmHg) and CON (128.49 ± 12.29 mmHg) groups. At 12-week, values remained similar for the NEX (126.26 ± 11.48 mmHg) and CON (128.72 ± 12.27 mmHg) groups, and this pattern continued at follow-up (NEX: 128.41 ± 11.64 mmHg; CON: 129.72 ± 13.47 mmHg). The mixed-design ANOVA revealed a non-significant time × intervention interaction ($F(2, 316) = 0.975$, $P = 0.378$, $\eta p^2 = 0.006$), with no significant main effects of time ($F(2, 316) = 2.618$, $P = 0.075$, $\eta p^2 = 0.016$) or group ($F(1, 158) = 0.678$, $P = 0.412$, $\eta p^2 = 0.004$).

These findings indicate that systolic blood pressure remained stable and comparable between groups across all time points. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on blood pressure outcomes, the intervention did not produce a differential effect on systolic blood pressure relative to the control condition over the 12-week RCT period. Therefore, the null hypothesis stating no significant difference in systolic blood pressure following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

4.3.9.2 Diastolic Blood Pressure

Group mean changes in diastolic blood pressure following the NEX and CON interventions are presented in Table 4.19. A mixed-design ANOVA with repeated measures revealed no significant intervention effect for diastolic blood pressure (P = 0.839), with consistently non-significant between-group comparisons and minimal within-group variations observed throughout the study period.

At baseline, diastolic blood pressure was comparable between the NEX (88.37 ± 9.12 mmHg) and CON (87.90 ± 8.99 mmHg) groups. At 12-week, values were similar for the NEX (87.22 ± 8.41 mmHg) and CON (87.18 ± 8.72 mmHg) groups, and this pattern continued at follow-up (NEX: 86.94 ± 8.51 mmHg; CON: 88.18 ± 8.53 mmHg). The mixed-design ANOVA revealed a non-significant time × intervention interaction ($F(2, 316) = 1.870$, $P = 0.156$, $\eta p^2 = 0.012$), with no significant main effects of time

(F(2, 316) = 2.118, P = 0.122, ηp^2 = 0.013) or group (F(1, 158) = 0.036, P = 0.850, ηp^2 = 0.000).

These findings indicate that diastolic blood pressure remained stable and comparable between groups across all time points. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on blood pressure outcomes, the intervention did not produce a differential effect on diastolic blood pressure relative to the control condition over the 12-week RCT period. Therefore, the null hypothesis stating no significant difference in diastolic blood pressure following the NEX intervention compared to CON in adults with overweight and obesity is accepted.

4.3.9.3 Mean Arterial Pressure

Group mean changes in mean arterial pressure (MAP) following the NEX and CON interventions are presented in Table 4.19. A mixed-design ANOVA with repeated measures revealed a significant time effect for MAP (P = 0.046). Follow-up tests indicated that this time effect was primarily driven by a small but significant reduction in the NEX group from baseline to 12-week (95% CI −2.36 to −0.34, P = 0.009). However, this change was not sustained at follow-up, and no significant between-group differences emerged at any assessment point.

At baseline, MAP was comparable between the NEX (101.54 ± 9.42 mmHg) and CON (101.42 ± 9.63 mmHg) groups. At 12-week, the NEX group showed lower MAP (100.20 ± 8.76 mmHg) compared to the CON group (101.07 ± 9.26 mmHg), and this pattern persisted at follow-up (NEX: 100.79 ± 8.87 mmHg; CON: 102.03 ± 9.53 mmHg). The mixed-design ANOVA confirmed a significant main effect of time (F(2, 316) = 3.119, P = 0.046, ηp^2 = 0.019), but no significant time × intervention interaction (F(2, 316) = 1.727, P = 0.180, ηp^2 = 0.011) or group effect (F(1, 158) = 0.225, P = 0.636, ηp^2 = 0.001). Between-group differences remained non-significant at all time points: baseline (95% CI −2.85 to 3.10, P = 0.934), 12-week (95% CI −3.68 to 1.95, P = 0.543), and follow-up (95% CI −4.11 to 1.63, P = 0.395). Within-group analyses revealed a significant reduction in the NEX group from baseline to 12-week (95% CI −2.36 to −0.34, P = 0.009), which was not sustained at follow-up (95% CI −1.84 to 0.33, P = 0.171). No significant within-group changes were observed in the CON group at 12-week (95% CI −1.32 to 0.61, P = 0.468) or follow-up (95% CI −0.53 to 1.75, P = 0.288).

The transient within-group reduction in MAP observed in the NEX group at 12-week was not sustained at follow-up, and no significant between-group differences emerged at any assessment point. In relation to the study objective of evaluating the effectiveness of NExGEN-ChatGPT on cardiovascular health outcomes, these findings indicate that the intervention did not produce a differential effect on mean arterial pressure relative to the control condition. Therefore, the null hypothesis stating no significant between-group difference in mean arterial pressure following the NEX intervention compared to CON is accepted.

Table 4.19
Blood Pressure Parameters Measured at Baseline and Across 12 Weeks of NEX Intervention and 3-Month Follow-Up in Intervention (NEX) and Control (CON) Groups (Intention-to-Treat Analysis)

Variable and group	Baseline (mean ± SD)	12 Week (mean ± SD)	Follow up (mean ± SD)
Systolic blood pressure (mmHg)			
NEX	127.9 ± 11.7	126.2 ± 11.5	128.4 ± 11.6
CON	128.5 ± 12.3	128.7 ± 12.3	129.7 ± 13.5
Diastolic blood pressure (mmHg)			
NEX	88.4 ± 9.2	87.2 ± 8.4	87.0 ± 8.5
CON	87.9 ± 9.0	87.2 ± 8.8	88.2 ± 8.5
Mean Arterial pressure (mmHg)			
NEX	101.5 ± 9.4	100.2 ± 8.8*	100.8 ± 8.9
CON	101.4 ± 9.6	101.1 ± 9.3	102.0 ± 9.5

Values are means ± SD. *Different from baseline, $P < 0.05$

4.3.10 Summary of Study Randomised Controlled Trial Findings

4.3.10.1 Primary Outcome (Body Weight)

The NEX intervention produced a statistically significant reduction in the primary outcome of body weight ($P < 0.001$). The NEX group lost 7.5% of baseline body weight at 12-week and maintained a 6.2% reduction at the 3-month follow-up, with the NEX group achieving 3.2 kg and 3.5 kg greater weight loss than the CON group at 12-week and follow-up, respectively. The sustained between-group difference at follow-up demonstrates the durability of the behavioural.

4.3.10.2 Secondary Outcomes with Significant Between-Group Differences

Significant time × intervention interaction effects favouring the NEX group were observed for waist circumference ($P < 0.001$), fat mass ($P = 0.02$), dietary energy density ($P = 0.0014$), energy intake ($P = 0.004$), fat intake ($P < 0.001$), accelerometer-

measured physical activity (P = 0.0017), and self-reported physical activity (P = 0.0021). All significant time × intervention interactions demonstrated sustained effects at the 3-month follow-up, indicating durability of the behavioural and dietary changes induced by the NExGEN-ChatGPT intervention.

4.3.10.3 Outcomes with No Significant Between-Group Differences

No significant between-group differences were observed for fat-free mass, carbohydrate intake, protein intake, fibre intake, HbA1c, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, systolic blood pressure, diastolic blood pressure, or mean arterial pressure. These outcomes were not differentially altered by the NEX intervention relative to the control condition within the 12-week intervention period and 3-month follow-up.

4.3.10.4 Clinical Interpretation of Body Composition and Dietary Changes

The preservation of fat-free mass alongside significant fat mass reduction indicates that the weight loss achieved in the NEX group was qualitatively favourable, being derived predominantly from adipose tissue. The selective dietary changes — with significant reductions in energy density, energy intake, and fat intake but not carbohydrate or protein intake, suggest that the NExGEN-ChatGPT intervention facilitated both quantitative and qualitative dietary modifications consistent with recommended weight loss strategies.

Taken together, the Randomised Controlled Trial findings provide evidence that the NExGEN-ChatGPT personalised AI prompt generator, developed and validated through the expert consensus process in Study 1, is effective in producing clinically meaningful weight loss and favourable body composition changes in adults with overweight and obesity. The sustained improvements observed at the 3-month follow-up suggest that the AI-driven personalised approach may promote durable behavioural changes that extend beyond the active intervention period.

CHAPTER 5
DISCUSSION AND CONCLUSION

5.1 Introduction

This chapter synthesizes the findings of two interconnected studies that collectively address the development, validation, and effectiveness testing of the NExGEN-ChatGPT personalized prompt generator system for weight management in adults with overweight and obesity. The research trajectory followed a logical two-phase progression: Study 1 established the validated foundation through multidisciplinary expert consensus using the Fuzzy Delphi Method, while Study 2 deployed this validated framework in a rigorous randomized controlled trial to evaluate its clinical effectiveness against structured manual ChatGPT guidance.

The discussion is organized to address each dimension of the research findings systematically. It begins with the expert validation outcomes from Study 1, examining the significance of consensus patterns across the six constructs, twelve elements, and 111 criteria that compose the NExGEN-ChatGPT framework. It then transitions to the clinical effectiveness findings from Study 2, interpreting the results for body weight, body composition, dietary behavior, physical activity, and cardiometabolic markers within the context of existing digital health and AI-driven intervention literature. Subsequent sections address the personalization mechanism that distinguishes the NExGEN system from generic AI approaches, the theoretical integration of behavior change frameworks with AI-driven technology, and the bridging evidence connecting expert validation to clinical outcomes.

The chapter concludes with sections addressing the theoretical, practical, and policy implications of both studies, the original contributions to knowledge, acknowledged limitations, and recommendations for future research directions. Throughout, the discussion contextualizes these findings against five identified research gaps: the absence of a validated prompt-generation mechanism for obesity-specific AI interventions, the inadequacy of existing evaluation frameworks for dynamic AI outputs, missing domain-specific quality criteria for AI-generated health recommendations, limited evidence comparing personalized versus generic AI-driven

weight management, and the need for comprehensive outcome assessment beyond body weight alone.

5.2 Expert Consensus of the NExGEN-ChatGPT Framework (Study 1)

5.2.1 Overall Expert Consensus and Frameworks Acceptance

The comprehensive acceptance of all 111 criteria across six constructs and twelve elements represents a significant methodological achievement in AI-driven health technology validation. Every criterion met the three Fuzzy Delphi Method acceptance thresholds: expert agreement of 75% or greater, threshold value d of 0.20 or less, and defuzzified alpha-cut score of 0.50 or above. To the best of the researcher's knowledge, this outcome suggests that the NExGEN-ChatGPT framework may represent the first expert-validated prompt engineering instrument designed specifically for AI-driven obesity management, directly addressing the first identified research gap concerning the absence of validated prompt-generation mechanisms for obesity-specific AI interventions.

The selection of the Fuzzy Delphi Method over conventional Delphi approaches was methodologically strategic. Traditional Delphi techniques require multiple iterative rounds of expert consultation, which can introduce attrition bias and logistical constraints that compromise panel quality (Ishikawa et al., 1993). The FDM accommodates expert uncertainty through triangular fuzzy numbers, enabling more nuanced quantification of agreement levels while requiring only a single round of data collection. This approach was particularly appropriate given the multidisciplinary panel composition spanning five professional domains, each bringing distinct epistemic frameworks to the evaluation of an inherently interdisciplinary tool.

The multidisciplinary composition of the 21-member panel, comprising five nutritionists, five exercise specialists, five medical practitioners, four psychologists, and two AI researchers, each with over a decade of professional experience, strengthens the validity of the consensus outcomes. Prior Delphi studies in health technology validation have frequently relied on narrower disciplinary representation, which can introduce systematic blind spots when evaluating tools that span multiple clinical domains (Boulkedid et al., 2011). The deliberate inclusion of AI researchers alongside clinical practitioners ensured that both technological feasibility and clinical appropriateness

were evaluated concurrently, a dual validation perspective not consistently documented in existing health technology assessment literature.

Instrument reliability, confirmed with Cronbach's alpha ranging from 0.714 to 0.891 across constructs, further supports the robustness of the consensus outcomes. These values exceed the commonly accepted threshold of 0.70 for internal consistency in expert panel research (Nunnally & Bernstein, 1994), indicating that items within each construct measured coherent underlying dimensions. The consistency of reliability across all six constructs suggests that the framework architecture itself, organized around clinically meaningful domains, facilitated structured and internally consistent expert evaluation.

5.2.2 Physical Activity Assessment, Goal Setting, and Stress Monitoring

The highest-ranked criterion, initial and ongoing physical activity assessment (fuzzy score 0.859, 100% consensus), reveals a strong expert conviction that baseline physical monitoring constitutes a non-negotiable prerequisite for safe and effective exercise prescription in populations with overweight and obesity. This unanimous agreement aligns with established clinical guidelines emphasizing that pre-exercise assessment is essential for identifying contraindications, calibrating exercise intensity, and preventing adverse events in higher-risk populations(American College of Sports Medicine, 2022) . The fact that this criterion achieved perfect consensus across all five disciplinary groups reinforces its status as a universal clinical principle rather than a domain-specific perspective.

Clear weight change goal definition ranked second (fuzzy score 0.806, 100% consensus), reflecting expert recognition that explicit, measurable targets serve as behavioral anchors in weight management interventions. This finding is consistent with Goal-Setting Theory (Locke & Latham, 2002), which demonstrates that specific and challenging goals produce higher performance than vague or easy goals. In the NExGEN framework, weight loss targets of 5 to 10 percent of baseline body weight over 12 weeks (0.5 to 1.0 kg per week) provide concrete anchoring that translates theoretical goal-setting principles into operationalized AI prompt parameters.

Stress consideration in health outcomes ranked third (fuzzy score 0.784, 95.2% consensus), a finding that carries particular significance for AI-driven interventions. The high consensus level suggests experts recognize that stress operates as both a direct

physiological driver of weight gain, through cortisol-mediated mechanisms affecting appetite regulation, fat storage, and metabolic function, and an indirect behavioral mediator that undermines dietary adherence and physical activity engagement (Tomiyama, 2019). The inclusion of stress assessment within a prompt-engineering framework acknowledges that AI-generated recommendations must account for psychosocial context, not merely physiological parameters, to produce actionable and safe guidance.

The close fuzzy scores between physical activity-goal alignment (0.765) and tracking systems (0.762) suggest these criteria function synergistically rather than hierarchically. This pattern indicates that experts perceive effective goal-setting as inseparable from integrated progress monitoring, a relationship well-documented in self-regulation theory (Carver & Scheier, 1998). For the NExGEN system, this synergistic relationship reinforces the design decision to embed both goal-setting parameters and progress tracking mechanisms within the prompt architecture, rather than treating them as independent modules.

5.2.3 Medication Interactions and Physiological Monitoring

Medication interaction avoidance achieved a fuzzy score of 0.698 with 95.2% consensus, while physiological monitoring reached 0.659 with 85.7% consensus. Although these scores were lower than the top-ranked criteria, their near-universal agreement levels reflect the clinical gravity these issues carry in unsupervised digital health contexts. The validation of safety-critical criteria directly addresses the third research gap, the absence of domain-specific quality criteria for AI-generated health recommendations that incorporate obesity-specific safety considerations.

The emphasis on medication interaction avoidance is particularly relevant given that individuals with obesity frequently present with comorbidities requiring pharmacological management, including antihypertensives, antidiabetic agents, and lipid-lowering medications (Gibbs et al., 2005). AI-generated dietary recommendations that fail to account for medication-nutrient interactions, such as the well-documented interaction between vitamin K-rich foods and warfarin, or grapefruit and statins, pose genuine safety risks. Similarly, exercise recommendations must consider medication-induced side effects such as hypoglycemia from sulfonylureas or postural hypotension from antihypertensives. The expert consensus confirms that any validated AI health tool

must embed these safety checks as foundational criteria rather than optional enhancements.

The physiological monitoring criterion, while achieving lower consensus than medication interactions, nonetheless reflects expert recognition that AI-driven exercise prescription for populations with overweight and obesity requires ongoing physiological feedback mechanisms. Current AI chatbot systems operate without real-time physiological data integration, relying instead on user self-report or initial assessment data to calibrate recommendations. The 85.7% consensus on this criterion suggests that while experts affirm its importance, they also acknowledge the current technological limitations in implementing continuous physiological monitoring within chatbot-based interventions. This tension between clinical ideal and technological reality represents a productive area for future system development.

5.2.4 Technology Dependence and Sleep-Exercise Interactions

Lower consensus on technology tools (76.19%, fuzzy score 0.665) compared to activity diversity (81%, fuzzy score 0.667) reveals an important tension within the expert panel. While all experts acknowledged technology's role in supporting behavior change, the relative reluctance to endorse technology dependence suggests genuine professional concern about over-reliance on digital tools for complex health behaviors. This finding echoes broader debates in digital health literature regarding whether technology-mediated interventions risk substituting surface-level engagement for deeper behavioral and cognitive change (Michie et al., 2017).

The borderline consensus on sleep's impact on fitness outcomes (76.2%, fuzzy score 0.665), hovering just above the 75% acceptance threshold, exposes genuine uncertainty about whether AI systems can adequately address complex psychophysiological mechanisms. The bidirectional relationship between sleep and exercise, mediated through hormonal cascades involving growth hormone, cortisol, leptin, and ghrelin, represents a level of physiological complexity that current large language models cannot dynamically model (Kline, 2014). This borderline acceptance does not indicate that experts dismiss the importance of sleep in obesity management, rather, it reflects skepticism about AI's current capacity to generate appropriately nuanced recommendations for this domain.

These areas of lower consensus serve a constructive function within the validation framework. Rather than undermining the NExGEN system's validity, they identify specific domains where prompt engineering must be particularly careful in its recommendation scope and where human clinical oversight may remain necessary. The FDM's capacity to quantify this uncertainty, rather than reducing it to binary accept-reject decisions, represents an advantage over conventional Delphi approaches that may mask meaningful variation in expert confidence.

5.2.5 NExGEN Versus Existing Health Technology Assessment Instruments

The NExGEN-ChatGPT framework distinguishes itself from existing health technology assessment instruments in three fundamental ways. First, it validates prompt engineering mechanisms rather than static application features. Current instruments such as the Quality Assessment of AI in Medicine (QAMAI) and the Mobile Application Rating Scale (MARS) were designed to evaluate fixed interface characteristics, usability metrics, and content accuracy of static applications (Stoyanov et al., 2015). These frameworks fail to account for how input quality and prompt engineering determine the relevance, safety, and personalization depth of AI-generated outputs, an evaluative gap that grows increasingly consequential as large language models become central to digital health delivery.

Second, the framework incorporates obesity-specific safety criteria that no prior evaluation instrument has established. Existing digital health evaluation tools apply generic quality metrics across health domains without accounting for the clinical complexity unique to obesity management, which spans physiological, psychological, behavioral, and environmental dimensions. The NExGEN framework's inclusion of medication interaction avoidance, comorbidity management, physiological monitoring thresholds, and culturally appropriate dietary guidance establishes domain-specific validation criteria that directly address the second and third research gaps.

Third, the framework achieves multidisciplinary consensus through fuzzy logic that accommodates expert uncertainty rather than forcing artificial agreement. This methodological characteristic is particularly important for evaluating AI health tools, where the rapid pace of technological advancement means that expert confidence levels naturally vary across different system capabilities. The graduated consensus scores produced by the FDM provide more actionable information for system refinement than

binary consensus outcomes, enabling developers to prioritize areas of strong agreement while directing additional development resources toward domains where expert confidence is lower

5.3 Effectiveness of the NExGEN-ChatGPT Intervention (Study 2)

5.3.1 Body Weight Reduction and Clinical Significance

The NExGEN personalized prompt generator integrated with ChatGPT demonstrated clinically and statistically significant superior efficacy for weight reduction compared to structured manual ChatGPT guidance. The NEX group achieved a weight loss of 6.6 kg (7.5%) at 12 weeks and 5.5 kg (6.2%) at 24 weeks, with both time points reaching statistical significance against baseline ($P < 0.001$). Between-group differences were substantial: the NEX group achieved 3.2 kg greater weight loss than the CON group at 12 weeks ($P = 0.016$) and 3.5 kg greater at 24 weeks ($P = 0.018$), with a significant group-by-time interaction ($P < 0.001$, effect size 0.14). These results directly address the fourth research gap, providing the first evidence from a randomized controlled trial comparing personalized versus generic AI-driven approaches for weight management.

The magnitude of weight loss achieved by the NEX group compares favorably to established benchmarks in digital health intervention literature. A systematic review by Kupila et al. (2023) reported that eHealth interventions for weight loss typically achieve mean differences of -1.07 to -3.10 kg versus control conditions, while digital self-monitoring interventions have been associated with mean weight reductions of approximately -2.87 kg (Cavero-Redondo et al., 2020). The 6.6 kg weight loss at 12 weeks in the NEX group substantially exceeds these benchmarks, suggesting that personalized prompt engineering adds meaningful clinical value beyond what generic digital health delivery achieves. Notably, an umbrella review of digital health interventions for noncommunicable disease prevention reported a pooled weight reduction of -1.99 kg (95% CI: -2.19 to -1.79), further contextualizing the superiority of the NExGEN (Willmott et al., 2019).

The clinical significance of these findings is reinforced by the proportion of NEX participants achieving clinically meaningful weight loss thresholds. The group mean of 7.5% body weight reduction at 12 weeks exceeds the 5% threshold widely

recognized as clinically significant for reducing obesity-related comorbidity risk (D. Ryan & Yockey, 2017). This threshold is associated with measurable improvements in visceral adipose tissue, glycemic control, lipid profiles, and blood pressure, though the magnitude of cardiometabolic benefit is dose-dependent and varies by baseline metabolic health status

A particularly important finding is the weight loss trajectory of the control group. Rather than achieving modest weight loss as might be expected from structured ChatGPT guidance, the CON group experienced gradual weight increases: 3.0% at 12 weeks and 1.7% at follow-up. This divergent trajectory suggests that, within this trial, generic AI-generated prompts delivered without individualized assessment data were insufficient for producing weight loss in this population. This finding should be interpreted cautiously, as the control condition was not designed to represent the full range of possible generic AI-delivered approaches, and the weight increase in the control group may partly reflect limitations of the comparator design rather than the inherent inadequacy of non-personalized AI guidance per se. Nevertheless, this provides preliminary evidence that prompt personalization quality warrants serious consideration in future AI health intervention design, a direction that warrants replication across different populations and comparator conditions.

Weight loss was sustained at 24 weeks (5.5 kg, 6.2%), demonstrating effectiveness beyond the 12-week active intervention period. While a modest degree of weight regain occurred between weeks 12 and 24, over 80% of the initial weight loss was maintained across the 12-week unsupervised follow-up period. This short-term maintenance is encouraging; however, direct comparisons with conventional programs that report regain over two to five years (Hall & Kahan, 2018) should be made cautiously, as the 12-week follow-up period examined here is substantially shorter than those timeframes. Whether these initial maintenance patterns persist beyond 24 weeks cannot be determined from the present data, and longer follow-up studies are needed to assess weight loss durability. The sustained behavioral changes documented through dietary and physical activity measures (discussed in subsequent sections) suggest that the personalized intervention instilled durable behavioral patterns rather than producing transient compliance-driven weight change.

5.3.2 Fat Mass Loss with Fat Free Mass Preservation

The body composition results reveal a qualitatively favorable pattern of weight loss that extends the clinical significance beyond raw weight change. Fat mass decreased significantly in the NEX group, with reductions of 3.7 kg (10.5%) at 12 weeks and 3.1 kg (8.8%) at follow-up (both $P < 0.001$), accompanied by a significant group-by-time interaction ($P = 0.02$, effect size 0.10). Critically, fat-free mass remained stable across all time points with no significant effects ($P > 0.05$), confirming that weight loss was primarily derived from adipose tissue rather than lean mass.

This preservation of fat-free mass during weight loss represents a clinically important outcome that is often overlooked in AI-based and digital health intervention studies. Low fat-free mass is independently associated with higher all-cause mortality, reduced resting energy expenditure, impaired functional capacity, and increased risk of sarcopenic obesity (Cava et al., 2017). Weight loss interventions that fail to protect lean mass can paradoxically worsen long-term metabolic outcomes by reducing basal metabolic rate and creating conditions favorable for subsequent weight regain with preferential fat deposition, a phenomenon sometimes termed "fat overshooting" (Dulloo et al., 2012). The NExGEN system's personalized exercise prescription, which included structured resistance training components tailored to individual fitness levels and equipment availability, likely contributed to this favorable body composition profile.

Waist circumference reductions provide additional evidence of favorable body composition change. The NEX group demonstrated significant reductions of 5.4 cm (5.3%) at 12 weeks and 5.0 cm (4.9%) at follow-up, with a significant time-by-intervention interaction ($P < 0.001$, effect size 0.13). Waist circumference serves as a validated proxy for visceral adipose tissue, the fat depot most closely linked to cardiometabolic risk (Ross et al., 2020). Clinical evidence supports that approximately 5% body weight loss can reduce visceral adipose tissue by 10% or more (Verheggen et al., 2016) and the magnitude of waist circumference reduction observed in the NEX group is consistent with clinically meaningful visceral fat mobilization.

The between-group differences in fat mass, while clinically meaningful, approached but did not reach conventional statistical significance at either time point (1.7 kg at 12 weeks, $P = 0.082$; 1.6 kg at follow-up, $P = 0.097$). This pattern suggests that the study may have been slightly underpowered for detecting between-group differences in this secondary outcome, as the trial was powered primarily on the basis

of body weight as the primary endpoint. The consistency of the direction and magnitude of between-group differences across both time points, combined with the significant group-by-time interaction, supports the clinical relevance of these findings despite the conventional significance threshold not being met for pairwise comparisons.

The comprehensive body composition assessment methodology, using the validated Seca mBCA 515 bioelectrical impedance analysis system under standardized conditions, directly addresses the fifth research gap concerning the need for outcome assessment beyond body weight alone. Few prior AI-based weight management interventions have performed validated bioelectrical impedance body composition analysis, and the inclusion of fat mass, fate free mass, and waist circumference alongside body weight provides a substantially more complete picture of intervention effects on body composition quality.

It is important to consider alternative explanations for the observed between-group differences before attributing effects exclusively to prompt personalisation. First, expectation and attention effects cannot be entirely ruled out; participants in the NEX group were aware they were receiving a more sophisticated personalised system, which may have increased motivation independent of the content quality of the prompts. Second, the NExGEN group's higher engagement levels may partly reflect novelty effects associated with the more interactive and tailored experience, rather than solely the clinical quality of the personalised recommendations. Third, the weekly feedback mechanism embedded within the NExGEN system introduced a structured self-monitoring component that the control group's static prompts did not replicate to the same degree, meaning that the between-group difference may partly reflect the benefit of structured self-monitoring rather than personalisation alone. These alternative pathways do not negate the primary interpretation but underscore that the observed effects are likely multifactorial, with personalised prompt engineering operating alongside engagement, self-monitoring, and expectation mechanisms to produce the observed outcomes.

5.3.3 Energy Density as the Primary Mechanism

Energy density reduction emerged as the primary dietary mechanism mediating weight loss in the NEX group. Significant group-by-time interactions were observed for dietary energy density ($P = 0.0014$, effect size 0.041) and fat intake ($P < 0.001$, effect

size 0.066), with the NEX group achieving significantly lower values than the CON group at both 12 weeks and follow-up. This pattern identifies dietary energy density as the primary behavioral pathway through which personalized AI-generated dietary recommendations translated into differential weight outcomes between groups.

The emphasis on energy density as a weight loss mechanism is well-supported by nutritional science. Reducing dietary energy density, defined as the energy content per unit weight of food, enables individuals to consume satisfying food volumes while reducing total energy intake (Rolls, 2009). Tailored feedback on energy density has been shown to enhance adherence to dietary strategies beyond what standardized digital content achieves (Sherrington et al., 2016). The NExGEN system's personalized dietary plans, which incorporated individual nutritional needs, caloric intake targets, macronutrient distribution, and culturally appropriate food recommendations, appear to have been more effective at facilitating energy density reduction than the generic dietary guidance provided to the control group.

The durability of dietary changes at follow-up carries important implications. Significant between-group differences in energy density persisted at 24 weeks ($P < 0.01$), twelve weeks after the active intervention period ended. This persistence suggests that the dietary changes achieved during the intervention were not entirely dependent on ongoing external guidance, a finding that is encouraging. However, the 12-week follow-up period is insufficient to characterize these changes as lasting dietary habits, and it remains unknown whether the improvements in energy density would be sustained at 12 or 24 months. Future studies with extended follow-up periods are needed to determine whether personalized AI-driven dietary guidance can support durable behavioral change beyond the timeframes examined in this trial. The three-phase intervention structure (exploration, consolidation, reinforcement) may have contributed to this durability by gradually building dietary self-management skills across the intervention period.

Fiber intake showed a significant time effect ($P < 0.001$, effect size 0.128) with both groups improving, though the time-by-intervention interaction did not reach significance ($P = 0.076$). This pattern suggests that both personalized and generic ChatGPT-delivered dietary guidance successfully communicated the importance of fiber intake, but personalization did not confer additional advantage for this specific dietary component. The broader dietary guidance literature supports that fiber intake recommendations are relatively straightforward to convey and implement regardless of

personalization level, in contrast to energy density management, which requires more individualized meal planning and food substitution strategies.

The cultural tailoring dimension of the NExGEN dietary recommendations warrants specific discussion. The system generated personalized grocery shopping lists and culturally appropriate recipes adapted for local Malaysian ingredients, addressing the practical implementation barriers that frequently undermine dietary adherence in digital health interventions. This approach aligns with evidence that culturally sensitive dietary guidance enhances engagement and adherence in diverse populations (Telle-Hjellset et al., 2013). In the Malaysian context, where dietary traditions involve specific cooking methods, ingredient preferences, and meal patterns that differ substantially from Western dietary frameworks, this cultural specificity likely contributed to the observed dietary improvements.

5.3.4 Physical Activity Measures

Physical activity improvements were confirmed through both objective and subjective measurement modalities. Accelerometry data from ActiGraph GT3X+ devices demonstrated a significant group-by-time interaction ($P = 0.0017$, effect size 0.040), with the NEX group achieving significantly higher activity levels than the CON group at 12 weeks ($P = 0.023$) and follow-up ($P = 0.024$). Self-reported physical activity measured by the IPAQ-LV confirmed this pattern, with a significant group-by-time interaction ($P = 0.0021$, effect size 0.038) and significant NEX increases at both time points (both $P < 0.001$).

The dual-modality measurement approach, combining objective accelerometry with validated self-report, provides substantially stronger evidence than either measure alone. Objective accelerometry eliminates social desirability and recall bias that commonly inflate self-reported physical activity estimates, while the IPAQ-LV captures activity domains such as household and occupational activities that waist-worn accelerometers may underestimate (P. Lee et al., 2011). The convergence of significant findings across both measures strengthens confidence in the physical activity improvements attributed to the NExGEN intervention.

The personalized exercise prescription generated by the NExGEN system, which tailored exercise modalities, duration targets, intensity levels, progression schedules, locations, timings, and instructions to individual fitness levels, preferences,

and equipment availability, appears to have been more effective at promoting physical activity than generic exercise recommendations. This finding is consistent with evidence that personalized exercise prescriptions account for individual barriers and facilitators of physical activity, producing higher adherence rates than standardized recommendations that may not align with individuals' practical circumstances (Burgess et al., 2017).

Both physical activity measures demonstrated sustained improvements at follow-up, a finding that carries particular significance for weight loss maintenance. The persistence of elevated physical activity levels twelve weeks after the active intervention ended suggests that the NExGEN system's progressive approach, moving through exploration, consolidation, and reinforcement phases, successfully established exercise habits that outlasted the intervention period. Exercise behavior maintenance is a well-recognized challenge in lifestyle interventions, and the sustained improvements observed here compare favorably to the typical pattern of physical activity decline following intervention cessation documented in the broader literature (Fjeldsoe et al., 2011).

It is worth noting that the effect sizes for physical activity outcomes (0.038 to 0.040) were smaller than those observed for body weight (0.14) and waist circumference (0.13). This differential suggests that while personalized exercise prescription conferred a statistically significant advantage over generic guidance, the magnitude of this advantage was more modest than the differences observed in dietary and weight outcomes. One plausible interpretation is that dietary behavior changes were the primary drivers of the observed weight differential, with exercise prescription contributing a complementary but smaller incremental effect, a pattern consistent with evidence that dietary modification produces larger caloric deficits than exercise prescription in short-term weight loss interventions (Johns et al., 2014).

5.3.5 Cardiometabolic Markers

Cardiometabolic outcomes presented a nuanced pattern that requires careful interpretation. Total cholesterol showed a significant time effect ($P < 0.0001$), with the NEX group achieving significant reductions at both 12 weeks and follow-up, and LDL cholesterol similarly demonstrated a significant time effect ($P = 0.001$) with NEX reductions at 12 weeks ($P = 0.0035$) and follow-up ($P = 0.023$). However, no significant

between-group differences emerged for any cardiometabolic marker, including HbA1c (P = 0.82 for intervention effect), HDL cholesterol, triglycerides, or blood pressure parameters.

The absence of significant between-group differences in cardiometabolic markers, despite significant between-group differences in weight and behavioral outcomes, most plausibly reflects the metabolically healthy baseline profile of study participants. Participants in this trial had relatively normal blood pressure, lipid, and glucose levels despite excess weight, a phenotype variously described as metabolically healthy obesity (MHO). In metabolically healthy individuals, weight loss interventions frequently demonstrate attenuated cardiometabolic improvements because there is less metabolic abnormality to correct (Arsenault et al., 2009). Put directly, ceiling effects in cardiometabolic parameters constrain the detectable benefit of weight loss when baseline values are already within or near normal ranges.

Additionally, less than half of NEX participants achieved more than 5% body weight loss, the threshold generally considered necessary for clinically relevant cardiometabolic effects (D. Ryan & Yockey, 2017). The dose-response relationship between weight loss magnitude and cardiometabolic improvement is well-established: approximately 5% weight loss is required to reduce visceral adipose tissue sufficiently to influence lipid metabolism, insulin sensitivity, and vascular function (Chaston & Dixon, 2008). A study population with a higher prevalence of metabolic abnormalities at baseline, or one in which a greater proportion of participants achieved the 5% weight loss threshold, would likely reveal more pronounced cardiometabolic effects.

The modest but significant improvement in mean arterial pressure, driven by a small NEX reduction at 12 weeks (P = 0.009) that was not sustained at follow-up, aligns with evidence that blood pressure responses to lifestyle intervention are among the earliest cardiometabolic changes to manifest but require sustained behavioral adherence for long-term maintenance (Neter et al., 2003). The transient nature of this improvement suggests that the 12-week active intervention was sufficient to produce initial hemodynamic benefit, but the maintenance phase did not provide adequate ongoing behavioral reinforcement to sustain it.

An important consideration is that cardiometabolic benefits may accrue gradually and could become statistically significant with longer follow-up periods, larger sample sizes, or study populations selected for metabolic abnormalities. The 24-week total study duration may have been insufficient to capture the full trajectory of

lipid and glycemic adaptations to weight loss, particularly given that some cardiometabolic improvements require sustained weight maintenance over several months to manifest fully (Wing et al., 2010). Future studies targeting metabolically unhealthy populations with extended follow-up periods would be better positioned to detect the cardiometabolic benefits of personalized AI-driven weight management.

5.4 Personalized Versus Generic AI-Driven Weight Management

5.4.1 The Role of Input Quality and Prompt Personalization

The divergent outcomes between the NEX and CON groups provide direct empirical evidence that the quality and personalization of input prompts determine the clinical effectiveness of AI-generated health interventions. Both groups accessed the same ChatGPT platform with identical underlying large language model capabilities, received the same weight loss targets (5 to 10% of baseline body weight), followed equivalent physical activity goals aligned with WHO guidelines, and received complimentary ChatGPT Business subscriptions (worth approximately ~RM830; sponsored by *Pragraph Digital Resources, SA0614953-A*). The critical distinguishing variable was data utilization: the NExGEN system algorithmically synthesized all 111 assessment variables to generate personalized prompts, while the control group received uniform prompts regardless of individual profile.

This finding carries substantial implications for the burgeoning field of AI-driven health interventions. It suggests that the therapeutic value of large language model chatbots for health behavior change may not be inherent to the technology itself but could be substantially shaped by the prompt engineering architecture that mediates between individual patient data and AI output. In this trial, generic prompts delivered without individualized assessment data did not produce weight loss, suggesting that prompt personalization quality may be an important, if previously underappreciated, factor in AI health intervention design. These findings should be interpreted as preliminary evidence from a single-site controlled trial, and broader conclusions about generic versus personalized AI guidance await replication in different populations and using different AI platforms. Personalized prompts, built from comprehensive individual assessment data, may enhance the clinical effectiveness of AI systems for health behavior change.

This conclusion finds support in the broader AI personalization evidence base. Graham et al. (2022) documented that AI-guided weight loss programs utilizing individualized data achieved approximately 12% body weight loss compared to 7% in physician-guided controls, suggesting that the degree of personalization correlates with intervention effectiveness. Similarly, recent findings from the AI-DPP trial demonstrated that a fully automated AI-powered lifestyle intervention was noninferior to human coach-led delivery for achieving weight loss and physical activity outcomes in adults with prediabetes (Mathioudakis et al., 2025). Collectively, these findings position prompt engineering quality as a critical, and previously underrecognized, determinant of AI health intervention effectiveness.

5.4.2 User Engagement and System Utilization

Usage analysis revealed higher engagement in the NEX group compared to the CON group, though suboptimal utilization of allocated intervention systems may have attenuated the true effectiveness of the NExGEN intervention. The personalized system's dynamic prompt-response cycles with ChatGPT, incorporating personalized feedback, adaptive recommendations, and ongoing dialogue, created a more engaging interaction paradigm than the static, non-adaptive content delivered to the control group. This engagement differential is consistent with evidence that personalized digital health interventions generate higher user retention and interaction frequency than generic alternatives.

The dose-response relationship between AI chatbot utilization and health outcomes documented in prior studies Auster-Gussman et al. (2022), suggests that the observed between-group differences may represent conservative estimates of the NExGEN system's potential effectiveness. If all NEX participants had utilized the system at optimal frequency, the weight loss differential between groups might have been larger. Conversely, the CON group received only structured manual prompts, so minimal engagement was expected relative to the personalized system. This asymmetry in expected engagement levels reflects a design feature inherent to the comparison of personalized versus generic interventions, rather than a methodological limitation.

The engagement findings also suggest the possibility of a self-reinforcing cycle that may characterize personalized AI health interventions: personalized content may drive higher engagement, higher engagement may produce better outcomes, and visible

progress may further reinforce continued use. This theorized positive feedback loop, which is consistent with Self-Determination Theory's competence-building mechanism, offers one plausible explanation for why between-group differences widened rather than converged over the intervention period. However, the present study was not designed to formally test this mediational pathway, and dedicated process evaluation would be required to confirm it.

5.5 Expert Validation to Clinical Effectiveness

The progression from Study 1 to Study 2 represents a logical and methodologically coherent research trajectory that addresses the primary limitation of Study 1 (absence of user trials) while leveraging its validated assessment architecture as the intervention foundation for Study 2. Study 1 established what should be measured and how those measurements should inform AI prompts through 111 expert-consensus criteria across six constructs. Study 2 tested whether that validated measurement-to-prompt pipeline produces superior real-world outcomes when deployed in a clinical trial context.

The bridging mechanism between the two studies is the 111-variable assessment instrument. Study 1 validated the comprehensiveness, appropriateness, and safety of these variables as inputs to AI prompt generation. Study 2 operationalized these variables as the data capture instrument that drives NExGEN's personalized prompt generation. The significant between-group differences in weight loss, dietary behavior, and physical activity confirm that the expert-validated personalization criteria identified in Study 1 capture clinically meaningful individual variability, as the superiority of the NEX group is most plausibly explained by the personalized prompts generated from these validated assessment variables, though differential engagement levels, expectation effects, and other unmeasured factors cannot be entirely excluded as contributing influences.

This validation-to-effectiveness trajectory aligns with established intervention development frameworks, particularly the Medical Research Council (MRC) framework for developing and evaluating complex interventions (Skivington et al., 2021). The MRC framework emphasizes the importance of systematic development (establishing theoretical basis and content validity) prior to effectiveness evaluation, a sequence that the Study 1 to Study 2 progression follows. The FDM validation

corresponds to the development and feasibility phases, while the RCT corresponds to the evaluation phase, with the pilot study providing an additional feasibility checkpoint between these stages.

The convergence of validation and effectiveness evidence also addresses a common criticism of expert consensus methodologies, namely that expert-validated criteria may not translate into clinically meaningful outcomes when implemented in real-world contexts (Murphy et al., 1998). The significant clinical outcomes in Study 2 provide supporting evidence that the expert panel's judgments about what constitutes appropriate personalization criteria were meaningful in a clinical context, though the specific contribution of individual criteria to the observed outcomes cannot be determined without dismantling studies. This two-stage validation, first through expert consensus and then through clinical outcomes, provides stronger evidence for the NExGEN framework's validity than either form of validation alone.

5.6 Implications of the Study

5.6.1 Practical and Clinical Implications

The practical implications of these findings span multiple domains of healthcare delivery. Most directly, the NExGEN-ChatGPT system demonstrates the potential for personalized AI-driven weight management to augment primary care capacity for obesity treatment. Primary care settings globally face well-documented limitations in their capacity to deliver intensive lifestyle therapy, constrained by appointment durations, practitioner training in behavioral counseling, and the sheer volume of patients requiring weight management support (Tsai & Wadden, 2005). A validated personalized prompt generator system could extend the reach of healthcare systems by providing individualized behavioral guidance at scale without proportional increases in clinical personnel.

The cost-effectiveness implications are similarly substantial. The cost implications of the NExGEN intervention appear potentially favorable at the platform level, as the primary recurring cost was a ChatGPT Business subscription (approximately USD 191 per participant for six months). However, a formal cost-effectiveness analysis was not conducted in this study, and this figure does not capture the full implementation costs, including system development, validation study

expenses, participant assessment resources, researcher oversight, and technical support. Direct cost comparisons with face-to-face interventions should therefore be treated as illustrative only. Dedicated health economic evaluations, incorporating all implementation costs and comparing cost per unit of clinically meaningful outcome across intervention modalities, are needed before conclusions about cost-effectiveness can be drawn.

For clinical practice, the findings underscore that the value of AI health tools depends critically on prompt engineering quality. Clinicians and health system administrators evaluating AI-based weight management tools should prioritize systems that incorporate comprehensive individual assessment and validated personalization algorithms over those that rely on generic conversational frameworks. The NExGEN validation framework itself offers a template for evaluating the personalization quality of competing AI health products.

5.6.2 Policy Implications

At the policy level, these findings inform several emerging regulatory and public health considerations. First, the demonstrated importance of prompt engineering quality for clinical outcomes suggests that regulatory frameworks for AI-generated health recommendations should include assessment of input quality and personalization mechanisms, not merely output accuracy. Current approaches to AI health tool regulation, which focus primarily on whether AI outputs are clinically accurate, may be insufficient if the inputs driving those outputs are not validated for comprehensiveness and appropriateness.

Second, the data privacy implications of comprehensive individual health assessment through AI systems require policy attention. The NExGEN system collects 111 variables spanning medical history, physiological measurements, dietary habits, psychological status, and environmental factors, a data richness that enables effective personalization but also creates substantial privacy obligations. The end-to-end encryption, GDPR-equivalent compliance protocols, and client-side data storage architecture implemented in the NExGEN system offer a model for data governance, but policy frameworks must establish minimum standards for personal health data protection in AI-driven health interventions.

Third, for national obesity strategies in Malaysia and comparable settings, these findings provide preliminary evidence relevant to discussions around AI-driven personalized interventions within national obesity management strategies. Malaysia's substantial obesity burden underscores the need for scalable intervention approaches, and the NExGEN system's generation of culturally appropriate dietary recommendations and grocery lists demonstrates contextual sensitivity to local dietary practices. However, given that this trial was conducted at a single location (Seremban, Negeri Sembilan) with a specific participant demographic, any consideration of broader deployment should be preceded by multinational and multi-site replication studies, cultural adaptation assessments for other regional contexts, and formal health technology assessment processes. The present findings represent a necessary first step toward, rather than sufficient justification for, national health system integration.

5.7 Contribution to Knowledge

This research makes several original contributions to the body of knowledge in digital health, AI-driven health interventions, and obesity management. First, it establishes the NExGEN-ChatGPT framework as, to the best of the author's knowledge, the first expert-validated prompt engineering instrument designed specifically for AI-driven obesity management within the published literature at the time of this study. No prior published instrument had provided validated criteria for evaluating and guiding the generation of personalized AI prompts for health behavior change, and the 111-criterion framework addresses this gap with multidisciplinary expert consensus across six constructs. This claim is made with recognition that the AI health tools field is advancing rapidly, and future systematic reviews may identify parallel or subsequent developments.

Second, to the best of the author's knowledge, it provides the first published randomized controlled trial evidence directly comparing personalized versus generic AI prompt-driven weight management within the same underlying AI platform. While prior studies have evaluated AI chatbot interventions for health behavior change, none identified in the systematic review of the literature had isolated prompt personalization quality as the primary independent variable within a controlled experimental design using the same AI platform for both groups. This distinction in research design, rather than merely in intervention content, constitutes the novel methodological contribution.

Third, this study provides evidence suggesting that prompt engineering quality may be an important determinant of AI health intervention effectiveness, contributing to the field's evolving understanding of what makes AI health tools therapeutically valuable. The divergent outcomes between the personalized and generic prompt groups, using the identical ChatGPT platform, suggest that the input architecture plays a critical role in shaping clinical outcomes, a relationship that warrants investigation across different platforms, populations, and intervention contexts.

Fourth, the comprehensive outcome assessment incorporating body composition analysis, objective physical activity measurement, detailed dietary intake assessment, and cardiometabolic biomarkers sets a methodological standard for future AI-based weight management intervention research. This multi-dimensional approach directly addresses the identified gap in outcome assessment beyond body weight alone.

Fifth, the sequential validation-to-effectiveness research design demonstrates a replicable methodological framework for developing and evaluating AI-driven health interventions, moving from expert consensus validation of the assessment and prompt architecture to clinical effectiveness testing of the complete system.

5.8 Limitations of the Study

5.8.1 Study 1 Limitations

Several limitations of Study 1 must be acknowledged. The validation was theoretical in nature, involving expert evaluation of framework criteria without direct user testing. While expert consensus provides strong content validity, the translation gap between expert-validated criteria and demonstrated user outcomes cannot be fully bridged without empirical user data, a limitation that Study 2 partially but not completely addresses, as Study 2 tested the system’s effectiveness rather than its usability and user experience specifically.

The expert panel was composed exclusively of Malaysian-based professionals, which limits the generalizability of consensus findings to other cultural and healthcare system contexts. While the panel’s multidisciplinary composition strengthens within-context validity, obesity management priorities, dietary norms, exercise culture, and technology acceptance patterns vary across populations and may produce different consensus outcomes in other settings.

The framework was validated using a single AI model (ChatGPT-o3 at the time of validation), and the criteria's applicability across different large language models has not been assessed. Given the rapid evolution of AI capabilities and the proliferation of competing language models, multi-model validation would strengthen the framework's claim to generalizability across AI platforms.

5.8.2 Study 2 Limitation

Complete participant blinding was not feasible due to the inherently different intervention modalities. Participants in the NEX group were aware they were receiving a personalized system, while CON participants received standardized prompts, potentially creating expectation bias. Although research personnel conducting outcome assessments and data analysts remained blinded to group allocation, participant awareness of their condition may have influenced self-reported outcomes and behavioral engagement through placebo-like expectation effects.

Both groups received complimentary six-month ChatGPT Business subscriptions (valued at approximately ~RM830; sponsored by *Pragraph Digital Resources, SA0614953-A*), which may have provided additional motivation beyond the interventions themselves. While this provision was essential for implementing both intervention conditions, the perceived value of receiving a premium subscription could have influenced engagement patterns in both groups, though this effect would apply equally to both conditions.

Self-reported dietary data, despite being collected through seven-day dietary records and analyzed using validated software with the Malaysia Food Composition Table, remain subject to reporting bias. Underreporting of energy intake is a well-documented phenomenon in dietary research, particularly among individuals with overweight and obesity (Livingstone & Black, 2003). While this bias likely affected both groups equally, it constrains the precision of dietary intake estimates and the confidence in between-group dietary comparisons.

The single geographic location Greater Kuala Lumpur (Klang Valley), Malaysia, limits the generalizability of findings to other populations. Demographic characteristics, dietary cultures, physical activity environments, technology access patterns, and healthcare system contexts vary substantially across regions, and the

intervention's effectiveness in Malaysian urban adults may not translate directly to other populations without cultural adaptation and retesting.

The 24-week total study duration (12-week intervention plus 12-week follow-up) may have been insufficient to capture long-term cardiometabolic benefits, which may require sustained weight maintenance over longer periods to manifest fully. Additionally, the follow-up period was relatively short for assessing long-term weight loss maintenance, as the well-documented pattern of post-intervention weight regain typically becomes most apparent at 12 to 24 months post-intervention.

The study did not include a no-intervention or attention-matched control condition. Both groups received active intervention components, meaning it is not possible to determine whether the NExGEN intervention's superiority reflects the benefit of personalization specifically, or whether the structured manual prompts provided to the control group represented a sufficiently active comparator. The control group's weight increase raises the possibility that the comparator condition was inadequate rather than that personalized AI is inherently superior to all forms of generic AI guidance. A three-arm design incorporating a waitlist or care-as-usual condition would have enabled clearer decomposition of these effects.

No formal process evaluation or intervention fidelity assessment was conducted. The mechanisms through which the NExGEN system produced its outcomes remain largely uncharacterized. Specifically, which of the 111 personalization variables contributed most to behavior change, the extent to which participants implemented the AI-generated recommendations in daily life, and how the quality of human-AI interaction varied across participants and over time are all unknown. Without process evaluation data, causal interpretation of the between-group differences is limited, and identification of the active intervention components is not possible.

The intervention was delivered exclusively through ChatGPT, a proprietary commercial platform subject to model version updates, usage policy changes, and pricing revisions outside the researcher's control. The performance observed in this trial is specific to the version of ChatGPT used and may not replicate with future model versions or alternative large language models. This platform dependence represents a practical constraint on the long-term reproducibility and scalability of the NExGEN system as described, and independent replication on open-source or alternative AI platforms cannot be assumed.

5.9 Recommendation for Future Research

Several directions for future research emerge from the current findings. First, multinational replication with diverse populations across multiple countries would establish the cross-cultural generalizability of the NExGEN system's effectiveness and identify any population-specific adaptations required for optimal outcomes. The cultural tailoring capability demonstrated in the Malaysian context should be systematically evaluated in other dietary and cultural traditions.

Second, long-term follow-up studies extending beyond 24 weeks, ideally to 12 months or longer, are needed to assess the durability of weight loss, behavioral changes, and any delayed cardiometabolic benefits. Such studies should include assessments of weight loss maintenance patterns, behavioral habit persistence, and the potential need for periodic re-engagement with the AI system to sustain outcomes.

Third, multi-model AI framework performance comparison across different large language models (such as GPT-4, Claude, Gemini, and open-source alternatives) would determine whether the NExGEN prompt engineering architecture produces consistent results across AI platforms or whether model-specific optimization is required.

Fourth, dismantling studies that systematically vary the number and type of personalization elements would identify which of the 111 assessment variables contribute most critically to intervention effectiveness. Such studies could inform the development of more efficient assessment instruments that capture the most impactful personalization variables while reducing participant burden.

Fifth, trials targeting metabolically unhealthy populations with elevated baseline cardiometabolic risk factors would provide a more sensitive context for detecting the cardiometabolic benefits of personalized AI-driven weight management. Such populations would also benefit most directly from the intervention's demonstrated effectiveness, aligning research design with clinical impact.

Sixth, cost-effectiveness analysis comparing the NExGEN system's cost per unit of weight loss against conventional face-to-face interventions, generic digital health programs, and pharmacological treatments would provide the economic evidence needed to inform health policy decisions about population-scale deployment.

Finally, investigation of the NExGEN system's applicability to related chronic disease management contexts, such as type 2 diabetes management, cardiovascular risk

reduction, and metabolic syndrome treatment, would test the generalizability of the personalized prompt engineering approach beyond obesity-specific applications.

5.10 Conclusion

This research, spanning from expert validation to clinical effectiveness testing, provides preliminary but encouraging evidence that personalized AI prompt engineering may represent a meaningful contribution to digital weight management interventions, subject to replication in broader populations and independent settings. Study 1 established the NExGEN-ChatGPT framework as, to the best of the author's knowledge, the first published expert-validated prompt engineering instrument for AI-driven obesity management, with all 111 criteria achieving multidisciplinary consensus through rigorous Fuzzy Delphi Method evaluation. Study 2 demonstrated that deploying this validated framework in a randomized controlled trial produced clinically significant weight loss (6.6 kg, 7.5% at 12 weeks), favorable body composition changes (fat mass loss with fat-free mass preservation), significant dietary behavior improvements (energy density reduction as the primary mechanism), and sustained physical activity increases, all significantly superior to structured manual ChatGPT guidance using the same AI platform.

The findings address all five identified research gaps. The NExGEN framework provides the first validated prompt-generation mechanism for obesity-specific AI interventions. Its evaluation criteria move beyond static feature assessment to evaluate dynamic, prompt-dependent AI outputs. The 111-criterion structure establishes domain-specific quality criteria incorporating obesity-specific safety considerations. The RCT provides direct comparative evidence between personalized and generic AI-driven approaches. And the comprehensive outcome battery, encompassing body composition, behavioral variables, and cardiometabolic markers, sets a standard for multi-dimensional outcome assessment in AI-based interventions.

The central insight from this research is that the therapeutic value of AI health interventions resides not in the AI model itself, but in the quality of the personalization architecture that mediates between individual patient data and AI outputs. The identical ChatGPT platform produced clinically significant weight loss when driven by personalized prompts and weight gain when driven by generic prompts. This finding advances the field's understanding of AI-driven health interventions by positioning

prompt engineering quality, validated through expert consensus and supported by clinical outcomes from this trial, as an important and previously underexamined factor in intervention effectiveness. Replication across diverse populations, AI platforms, and healthcare contexts will be needed to establish the generalizability of this principle. As AI technologies continue to evolve and integrate into healthcare delivery systems, the principles established by this research, comprehensive individual assessment, validated prompt generation, and theory-grounded personalization, offer a foundation for developing effective AI-driven interventions that translate technological capability into meaningful clinical outcomes.

REFERENCES

Abrahams, M., & Matusheski, N. V. (2020). Personalised nutrition technologies: A new paradigm for dietetic practice and training in a digital transformation era. Journal of Human Nutrition and Dietetics, 33(3). https://doi.org/10.1111/jhn.12746

Achamrah, N., Colange, G., Delay, J., Rimbert, A., Folope, V., Petit, A., Grigioni, S., Déchelotte, P., & Coëffier, M. (2018). Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures. PLOS ONE, 13(7), Article e0200465. https://doi.org/10.1371/journal.pone.0200465

Adilmetova, G., Nassyrov, R., Meyerbekova, A., Karabay, A., Varol, H. A., & Chan, M.-Y. (2025). Evaluating ChatGPT's multilingual performance in clinical nutrition advice using synthetic medical text: Insights from Central Asia. The Journal of Nutrition, 155(3), 729–735. https://doi.org/10.1016/j.tjnut.2024.12.018

Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. Journal of Medical Internet Research, 25, Article e40789. https://doi.org/10.2196/40789

Aggarwal, M., Devries, S., Freeman, A. M., Ostfeld, R., Gaggin, H., Taub, P., Rzeszut, A. K., Allen, K., & Conti, R. C. (2018). The deficit of nutrition education of physicians. The American Journal of Medicine, 131(4). https://doi.org/10.1016/j.amjmed.2017.11.036

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. Organizational Research Methods, 16(2), 270–301. https://doi.org/10.1177/1094428112470848

Al-Ashwal, F. Y., Zawiah, M., Gharaibeh, L., Abu-Farha, R., & Bitar, A. N. (2023). Evaluating the sensitivity, specificity, and accuracy of ChatGPT-3.5, ChatGPT-4, Bing AI, and Bard against conventional drug-drug interactions clinical tools. Drug, Healthcare and Patient Safety, 15, 137–147. https://doi.org/10.2147/DHPS.S425858

Alelyani, T. (2025). A validated framework for responsible AI in healthcare autonomous systems. Scientific Reports, 15, Article 44432. https://doi.org/10.1038/s41598-025-25266-z

Aljamaan, F., Temsah, M.-H., Altamimi, I., Al-Eyadhy, A., Jamal, A., Alhasan, K.,

Mesallam, T. A., Farahat, M., & Malki, K. H. (2024). Reference hallucination score for medical artificial intelligence chatbots: Development and usability study. JMIR Medical Informatics, 12, Article e54345. https://doi.org/10.2196/54345

Altman, D. G., & Bland, J. M. (1995). Statistics notes: The normal distribution. BMJ, 310(6975), 298. https://doi.org/10.1136/bmj.310.6975.298

American College of Physicians, American Telemedicine Association, & Organization for the Review of Care and Health Apps. (2026). Digital Health Assessment Framework.

American College of Sports Medicine. (2022). ACSM's guidelines for exercise testing and prescription (G. Liguori, Ed.; 11th ed.). Wolters Kluwer.

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). Sage.

Anh-Hoang, Tran, V., & Nguyen, L.-M. (2025). Survey and analysis of hallucinations in large language models: Attribution to prompting strategies or model behavior. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1622292

Arsenault, B., Côté, M., Cartier, A., Lemieux, I., Després, J.-P., Ross, R., Earnest, C. P., Blair, S. N., & Church, T. S. (2009). Effect of exercise training on cardiometabolic risk markers among sedentary, but metabolically healthy overweight or obese post-menopausal women with elevated blood pressure. Atherosclerosis, 207(2), 530–533. https://doi.org/10.1016/j.atherosclerosis.2009.05.009

Arslan, S. (2023). Exploring the potential of ChatGPT in personalized obesity treatment. Annals of Biomedical Engineering, 51(9). https://doi.org/10.1007/s10439-023-03227-9

Auster-Gussman, L. A., Lockwood, K. G., Graham, S. A., Pitter, V., & Branch, O. H. (2022). Engagement in digital health app-based prevention programs is associated with weight loss among adults age 65+. Frontiers in Digital Health, 4. https://doi.org/10.3389/fdgth.2022.886783

Azimi, I., Qi, M., Wang, L., Rahmani, A. M., & Li, Y. (2025). Evaluation of LLMs accuracy and consistency in the registered dietitian exam through prompt engineering and knowledge retrieval. Scientific Reports, 15, Article 1506. https://doi.org/10.1038/s41598-024-85003-w

Azmi, M. Y., Junidah, R., Siti Mariam, A., Safiah, M. Y., Fatimah, S., Norimah, A. K., Poh, B. K., Kandiah, M., Zalilah, M. S., Wan Abdul Manan, W. M., Siti Haslinda,

M. D., & Tahir, A. (2009). Body mass index (BMI) of adults: Findings of the Malaysian Adult Nutrition Survey (MANS). Malaysian Journal of Nutrition, 15(2).

Babak, A., Rouzbahani, R., Nejad, R. K., & Zadeh, A. R. (2019). Comparison of nutritional behaviors and physical activities between overweight/obese and normal-weight adults. Advanced Biomedical Research, 8(1). https://doi.org/10.4103/abr.abr_134_19

Bajramagic, M., Battelino, T., Cos, X., Cote, M., Cui, N., Forbes, A., Galderisi, A., Heinemann, L., Hussain, S., Imbert, J., Holm Jönsson, C., Joubert, M., Lalić, N. M., Phillip, M., Schwarz, P., Torbeyns, B., Wake, D. J., Zakrzewska, K., & Del Prato, S. (2026). Artificial intelligence-driven clinical decision support systems to assist healthcare professionals and people with diabetes in Europe at the point of care: A Delphi-based consensus roadmap. Diabetologia, 69, 259–273. https://doi.org/10.1007/s00125-025-06601-5

Baretta, D., Perski, O., & Steca, P. (2019). Exploring users' experiences of the uptake and adoption of physical activity apps: Longitudinal qualitative study. JMIR mHealth and uHealth, 7(2), Article e11636. https://doi.org/10.2196/11636

Barrett, D., & Heale, R. (2020). What are Delphi studies? Evidence Based Nursing, 23(3), 68–69. https://doi.org/10.1136/ebnurs-2020-103303

Barry, V. W., Baruth, M., Beets, M. W., Durstine, J. L., Liu, J., & Blair, S. N. (2014). Fitness vs. fatness on all-cause mortality: A meta-analysis. Progress in Cardiovascular Diseases, 56(4). https://doi.org/10.1016/j.pcad.2013.09.002

Barua, R., Sarkar, A., & Datta, S. (2023). Emerging advancement of 3D bioprinting technology in modern medical science and vascular tissue engineering education. In Handbook of research on instructional technologies in health education and allied disciplines. IGI Global. https://doi.org/10.4018/978-1-6684-7164-7.ch007

Basiotis, P. P., Welsh, S. O., Cronin, F. J., Kelsay, J. L., & Mertz, W. (1987). Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. The Journal of Nutrition, 117(9), 1638–1641. https://doi.org/10.1093/jn/117.9.1638

Bayram, B., & Arslan, S. (2025). Nutritional analysis of AI-generated diet plans based on popular online diet trends. Journal of Food Composition and Analysis, 145, Article 107850. https://doi.org/10.1016/j.jfca.2025.107850

Bays, H. E., Fitch, A., Cuda, S., Gonsahn-Bollie, S., Rickey, E., Hablutzel, J., Coy, R.,

& Censani, M. (2023). Artificial intelligence and obesity management: An Obesity Medicine Association (OMA) Clinical Practice Statement (CPS) 2023. Obesity Pillars, 6, Article 100065. https://doi.org/10.1016/j.obpill.2023.100065

Beleigoli, A. M., Andrade, A. Q., Cançado, A. G., Paulo, M. N. L., Diniz, M. F. H. S., & Ribeiro, A. L. (2019). Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: Systematic review and meta-analysis. Journal of Medical Internet Research, 21(1). https://doi.org/10.2196/jmir.9609

Bharti, A., Kulshrestha, K., & Kushwaha, A. (2021). Enhancing nutrition knowledge: A step towards childhood obesity prevention. The Pharma Innovation, 10(3). https://doi.org/10.22271/tpi.2021.v10.i3d.5785

Bhawra, J., Kirkpatrick, S., Hall, M. G., Vanderlee, L., White, C. M., & Hammond, D. (2023). Patterns and correlates of nutrition knowledge across five countries in the 2018 International Food Policy Study. Nutrition Journal, 22(1). https://doi.org/10.1186/s12937-023-00844-x

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema, 29(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Blüher, M. (2020). Metabolically healthy obesity. Endocrine Reviews, 41(3). https://doi.org/10.1210/endrev/bnaa004

Bodjanova, S. (2006). Median alpha-levels of a fuzzy number. Fuzzy Sets and Systems, 157(7), 879–891. https://doi.org/10.1016/j.fss.2005.10.015

Boit, S., & Patil, R. (2025). A prompt engineering framework for large language model-based mental health chatbots: Conceptual framework. JMIR Mental Health, 12, Article e75078. https://doi.org/10.2196/75078

Bosy-Westphal, A., Jensen, B., Braun, W., Pourhassan, M., Gallagher, D., & Müller, M. J. (2017). Quantification of whole-body and segmental skeletal muscle mass using phase-sensitive 8-electrode medical bioelectrical impedance devices. European Journal of Clinical Nutrition, 71, 1061–1067. https://doi.org/10.1038/ejcn.2017.27

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. PLOS ONE, 6(6), Article e20476. https://doi.org/10.1371/journal.pone.0020476

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008). Methods and processes of the CONSORT Group: Example of an extension for trials assessing nonpharmacologic treatments. Annals of Internal Medicine, 148(4), W60–W66. https://doi.org/10.7326/0003-4819-148-4-200802190-00008-w1

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33.

Brug, J., Oenema, A., & Campbell, M. (2003). Past, present, and future of computer-tailored nutrition education. The American Journal of Clinical Nutrition, 77(4 Suppl.), 1028S–1034S. https://doi.org/10.1093/ajcn/77.4.1028s

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R., Dempsey, P. C., DiPietro, L., Ekelund, U., Firth, J., Friedenreich, C. M., Garcia, L., Gichu, M., Jago, R., Katzmarzyk, P. T., & Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. British Journal of Sports Medicine, 54(24), 1451–1462. https://doi.org/10.1136/bjsports-2020-102955

Buratta, L., Reginato, E., Ranucci, C., Pippi, R., Aiello, C., Sbroma Tomaro, E., Perrone, C., Tirimagni, A., Russo, A., De Feo, P., & Mazzeschi, C. (2016). Stage of change and motivation to a healthier lifestyle before and after an intensive lifestyle intervention. Journal of Obesity, 2016, Article 6421265. https://doi.org/10.1155/2016/6421265

Burgess, E., Hassmén, P., & Pumpa, K. L. (2017). Determinants of adherence to lifestyle intervention in adults with obesity: A systematic review. Clinical Obesity, 7(3), 123–135. https://doi.org/10.1111/cob.12183

Buysse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research, 28(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Çaliş, H., Cüce, I., Polat, E., Hopean, S., Yaprak, E., Karabaş, Ç., Çelik, I., & Demir, F. G. Ü. (2023). An educational mobile health application for pulmonary rehabilitation in patients with mild to moderate COVID-19 pneumonia. In Handbook of research on instructional technologies in health education and allied

disciplines. IGI Global. https://doi.org/10.4018/978-1-6684-7164-7.ch010

Car, L. T., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y.-L., & Atun, R. (2020). Conversational agents in health care: Scoping review and conceptual analysis. Journal of Medical Internet Research, 22(8), Article e17158. https://doi.org/10.2196/17158

Carfora, V., Bertolotti, M., & Catellani, P. (2019). Informational and emotional daily messages to reduce red and processed meat consumption. Appetite, 141, Article 104331. https://doi.org/10.1016/j.appet.2019.104331

Carver, C. S., & Scheier, M. F. (1998). On the self-regulation of behavior. Cambridge University Press.

Casas, J., Mugellini, E., & Khaled, O. A. (2018). Food diary coaching chatbot. In UbiComp/ISWC 2018 Adjunct: Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers. https://doi.org/10.1145/3267305.3274191

Cava, E., Yeat, N. C., & Mittendorfer, B. (2017). Preserving healthy muscle during weight loss. Advances in Nutrition, 8(3), 511–519. https://doi.org/10.3945/an.116.014506

Cavero-Redondo, I., Martinez-Vizcaino, V., Fernandez-Rodriguez, R., Saz-Lara, A., Pascual-Morena, C., & Álvarez-Bueno, C. (2020). Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: A systematic review and meta-analysis. Nutrients, 12(7), Article 1977. https://doi.org/10.3390/nu12071977

Celis-Morales, C., Lara, J., & Mathers, J. C. (2015). Personalising nutritional guidance for more effective behaviour change. Proceedings of the Nutrition Society, 74(2). https://doi.org/10.1017/S0029665114001633

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (2018). Graphical methods for data analysis. Chapman and Hall/CRC. https://doi.org/10.1201/9781351072304

Chang, P.-L., Hsu, C.-W., & Chang, P.-C. (2011). Fuzzy Delphi method for evaluating hydrogen production technologies. International Journal of Hydrogen Energy, 36(21), 14172–14179. https://doi.org/10.1016/j.ijhydene.2011.05.045

Chang, P. T., Huang, L. C., & Lin, H. J. (2000). The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human

resources. Fuzzy Sets and Systems, 112(3), 511–520. https://doi.org/10.1016/S0165-0114(98)00067-0

Chaston, T. B., & Dixon, J. B. (2008). Factors associated with percent change in visceral versus subcutaneous abdominal fat during weight loss: Findings from a systematic review. International Journal of Obesity, 32(4), 619–628. https://doi.org/10.1038/sj.ijo.0803761

Chaston, T. B., Dixon, J. B., & O'Brien, P. E. (2007). Changes in fat-free mass during significant weight loss: A systematic review. International Journal of Obesity, 31(5). https://doi.org/10.1038/sj.ijo.0803483

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. Patterns, 6(6), Article 101260. https://doi.org/10.1016/j.patter.2025.101260

Cheng, C. H., & Lin, Y. (2002). Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation. European Journal of Operational Research, 142(1), 174–186. https://doi.org/10.1016/S0377-2217(01)00280-6

Chew, H. S. J. (2022). The use of artificial intelligence-based conversational agents (chatbots) for weight loss: Scoping review and practical recommendations. JMIR Medical Informatics, 10(4), Article e32578. https://doi.org/10.2196/32578

Chew, H. S. J., Ang, W. H. D., & Lau, Y. (2021). The potential of artificial intelligence in enhancing adult weight loss: A scoping review. Public Health Nutrition, 24(8). https://doi.org/10.1017/S1368980021000598

Choi, L., Ward, S. C., Schnelle, J. F., & Buchowski, M. S. (2012). Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Medicine & Science in Sports & Exercise, 44(10). https://doi.org/10.1249/MSS.0b013e318258cb36

Chu, H.-C., & Hwang, G.-J. (2008). A Delphi-based approach to developing expert systems with the cooperation of multiple experts. Expert Systems with Applications, 34(4), 2826–2840. https://doi.org/10.1016/j.eswa.2007.05.034

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART). American Journal of Preventive Medicine, 32(5 Suppl.), S112–S118. https://doi.org/10.1016/j.amepre.2007.01.022

Conroy, D. E., Hojjatinia, S., Lagoa, C. M., Yang, H., Lanza, S. T., & Smyth, J. M. (2019). Personalized models of physical activity responses to text message micro-interventions: A proof-of-concept application of control systems engineering methods. Psychology of Sport and Exercise, 41. https://doi.org/10.1016/j.psychsport.2018.06.011

Contreras, I., & Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: Literature review. Journal of Medical Internet Research, 20(5). https://doi.org/10.2196/10775

Corica, D., Aversa, T., Valenzise, M., Messina, M. F., Alibrandi, A., De Luca, F., & Wasniewska, M. (2018). Does family history of obesity, cardiovascular, and metabolic diseases influence onset and severity of childhood obesity? Frontiers in Endocrinology, 9, Article 187. https://doi.org/10.3389/fendo.2018.00187

Craig, H. C., Doran, Z. M., & le Roux, C. W. (2022). The concept of healthy behaviours in obesity may have unintended consequences. Nutrients, 15(1), Article 12. https://doi.org/10.3390/nu15010012

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical Activity Questionnaire: 12-country reliability and validity. Medicine & Science in Sports & Exercise, 35(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Dergaa, I., Fekih-Romdhane, F., Hallit, S., Loch, A. A., Glenn, J. M., Fessi, M. S., Ben Aissa, M., Souissi, N., Guelmami, N., Swed, S., El Omri, A., Bragazzi, N. L., & Ben Saad, H. (2024). ChatGPT is not ready yet for use in providing mental health assessment and interventions. Frontiers in Psychiatry, 14. https://doi.org/10.3389/fpsyt.2023.1277756

Dhurandhar, N. V., Schoeller, D., Brown, A. W., Heymsfield, S. B., Thomas, D., Sørensen, T. I. A., Speakman, J. R., Jeansonne, M., Allison, D. B., & Energy Balance Measurement Working Group. (2015). Energy balance measurement: When something is not better than nothing. International Journal of Obesity, 39(7), 1109–1113. https://doi.org/10.1038/ijo.2014.199

Di Bidino, R., Daugbjerg, S., Papavero, S., Haraldsen, I., Cicchetti, A., & Sacchini, D. (2024). Health technology assessment framework for artificial intelligence-based technologies. International Journal of Technology Assessment in Health Care, 40(1), Article e61. https://doi.org/10.1017/S0266462324000308

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. Journal of Clinical Epidemiology, 67(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Dobson, R., Whittaker, R., Jiang, Y., Maddison, R., Shepherd, M., McNamara, C., Cutfield, R., Khanolkar, M., & Murphy, R. (2018). Effectiveness of text message based, diabetes self management support programme (SMS4BG): Two arm, parallel randomised controlled trial. BMJ, 361, Article k1959. https://doi.org/10.1136/bmj.k1959

Donini, L. M., Busetto, L., Bischoff, S. C., Cederholm, T., Ballesteros-Pomar, M. D., Batsis, J. A., Bauer, J. M., Boirie, Y., Cruz-Jentoft, A. J., Dicker, D., Frara, S., Frühbeck, G., Genton, L., Gepner, Y., Giustina, A., Gonzalez, M. C., Han, H. S., Heymsfield, S. B., Higashiguchi, T., … Barazzoni, R. (2022). Definition and diagnostic criteria for sarcopenic obesity: ESPEN and EASO consensus statement. Obesity Facts, 15(3). https://doi.org/10.1159/000521241

Duffey, K. J., Gordon-Larsen, P., Jacobs, D. R., Jr., Williams, O. D., & Popkin, B. M. (2007). Differential associations of fast food and restaurant food consumption with 3-y change in body mass index: The Coronary Artery Risk Development in Young Adults Study. The American Journal of Clinical Nutrition, 85(1). https://doi.org/10.1093/ajcn/85.1.201

Dulloo, A. G., Jacquet, J., & Montani, J. P. (2012). How dieting makes some fatter: From a perspective of human body composition autoregulation. Proceedings of the Nutrition Society, 71(3). https://doi.org/10.1017/S0029665112000225

Efird, J. (2011). Blocked randomization with randomly selected block sizes. International Journal of Environmental Research and Public Health, 8(1), 15–20. https://doi.org/10.3390/ijerph8010015

Faccioli, N., Poitou, C., Clément, K., & Dubern, B. (2023). Current treatments for patients with genetic obesity. Journal of Clinical Research in Pediatric Endocrinology, 15(2). https://doi.org/10.4274/jcrpe.galenos.2023.2023-3-2

Fadhil, A., Wang, Y., & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. Methods of Information in Medicine, 58(1). https://doi.org/10.1055/s-0039-1688757

Farah, L., Borget, I., Martelli, N., & Vallee, A. (2024). Suitability of the current health

technology assessment of innovative artificial intelligence-based medical devices: Scoping literature review. Journal of Medical Internet Research, 26, Article e51514. https://doi.org/10.2196/51514

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Applied longitudinal analysis (2nd ed.). Wiley.

Fjeldsoe, B., Neuhaus, M., Winkler, E., & Eakin, E. (2011). Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychology, 30(1), 99–109. https://doi.org/10.1037/a0021974

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. In Lecture Notes in Computer Science (Vol. 11193). Springer. https://doi.org/10.1007/978-3-030-01437-7_16

Fukuoka, Y., Lindgren, T. G., Mintz, Y. D., Hooper, J., & Aswani, A. (2018). Applying natural language processing to understand motivational profiles for maintaining physical activity after a mobile app and accelerometer-based intervention: The mPED randomized controlled trial. JMIR mHealth and uHealth, 6(6), Article e10042. https://doi.org/10.2196/10042

Garcia, M. B. (2023). ChatGPT as a virtual dietitian: Exploring its potential as a tool for improving nutrition knowledge. Applied System Innovation, 6(5), Article 96. https://doi.org/10.3390/asi6050096

Garcia, M. B., & Garcia, P. (2023). Intelligent tutoring system as an instructional technology in learning basic nutrition concepts: An exploratory sequential mixed methods study. In Handbook of research on instructional technologies in health education and allied disciplines. IGI Global. https://doi.org/10.4018/978-1-6684-7164-7.ch012

Garcia, M. B., Mangaba, J. B., & Tanchoco, C. L. (2021). Virtual Dietitian: A nutrition knowledge-based system using forward chaining algorithm. In 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). IEEE. https://doi.org/10.1109/3ICT53449.2021.9581887

Garcia, M. B., Mangaba, J. B., & Vinluan, A. A. (2020). Towards the development of a personalized nutrition knowledge-based system: A mixed-methods needs analysis of Virtual Dietitian. International Journal of Scientific and Technology Research, 9(4).

Gattrell, W. T., Logullo, P., van Zuuren, E. J., Price, A., Hughes, E. L., Blazey, P., Winchester, C. C., Tovey, D., Goldman, K., Hungin, A. P., & Harrison, N. (2024). ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLOS Medicine, 21(1), Article e1004326. https://doi.org/10.1371/journal.pmed.1004326

Gemesi, K., Winkler, S., Schmidt-Tesch, S., Schederecker, F., Hauner, H., & Holzapfel, C. (2024). Efficacy of an app-based multimodal lifestyle intervention on body weight in persons with obesity: Results from a randomized controlled trial. International Journal of Obesity, 48(1), 118–126. https://doi.org/10.1038/s41366-023-01415-0

George, D., & Mallery, P. (2010). SPSS for Windows step by step: A simple guide and reference, 17.0 update (10th ed.). Pearson.

Gericke, F., Ebert, D. D., Breet, E., Auerbach, R. P., & Bantjes, J. (2021). A qualitative study of university students' experience of Internet-based CBT for depression. Counselling and Psychotherapy Research, 21(4). https://doi.org/10.1002/capr.12465

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism, 10(2), 486–489. https://doi.org/10.5812/ijem.3505

Gibbs, H., Broom, J., Brown, J., Laws, R., Reckless, J., Noble, P., Kumar, S., McCombie, L., Lean, M., Lyons, F., Frost, G., Quinn, M., Barth, J. H., Haynes, S., Finer, N., Ross, H., Hole, D., & Bray, C. A. (2005). The impact of obesity on drug prescribing in primary care. British Journal of General Practice, 55(519).

Gilbert, N., Gulati, M., Gluckman, T. J., Kittleson, M. M., Rikhi, R., Saseen, J. J., & Tchang, B. G. (2025). 2025 concise clinical guidance: An ACC expert consensus statement on medical weight management for optimization of cardiovascular health. Journal of the American College of Cardiology, 86(7). https://doi.org/10.1016/j.jacc.2025.05.024

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet

assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 42(3), 237–288. https://doi.org/10.3102/00346543042003237

Graham, S. A., Pitter, V., Hori, J. H., Stein, N., & Branch, O. H. (2022). Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. DIGITAL HEALTH, 8, Article 20552076221130619. https://doi.org/10.1177/20552076221130619

Guo, P., Liu, G., Xiang, X., & An, R. (2025). From AI to the table: A systematic review of ChatGPT's potential and performance in meal planning and dietary recommendations. Dietetics, 4(1), Article 7. https://doi.org/10.3390/dietetics4010007

Guttersrud, O., Dalane, J. O., & Pettersen, S. (2014). Improving measurement in nutrition literacy research using Rasch modelling: Examining construct validity of stage-specific "critical nutrition literacy" scales. Public Health Nutrition, 17(4). https://doi.org/10.1017/S1368980013000530

Hajduk, C., Roberts, S., & Saltzman, E. (2001). Dietary treatment of obesity. Current Opinion in Endocrinology and Diabetes, 8(5). https://doi.org/10.1097/00060793-200110000-00003

Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2021). Telemedicine for healthcare: Capabilities, features, barriers, and applications. Sensors International, 2, Article 100117. https://doi.org/10.1016/j.sintl.2021.100117

Hall, K. D., & Kahan, S. (2018). Maintenance of lost weight and long-term management of obesity. Medical Clinics of North America, 102(1), 183–197. https://doi.org/10.1016/j.mcna.2017.08.012

Han, M., & Rhee, S. Y. (2021). Effectiveness of mobile health applications for 5% body weight reduction in obese and overweight adults. Journal of Obesity & Metabolic Syndrome, 30(4). https://doi.org/10.7570/JOMES21014

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. International Journal of Behavioral Nutrition and Physical Activity, 16, Article 31. https://doi.org/10.1186/s12966-019-0792-7

Harlow, S. D., Gass, M., Hall, J. E., Lobo, R., Maki, P., Rebar, R. W., Sherman, S., Sluss, P. M., & de Villiers, T. J. (2012). Executive summary of the Stages of Reproductive Aging Workshop + 10: Addressing the unfinished agenda of staging

reproductive aging. The Journal of Clinical Endocrinology & Metabolism, 97(4), 1159–1168. https://doi.org/10.1210/jc.2011-3362

Hartmann-Boyce, J., Cobiac, L. J., Theodoulou, A., Oke, J. L., Butler, A. R., Scarborough, P., Bastounis, A., Dunnigan, A., Byadya, R., Hobbs, F. D. R., Sniehotta, F. F., Amies-Cull, B., Aveyard, P., & Jebb, S. A. (2023). Weight regain after behavioural weight management programmes and its impact on quality of life and cost effectiveness: Evidence synthesis and health economic analyses. Diabetes, Obesity and Metabolism, 25(2). https://doi.org/10.1111/dom.14895

Hauptmann, H., Leipold, N., Madenach, M., Wintergerst, M., Lurz, M., Groh, G., Böhm, M., Gedrich, K., & Krcmar, H. (2022). Effects and challenges of using a nutrition assistance system: Results of a long-term mixed-method study. User Modeling and User-Adapted Interaction, 32(5). https://doi.org/10.1007/s11257-021-09301-y

Havers, J., Jelonnek, C., Masur, L., Isenmann, E., Sperlich, B., Geisler, S., & Düking, P. (2025). A professional assessment of training plans for muscle hypertrophy and maximal strength developed by generative artificial intelligence. Biology of Sport, 42(4), 353–380. https://doi.org/10.5114/biolsport.2026.152350

Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. International Journal of Epidemiology, 24(2). https://doi.org/10.1093/ije/24.2.389

Hegde, N. L., Siddiqua, A., Shreyaswini, R., Kumar, S. D., Kumar, J. K., & Ravindra, S. L. (2023). Assessment of the need and effectiveness for nutritional counselling through Swasthya Poushan webportal for mothers of under five in Mysuru. International Journal of Community Medicine and Public Health, 10(3). https://doi.org/10.18203/2394-6040.ijcmph20230646

Herrmann, S. D., Willis, E. A., Ainsworth, B. E., Barreira, T. V., Hastert, M., Kracht, C. L., Schuna, J. M., Cai, Z., Quan, M., Tudor-Locke, C., Whitt-Glover, M. C., & Jacobs, D. R., Jr. (2024). 2024 Adult Compendium of Physical Activities: A third update of the energy costs of human activities. Journal of Sport and Health Science, 13(1), 6–12. https://doi.org/10.1016/j.jshs.2023.10.010

Heymsfield, S. B., Gonzalez, M. C., Shen, W., Redman, L., & Thomas, D. (2014). Weight loss composition is one-fourth fat-free mass: A critical review and critique of this widely cited rule. Obesity Reviews, 15(4).

https://doi.org/10.1111/obr.12143

Hicks, J., Adams-Campbell, L., Lu, J., Mills, M., & Dash, C. (2024). Validation study of physical activity and sedentary behavior in African-American men and women. Journal of Racial and Ethnic Health Disparities, 11(5). https://doi.org/10.1007/s40615-023-01734-y

Drabsch, T., & Holzapfel, C. (2019). A scientific perspective of personalised gene-based dietary recommendations for weight management. Nutrients, 11(3), Article 617. https://doi.org/10.3390/nu11030617

Horin, A., McNeely, M. E., Harrison, E. C., Myers, P. S., Sutter, E. N., Rawson, K. S., & Earhart, G. M. (2019). Usability of a daily mHealth application designed to address mobility, speech and dexterity in Parkinson's disease. Neurodegenerative Disease Management, 9(2). https://doi.org/10.2217/nmt-2018-0036

Howard, J. (2019). Artificial intelligence: Implications for the future of work. American Journal of Industrial Medicine, 62(11), 917–926. https://doi.org/10.1002/ajim.23037

Howell, D. C. (2013). Statistical methods for psychology (8th ed.). Wadsworth Cengage Learning.

Hsu, K.-J., Liao, C.-D., Tsai, M.-W., & Chen, C.-N. (2019). Effects of exercise and nutritional intervention on body composition, metabolic health, and physical performance in adults with sarcopenic obesity: A meta-analysis. Nutrients, 11(9), Article 2163. https://doi.org/10.3390/nu11092163

Huo, B., Collins, G. S., Chartash, D., Thirunavukarasu, A. J., Flanagin, A., Iorio, A., Cacciamani, G. E., Chen, X., Liu, N., Mathur, P., Chan, A.-W., Laine, C., Pacella, D., Berkwits, M., Antoniou, S. A., Camaradou, J. C., Canfield, C., Mittelman, M., Feeney, T., … Guyatt, G. (2025). Reporting guideline for chatbot health advice studies: The Chatbot Assessment Reporting Tool (CHART) statement. BMJ Medicine, 4(1). https://doi.org/10.1136/bmjmed-2025-001632

Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Daly, R., Frey, A.-L., & Leigh, S. (2023). Quality of digital health interventions across different health care domains: Secondary data analysis study. JMIR mHealth and uHealth, 11, Article e47043. https://doi.org/10.2196/47043

Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R., & Mieno, H. (1993). The max-min Delphi method and fuzzy Delphi method via fuzzy integration. Fuzzy Sets and Systems, 55(3), 241–253. https://doi.org/10.1016/0165-

0114(93)90251-C

Iyanna, S., Kaur, P., Ractham, P., Talwar, S., & Islam, A. K. M. N. (2022). Digital transformation of healthcare sector. What is impeding adoption and continued usage of technology-driven innovations by end-users? Journal of Business Research, 153. https://doi.org/10.1016/j.jbusres.2022.08.007

Jakicic, J. M., & Davis, K. K. (2011). Obesity and physical activity: Influence of physical activity on weight loss, physical activity with no prescribed reduction in energy intake. Psc, 34(4).

Jakicic, J. M., & Rogers, R. J. (2014). Physical activity as a weight management strategy. In Treatment of the obese patient. Springer. https://doi.org/10.1007/978-1-4939-2311-3_15

Jakubczyk, M., & Kamiński, B. (2017). Fuzzy approach to decision analysis with multiple criteria and uncertainty in health technology assessment. Annals of Operations Research, 251(1–2). https://doi.org/10.1007/s10479-015-1910-9

Jarchow, H., Bobrowski, C., Falk, S., Hermann, A., Kulaga, A., Põder, J.-C., Unfried, M., Usanov, N., Zendeh, B., Kennedy, B. K., Lobentanzer, S., & Fuellen, G. (2025). Benchmarking large language models for personalized, biomarker-based health intervention recommendations. npj Digital Medicine, 8, Article 631. https://doi.org/10.1038/s41746-025-01996-2

Jenkins, E. L., Ilicic, J., Barklamb, A. M., & McCaffrey, T. A. (2020). Assessing the credibility and authenticity of social media content for applications in health communication: Scoping review. Journal of Medical Internet Research, 22(7), Article e17296. https://doi.org/10.2196/17296

Jensen, M. D., Ryan, D. H., Apovian, C. M., Ard, J. D., Comuzzie, A. G., Donato, K. A., Hu, F. B., Hubbard, V. S., Jakicic, J. M., Kushner, R. F., Loria, C. M., Millen, B. E., Nonas, C. A., Pi-Sunyer, F. X., Stevens, J., Stevens, V. J., Wadden, T. A., Wolfe, B. M., & Yanovski, S. Z. (2014). 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation, 129(25 Suppl. 2). https://doi.org/10.1161/01.cir.0000437739.71477.ee

Johns, D. J., Hartmann-Boyce, J., Jebb, S. A., & Aveyard, P. (2014). Diet or exercise interventions vs combined behavioral weight management programs: A systematic review and meta-analysis of direct comparisons. Journal of the

Academy of Nutrition and Dietetics, 114(10), 1557–1568. https://doi.org/10.1016/j.jand.2014.07.005

Jones, H., & Twiss, B. C. (1978). Forecasting technology for planning decisions. Springer. https://doi.org/10.1007/978-1-349-03134-4

Kankanhalli, A., Shin, J., & Oh, H. (2019). Mobile-based interventions for dietary behavior change and health outcomes: Scoping review. JMIR mHealth and uHealth, 7(1). https://doi.org/10.2196/11312

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, 62(1). https://doi.org/10.1016/j.bushor.2018.08.004

Kaya Kaçar, H., Kaçar, Ö. F., & Avery, A. (2025). Diet quality and caloric accuracy in AI-generated diet plans: A comparative study across chatbots. Nutrients, 17(2), Article 206. https://doi.org/10.3390/nu17020206

Keaver, L., Richmond, J., Rafferty, F., & Douglas, P. (2023). Sources of nutrition advice and desired nutrition guidance in oncology care: Patient's perspectives. Journal of Human Nutrition and Dietetics, 36(2). https://doi.org/10.1111/jhn.13111

Keppel, G., & Wickens, T. D. (2004). Design and analysis: A researcher's handbook (4th ed.). Pearson Prentice Hall.

Khalafi, M., Sakhaei, M. H., Kazeminasab, F., Rosenkranz, S. K., & Symonds, M. E. (2023). Exercise training, dietary intervention, or combined interventions and their effects on lipid profiles in adults with overweight and obesity: A systematic review and meta-analysis of randomized clinical trials. Nutrition, Metabolism and Cardiovascular Diseases, 33(9). https://doi.org/10.1016/j.numecd.2023.05.024

Kim, Haam, J., Kim, B., Kim, E., Park, J., Rhee, S., Jeon, E., Kang, E., Nam, G. E., Koo, H. Y., Lim, J. H., Jeong, J. E., Kim, J. H., Kim, J. W., Park, J. H., Hong, J. H., Lee, S. E., Min, S. H., Kim, S. J., … Lee, C. B. (2023). Evaluation and treatment of obesity and its comorbidities: 2022 update of Clinical Practice Guidelines for Obesity by the Korean Society for the Study of Obesity. Journal of Obesity & Metabolic Syndrome, 32(1). https://doi.org/10.7570/jomes23016

Kirkpatrick, S. I., Baranowski, T., Subar, A. F., Tooze, J. A., & Frongillo, E. A. (2019). Best practices for conducting and interpreting studies to validate self-report dietary assessment methods. Journal of the Academy of Nutrition and Dietetics, 119(11). https://doi.org/10.1016/j.jand.2019.06.010

Klassen, K. M., Douglass, C. H., Brennan, L., Truby, H., & Lim, M. S. C. (2018). Social media use for nutrition outcomes in young adults: A mixed-methods systematic review. International Journal of Behavioral Nutrition and Physical Activity, 15, Article 70. https://doi.org/10.1186/s12966-018-0696-y

Klein, H., Zelicha, H., Yaskolka Meir, A., Rinott, E., Tsaban, G., Kaplan, A., Chassidim, Y., Gepner, Y., Blüher, M., Ceglarek, U., Isermann, B., Stumvoll, M., Shelef, I., Qi, L., Li, J., Hu, F. B., Stampfer, M. J., & Shai, I. (2025). Visceral adipose tissue area and proportion provide distinct reflections of cardiometabolic outcomes in weight loss; pooled analysis of MRI-assessed CENTRAL and DIRECT PLUS dietary randomized controlled trials. BMC Medicine, 23, Article 57. https://doi.org/10.1186/s12916-025-03891-9

Kline, C. E. (2014). The bidirectional relationship between exercise and sleep. American Journal of Lifestyle Medicine, 8(6), 375–379. https://doi.org/10.1177/1559827614544437

Kline, C. E., Burke, L. E., Sereika, S. M., Imes, C. C., Bonny, A. E., Mendez, D. D., Strollo, P. J., Zheng, Y., Rathbun, S. L., & Chasens, E. R. (2018). Bidirectional relationships between weight change and sleep apnea in a behavioral weight loss intervention. Mayo Clinic Proceedings, 93(9). https://doi.org/10.1016/j.mayocp.2018.04.026

Kocielnik, R., Xiao, L., Avrahami, D., & Hsieh, G. (2018). Reflection Companion: A conversational system for engaging users in reflection on physical activity. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(2).

Kokura, Y., Ueshima, J., Saino, Y., & Maeda, K. (2024). Enhanced protein intake on maintaining muscle mass, strength, and physical function in adults with overweight/obesity: A systematic review and meta-analysis. Clinical Nutrition ESPEN, 63. https://doi.org/10.1016/j.clnesp.2024.06.030

Kołłajtis, A., & Żamojcin, K. (2016). The level of knowledge on nutrition and its relation to health among Polish young men. Roczniki Państwowego Zakładu Higieny, 67(2).

Kramer, J. N., Künzler, F., Mishra, V., Smith, S. N., Kotz, D., Scholz, U., Fleisch, E., & Kowatsch, T. (2020). Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Annals of Behavioral Medicine, 54(7). https://doi.org/10.1093/abm/kaaa002

Künzler, F., Mishra, V., Kramer, J. N., Kotz, D., Fleisch, E., & Kowatsch, T. (2019). Exploring the state-of-receptivity for mHealth interventions. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(4). https://doi.org/10.1145/3369805

Kupila, S. K. E., Joki, A., Suojanen, L.-U., & Pietiläinen, K. H. (2023). The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: A systematic review of systematic reviews. Current Obesity Reports, 12(3), 371–394. https://doi.org/10.1007/s13679-023-00515-2

Lake, J. K., Power, C., & Cole, T. J. (1997). Child to adult body mass index in the 1958 British birth cohort: Associations with parental obesity. Archives of Disease in Childhood, 77(5). https://doi.org/10.1136/adc.77.5.376

Ledikwe, J. H., Blanck, H. M., Kettel Khan, L., Serdula, M. K., Seymour, J. D., Tohill, B. C., & Rolls, B. J. (2006). Dietary energy density is associated with energy intake and weight status in US adults. The American Journal of Clinical Nutrition, 83(6), 1362–1368. https://doi.org/10.1093/ajcn/83.6.1362

Lee, H., Hwang, J., Yon, D. K., & Rhee, S. Y. (2025). Multimodal and multidimensional artificial intelligence technology in obesity. Journal of Obesity & Metabolic Syndrome, 34(4), 394–404. https://doi.org/10.7570/jomes25035

Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. International Journal of Behavioral Nutrition and Physical Activity, 8, Article 115. https://doi.org/10.1186/1479-5868-8-115

Lemos, T., & Gallagher, D. (2017). Current body composition measurement techniques. Current Opinion in Endocrinology, Diabetes and Obesity, 24(5). https://doi.org/10.1097/MED.0000000000000360

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 49(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Li, B., Kumar, A., Finlay, C., van Drimmelen, M., Barnes, E., Southby, S., & Florkowski, C. (2023). Verification of point-of-care analysers for C-reactive protein, lipid studies and glycated haemoglobin. Pathology, 55(7), 989–999. https://doi.org/10.1016/j.pathol.2023.07.004

Liao, Y.-C., Chang, L.-C., & Lai, I.-J. (2024). Assessing the quality of ChatGPT's dietary advice for college students from dietitians' perspectives. Nutrients, 16(12), Article 1939. https://doi.org/10.3390/nu16121939

Liao, H., Li, M., Ma, C., Han, Y., Wang, D., Liu, H., Wang, Y., Feng, Z., Wang, H., & Guan, Y. (2026). Developing a quality evaluation index system for health conversational artificial intelligence: Mixed methods study. Journal of Medical Internet Research, 28, Article e83188. https://doi.org/10.2196/83188

Lin, G. S. S., Chin, Y. J., Chong, R. S., Baharin, F., Syed Saadun Tarek Wafa, S. W. W., & Dziaruddin, N. (2023). Expert consensus on relevant topics for undergraduate paediatric dental curriculum using the fuzzy Delphi method: A new direction for Malaysian dental education. BMC Oral Health, 23, Article 452. https://doi.org/10.1186/s12903-023-03130-8

Lin, P.-H., Wang, Y., Levine, E., Askew, S., Lin, S., Chang, C., Sun, J., Foley, P., Wang, H., Li, X., & Bennett, G. G. (2014). A text messaging-assisted randomized lifestyle weight loss clinical trial among overweight adults in Beijing. Obesity, 22(5). https://doi.org/10.1002/oby.20686

Lin, H., Zhang, L., Zheng, R., & Zheng, Y. (2017). The prevalence, metabolic risk and effects of lifestyle intervention for metabolically healthy obesity: A systematic review and meta-analysis. Medicine, 96(47). https://doi.org/10.1097/MD.0000000000008838

Livingstone, M. B. E., & Black, A. E. (2003). Markers of the validity of reported energy intake. The Journal of Nutrition, 133(3), 895S–920S. https://doi.org/10.1093/jn/133.3.895S

Lobstein, T., Jackson-Leach, R., Powis, J., Brinsden, H., & Gray, M. (2023). World obesity atlas 2023. World Obesity Federation. https://www.worldobesity.org/resources/resource-library/world-obesity-atlas-2023

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Lombardo, T., Duquesnoy, M., El-Bouysidy, H., Årén, F., Gallo-Bueno, A., Jørgensen, P. B., Bhowmik, A., Demortière, A., Ayerbe, E., Alcaide, F., Reynaud, M., Carrasco, J., Grimaud, A., Zhang, C., Vegge, T., Johansson, P., & Franco, A. A. (2022). Artificial intelligence applied to battery research: Hype or reality?

Chemical Reviews, 122(12). https://doi.org/10.1021/acs.chemrev.1c00108

López-Hernández, L., Martínez-Arnau, F. M., Pérez-Ros, P., Drehmer, E., & Pablos, A. (2020). Improved nutritional knowledge in the obese adult population modifies eating habits and serum and anthropometric markers. Nutrients, 12(11), Article 3355. https://doi.org/10.3390/nu12113355

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. Annual Review of Public Health, 23(1), 151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News, 40(3). https://doi.org/10.1108/LHTN-01-2023-0009

Maaliw, R. R., Susa, J. A. B., Alon, A. S., Lagman, A. C., Ambat, S. C., Garcia, M. B., Piad, K. C., & Fernando-Raguro, M. C. (2022). A deep learning approach for automatic scoliosis Cobb angle identification. In 2022 IEEE World AI IoT Congress (AIIoT 2022). IEEE. https://doi.org/10.1109/AIIoT54504.2022.9817290

Mabry-Hernandez, I., & Ojeda, L. (2019). Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults. American Family Physician, 99(8).

Machado, A. M., Guimarães, N. S., Bocardi, V. B., da Silva, T. P. R., Carmo, A. S. do, Menezes, M. C. de, & Duarte, C. K. (2022). Understanding weight regain after a nutritional weight loss intervention: Systematic review and meta-analysis. Clinical Nutrition ESPEN, 49. https://doi.org/10.1016/j.clnesp.2022.03.020

Magkos, F., Fraterrigo, G., Yoshino, J., Luecking, C., Kirbach, K., Kelly, S. C., de Las Fuentes, L., He, S., Okunade, A. L., Patterson, B. W., & Klein, S. (2016). Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metabolism, 23(4). https://doi.org/10.1016/j.cmet.2016.02.005

Maher, C. A., Davis, C. R., Curtis, R. G., Short, C. E., & Murphy, K. J. (2020). A physical activity and diet program delivered by artificially intelligent virtual health coach: Proof-of-concept study. JMIR mHealth and uHealth, 8(7), Article e17558. https://doi.org/10.2196/17558

Mainardi, F., Campos, V. C., Côté, R. G., Silber, N. K., Plestina, R., & Angeles-Agdeppa, I. (2023). Estimation of free sugars in the Filipino food composition

table and evaluation of population-level intake. Nutrients, 15(6), Article 1343. https://doi.org/10.3390/nu15061343

Maknickiene, N. (2022). Application of evolutionary artificial intelligence: An exploratory literature review. Applied Business: Issues & Solutions, 1. https://doi.org/10.57005/ab.2022.1.4

Mallinar, N., Heydari, A. A., Liu, X., Faranesh, A. Z., Winslow, B., Hammerquist, N., Graef, B., Speed, C., Malhotra, M., Patel, S., Prieto, J. L., McDuff, D., & Metwally, A. A. (2026). A scalable framework for evaluating health language models. arXiv. https://doi.org/10.48550/arXiv.2503.23339

Manakandan, S. K., Ismail, R., Jamil, M. R. M., & Ragunath, P. (2017). Pesticide applicators questionnaire content validation: A fuzzy Delphi method. Medical Journal of Malaysia, 72(4).

Markert, C., Sasangohar, F., Mortazavi, B. J., & Fields, S. (2021). The use of telehealth technology to support health coaching for older adults: Literature review. JMIR Human Factors, 8(1), Article e23796. https://doi.org/10.2196/23796

Mathioudakis, N., Lalani, B., Abusamaan, M. S., Alderfer, M., Alver, D., Dobs, A., Kane, B., McGready, J., Riekert, K., Ringham, B., Shehadeh, A., Vandi, F., Wanigatunga, A. A., Zade, D., Maruthur, N. M., Mathioudakis, N. N., Lalani, B., Abusamaan, M. S., Dobs, A., … Benjamin-Neelon, S. (2025). An AI-powered lifestyle intervention vs human coaching in the diabetes prevention program. JAMA, 334(23), 2079. https://doi.org/10.1001/jama.2025.19563

McMinn, D., Acharya, R., Rowe, D. A., Gray, S. R., & Allan, J. L. (2013). Measuring activity energy expenditure: Accuracy of the GT3X+ and Actiheart monitors. International Journal of Exercise Science, 6(3).

Merrigan, J. J., Stute, N. L., Eckerle, J. J., Mackowski, N., Walters, J., O'Connor, M., Barrett, K., Robert, R., Strang, A. J., & Hagen, J. A. (2022). Reliability and validity of contemporary bioelectrical impedance analysis devices for body composition assessment. Journal of Exercise and Nutrition, 5(4). https://doi.org/10.53520/jen2022.103133

Messner, E.-M., Terhorst, Y., Barke, A., Baumeister, H., Stoyanov, S., Hides, L., Kavanagh, D. J., Pryss, R., Sander, L., & Probst, T. (2020). The German version of the Mobile App Rating Scale (MARS-G): Development and validation study. JMIR mHealth and uHealth, 8(3), Article e14479. https://doi.org/10.2196/14479

Michaelides, A., Raby, C., Wood, M., Farr, K., & Toro-Ramos, T. (2016). Weight loss

efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. BMJ Open Diabetes Research & Care, 4(1). https://doi.org/10.1136/bmjdrc-2016-000264

Michie, S., Yardley, L., West, R., Patrick, K., & Greaves, F. (2017). Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop. Journal of Medical Internet Research, 19(6), Article e232. https://doi.org/10.2196/jmir.7126

Mikkelsen, B. E., Engesveen, K., Afflerbach, T., & Barnekow, V. (2016). The human rights framework, the school and healthier eating among young people: A European perspective. Public Health Nutrition, 19(1). https://doi.org/10.1017/S1368980015001627

Ministry of Health Malaysia. (2023). Malaysian clinical practice guidelines for the management of obesity (2nd ed.).

Mohd Razali, N., & Bee Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics, 2(1).

Mohd Yusoff, H., Heng, P. P., Hj Illias, M. R., Karrupayah, S., Fadhli, M. A., & Hod, R. (2024). A qualitative exploration and a fuzzy Delphi validation of high-risk scaffolding tasks and fatigue-related safety behavioural deviation among scaffolders. Heliyon, 10(15), Article e34599. https://doi.org/10.1016/j.heliyon.2024.e34599

Mohseni, M., Kuckuck, S., Meeusen, R. E. H., Jiskoot, G., Lengton, R., Savas, M., Berk, K. A. C., Van der Valk, E. S., Van der Voorn, B., Van den Berg, S. A. A., Iyer, A. M., Bussmann, J. B. J., Leenen, P. J. M., Dik, W. A., de Groot, C. J., Van den Akker, E. L. T., & Van Rossum, E. F. C. (2023). Improved physical and mental health after a combined lifestyle intervention with cognitive behavioural therapy for obesity. International Journal of Endocrinology and Metabolism, 21(1). https://doi.org/10.5812/ijem-129906

Morgan-Bathke, M., Baxter, S. D., Halliday, T. M., Lynch, A., Malik, N., Raynor, H. A., Garay, J. L., & Rozga, M. (2023). Weight management interventions provided by a dietitian for adults with overweight or obesity: An Evidence Analysis Center systematic review and meta-analysis. Journal of the Academy of Nutrition and Dietetics, 123(11). https://doi.org/10.1016/j.jand.2022.03.014

Morris, E., Jebb, S. A., Oke, J., Nickless, A., Ahern, A., Boyland, E., Caterson, I.,

Halford, J., Hauner, H., & Aveyard, P. (2021). Effect of weight loss on cardiometabolic risk: Observational analysis of two randomised controlled trials of community weight-loss programmes. British Journal of General Practice, 71(705). https://doi.org/10.3399/BJGP20X714113

Motulsky, H. J., & Brown, R. E. (2006). Detecting outliers when fitting data with nonlinear regression: A new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics, 7, Article 123. https://doi.org/10.1186/1471-2105-7-123

Muntner, P., Shimbo, D., Carey, R. M., Charleston, J. B., Gaillard, T., Misra, S., Myers, M. G., Ogedegbe, G., Schwartz, J. E., Townsend, R. R., Urbina, E. M., Viera, A. J., White, W. B., & Wright, J. T., Jr. (2019). Measurement of blood pressure in humans: A scientific statement from the American Heart Association. Hypertension, 73(5). https://doi.org/10.1161/HYP.0000000000000087

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. Health Technology Assessment, 2(3). https://doi.org/10.3310/hta2030

Murry, J. W., Jr., & Hammons, J. O. (1995). Delphi: A versatile methodology for conducting qualitative research. The Review of Higher Education, 18(4). https://doi.org/10.1353/rhe.1995.0008

Mustafa, N., Safii, N., Mohamad, M., Jamil, S., Rahman, A., Sani, N., Ahmad, A., Yatiman, N., Zakaria, A., Talib, R., & Koon, P. (2022). Sports nutritionists adopt the nutrition care process road map to develop individualized meal planning for athletes. Malaysian Journal of Movement, Health & Exercise, 11(2). https://doi.org/10.4103/mohe.mohe_27_22

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. World Journal of Methodology, 11(4), 116–129. https://doi.org/10.5662/wjm.v11.i4.116

Natalucci, V., Marmondi, F., Biraghi, M., & Bonato, M. (2023). The effectiveness of wearable devices in non-communicable diseases to manage physical activity and nutrition: Where we are? Nutrients, 15(4), Article 913. https://doi.org/10.3390/nu15040913

Institute for Public Health. (2023). National Health and Morbidity Survey (NHMS) 2023: Non-communicable diseases and healthcare demand. Ministry of Health

Malaysia.

Neter, J. E., Stam, B. E., Kok, F. J., Grobbee, D. E., & Geleijnse, J. M. (2003). Influence of weight reduction on blood pressure. Hypertension, 42(5), 878–884. https://doi.org/10.1161/01.HYP.0000094221.86888.AE

Nicklas, J. M., Huskey, K. W., Davis, R. B., & Wee, C. C. (2012). Successful weight loss among obese U.S. adults. American Journal of Preventive Medicine, 42(5). https://doi.org/10.1016/j.amepre.2012.01.005

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Niszczota, P., & Rybicka, I. (2023). The credibility of dietary advice formulated by ChatGPT: Robo-diets for people with food allergies. Nutrition, 112, Article 112076. https://doi.org/10.1016/j.nut.2023.112076

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. Advances in Health Sciences Education, 15(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

Oh, Y. J., Zhang, J., Fang, M.-L., & Fukuoka, Y. (2021). A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. International Journal of Behavioral Nutrition and Physical Activity, 18. https://doi.org/10.1186/s12966-021-01224-6

Omar, M., Sorin, V., Collins, J. D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N. L., Nadkarni, G. N., & Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. Communications Medicine, 5, Article 330. https://doi.org/10.1038/s43856-025-01021-3

Ong, J. C. L., Chen, M. H., Ng, N., Elangovan, K., Tan, N. Y. T., Jin, L., Xie, Q., Ting, D. S. W., Rodriguez-Monguio, R., Bates, D. W., & Liu, N. (2025). A scoping review on generative AI and large language models in mitigating medication related harm. npj Digital Medicine, 8, Article 182. https://doi.org/10.1038/s41746-025-01565-7

O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and

ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. International Journal of Medical Robotics and Computer Assisted Surgery, 15(1). https://doi.org/10.1002/rcs.1968

Otten, J. J., Jones, K. E., Littenberg, B., & Harvey-Berino, J. (2009). Effects of television viewing reduction on energy intake and expenditure in overweight and obese adults: A randomized controlled trial. Archives of Internal Medicine, 169(22). https://doi.org/10.1001/archinternmed.2009.430

Ozlu Karahan, T., & Kenger, E. B. (2025). ChatGPT-4o for weight management: Comparison of different diet models. Food Science & Nutrition, 13(7). https://doi.org/10.1002/fsn3.70639

Pantoja, T., Opiyo, N., Lewin, S., Paulsen, E., Ciapponi, A., Wiysonge, C. S., Herrera, C. A., Rada, G., Peñaloza, B., Dudley, L., Gagnon, M. P., Marti, S. G., & Oxman, A. D. (2017). Implementation strategies for health systems in low-income countries: An overview of systematic reviews. Cochrane Database of Systematic Reviews, 2017(9). https://doi.org/10.1002/14651858.CD011086.pub2

Parry, S., Piché, M., Cornfield, T., Dyson, P., Francis, J., Mozes, F., & Hodson, L. (2018). The influence of dietary fat and free sugars on liver fat content and metabolism. Proceedings of the Nutrition Society, 77(OCE4). https://doi.org/10.1017/S0029665118001234

Pepe, R. B., Lottenberg, A. M., Fujiwara, C. T. H., Beyruti, M., Cintra, D. E., Machado, R. M., Rodrigues, A., Jensen, N. S. O., Caldas, A. P. S., Fernandes, A. E., Rossoni, C., Mattos, F., Motarelli, J. H. F., Bressan, J., Saldanha, J., Beda, L. M. M., Lavrador, M. S. F., Del Bosco, M., Cruz, P., … Piovacari, S. M. F. (2023). Position statement on nutrition therapy for overweight and obesity: Nutrition department of the Brazilian association for the study of obesity and metabolic syndrome (ABESO—2022). Diabetology and Metabolic Syndrome, 15(1). https://doi.org/10.1186/s13098-023-01037-6

NCD Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. The Lancet, 403(10431), 1027–1050. https://doi.org/10.1016/S0140-6736(23)02750-2

Piao, M., Ryu, H., Lee, H., & Kim, J. (2020). Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: Exploratory

randomized controlled trial. JMIR mHealth and uHealth, 8(5), Article e15085. https://doi.org/10.2196/15085

Pidd, K., Breeze, P., Ahern, A., Griffin, S., & Brennan, A. (2025). Effects of weight loss and weight gain on HbA1c, systolic blood pressure and total cholesterol in three subgroups defined by blood glucose: A pooled analysis of two behavioural weight management trials in England. BMJ Open, 15(4). https://doi.org/10.1136/bmjopen-2024-095046

Piette, J. D., List, J., Rana, G. K., Townsend, W., Striplin, D., & Heisler, M. (2015). Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation, 132(21). https://doi.org/10.1161/CIRCULATIONAHA.114.008723

Pinho, C. P. S., Diniz, A. da S., Arruda, I. K. G. de, Leite, A. P. D. L., & Rodrigues, I. G. (2018). Effects of weight loss on adipose visceral and subcutaneous tissue in overweight adults. Clinical Nutrition, 37(4). https://doi.org/10.1016/j.clnu.2017.05.011

Ponzo, V., Goitre, I., Favaro, E., Merlo, F. D., Mancino, M. V., Riso, S., & Bo, S. (2024). Is ChatGPT an effective tool for providing dietary advice? Nutrients, 16(4), Article 469. https://doi.org/10.3390/nu16040469

Poslusna, K., Ruprich, J., de Vries, J. H. M., Jakubikova, M., & Van't Veer, P. (2009). Misreporting of energy and micronutrient intake estimated by food records and 24-hour recalls, control and adjustment methods in practice. British Journal of Nutrition, 101(Suppl. 2). https://doi.org/10.1017/S0007114509990602

Powell, K. E., King, A. C., Buchner, D. M., Campbell, W. W., DiPietro, L., Erickson, K. I., Hillman, C. H., Jakicic, J. M., Janz, K. F., Katzmarzyk, P. T., Kraus, W. E., Macko, R. F., Marquez, D. X., McTiernan, A., Pate, R. R., Pescatello, L. S., & Whitt-Glover, M. C. (2019). The scientific foundation for the physical activity guidelines for Americans, 2nd edition. Journal of Physical Activity and Health, 16(1). https://doi.org/10.1123/jpah.2018-0618

Prentice, R. L., Mossavar-Rahmani, Y., Huang, Y., Van Horn, L., Beresford, S. A., Caan, B., Tinker, L. F., Schoeller, D., Bingham, S., Eaton, C. B., Thomson, C., Johnson, K. C., Ockene, J., Sarto, G., Heiss, G., & Neuhouser, M. L. (2011). Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. American Journal of Epidemiology, 174(5). https://doi.org/10.1093/aje/kwr140

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. International Journal of Behavioral Nutrition and Physical Activity, 5, Article 56. https://doi.org/10.1186/1479-5868-5-56

Quaidoo, E. Y., Ohemeng, A., & Amankwah-Poku, M. (2018). Sources of nutrition information and level of nutrition knowledge among young adults in the Accra metropolis. BMC Public Health, 18(1). https://doi.org/10.1186/s12889-018-6159-1

Robles, N., Puigdomènech Puig, E., Gómez-Calderón, C., Saigí-Rubió, F., Cuatrecasas Cambra, G., Zamora, A., Moharra, M., Paluzié, G., Balfegó, M., & Carrion, C. (2020). Evaluation criteria for weight management apps: Validation using a modified Delphi process. JMIR mHealth and uHealth, 8(7), Article e16899. https://doi.org/10.2196/16899

Roldán López, A. F., Sánchez, M., Puente-Fernández, D., Montoya-Juárez, R., & Roldán, C. (2021). A fuzzy Delphi consensus methodology based on fuzzy rankings. Mathematics, 9(18), Article 2323. https://doi.org/10.3390/math9182323

Rolls, B. J. (2009). The relationship between dietary energy density and energy intake. Physiology & Behavior, 97(5). https://doi.org/10.1016/j.physbeh.2009.03.011

Ronquillo, C. E., Peltonen, L.-M., Pruinelli, L., Chu, C. H., Bakken, S., Beduschi, A., Cato, K., Hardiker, N., Junger, A., Michalowski, M., Nyrup, R., Rahimi, S., Reed, D. N., Salakoski, T., Salanterä, S., Walton, N., Weber, P., Wiegand, T., & Topaz, M. (2021). Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. Journal of Advanced Nursing, 77(9), 3707–3717. https://doi.org/10.1111/jan.14855

Ross, R., Neeland, I. J., Yamashita, S., Shai, I., Seidell, J., Magni, P., Santos, R. D., Arsenault, B., Cuevas, A., Hu, F. B., Griffin, B. A., Zambon, A., Barter, P., Fruchart, J. C., Eckel, R. H., Matsuzawa, Y., & Després, J.-P. (2020). Waist circumference as a vital sign in clinical practice: A consensus statement from the IAS and ICCR Working Group on Visceral Obesity. Nature Reviews Endocrinology, 16(3). https://doi.org/10.1038/s41574-019-0310-7

Rusali, R., Manaf, Z. A., Shahar, S., Mazri, F. H., Ibrahim, N., Ludin, A. F. M., Singh,

D. K. A., & Ali, N. M. (2018). Comparison of the effectiveness of online and face-to-face weight-loss interventions in the workplace: Evidence from Malaysia. Sains Malaysiana, 47(10). https://doi.org/10.17576/jsm-2018-4710-20

WHO Expert Consultation. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. The Lancet, 363(9403), 157–163. https://doi.org/10.1016/S0140-6736(03)15268-3

Ryan, D. H., & Yockey, S. R. (2017). Weight loss and improvement in comorbidity: Differences at 5%, 10%, 15%, and over. Current Obesity Reports, 6(2), 187–194. https://doi.org/10.1007/s13679-017-0262-y

Salinari, A., Machì, M., Armas Diaz, Y., Cianciosi, D., Qi, Z., Yang, B., Ferreiro Cotorruelo, M. S., Villar, S. G., Dzul Lopez, L. A., Battino, M., & Giampieri, F. (2023). The application of digital technologies and artificial intelligence in healthcare: An overview on nutrition assessment. Diseases, 11(3), Article 97. https://doi.org/10.3390/diseases11030097

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare, 11(6), Article 887. https://doi.org/10.3390/healthcare11060887

Schachner, T., Keller, R., & von Wangenheim, F. (2020). Artificial intelligence-based conversational agents for chronic conditions: Systematic literature review. Journal of Medical Internet Research, 22(9). https://doi.org/10.2196/20701

Schifano, J., & Niederberger, M. (2025). How Delphi studies in the health sciences find consensus: A scoping review. Systematic Reviews, 14, Article 14. https://doi.org/10.1186/s13643-024-02738-3

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Methodology, 6(4), 147–151. https://doi.org/10.1027/1614-2241/a000016

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., … Resnik, P. (2025). The prompt report: A systematic survey of prompt engineering techniques. arXiv. https://doi.org/10.48550/arXiv.2406.06608

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ, 340, Article c332. https://doi.org/10.1136/bmj.c332

Selvaraju, J. R., & Nik, N. H. (2024). Determining the new norm elements in emergency departments in Malaysia during a pandemic: A fuzzy Delphi method. Malaysian Journal of Medical Sciences, 31(5). https://doi.org/10.21315/mjms2024.31.5.17

Sheikh, A., Anderson, M., Albala, S., Casadei, B., Franklin, B. D., Richards, M., Taylor, D., Tibble, H., & Mossialos, E. (2021). Health information technology and digital innovation for national learning health and care systems. The Lancet Digital Health, 3(6). https://doi.org/10.1016/S2589-7500(21)00005-4

Sherrington, A., Newham, J. J., Bell, R., Adamson, A., McColl, E., & Araújo-Soares, V. (2016). Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obesity Reviews, 17(6). https://doi.org/10.1111/obr.12396

Shi, Y., Ren, P., Wang, J., Han, B., ValizadehAslani, T., Agbavor, F., Zhang, Y., Hu, M., Zhao, L., & Liang, H. (2023). Leveraging GPT-4 for food effect summarization to enhance product-specific guidance development via iterative prompting. Journal of Biomedical Informatics, 148, Article 104533. https://doi.org/10.1016/j.jbi.2023.104533

Shiferaw, K. B., Roloff, M., Balaur, I., Welter, D., Waltemath, D., & Zeleke, A. A. (2025). Guidelines and standard frameworks for artificial intelligence in medicine: A systematic review. JAMIA Open, 8(1), Article ooae155. https://doi.org/10.1093/jamiaopen/ooae155

Silva, D. A. S., Petroski, E. L., & Pelegrini, A. (2014). Effects of aerobic exercise on the body composition and lipid profile of overweight adolescents. Revista Brasileira de Ciências do Esporte, 36(2). https://doi.org/10.1590/S0101-32892014000200002

Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., Watson, A., Szeto, K., Eglitis, E., Miatke, A., Simpson, C. E. M., Vandelanotte, C., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. npj Digital Medicine, 6, Article 118. https://doi.org/10.1038/s41746-023-00856-1

Singh, K., Siddhanta, A., & Goswami, L. (2021). Improving tuberculosis treatment success rate through nutrition supplements and counselling: Findings from a pilot intervention in India. Clinical Epidemiology and Global Health, 11, Article 100782. https://doi.org/10.1016/j.cegh.2021.100782

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N.,

Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., … Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

Sivarajkumar, S., Kelley, M., Samolyk-Mazzanti, A., Visweswaran, S., & Wang, Y. (2024). An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: Algorithm development and validation study. JMIR Medical Informatics, 12, Article e55318. https://doi.org/10.2196/55318

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. Journal of Information Technology Education: Research, 6. https://doi.org/10.28945/199

Smart, N. A., Downes, D., van der Touw, T., Hada, S., Dieberg, G., Pearson, M. J., Wolden, M., King, N., & Goodman, S. P. J. (2025). The effect of exercise training on blood lipids: A systematic review and meta-analysis. Sports Medicine, 55, 67–78. https://doi.org/10.1007/s40279-024-02115-z

Solanki, K., Rajawat, S., Gadekar, R., & Patil, M. E. (2023). Building a conversational chatbot using machine learning: Towards a more intelligent healthcare application. In Handbook of research on instructional technologies in health education and allied disciplines. IGI Global. https://doi.org/10.4018/978-1-6684-7164-7.ch013

Sorgente, A., Pietrabissa, G., Manzoni, G. M., Re, F., Simpson, S., Perona, S., Rossi, A., Cattivelli, R., Innamorati, M., Jackson, J. B., & Castelnuovo, G. (2017). Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: A systematic review of systematic reviews. Journal of Medical Internet Research, 19(6). https://doi.org/10.2196/jmir.6972

Spronk, I., Kullen, C., Burdon, C., & O'Connor, H. (2014). Relationship between nutrition knowledge and dietary intake. British Journal of Nutrition, 111(10), 1713–1726. https://doi.org/10.1017/S0007114514000087

Stephens, T. N., Joerin, A., Rauws, M., & Werk, L. N. (2019). Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Translational Behavioral Medicine, 9(3). https://doi.org/10.1093/tbm/ibz043

Stewart-Knox, B., Markovina, J., Rankin, A., Bunting, B., Kuznesof, S., Fischer, A. R. H., van der Lans, I. A., Poínhos, R., de Almeida, M. D. V., Panzone, L., Gibney, M., & Frewer, L. J. (2016). Making personalised nutrition the easy choice: Creating policies to break down the barriers and reap the benefits. Food Policy, 63. https://doi.org/10.1016/j.foodpol.2016.08.001

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile App Rating Scale: A new tool for assessing the quality of health mobile apps. JMIR mHealth and uHealth, 3(1), Article e27. https://doi.org/10.2196/mhealth.3422

Suenghataiphorn, T., Tribuddharat, N., Danpanichkul, P., & Kulthamrongsri, N. (2025). Large language models in obesity: A systematic review. International Journal of Obesity. https://doi.org/10.1038/s41366-025-01992-2

Sulaiman, H., Ismail, R., Mohd Yusoff, H., Anuar, N., Mohd Jamil, M. R., & Daud, F. (2020). Validation of occupational zoonotic disease questionnaire using fuzzy Delphi method. Journal of Agromedicine, 25(2), 166–172. https://doi.org/10.1080/1059924X.2019.1666763

Dan, A. S. M., Linoby, A., Kasim, S. S., Zaki, S., Sazali, R., Yusoff, Y., Nasir, Z., & Abidin, A. H. (2025). Validation of a personalized AI prompt generator (NExGEN-ChatGPT) for obesity management using fuzzy Delphi method. Biology Methods and Protocols, 10(1), Article bpaf085. https://doi.org/10.1093/biomethods/bpaf085

Tabachnick, B. G., & Fidell, L. S. (2019). Using multivariate statistics (7th ed.). Pearson.

Tam, T. Y. C., Sivarajkumar, S., Kapoor, S., Stolyar, A. V., Polanska, K., McCarthy, K. R., Osterhoudt, H., Wu, X., Visweswaran, S., Fu, S., Mathur, P., Cacciamani, G. E., Sun, C., Peng, Y., & Wang, Y. (2024). A framework for human evaluation of large language models in healthcare derived from literature review. npj Digital Medicine, 7, Article 258. https://doi.org/10.1038/s41746-024-01258-7

Tang, X. (2020). The role of artificial intelligence in medical imaging research. BJR|Open, 2(1). https://doi.org/10.1259/bjro.20190031

Tavares, D., Lopes, A. I., Castro, C., Maia, G., Leite, L., & Quintas, M. (2023). The intersection of artificial intelligence, telemedicine, and neurophysiology: Opportunities and challenges. In Handbook of research on instructional technologies in health education and allied disciplines. IGI Global.

https://doi.org/10.4018/978-1-6684-7164-7.ch006

Tay, J. E. F., Tung, S. E. H., Kaur, S., Gan, W. Y., Che'Ya, N. N., & Tan, C. H. (2023). Seasonality, food security, diet quality and nutritional status in urban poor adolescents in Malaysia. Scientific Reports, 13, Article 15067. https://doi.org/10.1038/s41598-023-42394-6

Teh, W. L., Abdin, E., P. V., A., Siva Kumar, F. D., Roystonn, K., Wang, P., Shafie, S., Chang, S., Jeyagurunathan, A., Vaingankar, J. A., Sum, C. F., Lee, E. S., van Dam, R. M., & Subramaniam, M. (2023). Measuring social desirability bias in a multi-ethnic cohort sample: Its relationship with self-reported physical activity, dietary habits, and factor structure. BMC Public Health, 23, Article 415. https://doi.org/10.1186/s12889-023-15309-3

Telle-Hansen, V. H., Christensen, J. J., Formo, G. A., Holven, K. B., & Ulven, S. M. (2020). A comprehensive metabolic profiling of the metabolically healthy obesity phenotype. Lipids in Health and Disease, 19(1). https://doi.org/10.1186/s12944-020-01273-z

Telle-Hjellset, V., Kjøllesdal, M. K. R., Bjørge, B., Holmboe-Ottesen, G., Wandel, M., Birkeland, K. I., Eriksen, H. R., & Høstmark, A. T. (2013). The InnvaDiab-DE-PLAN study: A randomised controlled trial with a culturally adapted education programme improved the risk profile for type 2 diabetes in Pakistani immigrant women. British Journal of Nutrition, 109(3), 529–538. https://doi.org/10.1017/S000711451200133X

Terhorst, Y., Philippi, P., Sander, L. B., Schultchen, D., Paganini, S., Bardus, M., Santo, K., Knitza, J., Machado, G. C., Schoeppe, S., Bauereiß, N., Portenhauser, A., Domhardt, M., Walter, B., Krusche, M., Baumeister, H., & Messner, E.-M. (2020). Validation of the Mobile Application Rating Scale (MARS). PLOS ONE, 15(11), Article e0241480. https://doi.org/10.1371/journal.pone.0241480

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. BMC Medical Research Methodology, 10, Article 1. https://doi.org/10.1186/1471-2288-10-1

Tham, K. W., Abdul Ghani, R., Cua, S. C., Deerochanawong, C., Fojas, M., Hocking, S., Lee, J., Nam, T. Q., Pathan, F., Saboo, B., Soegondo, S., Somasundaram, N., Yong, A. M. L., Ashkenas, J., Webster, N., & Oldfield, B. (2023). Obesity in South and Southeast Asia-A new consensus on care and management. Obesity

Reviews, 24(2). https://doi.org/10.1111/obr.13520

Thomaidis, N. S., Nikitakos, N., & Dounias, G. D. (2006). The evaluation of information technology projects: A fuzzy multicriteria decision-making approach. International Journal of Information Technology & Decision Making, 5(1). https://doi.org/10.1142/S0219622006001897

Thompson, F. E., & Subar, A. F. (2017). Dietary assessment methodology. In Nutrition in the prevention and treatment of disease (pp. 5–48). Elsevier. https://doi.org/10.1016/B978-0-12-802928-2.00001-1

Tian, Y. (2020). Artificial intelligence image recognition method based on convolutional neural network algorithm. IEEE Access, 8. https://doi.org/10.1109/ACCESS.2020.3006097

Tokgöz, P., Albrecht, J., & Dockweiler, C. (2025). Implementation of artificial intelligence-based decision support systems for antibiotic prescribing in hospitals: A Delphi study. Frontiers in Digital Health, 7, Article 1555042. https://doi.org/10.3389/fdgth.2025.1555042

Tomiyama, A. J. (2019). Stress and obesity. Annual Review of Psychology, 70(1), 703–718. https://doi.org/10.1146/annurev-psych-010418-102936

Tonkin, E., Brimblecombe, J., & Wycherley, T. P. (2017). Characteristics of smartphone applications for nutrition improvement in community settings: A scoping review. Advances in Nutrition, 8(2). https://doi.org/10.3945/an.116.013748

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. Medicine & Science in Sports & Exercise, 40(1). https://doi.org/10.1249/mss.0b013e31815a51b3

Tsai, A. G., & Wadden, T. A. (2005). Systematic review: An evaluation of major commercial weight loss programs in the United States. Annals of Internal Medicine, 142(1), 56–66. https://doi.org/10.7326/0003-4819-142-1-200501040-00012

Tu, C., & Benn, E. K. T. (2017). RRApp, a robust randomization app, for clinical and translational research. Journal of Clinical and Translational Science, 1(6), 323–327. https://doi.org/10.1017/cts.2017.310

Tustumi, F., Andreollo, N. A., & de Aguilar-Nascimento, J. E. (2023). Future of the language models in healthcare: The role of ChatGPT. Arquivos Brasileiros de

Cirurgia Digestiva, 36, Article e1727. https://doi.org/10.1590/0102-672020230002e1727

Ufholz, K., & Werner, J. (2023). The efficacy of mobile applications for weight loss. Current Cardiovascular Risk Reports, 17(4). https://doi.org/10.1007/s12170-023-00717-2

Ulker, I., & Ayyildiz, F. (2021). Artificial intelligence applications in nutrition and dietetics. Journal of Intelligent Systems with Applications. https://doi.org/10.54856/jiswa.202112175

Vaira, L. A., Lechien, J. R., Abbate, V., Allevi, F., Audino, G., Beltramini, G. A., Bergonzani, M., Boscolo-Rizzo, P., Califano, G., Cammaroto, G., Chiesa-Estomba, C. M., Committeri, U., Crimi, S., Curran, N. R., di Bello, F., di Stadio, A., Frosolini, A., Gabriele, G., Gengler, I. M., … De Riu, G. (2024). Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: A new tool to assess the quality of health information provided by AI platforms. European Archives of Oto-Rhino-Laryngology, 281(11), 6123–6131. https://doi.org/10.1007/s00405-024-08710-0

Verheggen, R. J. H. M., Maessen, M. F. H., Green, D. J., Hermus, A. R. M. M., Hopman, M. T. E., & Thijssen, D. H. T. (2016). A systematic review and meta-analysis on the effects of exercise training versus hypocaloric diet: Distinct effects on body weight and visceral adipose tissue. Obesity Reviews, 17(8). https://doi.org/10.1111/obr.12406

Vilar, M., Burrola-Méndez, S., Lozano-Marrufo, A., Ferré-Eguiluz, I., Flores, D., Gaitán-Rossi, P., Teruel, G., & Pérez-Escamilla, R. (2021). Urban poverty and nutrition challenges associated with accessibility to a healthy diet: A global systematic literature review. International Journal for Equity in Health, 20(1). https://doi.org/10.1186/s12939-020-01330-0

Vrkatić, A., Grujičić, M., Jovičić-Bata, J., & Novaković, B. (2022). Nutritional knowledge, confidence, attitudes towards nutritional care and nutrition counselling practice among general practitioners. Healthcare, 10(11), Article 2222. https://doi.org/10.3390/healthcare10112222

Wadden, T. A., Webb, V. L., Moran, C. H., & Bailer, B. A. (2012). Lifestyle modification for obesity: New developments in diet, physical activity, and behavior therapy. Circulation, 125(9). https://doi.org/10.1161/CIRCULATIONAHA.111.039453

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. arXiv. https://arxiv.org/abs/2201.11903

Wharton, S., Lau, D. C. W., Vallis, M., Sharma, A. M., Biertho, L., Campbell-Scherer, D., Adamo, K., Alberga, A., Bell, R., Boulé, N., Boyling, E., Brown, J., Calam, B., Clarke, C., Crowshoe, L., Divalentino, D., Forhan, M., Freedhoff, Y., Gagner, M., … Wicklum, S. (2020). Obesity in adults: A clinical practice guideline. Canadian Medical Association Journal, 192(31), E875–E891. https://doi.org/10.1503/cmaj.191707

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv. https://arxiv.org/abs/2302.11382

Widmer, R. J., Flammer, A. J., Lerman, L. O., & Lerman, A. (2015). The Mediterranean diet, its components, and cardiovascular disease. The American Journal of Medicine, 128(3). https://doi.org/10.1016/j.amjmed.2014.10.014

Willmott, T., Pang, B., Rundle-Thiele, S., & Badejo, A. (2019). Weight management in young adults: Systematic review of electronic health intervention components and outcomes. Journal of Medical Internet Research, 21(2), Article e10265. https://doi.org/10.2196/10265

Wing, R. R., Bahnson, J. L., Bray, G. A., Clark, J. M., Coday, M., Egan, C., Espeland, M. A., Foreyt, J. P., Gregg, E. W., Goldman, V., Haffner, S. M., Hazuda, H., Hill, J. O., Horton, E. S., Hubbard, V. S., Jakicic, J., Jeffery, R. W., Johnson, K. C., Kahn, S., … Yanovski, S. Z. (2010). Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus. Archives of Internal Medicine, 170(17). https://doi.org/10.1001/archinternmed.2010.334

Wing, R. R., Lang, W., Wadden, T. A., Safford, M., Knowler, W. C., Bertoni, A. G., Hill, J. O., Brancati, F. L., Peters, A., & Wagenknecht, L. (2011). Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care, 34(7). https://doi.org/10.2337/dc10-2415

Wong, I. A., Huang, J., Lin, Z. (C. J.), & Jiao, H. (2022). Smart dining, smart restaurant, and smart service quality (SSQ). International Journal of Contemporary Hospitality Management, 34(6). https://doi.org/10.1108/IJCHM-10-2021-1207

World Health Organization. (2025). Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Worsley, A. (2002). Nutrition knowledge and food consumption: Can nutrition knowledge change food behaviour? Asia Pacific Journal of Clinical Nutrition, 11(Suppl. 3). https://doi.org/10.1046/j.1440-6047.11.supp3.7.x

Yahia, N., Wang, D., Rapley, M., & Dey, R. (2016). Assessment of weight status, dietary habits and beliefs, physical activity, and nutritional knowledge among university students. Perspectives in Public Health, 136(4). https://doi.org/10.1177/1757913915609945

Yakan, S. A. (2022). Analysis of development of artificial intelligence in the game industry. International Journal of Cyber and IT Service Management, 2(2). https://doi.org/10.34306/ijcitsm.v2i2.100

Yin, T. T., & Hanif, H. (2024). Fuzzy Delphi method: A step-by-step guide to obtain expert consensus on MUETBot functionalities. International Journal of Academic Research in Business and Social Sciences, 14(4). https://doi.org/10.6007/IJARBSS/v14-i4/21307

Zaghir, J., Naguib, M., Bjelogrlic, M., Névéol, A., Tannier, X., & Lovis, C. (2024). Prompt engineering paradigms for medical applications: Scoping review. Journal of Medical Internet Research, 26. https://doi.org/10.2196/60501

Zaleski, A. L., Berkowsky, R. W., Craig, K. J. T., & Pescatello, L. S. (2024). Comprehensiveness, accuracy, and readability of exercise recommendations provided by an AI-based chatbot: Mixed methods study. JMIR Medical Education, 10(1), Article e51308. https://doi.org/10.2196/51308

Zhang, Y., & Lam, J. S. L. (2019). A fuzzy Delphi-AHP-TOPSIS framework to identify barriers in big data analytics adoption: Case of maritime organizations. Maritime Policy & Management, 46(7). https://doi.org/10.1080/03088839.2019.1628318

APPENDICES

APPENDIX 1

Ethics Approval Letter

www.uitm.edu.my

UNIVERSITI TEKNOLOGI MARA | Pejabat Timbalan Naib Canselor (Penyelidikan dan Inovasi)

Reference : 600-TNCPI (5/1/6)
Our reference : REC/09/2024 (PG/FB/30)
Date : 1445H 20 ربيع الأول
24 September 2024

Ms Azwa Suraya Mohd Dan - 2023653718
(Supervisor: Adam Feizrel Linoby Ronny Linoby)
Faculty of Sports Science and Recreation
UiTM Negeri Sembilan Branch
Seremban Campus
Persiaran Seremban Tiga/1, Seremban 3
70300 Seremban
NEGERI SEMBILAN

سلام عليكم ورحمة الله وبركاته and Greetings

Mr

APPROVAL LETTER - UiTM RESEARCH ETHICS COMMITTEE

Thank you for submitting your research proposal to the Research Ethics Committee (REC). After reviewing the list of documents as attached the Committee approved your proposal titled "Pilot Evaluation of Integrating A Personalized Prompt Generator with an Artificial Intelligence Chatbot for Dietary and Exercise Planning: Feasibility and Outcomes in Weight Management for Obese Adults" at UiTM Seremban Campus, Negeri Sembilan. Details of the approval are as follows:

Approval number:	REC/09/2024 (PG/FB/30)
Approval Period:	24 September 2024 until 31 December 2024
Authorised personnel:	1. Azwa Suraya Mohd Dan 2. Adam Feizrel Linoby Ronny Linoby

The UiTM Research Ethics Committee operates in accordance to the ICH Good Clinical Practice Guidelines, Malaysian Good Clinical Practice Guidelines and the Declaration of Helsinki. The approval of this project is conditional upon your continuing compliance with these guidelines and declaration.

Failure to submit reports will result in withdrawal of consent for the project to proceed. Amendments, if any, to the study documents are to be submitted to the REC for approval.

If you require further information, please contact the REC Secretariat at 03-55448069/03-55442794 or email at recsecretariat@uitm.edu.my.

اوسها، تقوى، مولیا
"MALAYSIA MADANI"
"BERKHIDMAT UNTUK NEGARA"

Yours sincerely,

(EMERITUS PROFESSOR DATO' DR RAYMOND AZMAN ALI)
Chairman
UiTM Research Ethics Committee

c.c.: Deputy Rector (PJIM&A), UiTM Sarawak Branch

Universiti Teknologi MARA
Aras 3, Bangunan Wawasan
40450 Shah Alam, Selangor, MALAYSIA
Tel: (+603) 5544 2004/2255
Faks: (+603) 5544 2070

ISO 9001:2015 No. Sijil: 10120156
UiTM di hatiku

APPENDIX 2

Subject Information Sheet for Study 1 and Study 2 (English Version)

REC 4 / 2020/BM Pind. 2(2020)

Research Ethics Committee
Research Management Centre
Universiti Teknologi MARA
40450 SHAH ALAM
Tel: 03-5544-8069, Faks: 03-5544-2096/2767

Subject Information Sheet

Research Title
Pilot Evaluation of Integrating a Personalized Prompt Generator with an Artificial Intelligence Chatbot for Dietary and Exercise Planning: Feasibility and Outcomes in Weight Management for Obese Adults

Introduction of Research
Obesity is a global health crisis significantly affecting health. High obesity rates in Malaysia impact public health and involve considerable government healthcare spending. Advanced chatbot technology, powered by artificial intelligence, can assist in addressing obesity-related issues. This study aims to evaluate the effectiveness of a directive generator system named NExGEN integrated with the renowned chatbot, ChatGPT, in managing effective weight management among adult obese citizens.

Purpose of Research
This study aims to assess the effectiveness of the NExGEN directive generator system integrated with ChatGPT in producing nutrition and exercise plans that consider individual factors and personal weight management goals.

Research Procedure
Participants will be asked to attend the laboratory at UITM Seremban 3. They will be introduced to the NExGEN system and guided to complete a survey designed to collect comprehensive information and to use ChatGPT. Subsequently, participants will be requested to follow a nutrition and exercise plan generated by the NExGEN-ChatGPT system for three (3) months.

Participation in Research
Your participation in this research is voluntary. You have the right not to participate or to withdraw from the study at any time without penalty.

Benefit of Research
The information obtained from this research will benefit active individuals, obese individuals, health practitioners, researchers, the community, and the Government of Malaysia to enhance the knowledge of artificial intelligence technology and the practice of health policy in the future.

Research Risk
None

Confidentiality
Your information will be kept confidential by the investigators and will not be made public unless disclosure is required by law. By signing this consent form, you will authorize the review of records, analysis and use of the data arising from this research. If you have any question about this research or your rights, please contact the investigator at telephone number 012-2655934 (Miss Azwa Suraya) or 013-3632005 (Mr Adam).

APPENDIX 3

Subject Information Sheet for Study 1 and Study 2 (Malay Version)

REC 4 / 2020/BM Pind. 2(2020)

Jawatankuasa Etika Penyelidikan
Pusat Pengurusan Peyelidikan
Universiti Teknologi MARA
40450 SHAH ALAM
Tel: 03-5544-8069, Faks: 03-5544-2096/2767

Borang Maklumat Subjek

Tajuk Penyelidikan
Penilaian Rintis Pengintegrasian Penjana Arahan Peribadi dengan Chatbot Kecerdasan Buatan untuk Perancangan Pemakanan dan Senaman: Kajian Kebolehlaksanaan dan Hasil dalam Pengurusan Berat Badan di Kalangan Warga Obes Dewasa.

Pengenalan penyelidikan
Obesiti adalah krisis kesihatan global yang memberi kesan ketara terhadap kesihatan. Kadar obesiti tinggi di di Malaysia memberikan kesan terhadap kesihatan umum dan melibatkan perbelanjaan rawatan kerajaan yang besar. Perkembangan teknologi chatbot canggih yang menggunakan kecerdasan buatan mampu membantu dalam mengatasi masalah berkaitan obesiti. Kajian ini ingin menilai keberkesanan sistem penjana arahan yang dinamakan NExGEN dengan penyepaduan chatbot terkenal iaitu ChatGPT dalam aspek pengurusan berat badan efektif di kalangan warga obes dewasa.

Tujuan penyelidikan
Kajian ini bertujuan menilai keberkesanan sistem penjana arahan NExGEN dengan penyepaduan ChatGPT dalam penghasilan rancangan pemakanan dan senaman yang mengambil kira faktor peribadi dan sasaran pengurusan berat badan individu.

Prosedur penyelidikan
Peserta akan diminta hadir di makmal UITM Seremban 3. Peserta akan diperkenalkan kepada sistem NExGEN dan para peserta akan dibimbing untuk melengkapkan soal selidik yang direka bentuk untuk mengumpul maklumat komprehensif dan penggunaan ChatGPT. Seterusnya peserta akan diminta untuk untuk mengikuti program perancangan pemakanan dan senaman yang dihasilkan oleh sistem NExGEN-ChatGPT selama tiga (3) bulan.

Penyertaan dalam penyelidikan
Penyertaan anda dalam penyelidikan ini adalah secara sukarela. Anda mempunyai hak untuk tidak menyertai atau mengundurkan diri daripada kajian bila-bila masa tanpa penalti.

Manfaat penyelidikan
Maklumat yang diperolehi daripada penyelidikan ini akan memberi manfaat kepada individu aktif, individu obes, pengamal kesihatan, penyelidik, masyarakat, dan Kerajaan Malaysia untuk mengembangkan ilmu teknologi kecerdasan buatan dan amalan polisi kesihatan di masa hadapan.

Risiko penyelidikan
Tiada

Kerahsiaan
Maklumat anda akan dirahsiakan oleh penyelidik dan tidak akan didedahkan melainkan jika ia dikehendaki oleh undang-undang. Dengan menandatangani borang persetujuan ini, anda membenarkan penelitian rekod, penganalisaan dan penggunaan data hasil daripada penyelidikan ini. Sekiranya anda mempunyai sebarang pertanyaan mengenai penyelidikan ini atau hakhak anda, sila hubungi penyelidik di talian 012-2655934 (Cik Azwa Suraya) atau 0133632005 (En. Adam).

APPENDIX 4

Participant Consent Form (English Version)

Consent Form[1]

To become a subject in the research, you or your legal guardian are required to sign this Consent Form
I herewith confirm that I have met the requirement of age and am capable of acting on behalf of myself / as
2 a legal guardian as follow:
1. I understand the nature and scope of the research being undertaken.
2. I have read and understood all the terms and conditions of my participation in the research.
3. All my questions relating to this research and my participation therein have been answered to my satisfaction.
4. I voluntarily agree to take part in this research, to follow the study procedures and to provide all necessary information to the investigators as requested.
5. I may at any time choose to withdraw from this research without giving any reason.
6. I have received a copy of the Subjects Information Sheet and Consent Form.
7. Except for damages resulting from negligent or malicious conduct of the researcher(s), I hereby release and discharge UiTM and all participating researchers from all liability associated with, arising out of, or related to my participation. I agree to hold them harmless from any harm or loss that may be incurred by me due to my participation in the research.

Name of Subject/Legally authorized representative (LAR)	Signature
I.C No	Date
Name of Witness[3]	Signature
I.C No	Date
Name of Consent Taker	Signature
I.C No	Date

[1] Original signed copy is to be retained by the Principal Investigator.
[2] Delete whichever is not applicable.
[3] A witness is only required for oral consent.

APPENDIX 5

Participant Consent Form (Malay Version)

Borang Izin[1]

Untuk menyertai penyelidikan ini, anda atau penjaga sah perlu menandatangani Borang Izin ini.
Saya dengan ini mengesahkan bahawa saya telah memenuhi syarat umur dan berupaya bertindak bagi pihak saya sendiri/ sebagai 2 penjaga yang sah dalam perkara-perkara berikut:

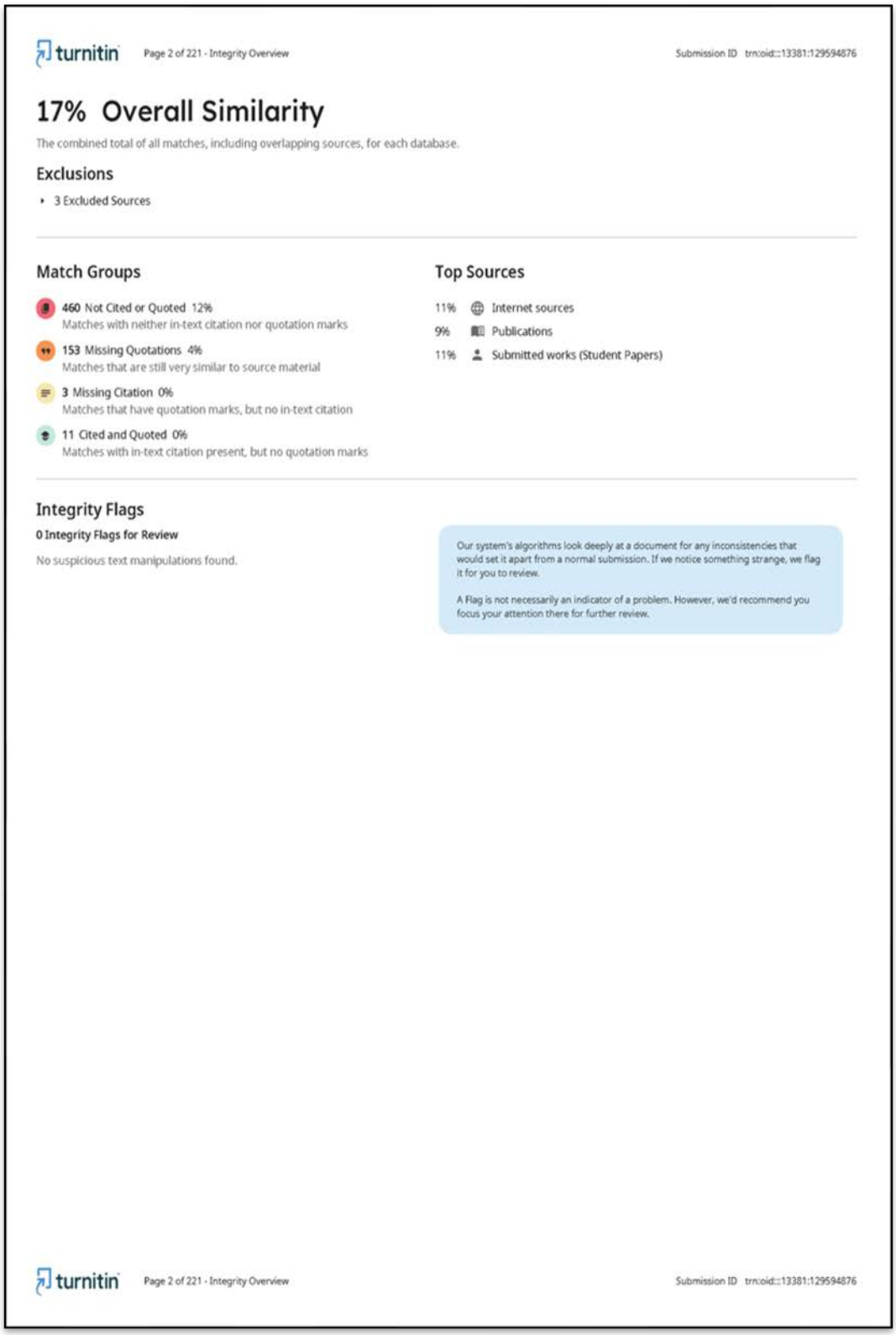
1. Saya memahami ciri-ciri dan skop penyelidikan ini.
2. Saya telah membaca dan memahami semua syarat penyertaan penyelidikan ini.
3. Saya berpuas hati dengan jawapan pada kemusykilan saya tentang penyelidikan ini.
4. Saya secara sukarela bersetuju menyertai penyelidikan ini dan mengikuti segala atur cara dan memberi maklumat yang diperlukan kepada penyelidik seperti yang dikehendaki.
5. Saya boleh menarik diri daripada penyelidikan ini pada bila-bila masa tanpa memberi sebab.
6. Saya telah pun menerima satu salinan Borang Maklumat Subjek dan Borang Izin.
7. Selain daripada kecederaan yang disebabkan oleh kelalaian dan kecuaian penyelidik, saya dengan ini melepaskan dan menggugurkan UiTM dan semua penyelidik dari semua liabiliti berhubung dengan, wujud dari atau berkaitan dengan penyertaan saya. Saya bersetuju untuk menjadikan mereka tidak bertanggunggjawab terhadap apa-apa kemudaratan atau kerugian yang mungkin akan saya tanggung disebabkan oleh penyertaan saya.

Nama Subjek/ Wakil Sah yang berkuatkuasa	Tandatangan
No. Kad Pengenalan	Tarikh
Nama Saksi [3]	Tandatangan
No. Kad Pengenalan	Tarikh
Nama Penyelidik/Pengambil Izin	Tandatangan
No. Kad Pengenalan	Tarikh

[1] Salinan asal disimpan oleh Penyelidik Utama dan satu salinan diserahkan kepada subjek.
[2] Potong mana yang tidak berkenaan.
[3] Saksi dimestikan bagi izin secara lisan.

APPENDIX 6

Thesis Originality Report



turnitin Page 2 of 221 - Integrity Overview Submission ID trn:oid:::13381:129594876

17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Exclusions

▸ 3 Excluded Sources

Match Groups

460 Not Cited or Quoted 12%
Matches with neither in-text citation nor quotation marks

153 Missing Quotations 4%
Matches that are still very similar to source material

3 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

11 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

11% Internet sources
9% Publications
11% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

turnitin Page 2 of 221 - Integrity Overview Submission ID trn:oid:::13381:129594876

APPENDIX 7

Subject Information Form

UNIVERSITI TEKNOLOGI MARA

Subject Information Form

Personal Information Form			
Name			
Phone number			
Age			
Marital status	Single	Married	Divorce
Address			
Weight (kg)			
Height (kg)			
BMI (kg/m^2)			
Type of medication (If any)			
Emergency Contact			
Name			
Phone number			
Relationship			

This form contained personal and confidential information. If found, please immediately refers to Azwa Suraya +0122655934 / linoby@uitm.edu.my

APPENDIX 8

Invitation Email to Expert Penalist (Study 1)



4/27/24, 10:18 AM Universiti Teknologi MARA Mail - Invitation to Participate as Expert Panelist: Validation of a Personalized Weight Management Framework...

UNIVERSITI TEKNOLOGI MARA

AZWA SURAYA MOHD DAN <2023653718@student.uitm.edu.my>

Invitation to Participate as Expert Panelist: Validation of a Personalized Weight Management Framework for Adults with Overweight and Obesity

1 message

AZWA SURAYA MOHD DAN <2023653718@student.uitm.edu.my> 27 April 2024 at 09:17
To: [redacted]

Dear [redacted],

I am Azwa Suraya, a Master's candidate at Universiti Teknologi MARA. I am reaching out to invite you to participate as an **expert panelist in a Fuzzy Delphi Method validation study for my thesis research**.

The study validates a **personalized weight management framework comprising 111 criteria across 6 constructs, covering dietary, exercise, and behavioural recommendations for adults with overweight and obesity**. We are recruiting multidisciplinary experts with 10+ years of experience in [nutritional sciences / exercise science / clinical medicine / psychology / health technology].

Please note that the development origin of the framework is intentionally withheld during evaluation. You are asked to assess its content based solely on scientific merit, clinical appropriateness, and safety. Full development details will be disclosed after the study concludes.

Participation details:

- Complete a bilingual questionnaire (Malay/English) with 111 items on a 7-point Likert scale
- Estimated completion time: [30–60] minutes
- Response window: 2 weeks
- Distribution: electronic or face-to-face (your preference)

The study is approved by UiTM's Institutional Review Board (Ethics REC/09/2024) and registered at UMIN Clinical Trials Registry (UMIN000053570). Participation is voluntary and all responses remain confidential.

Could you confirm your availability by **22/5/2024**? I am happy to provide additional details or the full research proposal upon request.

Thank you for your time and consideration.

Best regards,

Azwa Suraya Binti Mohd Dan
Master of Science (Sport Science and Recreation)
Faculty of Sport Science and Recreation, UiTM
Supervisor: Adam Linoby
linoby@uitm.edu.my | 013-3632005

PENAFIAN: E-mel ini dan apa-apa fail yang dihantar bersama-samanya ("Mesej") adalah dihasratkan hanya untuk kegunaan penerima yang dinyatakan di atas dan mungkin mengandungi maklumat yang tidak umum, bermilik, istimewa, sulit dan dikecualikan dari penzahiran di bawah undang-undang yang terpakai termasuklah Akta Rahsia Rasmi 1972. BACA SELANJUTNYA...

DISCLAIMER : *This e-mail and any files transmitted with it ("Message") is intended only for the use of the recipient(s) named above and may contain information that is non-public, proprietary, privileged, confidential and exempt from disclosure under applicable law including the Official Secrets Act 1972.* READ MORE...

APPENDIX 9

Expert Consent Form (English Version) (Study 1)

Consent Form

To become an expert panelist in this research, you are required to sign this Consent Form. I herewith confirm that I have met the eligibility criteria for expert panelist selection, including a minimum of 10 years of professional experience in a relevant discipline (nutrition, exercise science, medicine, psychology, or artificial intelligence research), and am capable of acting on behalf of myself as follows:

1. I understand the nature and scope of the research being undertaken.
2. I have read and understood all the terms and conditions of my participation in the research.
3. All my questions relating to this research and my participation therein have been answered to my satisfaction.
4. I voluntarily agree to take part in this research as an expert panelist, to follow the study procedures and to provide all necessary information to the investigators as requested.
5. I may at any time choose to withdraw from the research without giving any reason.
6. I have received a copy of the Subjects Information Sheet and Consent Form.
7. Except for damages resulting from negligent or malicious conduct of the researcher(s), I hereby release and discharge UiTM and all participating researchers from all liability associated with, arising out of, or related to my participation. I agree to hold them harmless from any harm or loss that may be incurred by me due to my participation in in

Name of Expert Panelist / Legally Authorized Representative (LAR)	Signature
I.C No	Date
Name of Witness	Signature
I.C No	Date
Name of Consent Taker	Signature

[1] Original signed copy is to be retained by the Principal Investigator.
[2] Delete whichever is not applicable.
[3] A witness is only required for oral consent.

APPENDIX 10

Expert Consent Form (Malay Version) (Study 1)

Borang Izin

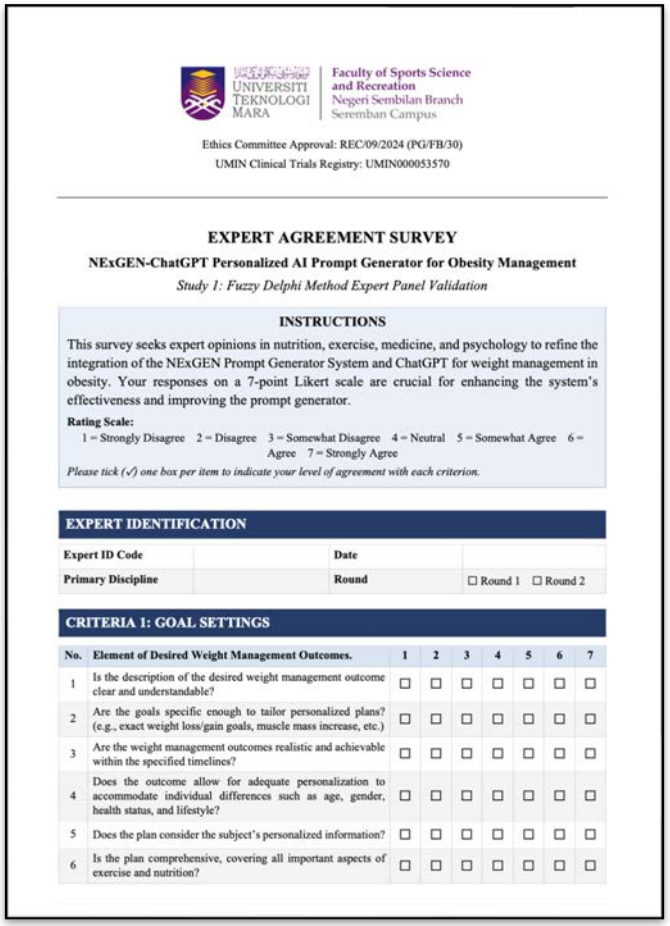
Untuk menjadi ahli panel pakar dalam penyelidikan ini, anda dikehendaki menandatangani Borang Persetujuan ini. Saya dengan ini mengesahkan bahawa saya telah memenuhi kriteria kelayakan untuk pemilihan ahli panel pakar, termasuk pengalaman profesional sekurang-kurangnya 10 tahun dalam disiplin yang berkaitan (pemakanan, sains senaman, perubatan, psikologi, atau penyelidikan kecerdasan buatan), dan mampu bertindak bagi pihak diri saya sendiri seperti berikut:

1. Saya Saya memahami sifat dan skop penyelidikan yang sedang dijalankan.
2. Saya telah membaca dan memahami semua terma dan syarat penyertaan saya dalam penyelidikan ini.
3. Semua soalan saya berkaitan penyelidikan ini dan penyertaan saya di dalamnya telah dijawab dengan memuaskan.
4. Saya secara sukarela bersetuju untuk mengambil bahagian dalam penyelidikan ini sebagai ahli panel pakar, untuk mengikuti prosedur kajian dan memberikan semua maklumat yang diperlukan kepada penyelidik seperti yang diminta.
5. Saya boleh pada bila-bila masa memilih untuk menarik diri daripada penyelidikan tanpa perlu memberikan sebarang alasan.
6. Saya telah menerima salinan Helaian Maklumat Subjek dan Borang Persetujuan.
7. Kesuali bagi, kerosakan yang disebabkan oleh kecuaian atau kelakuan jahat penyelidik, saya dengan ini melepaskan dan membebaskan UiTM serta semua penyelidik yang terlibat daripada semua liabiliti yang berkaitan dengan, timbul daripada, atau berhubung dengan penyertaan saya. Saya bersetuju untuk membebaskan mereka daripada sebarang kemudaratan atau kerugian yang mungkin ditanggung oleh saya akibat penyertaan saya dalam penyelidikan ini.

Nama Ahli Panel Pakar / Wakil Sah yang Diberi Kuasa (LAR)	Tandatangan
No. K/P	Tarikh
Nama Saksi	Tandatangan
No. K/P	Tarikh
Nama Pengambil	Tandatangan
No. K/P	Traikh

APPENDIX 11

Expert Aggreement Survey (Study 1)



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

EXPERT AGREEMENT SURVEY

NExGEN-ChatGPT Personalized AI Prompt Generator for Obesity Management

Study 1: Fuzzy Delphi Method Expert Panel Validation

INSTRUCTIONS

This survey seeks expert opinions in nutrition, exercise, medicine, and psychology to refine the integration of the NExGEN Prompt Generator System and ChatGPT for weight management in obesity. Your responses on a 7-point Likert scale are crucial for enhancing the system's effectiveness and improving the prompt generator.

Rating Scale:
1 = Strongly Disagree 2 = Disagree 3 = Somewhat Disagree 4 = Neutral 5 = Somewhat Agree 6 = Agree 7 = Strongly Agree

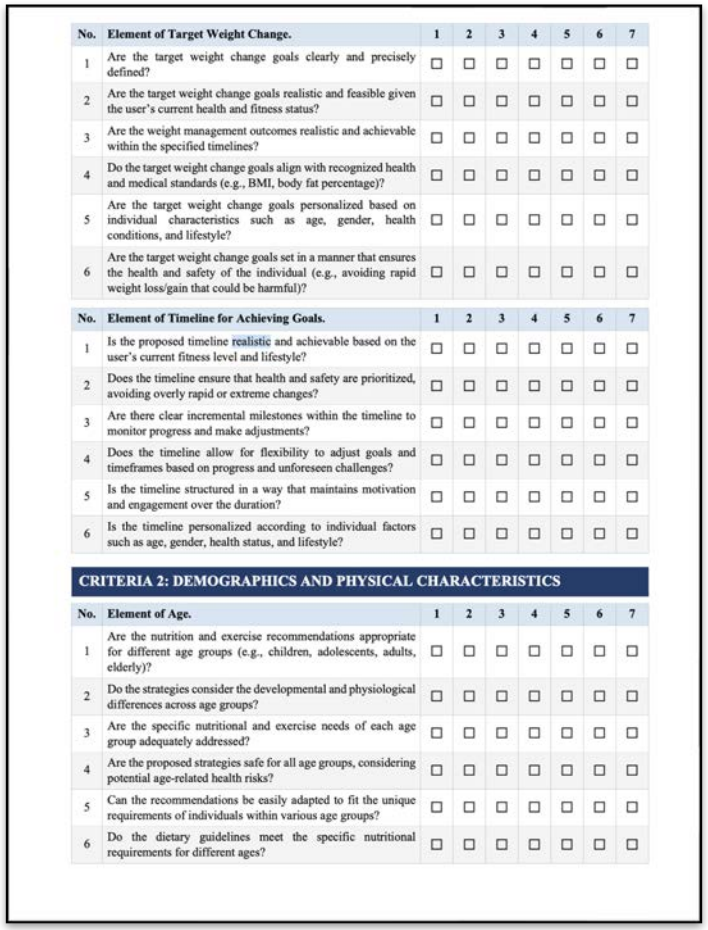
Please tick (✓) one box per item to indicate your level of agreement with each criterion.

EXPERT IDENTIFICATION

Expert ID Code		**Date**	
Primary Discipline		**Round**	☐ Round 1 ☐ Round 2

CRITERIA 1: GOAL SETTINGS

No.	Element of Desired Weight Management Outcomes.	1	2	3	4	5	6	7
1	Is the description of the desired weight management outcome clear and understandable?	☐	☐	☐	☐	☐	☐	☐
2	Are the goals specific enough to tailor personalized plans? (e.g., exact weight loss/gain goals, muscle mass increase, etc.)	☐	☐	☐	☐	☐	☐	☐
3	Are the weight management outcomes realistic and achievable within the specified timelines?	☐	☐	☐	☐	☐	☐	☐
4	Does the outcome allow for adequate personalization to accommodate individual differences such as age, gender, health status, and lifestyle?	☐	☐	☐	☐	☐	☐	☐
5	Does the plan consider the subject's personalized information?	☐	☐	☐	☐	☐	☐	☐
6	Is the plan comprehensive, covering all important aspects of exercise and nutrition?	☐	☐	☐	☐	☐	☐	☐



No.	Element of Target Weight Change.	1	2	3	4	5	6	7
1	Are the target weight change goals clearly and precisely defined?	☐	☐	☐	☐	☐	☐	☐
2	Are the target weight change goals realistic and feasible given the user's current health and fitness status?	☐	☐	☐	☐	☐	☐	☐
3	Are the weight management outcomes realistic and achievable within the specified timelines?	☐	☐	☐	☐	☐	☐	☐
4	Do the target weight change goals align with recognized health and medical standards (e.g., BMI, body fat percentage)?	☐	☐	☐	☐	☐	☐	☐
5	Are the target weight change goals personalized based on individual characteristics such as age, gender, health conditions, and lifestyle?	☐	☐	☐	☐	☐	☐	☐
6	Are the target weight change goals set in a manner that ensures the health and safety of the individual (e.g., avoiding rapid weight loss/gain that could be harmful)?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Timeline for Achieving Goals.	1	2	3	4	5	6	7
1	Is the proposed timeline realistic and achievable based on the user's current fitness level and lifestyle?	☐	☐	☐	☐	☐	☐	☐
2	Does the timeline ensure that health and safety are prioritized, avoiding overly rapid or extreme changes?	☐	☐	☐	☐	☐	☐	☐
3	Are there clear incremental milestones within the timeline to monitor progress and make adjustments?	☐	☐	☐	☐	☐	☐	☐
4	Does the timeline allow for flexibility to adjust goals and timeframes based on progress and unforeseen challenges?	☐	☐	☐	☐	☐	☐	☐
5	Is the timeline structured in a way that maintains motivation and engagement over the duration?	☐	☐	☐	☐	☐	☐	☐
6	Is the timeline personalized according to individual factors such as age, gender, health status, and lifestyle?	☐	☐	☐	☐	☐	☐	☐

CRITERIA 2: DEMOGRAPHICS AND PHYSICAL CHARACTERISTICS

No.	Element of Age.	1	2	3	4	5	6	7
1	Are the nutrition and exercise recommendations appropriate for different age groups (e.g., children, adolescents, adults, elderly)?	☐	☐	☐	☐	☐	☐	☐
2	Do the strategies consider the developmental and physiological differences across age groups?	☐	☐	☐	☐	☐	☐	☐
3	Are the specific nutritional and exercise needs of each age group adequately addressed?	☐	☐	☐	☐	☐	☐	☐
4	Are the proposed strategies safe for all age groups, considering potential age-related health risks?	☐	☐	☐	☐	☐	☐	☐
5	Can the recommendations be easily adapted to fit the unique requirements of individuals within various age groups?	☐	☐	☐	☐	☐	☐	☐
6	Do the dietary guidelines meet the specific nutritional requirements for different ages?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Height and Weight.	1	2	3	4	5	6	7
1	Are the nutrition and exercise recommendations tailored to different Body Mass Index (BMI) categories (e.g., underweight, normal weight, overweight, obese)?	☐	☐	☐	☐	☐	☐	☐
2	Are the recommendations personalized based on individual height and weight measurements?	☐	☐	☐	☐	☐	☐	☐
3	Are the proposed strategies safe for individuals with different height and weight profiles, considering potential health risks?	☐	☐	☐	☐	☐	☐	☐
4	Are the specific nutritional needs of individuals with different body sizes adequately addressed?	☐	☐	☐	☐	☐	☐	☐
5	Is the intensity and type of exercise recommended appropriate for different height and weight categories?	☐	☐	☐	☐	☐	☐	☐
6	Are weight management goals realistic and achievable for individuals with different height and weight profiles?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Physiological Measurements.	1	2	3	4	5	6	7
1	Do the recommendations personalize strategies based on individual physiological measurements?	☐	☐	☐	☐	☐	☐	☐
2	Are the methods and tools used for measuring physiological metrics accurate and reliable?	☐	☐	☐	☐	☐	☐	☐
3	Are the physiological measurements used to ensure the safety and monitor the health of participants during exercise and nutrition programs?	☐	☐	☐	☐	☐	☐	☐
4	Can fitness and nutrition plans be adjusted based on changes in physiological measurements over time?	☐	☐	☐	☐	☐	☐	☐
5	Are there effective systems in place for tracking and reporting physiological measurements over time?	☐	☐	☐	☐	☐	☐	☐
6	Are modern technologies employed to enhance the collection, analysis, and feedback of physiological measurements?	☐	☐	☐	☐	☐	☐	☐

CRITERIA 3: HEALTH AND MEDICAL HISTORY

No.	Element of Known Medical Conditions.	1	2	3	4	5	6	7
1	Are a wide range of known medical conditions considered in the fitness and nutrition recommendations (e.g., diabetes, heart disease, arthritis)?	☐	☐	☐	☐	☐	☐	☐
2	How effectively are the exercise and diet plans customized to accommodate specific medical conditions?	☐	☐	☐	☐	☐	☐	☐
3	Are the recommendations safe for individuals with different medical conditions, minimizing health risks?	☐	☐	☐	☐	☐	☐	☐
4	Are there mechanisms in place to monitor the impact of the fitness and diet plans on the medical conditions and adjust them as necessary?	☐	☐	☐	☐	☐	☐	☐
5	Are the programs accessible and feasible for individuals with varying degrees of medical limitations?	☐	☐	☐	☐	☐	☐	☐
6	Is there a feedback system to gather input from participants on the effectiveness of the programs in managing their medical conditions?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Current Medications.	1	2	3	4	5	6	7
1	Are the exercise and nutrition recommendations designed to avoid negative interactions with commonly prescribed medications?	☐	☐	☐	☐	☐	☐	☐
2	How effectively are health and fitness plans customized to accommodate the effects of current medications on individuals?	☐	☐	☐	☐	☐	☐	☐
3	Can the fitness and nutrition plans be easily adjusted based on changes in medication regimes?	☐	☐	☐	☐	☐	☐	☐
4	Are medical records that include medication histories integrated into the fitness and nutrition planning process?	☐	☐	☐	☐	☐	☐	☐
5	Are considerations for medication adjustments included if fitness and nutritional changes necessitate such adjustments?	☐	☐	☐	☐	☐	☐	☐
6	Are ethical and legal considerations regarding medication information and privacy maintained throughout the program?	☐	☐	☐	☐	☐	☐	☐

CRITERIA 4: LIFESTYLE AND BEHAVIOUR

No.	Element of Sleep Pattern.	1	2	3	4	5	6	7
1	How well are sleep patterns considered and integrated into personalized fitness and nutrition plans?	☐	☐	☐	☐	☐	☐	☐
2	Are the effects of sleep quality and duration on health and fitness outcomes sufficiently addressed in the program?	☐	☐	☐	☐	☐	☐	☐
3	How effectively are recommendations customized based on individual sleep quality and patterns?	☐	☐	☐	☐	☐	☐	☐
4	Are participants educated about the importance of sleep and its impact on weight management and overall health?	☐	☐	☐	☐	☐	☐	☐
5	Are there effective systems in place for monitoring and tracking changes in sleep patterns as part of the fitness and nutrition program?	☐	☐	☐	☐	☐	☐	☐
6	Are adjustments made for participants with diagnosed sleep disorders or irregular sleep patterns?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Stress Level.	1	2	3	4	5	6	7
1	Are initial and ongoing assessments of stress levels included in the program to tailor recommendations?	☐	☐	☐	☐	☐	☐	☐
2	How are stress levels considered in relation to their impact on health outcomes, including weight management?	☐	☐	☐	☐	☐	☐	☐
3	How effectively are stress management strategies integrated into fitness and nutrition plans?	☐	☐	☐	☐	☐	☐	☐
4	Are stress management techniques customized based on individual stress triggers and responses?	☐	☐	☐	☐	☐	☐	☐
5	Are the recommended stress management techniques safe and proven effective?	☐	☐	☐	☐	☐	☐	☐
6	Are relaxation techniques (e.g., meditation, deep breathing, yoga) incorporated into the program?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Physical Activity Habits.	1	2	3	4	5	6	7
1	How thoroughly are initial and ongoing assessments of physical activity levels integrated into the program?	☐	☐	☐	☐	☐	☐	☐
2	How effectively are physical activity habits integrated with overall health and weight management goals?	☐	☐	☐	☐	☐	☐	☐
3	Does the program include a wide range of physical activities to cater to different interests and capabilities?	☐	☐	☐	☐	☐	☐	☐
4	Are there effective systems in place for tracking physical activity and providing feedback to participants?	☐	☐	☐	☐	☐	☐	☐
5	How is technology (e.g., apps, wearable devices) used to support and enhance engagement in physical activities?	☐	☐	☐	☐	☐	☐	☐
6	Are specific metrics used to measure improvements or changes in physical activity habits?	☐	☐	☐	☐	☐	☐	☐

No.	Element of Dietary Habits.	1	2	3	4	5	6	7
1	How thoroughly are initial and ongoing assessments of dietary habits integrated into the program?	☐	☐	☐	☐	☐	☐	☐
2	How effectively are dietary habits aligned with and supportive of individual fitness and weight management goals?	☐	☐	☐	☐	☐	☐	☐
3	Are the recommended diets safe and suitable for individuals based on their medical history and current health status?	☐	☐	☐	☐	☐	☐	☐
4	How is technology (e.g., meal tracking apps, online nutrition counseling) used to support and enhance adherence to dietary recommendations?	☐	☐	☐	☐	☐	☐	☐
5	Are the effects of dietary habits on psychological well-being considered and addressed?	☐	☐	☐	☐	☐	☐	☐
6	Are environmental impacts and sustainability considered in the dietary recommendations?	☐	☐	☐	☐	☐	☐	☐

CRITERIA 5: MOTIVATION AND SOCIAL SUPPORT

No.	Element of Motivations for Weight Management.	1	2	3	4	5	6	7
1	Are the motivational strategies clear and directly relevant to the individual's weight management goals?	☐	☐	☐	☐	☐	☐	☐
2	How well are motivational strategies aligned with personal health and fitness objectives?	☐	☐	☐	☐	☐	☐	☐
3	How personalized are the motivational strategies based on individual needs and backgrounds?	☐	☐	☐	☐	☐	☐	☐
4	Are regular feedback and positive reinforcement used as part of the motivational strategies?	☐	☐	☐	☐	☐	☐	☐
5	How well are motivational strategies integrated into participants' daily routines and lifestyles?	☐	☐	☐	☐	☐	☐	☐
6	Are motivational strategies flexible and adjustable based on feedback and changing needs?	☐	☐	☐	☐	☐	☐	☐

CRITERIA 6: ENVIRONMENTAL FACTORS

No.	Element of Grocery Shopping Habits.	1	2	3	4	5	6	7
1	How well do grocery shopping habits support the nutritional goals set in the program?	☐	☐	☐	☐	☐	☐	☐
2	How significantly do grocery shopping habits influence the dietary choices of participants?	☐	☐	☐	☐	☐	☐	☐
3	Does the program provide adequate education on how to make healthy choices while grocery shopping?	☐	☐	☐	☐	☐	☐	☐
4	Are grocery shopping strategies personalized based on individual dietary needs and preferences?	☐	☐	☐	☐	☐	☐	☐
5	How effectively are grocery shopping habits integrated with meal planning within the program?	☐	☐	☐	☐	☐	☐	☐
6	Are there feedback mechanisms in place to assess the effectiveness of changes in grocery shopping habits?	☐	☐	☐	☐	☐	☐	☐

EXPERT COMMENTS AND RECOMMENDATIONS

Please provide any additional comments, suggestions, or recommendations for improving the NExGEN-ChatGPT prompt generator system for obesity management. Your qualitative feedback is valuable for refining the framework.

EXPERT DECLARATION AND SIGN-OFF

I confirm that I have reviewed each criterion in this survey independently and that my responses reflect my professional judgment based on my expertise in the relevant domain(s). I understand that my responses will be analyzed using the Fuzzy Delphi Method and reported in aggregate form only.

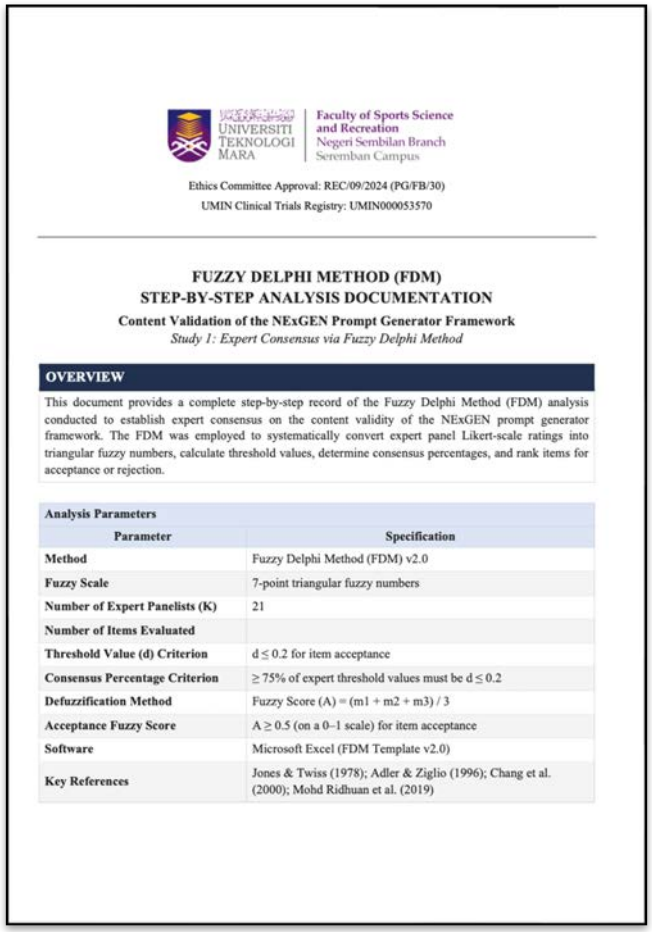
Name : _____ *Signature* _____ *Date:* _____

For Research Team Use Only:

Received by: _____ Date Received: _____ Data Entry Completed: ☐

APPENDIX 12

Fuzzy Delphi Method Analysis (Study 1)



Faculty of Sports Science and Recreation
Negeri Sembilan Branch
Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

FUZZY DELPHI METHOD (FDM)
STEP-BY-STEP ANALYSIS DOCUMENTATION

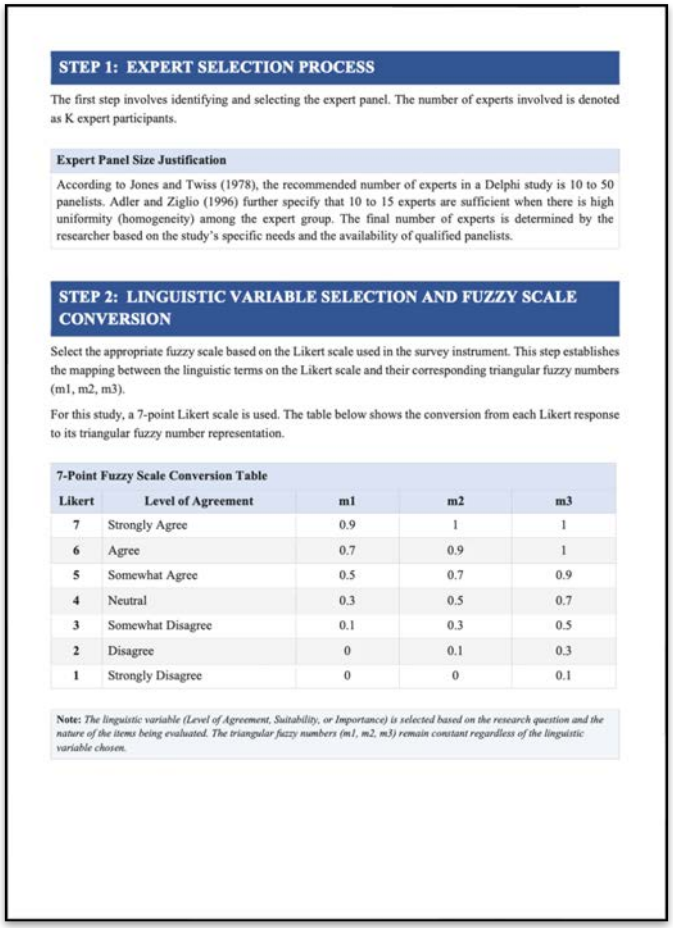
Content Validation of the NExGEN Prompt Generator Framework
Study 1: Expert Consensus via Fuzzy Delphi Method

OVERVIEW

This document provides a complete step-by-step record of the Fuzzy Delphi Method (FDM) analysis conducted to establish expert consensus on the content validity of the NExGEN prompt generator framework. The FDM was employed to systematically convert expert panel Likert-scale ratings into triangular fuzzy numbers, calculate threshold values, determine consensus percentages, and rank items for acceptance or rejection.

Analysis Parameters

Parameter	Specification
Method	Fuzzy Delphi Method (FDM) v2.0
Fuzzy Scale	7-point triangular fuzzy numbers
Number of Expert Panelists (K)	21
Number of Items Evaluated	
Threshold Value (d) Criterion	$d \leq 0.2$ for item acceptance
Consensus Percentage Criterion	$\geq 75\%$ of expert threshold values must be $d \leq 0.2$
Defuzzification Method	Fuzzy Score (A) = (m1 + m2 + m3) / 3
Acceptance Fuzzy Score	$A \geq 0.5$ (on a 0–1 scale) for item acceptance
Software	Microsoft Excel (FDM Template v2.0)
Key References	Jones & Twiss (1978); Adler & Ziglio (1996); Chang et al. (2000); Mohd Ridhuan et al. (2019)



STEP 1: EXPERT SELECTION PROCESS

The first step involves identifying and selecting the expert panel. The number of experts involved is denoted as K expert participants.

Expert Panel Size Justification

According to Jones and Twiss (1978), the recommended number of experts in a Delphi study is 10 to 50 panelists. Adler and Ziglio (1996) further specify that 10 to 15 experts are sufficient when there is high uniformity (homogeneity) among the expert group. The final number of experts is determined by the researcher based on the study's specific needs and the availability of qualified panelists.

STEP 2: LINGUISTIC VARIABLE SELECTION AND FUZZY SCALE CONVERSION

Select the appropriate fuzzy scale based on the Likert scale used in the survey instrument. This step establishes the mapping between the linguistic terms on the Likert scale and their corresponding triangular fuzzy numbers (m1, m2, m3).

For this study, a 7-point Likert scale is used. The table below shows the conversion from each Likert response to its triangular fuzzy number representation.

7-Point Fuzzy Scale Conversion Table

Likert	Level of Agreement	m1	m2	m3
7	Strongly Agree	0.9	1	1
6	Agree	0.7	0.9	1
5	Somewhat Agree	0.5	0.7	0.9
4	Neutral	0.3	0.5	0.7
3	Somewhat Disagree	0.1	0.3	0.5
2	Disagree	0	0.1	0.3
1	Strongly Disagree	0	0	0.1

Note: *The linguistic variable (Level of Agreement, Suitability, or Importance) is selected based on the research question and the nature of the items being evaluated. The triangular fuzzy numbers (m1, m2, m3) remain constant regardless of the linguistic variable chosen.*

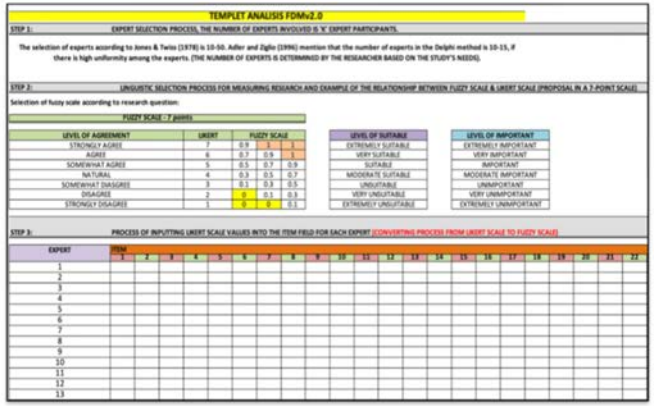


Figure 1: FDM Template v2.0 — Steps 1–3 (Expert Selection, Fuzzy Scale, and Likert-to-Fuzzy Conversion)

STEP 3: DATA INPUT: CONVERTING LIKERT RESPONSES TO FUZZY SCALE VALUES

In this step, each expert's Likert-scale response for each item is converted to its corresponding triangular fuzzy number using the conversion table from Step 2. This produces a matrix of fuzzy values with dimensions K (experts) × N (items) × 3 (m1, m2, m3).

Procedure
1. Tabulate all expert Likert responses in an Expert × Item matrix.
2. For each response, replace the Likert value with its corresponding triangular fuzzy number triplet (m1, m2, m3).
3. Example: If Expert 1 gave a Likert rating of 7 ("Strongly Agree") for Item 1, this is converted to the fuzzy triplet (0.9, 1, 1).
4. Repeat for all K experts across all N items.
5. The resulting matrix contains 3 values (m1, m2, m3) per cell, organized by item.

STEP 4: CALCULATING AVERAGE FUZZY VALUES (m1, m2, m3)

For each item, calculate the average of each fuzzy component (m1, m2, m3) across all K experts. These averaged values represent the collective expert judgment for each item.

Average m1 for Item j = $\Sigma\, m1_{ij} / K$
Average m2 for Item j = $\Sigma\, m2_{ij} / K$
Average m3 for Item j = $\Sigma\, m3_{ij} / K$
Where: K = total number of experts; j = item number; i = expert number



Figure 2: FDM Template v2.0 — Step 4 (Calculating Average Fuzzy Values Across All Experts)

Interpretation: *The bottom row of the expert matrix shows the averaged m1, m2, and m3 values for each item. These values will be used in subsequent steps for threshold calculation, defuzzification, and ranking.*

STEP 5: DETERMINING THE THRESHOLD VALUE (d)

The threshold value d measures the distance between each individual expert's fuzzy evaluation and the group average for each item. It indicates how far each expert's response deviates from the collective judgment. The threshold d is calculated for every Expert × Item pair using the Euclidean distance formula for triangular fuzzy numbers.

$$d(\tilde{m}, \tilde{m}) = \sqrt{[\,\tfrac{1}{3} \times ((m1_i - m1)^2 + (m2_i - m2)^2 + (m3_i - m3)^2)\,]}$$

Where:

m1i, m2i, m3i = individual expert's fuzzy values for the item

m1, m2, m3 = average fuzzy values for the item (from Step 4)

Acceptance Criterion for Threshold Value

If d ≤ 0.2: The expert's response for that item is considered to be in agreement with the group consensus.

If d > 0.2: The expert's response deviates substantially from the group consensus (highlighted in red in the template).

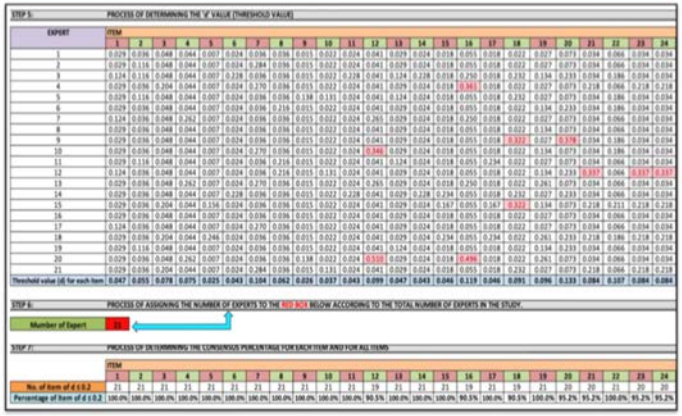


Figure 3: FDM Template v2.0 — Steps 5–7 (Threshold Values, Expert Count, and Consensus Percentage)

STEP 6: COUNTING EXPERTS WITHIN THRESHOLD (d ≤ 0.2)

For each item, count the number of experts whose threshold value d is less than or equal to 0.2. This count represents the number of experts who are in agreement with the group consensus for that specific item.

Procedure

1. Examine the threshold value (d) calculated in Step 5 for each expert on each item.
2. For each item, count the number of experts with d ≤ 0.2.
3. Record this count in the "No. of item of d ≤ 0.2" row.
4. The total number of experts (K) must be entered into the template to enable percentage calculation.

STEP 7: DETERMINING THE CONSENSUS PERCENTAGE

Calculate the consensus percentage for each item by dividing the number of experts within the threshold (d ≤ 0.2) by the total number of experts (K), then multiplying by 100.

Consensus Percentage for Item j = (Number of experts with d ≤ 0.2 / K) × 100%

Acceptance Criterion for Consensus Percentage

≥ 75% expert consensus: The item meets the minimum consensus threshold and is considered to have adequate expert agreement.

< 75% expert consensus: The item fails to reach adequate consensus. A second round of Delphi may be warranted, or the item should be removed or revised.

STEP 8: DEFUZZIFICATION: CALCULATING THE FUZZY SCORE (A)

Defuzzification converts the averaged triangular fuzzy numbers back into a single crisp value. The fuzzy score (A) is calculated using the simple average of the three fuzzy components for each item.

Fuzzy Score (A) for Item j = (m1 + m2 + m3) / 3

Where m1, m2, m3 are the averaged fuzzy values from Step 4

Acceptance Criterion for Fuzzy Score

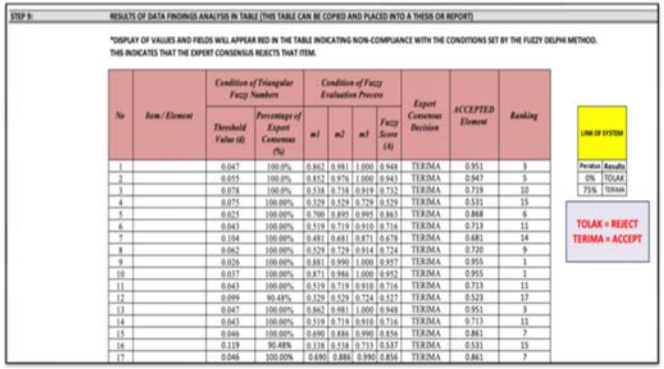
A ≥ 0.5: Item accepted (TERIMA / ACCEPT). Indicates a positive consensus that the item is suitable, important, or agreed upon by the expert panel.

A < 0.5: Item rejected (TOLAK / REJECT). Indicates insufficient expert endorsement of the item.

Note: *The fuzzy score also serves as the basis for ranking items. Higher fuzzy scores indicate stronger expert consensus and endorsement. Items can be ranked from highest to lowest A value to determine priority.*

STEP 9: RESULTS SUMMARY TABLE

The final step consolidates all computed values into a single summary table suitable for inclusion in a thesis or research report. This table presents the complete FDM analysis results, including the threshold value, consensus percentage, average fuzzy numbers, fuzzy score, expert consensus decision (Accept / Reciect), accepted element score, and item ranking.



STEP 9: RESULTS OF DATA FINDINGS ANALYSIS IN TABLE [THIS TABLE CAN BE COPIED AND PLACED INTO A THESIS OR REPORT]

*DISPLAY OF VALUES AND FIELDS WILL APPEAR RED IN THE TABLE INDICATING NON-COMPLIANCE WITH THE CONDITIONS SET BY THE FUZZY DELPHI METHOD. THIS INDICATES THAT THE EXPERT CONSENSUS REJECTS THAT ITEM.

No	Item / Element	Condition of Triangular Fuzzy Numbers		Condition of Fuzzy Evaluation Process				Expert Consensus Decision	ACCEPTED Element	Ranking
		Threshold Value (d)	Percentage of Expert Consensus (%)	m1	m2	m3	Fuzzy Score (A)			
1		0.047	100.0%	0.862	0.981	1.000	0.948	TERIMA	0.951	3
2		0.055	100.0%	0.852	0.976	1.000	0.943	TERIMA	0.947	5
3		0.078	100.0%	0.538	0.738	0.919	0.732	TERIMA	0.719	10
4		0.075	100.00%	0.329	0.529	0.729	0.529	TERIMA	0.531	15
5		0.025	100.00%	0.700	0.895	0.995	0.863	TERIMA	0.868	6
6		0.043	100.00%	0.519	0.719	0.910	0.716	TERIMA	0.713	11
7		0.104	100.00%	0.481	0.681	0.871	0.678	TERIMA	0.681	14
8		0.062	100.00%	0.529	0.729	0.914	0.724	TERIMA	0.720	9
9		0.026	100.00%	0.881	0.990	1.000	0.957	TERIMA	0.955	1
10		0.037	100.00%	0.871	0.986	1.000	0.952	TERIMA	0.955	1
11		0.043	100.00%	0.519	0.719	0.910	0.716	TERIMA	0.713	11
12		0.099	90.48%	0.329	0.529	0.724	0.527	TERIMA	0.523	17
13		0.047	100.00%	0.862	0.981	1.000	0.948	TERIMA	0.951	3
14		0.043	100.00%	0.519	0.719	0.910	0.716	TERIMA	0.713	11
15		0.046	100.00%	0.690	0.886	0.990	0.856	TERIMA	0.861	7
16		0.119	90.48%	0.338	0.538	0.733	0.537	TERIMA	0.531	15
17		0.046	100.00%	0.690	0.886	0.990	0.856	TERIMA	0.861	7

LINK OF SYSTEM

Peratus	Results
0%	TOLAK
75%	TERIMA

TOLAK = REJECT
TERIMA = ACCEPT

Figure 4: FDM Template v2.0 — Step 9 (Results Summary Table for Thesis Reporting)

FDM ACCEPTANCE DECISION RULES (SUMMARY)

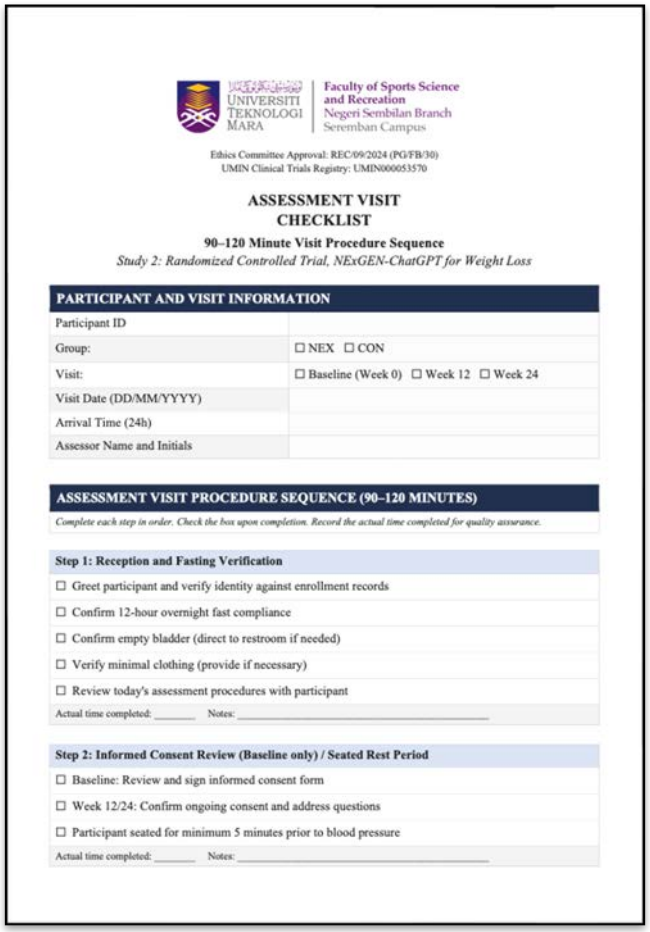
Condition	Decision Rule
Condition 1: Threshold Value (d)	$d \leq 0.2$ for each expert-item pair indicates agreement
Condition 2: Expert Consensus (%)	$\geq 75\%$ of experts must have $d \leq 0.2$ for the item to pass
Condition 3: Fuzzy Score (A)	$A \geq 0.5$ for item acceptance (TERIMA / ACCEPT)
All 3 conditions met	TERIMA (ACCEPT) — Item retained in the framework
Any condition not met	TOLAK (REJECT) — Item removed, revised, or subjected to Round 2

Item Ranking

Items are ranked by their fuzzy score (A) in descending order. A higher fuzzy score indicates stronger expert agreement and endorsement. When two items share the same fuzzy score, the item with the lower threshold value (d) is ranked higher, as it demonstrates tighter consensus among experts.

APPENDIX 13

Assessment Visit Checklist



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

ASSESSMENT VISIT CHECKLIST

90–120 Minute Visit Procedure Sequence

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

PARTICIPANT AND VISIT INFORMATION	
Participant ID	
Group:	☐ NEX ☐ CON
Visit:	☐ Baseline (Week 0) ☐ Week 12 ☐ Week 24
Visit Date (DD/MM/YYYY)	
Arrival Time (24h)	
Assessor Name and Initials	

ASSESSMENT VISIT PROCEDURE SEQUENCE (90–120 MINUTES)

Complete each step in order. Check the box upon completion. Record the actual time completed for quality assurance.

Step 1: Reception and Fasting Verification

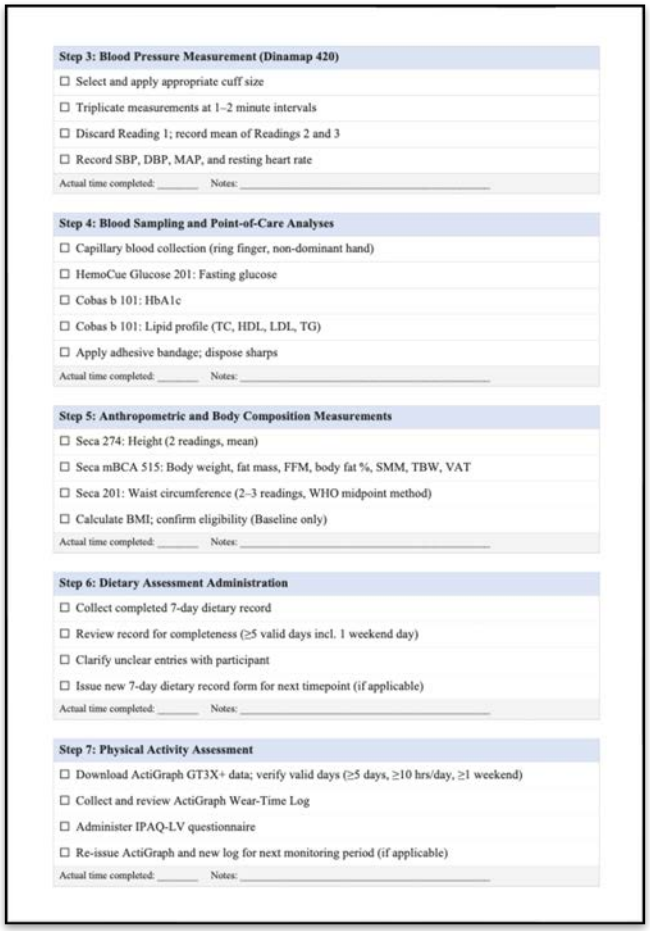
- ☐ Greet participant and verify identity against enrollment records
- ☐ Confirm 12-hour overnight fast compliance
- ☐ Confirm empty bladder (direct to restroom if needed)
- ☐ Verify minimal clothing (provide if necessary)
- ☐ Review today's assessment procedures with participant

Actual time completed: _____ Notes: _____

Step 2: Informed Consent Review (Baseline only) / Seated Rest Period

- ☐ Baseline: Review and sign informed consent form
- ☐ Week 12/24: Confirm ongoing consent and address questions
- ☐ Participant seated for minimum 5 minutes prior to blood pressure

Actual time completed: _____ Notes: _____



Step 3: Blood Pressure Measurement (Dinamap 420)

- ☐ Select and apply appropriate cuff size
- ☐ Triplicate measurements at 1–2 minute intervals
- ☐ Discard Reading 1; record mean of Readings 2 and 3
- ☐ Record SBP, DBP, MAP, and resting heart rate

Actual time completed: _____ Notes: _____

Step 4: Blood Sampling and Point-of-Care Analyses

- ☐ Capillary blood collection (ring finger, non-dominant hand)
- ☐ HemoCue Glucose 201: Fasting glucose
- ☐ Cobas b 101: HbA1c
- ☐ Cobas b 101: Lipid profile (TC, HDL, LDL, TG)
- ☐ Apply adhesive bandage; dispose sharps

Actual time completed: _____ Notes: _____

Step 5: Anthropometric and Body Composition Measurements

- ☐ Seca 274: Height (2 readings, mean)
- ☐ Seca mBCA 515: Body weight, fat mass, FFM, body fat %, SMM, TBW, VAT
- ☐ Seca 201: Waist circumference (2–3 readings, WHO midpoint method)
- ☐ Calculate BMI; confirm eligibility (Baseline only)

Actual time completed: _____ Notes: _____

Step 6: Dietary Assessment Administration

- ☐ Collect completed 7-day dietary record
- ☐ Review record for completeness (≥5 valid days incl. 1 weekend day)
- ☐ Clarify unclear entries with participant
- ☐ Issue new 7-day dietary record form for next timepoint (if applicable)

Actual time completed: _____ Notes: _____

Step 7: Physical Activity Assessment

- ☐ Download ActiGraph GT3X+ data; verify valid days (≥5 days, ≥10 hrs/day, ≥1 weekend)
- ☐ Collect and review ActiGraph Wear-Time Log
- ☐ Administer IPAQ-LV questionnaire
- ☐ Re-issue ActiGraph and new log for next monitoring period (if applicable)

Actual time completed: _____ Notes: _____

Step 8: Intervention Check and Scheduling
☐ Review ChatGPT usage and adherence log
☐ Address participant questions regarding the intervention
☐ Schedule next assessment visit (if applicable)
☐ Provide new ChatGPT Business subscription credentials (if needed)
Actual time completed: _____ Notes: _____

Step 9: Data Verification and Wrap-Up
☐ Review all CRF entries for completeness and accuracy
☐ Ensure all data collection forms are signed
☐ Thank participant and confirm next visit details
☐ Complete assessor sign-off on visit checklist
Actual time completed: _____ Notes: _____

Visit Summary	
Visit Start Time (24h)	
Visit End Time (24h)	
Total Visit Duration (minutes)	
All procedures completed:	☐ Yes ☐ No (specify below)
Incomplete procedures / deviations: _____	

VERIFICATION AND SIGN-OFF

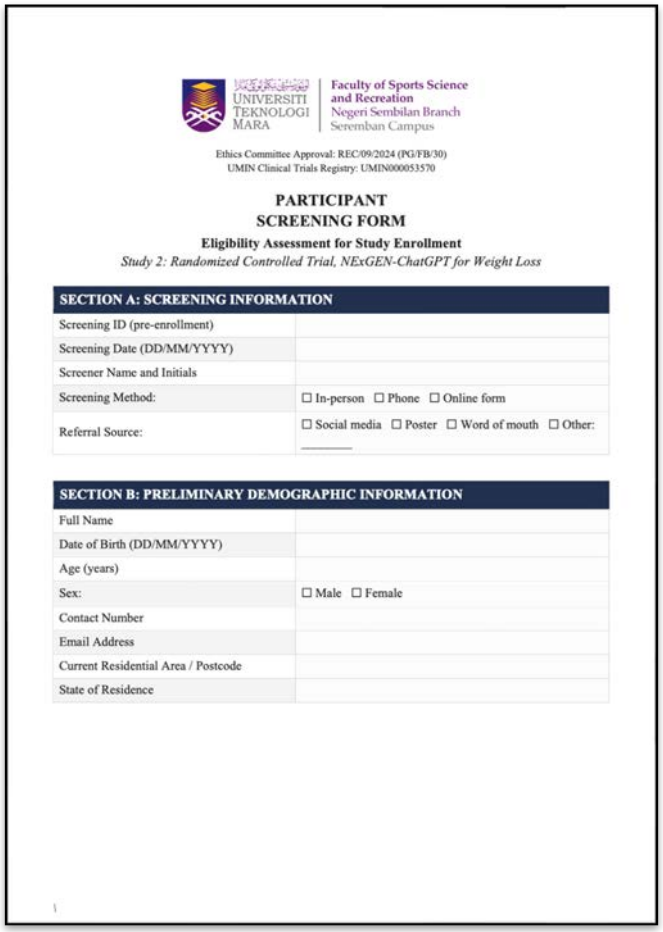
I confirm that all assessment procedures listed above were completed in the prescribed sequence, that all data collection forms were reviewed for completeness, and that any deviations have been documented.

Principal Investigator / Assessor	Independent Reviewer / Supervisor
_____ Name	_____ Name
_____ Signature	_____ Signature
_____ Date	_____ Date

CONFIDENTIAL: This form contains protected health information. Handle in accordance with the approved data management protocol (Ethics REC/09/2024 (PG/FB/30)). Store securely and restrict access to authorized research personnel only.

APPENDIX 14

Participant Screening Form (Study 2)



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

PARTICIPANT SCREENING FORM

Eligibility Assessment for Study Enrollment

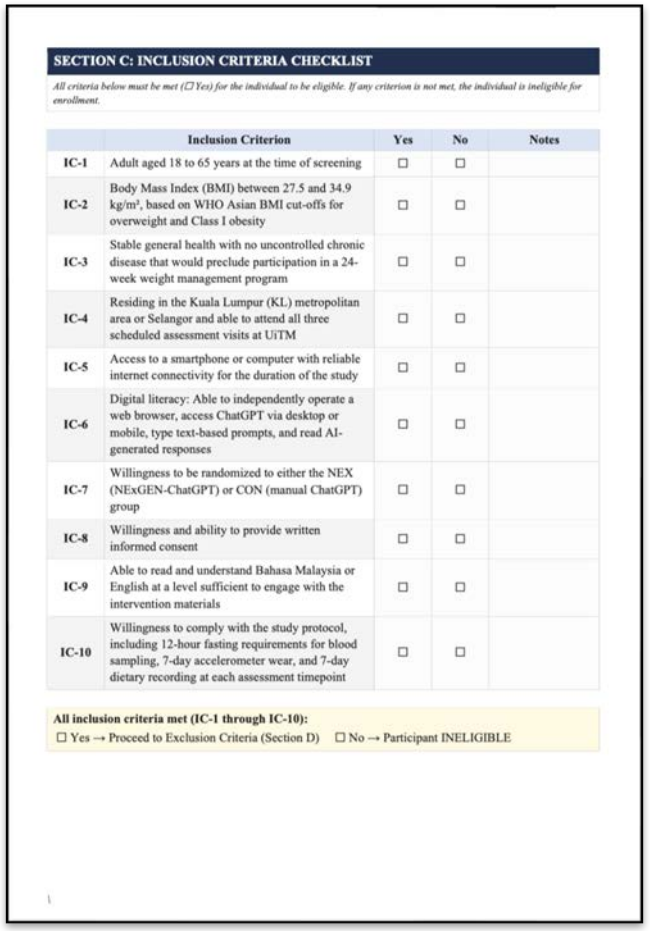
Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: SCREENING INFORMATION

Screening ID (pre-enrollment)	
Screening Date (DD/MM/YYYY)	
Screener Name and Initials	
Screening Method:	☐ In-person ☐ Phone ☐ Online form
Referral Source:	☐ Social media ☐ Poster ☐ Word of mouth ☐ Other: _____

SECTION B: PRELIMINARY DEMOGRAPHIC INFORMATION

Full Name	
Date of Birth (DD/MM/YYYY)	
Age (years)	
Sex:	☐ Male ☐ Female
Contact Number	
Email Address	
Current Residential Area / Postcode	
State of Residence	



SECTION C: INCLUSION CRITERIA CHECKLIST

All criteria below must be met (☐ Yes) for the individual to be eligible. If any criterion is not met, the individual is ineligible for enrollment.

	Inclusion Criterion	Yes	No	Notes
IC-1	Adult aged 18 to 65 years at the time of screening	☐	☐	
IC-2	Body Mass Index (BMI) between 27.5 and 34.9 kg/m², based on WHO Asian BMI cut-offs for overweight and Class I obesity	☐	☐	
IC-3	Stable general health with no uncontrolled chronic disease that would preclude participation in a 24-week weight management program	☐	☐	
IC-4	Residing in the Kuala Lumpur (KL) metropolitan area or Selangor and able to attend all three scheduled assessment visits at UiTM	☐	☐	
IC-5	Access to a smartphone or computer with reliable internet connectivity for the duration of the study	☐	☐	
IC-6	Digital literacy: Able to independently operate a web browser, access ChatGPT via desktop or mobile, type text-based prompts, and read AI-generated responses	☐	☐	
IC-7	Willingness to be randomized to either the NEX (NExGEN-ChatGPT) or CON (manual ChatGPT) group	☐	☐	
IC-8	Willingness and ability to provide written informed consent	☐	☐	
IC-9	Able to read and understand Bahasa Malaysia or English at a level sufficient to engage with the intervention materials	☐	☐	
IC-10	Willingness to comply with the study protocol, including 12-hour fasting requirements for blood sampling, 7-day accelerometer wear, and 7-day dietary recording at each assessment timepoint	☐	☐	

All inclusion criteria met (IC-1 through IC-10):
☐ Yes → Proceed to Exclusion Criteria (Section D) ☐ No → Participant INELIGIBLE

SECTION D: ANTHROPOMETRIC SCREENING MEASUREMENTS

Measured at screening to confirm BMI eligibility (IC-2). If BMI falls outside the eligible range, the individual is ineligible.

Height (cm) — Seca 274 or equivalent	
Weight (kg) — Seca mBCA 515 or equivalent	
Calculated BMI (kg/m²)	

BMI Classification (WHO Asian Cut-offs)

Classification	BMI Range (kg/m²)
Underweight	< 18.5
Normal weight	18.5 – 22.9
Overweight (at risk)	23.0 – 27.4
Eligible: Overweight (obese)	**27.5 – 34.9**
Obese Class II	35.0 – 39.9
Obese Class III	≥ 40.0

BMI within eligible range (27.5 – 34.9 kg/m²): ☐ Yes → Proceed ☐ No → Participant INELIGIBLE

SECTION E: DIGITAL LITERACY AND DEVICE ACCESS ASSESSMENT

Verify participant capacity to engage with the ChatGPT-based intervention.

Device Access

☐ Participant owns or has regular access to a smartphone (Android or iOS)

☐ Participant owns or has regular access to a laptop or desktop computer

☐ Participant has reliable internet access at home or workplace

Digital Skills Verification

☐ Can independently open a web browser and navigate to a specified URL

☐ Can independently create or log into an online account (e.g., email, social media)

☐ Can type text-based messages or queries into a chat interface

☐ Can read and comprehend English or Bahasa Malaysia text displayed on a screen

☐ Can copy, paste, and share text between applications (e.g., ChatGPT to WhatsApp)

☐ Has used or is willing to learn to use ChatGPT or similar AI chatbot platforms

Digital literacy confirmed:
☐ Yes ☐ No → Participant requires digital training prior to enrollment / INELIGIBLE

SECTION F: SCREENING OUTCOME

Final Screening Decision:

☐ **ELIGIBLE** — All inclusion criteria met. Proceed to Exclusion Criteria Checklist (Appendix 14) and Informed Consent.

☐ **INELIGIBLE** — One or more inclusion criteria not met. Reason(s) for ineligibility:

_____

☐ **DEFERRED** — Requires further information or assessment. Rescreening date: _____

Participant informed of screening outcome: ☐ Yes ☐ No	
Date of outcome communication (DD/MM/YYYY)	

SECTION G: SCREENER VERIFICATION AND SIGN-OFF

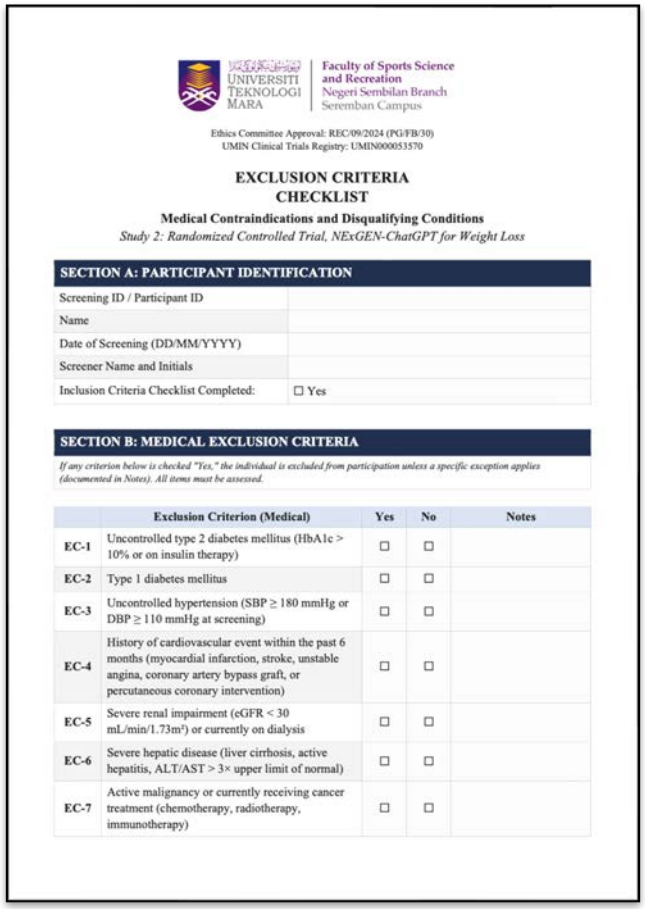
I confirm that this screening form was completed in full, that all eligibility criteria were assessed accurately against the study protocol, and that the screening outcome recorded above is correct.

Principal Investigator / Assessor	Independent Reviewer / Supervisor
_____	_____
Name	Name
_____	_____
Signature	Signature
_____	_____
Date	Date

CONFIDENTIAL: This form contains protected health information. Handle in accordance with the approved data management protocol (Ethics REC/09/2024 (PG/FB/30)). Store securely and restrict access to authorized research personnel only.

APPENDIX 15

Exclusion Criteria Checklist (Study 2)



UNIVERSITI TEKNOLOGI MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

EXCLUSION CRITERIA CHECKLIST

Medical Contraindications and Disqualifying Conditions

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: PARTICIPANT IDENTIFICATION

Screening ID / Participant ID	
Name	
Date of Screening (DD/MM/YYYY)	
Screener Name and Initials	
Inclusion Criteria Checklist Completed:	☐ Yes

SECTION B: MEDICAL EXCLUSION CRITERIA

If any criterion below is checked "Yes," the individual is excluded from participation unless a specific exception applies (documented in Notes). All items must be assessed.

	Exclusion Criterion (Medical)	Yes	No	Notes
EC-1	Uncontrolled type 2 diabetes mellitus (HbA1c > 10% or on insulin therapy)	☐	☐	
EC-2	Type 1 diabetes mellitus	☐	☐	
EC-3	Uncontrolled hypertension (SBP ≥ 180 mmHg or DBP ≥ 110 mmHg at screening)	☐	☐	
EC-4	History of cardiovascular event within the past 6 months (myocardial infarction, stroke, unstable angina, coronary artery bypass graft, or percutaneous coronary intervention)	☐	☐	
EC-5	Severe renal impairment (eGFR < 30 mL/min/1.73m²) or currently on dialysis	☐	☐	
EC-6	Severe hepatic disease (liver cirrhosis, active hepatitis, ALT/AST > 3× upper limit of normal)	☐	☐	
EC-7	Active malignancy or currently receiving cancer treatment (chemotherapy, radiotherapy, immunotherapy)	☐	☐	
EC-8	Diagnosed eating disorder (anorexia nervosa, bulimia nervosa, binge eating disorder) currently under treatment	☐	☐	
EC-9	Severe or uncontrolled psychiatric condition that would impair capacity to engage with the intervention (e.g., active psychosis, severe depression requiring hospitalization, substance use disorder)	☐	☐	
EC-10	Currently pregnant, breastfeeding, or planning pregnancy during the 24-week study period	☐	☐	

SECTION C: MEDICATION AND CONCURRENT INTERVENTION EXCLUSIONS

	Exclusion Criterion (Medication / Intervention)	Yes	No	Notes
EC-11	Currently taking prescription weight-loss medication (e.g., orlistat, liraglutide, semaglutide, phentermine)	☐	☐	
EC-12	Currently using over-the-counter weight-loss supplements with known pharmacological effects on metabolism or appetite	☐	☐	
EC-13	History of bariatric surgery (gastric bypass, sleeve gastrectomy, gastric banding, or any surgical weight-loss procedure)	☐	☐	
EC-14	Currently enrolled in another structured weight-loss program, clinical trial, or dietary intervention	☐	☐	
EC-15	Systemic corticosteroid use for ≥ 14 consecutive days within the past 3 months (excluding inhaled or topical corticosteroids)	☐	☐	
EC-16	Use of medications known to cause significant weight gain (e.g., atypical antipsychotics, certain antidepressants, valproic acid) with dose changes in the past 3 months	☐	☐	

SECTION D: DIGITAL AND LOGISTICAL EXCLUSIONS

	Exclusion Criterion (Digital / Logistical)	Yes	No	Notes
EC-17	Unable to access or operate ChatGPT independently despite assistance and digital training offered	☐	☐	
EC-18	No access to a smartphone, computer, or reliable internet connection for the study duration	☐	☐	

EC-19	Unable to read and understand Bahasa Malaysia or English at a functional level required for the intervention	☐	☐	
EC-20	Unable to attend the three required in-person assessment visits at UiTM (Baseline, Week 12, Week 24)	☐	☐	
EC-21	Planning to relocate outside the Kuala Lumpur / Selangor metropolitan area during the 24-week study period	☐	☐	

SECTION E: EXCLUSION CRITERIA OUTCOME

Total exclusion criteria triggered ("Yes" responses): _____

☐ **NO exclusion criteria met** — Participant ELIGIBLE for enrollment. Proceed to Informed Consent and Randomization.

☐ **One or more exclusion criteria met** — Participant EXCLUDED. Specify criterion code(s):

_____

Participant informed of exclusion decision:	☐ Yes ☐ No
Date informed (DD/MM/YYYY)	

Additional notes / clinical justification for borderline cases:

_____

_____

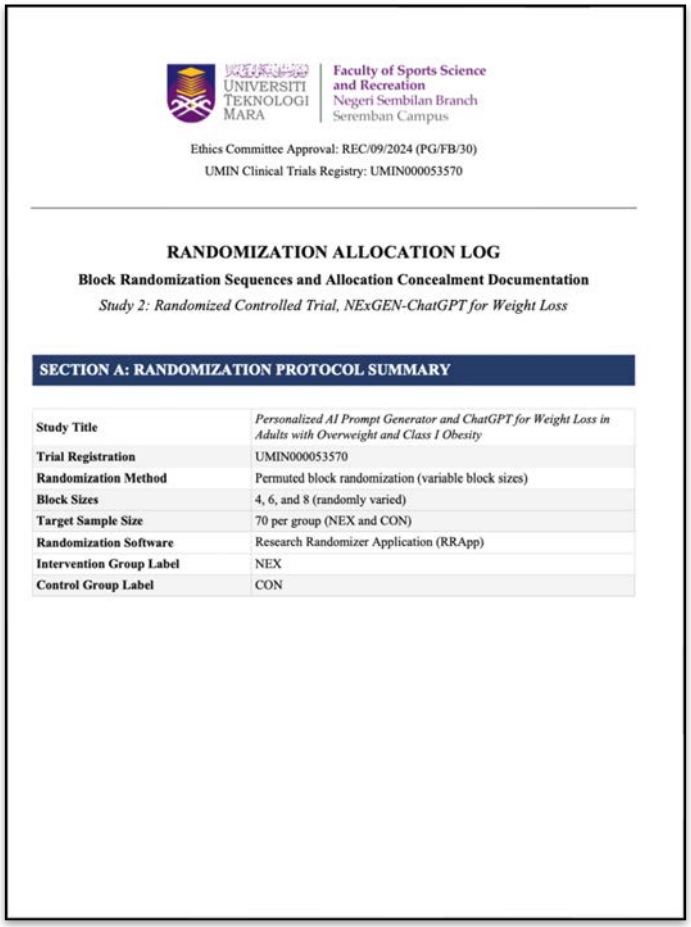
SECTION F: SCREENER VERIFICATION AND SIGN-OFF

I confirm that all exclusion criteria (EC-1 through EC-21) were systematically assessed, that the screening outcome recorded above is accurate, and that any borderline or uncertain cases have been discussed with the Principal Investigator prior to final determination.

Principal Investigator / Assessor	**Independent Reviewer / Supervisor**
_____	_____
Name	Name
_____	_____
Signature	Signature
_____	_____
Date	Date

APPENDIX 16

Randomization Allocation Log (Study 2)



UNIVERSITI TEKNOLOGI MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)

UMIN Clinical Trials Registry: UMIN000053570

RANDOMIZATION ALLOCATION LOG

Block Randomization Sequences and Allocation Concealment Documentation

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: RANDOMIZATION PROTOCOL SUMMARY

Study Title	*Personalized AI Prompt Generator and ChatGPT for Weight Loss in Adults with Overweight and Class I Obesity*
Trial Registration	UMIN000053570
Randomization Method	Permuted block randomization (variable block sizes)
Block Sizes	4, 6, and 8 (randomly varied)
Target Sample Size	70 per group (NEX and CON)
Randomization Software	Research Randomizer Application (RRApp)
Intervention Group Label	NEX
Control Group Label	CON



SECTION B: BLOCK RANDOMIZATION SEQUENCE STRUCTURE

Block No.	Block Size	Sequence Position(s)	Allocation Sequence within Block	Verified	Initials
1				☐	
2				☐	
3				☐	
4				☐	
5				☐	
6				☐	
7				☐	
8				☐	
9				☐	
10				☐	
11				☐	
12				☐	
13				☐	
14				☐	
15				☐	
16				☐	
17				☐	
18				☐	
19				☐	
20				☐	
21				☐	
22				☐	
23				☐	
24				☐	
25				☐	

[Continue until target sample size met]

Total blocks: _____ Total sequence positions generated: _____ Ratio check (NEX:CON): _____

SECTION C: PARTICIPANT ALLOCATION LOG

Seq. No.	Participant ID	Date Randomized	Block No.	Group Allocated	Envelope No. / RRApp Ref.	Allocated by (Initials)	Concealment Intact ✓	Notes
1							☐	
2							☐	
3							☐	
4							☐	
5							☐	
6							☐	
7							☐	
8							☐	
9							☐	
10							☐	
11							☐	
12							☐	
13							☐	
14							☐	
15							☐	
16							☐	
17							☐	
18							☐	
19							☐	
20							☐	
21							☐	
22							☐	
23							☐	
24							☐	
25							☐	

[Continue until target sample size met]

SECTION D: ALLOCATION CONCEALMENT PROCEDURES

D1. CONCEALMENT METHOD

Indicate the method(s) used to conceal the allocation sequence from research personnel involved in screening and enrollment:

- ☐ Sequentially numbered, opaque, sealed envelopes (SNOSE)
- ☐ Central randomization via RRApp (web-based, accessed only at point of allocation)
- ☐ Password-protected electronic file accessible only to the allocation administrator
- ☐ Independent third-party allocation (specify): _____
- ☐ Other (specify): _____

Role	Details
D2a. Sequence Generator	Name: _____ Role: _____ Independent of enrollment: ☐ Yes ☐ No
D2b. Allocation Administrator	Name: _____ Role: _____ Independent of enrollment: ☐ Yes ☐ No
D2c. Enrollment Assessor	Name: _____ Blinded to upcoming allocation: ☐ Yes ☐ No
D2d. Outcome Assessor	Name: _____ Blinded to group allocation: ☐ Yes ☐ No ☐ Partial (specify): _____

D3. CONCEALMENT INTEGRITY CHECKLIST

Confirm each item by ticking (✓) the appropriate column. All items should be confirmed "Yes" for adequate concealment.

No.	Concealment Integrity Item	Yes	No	N/A
1	The randomization sequence was generated before the first participant was enrolled.	☐	☐	☐
2	The allocation sequence is stored in a secure location inaccessible to the enrollment assessor.	☐	☐	☐
3	The enrollment assessor cannot predict the next allocation before a participant is registered.	☐	☐	☐
4	Allocation is revealed only after the participant has completed screening, met all eligibility criteria, and provided written informed consent.	☐	☐	☐
5	Sequentially numbered envelopes (if used) are opaque, tamper-evident, and opened only in sequence.	☐	☐	☐
6	No participant has been allocated out of sequence.	☐	☐	☐
7	The allocation administrator has no role in participant screening or eligibility assessment.	☐	☐	☐
8	No research personnel involved in outcome assessment have access to the allocation log during the study.	☐	☐	☐
9	The randomization sequence was verified against the RRApp output by an independent person.	☐	☐	☐
10	Any protocol deviation in the allocation process has been documented in Section E.	☐	☐	☐

If any item is marked "No," document the circumstances and corrective actions in Section E (Protocol Deviations).

SECTION E: RANDOMIZATION PROTOCOL DEVIATIONS

Document any deviation from the randomization or allocation concealment protocol. Examples include: out-of-sequence allocation, accidental unblinding of upcoming allocation, envelope opened prematurely, participant allocated before eligibility confirmed, or allocation sequence accessed by unauthorized personnel.

Dev. No.	Date Identified	Sequence / Participant Affected	Description of Deviation	Corrective Action Taken	Reported to PI
1					☐ Yes ☐ No
2					☐ Yes ☐ No
3					☐ Yes ☐ No
4					☐ Yes ☐ No
5					☐ Yes ☐ No

SECTION F: FINAL ENROLLMENT AND ALLOCATION SUMMARY

To be completed after the final participant has been randomized.

Parameter	Value
Enrollment period (first to last randomization)	From: _____ To: _____
Total participants screened	
Total participants eligible	
Total participants randomized	
Allocated to NEX (Intervention) Group	n = _____
Allocated to CON (Control) Group	n = _____
Allocation ratio achieved	_____ : _____ (Target: 1:1)
Number of blocks used	_____ of _____ generated
Unused sequence positions remaining	
Total protocol deviations (randomization)	
Allocation concealment maintained throughout	☐ Yes ☐ No (document in Section E)

SECTION G: VERIFICATION AND SIGN-OFF

SEQUENCE VERIFICATION

I confirm that I have independently verified the randomization sequence recorded in this log against the original RRApp output, and that all entries in the Participant Allocation Log (Section C) are consistent with the sequence documented in Section B.

Independent Verifier Name	*Signature*	*Date*

Designation: _____

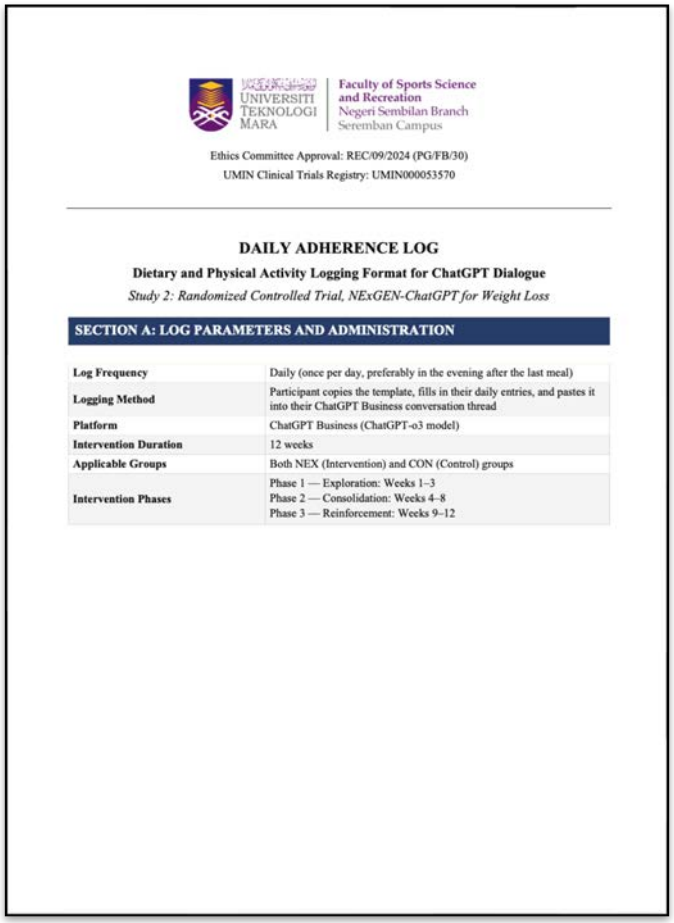
PRINCIPAL INVESTIGATOR APPROVAL

I confirm that the randomization and allocation concealment procedures documented in this log were conducted in accordance with the approved study protocol (REC/09/2024, PG/FB/30). I take responsibility for the integrity of the randomization process and confirm that this log is a complete and accurate record of participant allocation for the NExGEN-ChatGPT randomized controlled trial.

Principal Investigator Name	*Signature*	*Date*

APPENDIX 17

Daily Adherence Log (Study 2)



Universiti Teknologi MARA | Faculty of Sports Science and Recreation
Negeri Sembilan Branch
Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

DAILY ADHERENCE LOG

Dietary and Physical Activity Logging Format for ChatGPT Dialogue

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: LOG PARAMETERS AND ADMINISTRATION

Log Frequency	Daily (once per day, preferably in the evening after the last meal)
Logging Method	Participant copies the template, fills in their daily entries, and pastes it into their ChatGPT Business conversation thread
Platform	ChatGPT Business (ChatGPT-o3 model)
Intervention Duration	12 weeks
Applicable Groups	Both NEX (Intervention) and CON (Control) groups
Intervention Phases	Phase 1 — Exploration: Weeks 1–3 Phase 2 — Consolidation: Weeks 4–8 Phase 3 — Reinforcement: Weeks 9–12



SECTION B: DAILY ADHERENCE LOG TEMPLATE

B1. PARTICIPANT IDENTIFICATION HEADER

DAILY LOG
Date:
Day of Week: [Monday / Tuesday / Wednesday / Thursday / Friday / Saturday / Sunday]
Intervention Week: [1–12]
Phase: [Exploration / Consolidation / Reinforcement]

Date	
Day of Week	
Intervention Week	
Intervention Duration	
Phase	

B2. DIETARY INTAKE LOG

INSTRUCTIONS FOR DIETARY LOGGING
Record everything you ate and drank today, including all meals, snacks, and beverages. For each item, describe what you consumed and estimate the portion size using household measures (e.g., 1 cup, 1 plate, palm-sized, 2 tablespoons) or familiar references. Include cooking methods where relevant (e.g., fried, steamed, grilled). Be as specific as possible with food descriptions (e.g., "brown rice" instead of "rice," "chicken breast grilled" instead of "chicken").

MEALS AND FOOD INTAKE

BREAKFAST (Time: ___:___ AM/PM)
1. [Food/drink item] — [Portion size] — [Cooking method if applicable]
2. [Food/drink item] — [Portion size]
3. ...

MORNING SNACK (Time: ___:___ AM/PM)
1. [Food/drink item] — [Portion size]
2. ...
(If no snack, write: None)

LUNCH (Time: ___:___ PM)
1. [Food/drink item] — [Portion size] — [Cooking method if applicable]
2. [Food/drink item] — [Portion size]
3. ...

AFTERNOON SNACK (Time: ___:___ PM)
1. [Food/drink item] — [Portion size]
2. ...
(If no snack, write: None)

DINNER (Time: ___:___ PM)

1. [Food/drink item] — [Portion size] — [Cooking method if applicable]
2. [Food/drink item] — [Portion size]
3. ...

EVENING SNACK / SUPPER (Time: ___:___ PM)
1. [Food/drink item] — [Portion size]
(If no snack, write: None)

BEVERAGES (throughout the day)
Water: _____ glasses (approximately _____ mL total)
Other beverages:
1. [Beverage] — [Amount] — [Sweetened? Yes/No]
2. ...

DIETARY NOTES
Ate out or home-cooked? [Home / Eating out / Mixed]
Followed planned meals? [Yes / Partially / No]
Any unplanned eating or cravings? [Describe briefly or write None]

BREAKFAST (Time: ___:___ AM/PM)	
MORNING SNACK (Time: ___:___ AM/PM)	
LUNCH (Time: ___:___ AM/PM)	
AFTERNOON SNACK (Time: ___:___ AM/PM)	
DINNER (Time: ___:___ AM/PM)	
EVENING SNACK / SUPPER (Time: ___:___ AM/PM)	

BEVERAGES (throughout the day) (Time: ___:___ AM/PM)	
BEVERAGES (throughout the day) (Time: ___:___ AM/PM)	
DIETARY NOTES (Time: ___:___ AM/PM)	

B3. PHYSICAL ACTIVITY LOG

INSTRUCTIONS FOR PHYSICAL ACTIVITY LOGGING
Record all planned exercise sessions and incidental physical activity for today. For each activity, specify the type, duration, and your perceived intensity. Use the talk test as a guide: moderate intensity means you can talk but not sing during the activity; vigorous intensity means you can only say a few words before needing to catch your breath.

PHYSICAL ACTIVITY

PLANNED EXERCISE SESSION(S)
Activity 1:
Type: [e.g., brisk walking, jogging, cycling, swimming, resistance training, group fitness class]
Duration: _____ minutes
Intensity: [Light / Moderate / Vigorous]
Location: [e.g., home, gym, park, neighborhood]
Notes: [any relevant details, e.g., completed full session, modified due to fatigue]

Activity 2 (if applicable):
Type: [e.g., resistance training, yoga, stretching]
Duration: _____ minutes
Intensity: [Light / Moderate / Vigorous]
Notes: [e.g., exercises performed, sets/reps if resistance training]

(If no planned exercise today, write: Rest day)

INCIDENTAL ACTIVITY
Steps (if tracked): approximately _____ steps
Other activity: [e.g., walked to work (20 min), took stairs, housework (30 min), gardening]
(If none beyond normal routine, write: Routine daily activity only)

WEEKLY EXERCISE PROGRESS (running total)
Aerobic minutes this week so far: _____ / 150 minutes
Resistance training sessions this week so far: _____ / 2 sessions

PLANNED EXERCISE SESSION(S)	
INCIDENTAL ACTIVITY	
WEEKLY EXERCISE PROGRESS (running total)	

B4. BEHAVIORAL AND WELLNESS INDICATORS

INSTRUCTIONS FOR BEHAVIORAL LOGGING
These indicators help track overall wellness patterns that influence weight management outcomes. Complete briefly each day. Honest reporting supports better AI-driven feedback and long-term progress tracking.

DAILY WELLNESS CHECK

Sleep last night: _____ hours _____ minutes
Sleep quality: [Poor / Fair / Good / Very Good]
Energy level today: [Low / Moderate / High]
Stress level today: [Low / Moderate / High]
Overall mood: [Positive / Neutral / Negative]

ADHERENCE SELF-RATING

How well did you follow your dietary plan today? [1–10 scale]
Rating: _____ / 10
How well did you follow your exercise plan today? [1–10 scale]
Rating: _____ / 10

COMMENTS / CHALLENGES / QUESTIONS

[Write any comments, challenges faced today, questions for ChatGPT, or anything else you want to share]

DAILY WELLNESS CHECK	
ADHERENCE SELF-RATING	
COMMENTS / CHALLENGES / QUESTIONS	

SECTION C: SAMPLE COMPLETED DAILY LOG

The following example illustrates a properly completed daily log entry to serve as a reference for participants.

DAILY LOG
Date: 15/08/2024
Day of Week: Thursday
Intervention Week: 3
Phase: Exploration

MEALS AND FOOD INTAKE

BREAKFAST (Time: 7:30 AM)
1. Wholemeal bread — 2 slices — toasted
2. Peanut butter — 1 tablespoon
3. Banana — 1 medium
4. Teh O kosong (unsweetened tea) — 1 cup

MORNING SNACK (Time: 10:30 AM)
1. Apple — 1 medium

LUNCH (Time: 1:00 PM)
1. Brown rice — 1 cup
2. Ikan tenggiri (mackerel) — 1 palm-sized piece — grilled
3. Stir-fried kangkung with garlic — 1 cup — minimal oil
4. Cucumber slices — half cup
5. Plain water — 1 glass

AFTERNOON SNACK (Time: 4:00 PM)
1. Low-fat yogurt — 1 small tub (125g)
2. Mixed nuts (almonds, cashews) — 1 small handful (about 15g)

DINNER (Time: 7:30 PM)
1. Chicken breast — 1 palm-sized piece — steamed
2. White rice — 3/4 cup
3. Sayur campur (mixed vegetable soup) — 1 bowl
4. Tauhu (tofu) — 2 small pieces — in soup

EVENING SNACK / SUPPER
None

BEVERAGES (throughout the day)
Water: 8 glasses (approximately 2000 mL total)
Other beverages:
1. Teh O kosong — 1 cup — Sweetened? No
2. Black coffee — 1 cup — Sweetened? No

DIETARY NOTES
Ate out or home-cooked? Home (breakfast, dinner) / Eating out (lunch at office cafeteria)
Followed planned meals? Partially — lunch was slightly different from plan, substituted fish type
Any unplanned eating or cravings? Had a mild craving for fried snacks at 3 PM but had yogurt instead

PHYSICAL ACTIVITY

PLANNED EXERCISE SESSION(S)
Activity 1:
Type: Brisk walking
Duration: 35 minutes
Intensity: Moderate (could talk but not sing)
Location: Neighborhood park
Notes: Completed full session, felt good

Activity 2: None today

INCIDENTAL ACTIVITY
Steps (if tracked): approximately 7,200 steps
Other activity: Walked to LRT station and back (15 min each way), took stairs at office (3 floors)

WEEKLY EXERCISE PROGRESS (running total)
Aerobic minutes this week so far: 95 / 150 minutes
Resistance training sessions this week so far: 1 / 2 sessions

DAILY WELLNESS CHECK

Sleep last night: 7 hours 15 minutes
Sleep quality: Good
Energy level today: Moderate
Stress level today: Moderate (busy work day)
Overall mood: Positive

ADHERENCE SELF-RATING

How well did you follow your dietary plan today?
Rating: 7 / 10
How well did you follow your exercise plan today?
Rating: 8 / 10

COMMENTS / CHALLENGES / QUESTIONS
Lunch options at the cafeteria were limited today. I tried to choose the healthiest option available. Should I bring lunch from home on days when the cafeteria menu is mostly fried foods?

---END OF DAILY LOG---

SECTION D: WEEKLY SUMMARY LOG (END-OF-WEEK ENTRY)

In addition to the daily log, participants complete a brief weekly summary every Sunday (or on their last day of the week). This summary is pasted into the ChatGPT conversation after the daily log for that day.

WEEKLY SUMMARY — WEEK __
Date range: [DD/MM/YYYY] to [DD/MM/YYYY]
Phase: [Exploration / Consolidation / Reinforcement]

EXERCISE SUMMARY
Total aerobic minutes completed: _____ / 150 minutes target
Resistance training sessions completed: _____ / 2 sessions target
Total days with planned exercise: _____ / 7 days

DIETARY SUMMARY
Days I followed the meal plan well (rating ≥7): _____ / 7 days
Average daily water intake: _____ glasses
Biggest dietary challenge this week: [describe briefly]
Biggest dietary success this week: [describe briefly]

BODY WEIGHT (if self-monitoring)
Weight this week: _____ kg (measured on: [day], [time], [condition: e.g., morning, before breakfast])
(Self-weighing is optional and not required by the study protocol)

WELLNESS SUMMARY
Average sleep duration this week: approximately _____ hours/night
Overall energy level this week: [Low / Moderate / High]
Overall stress level this week: [Low / Moderate / High]

SELF-REFLECTION
What went well this week?
[Write 1–2 sentences]

What could be improved next week?
[Write 1–2 sentences]

Any questions or support needed from ChatGPT?
[Write your questions or type None]

EXERCISE SUMMARY	
DIETARY SUMMARY	
BODY WEIGHT (if self-monitoring)	
WELLNESS SUMMARY	
SELF-REFLECTION	

SECTION E: PHASE-SPECIFIC LOGGING EXPECTATIONS

Logging detail and focus shift across the three intervention phases to align with the progressive skill-building framework.

Phase	Logging Focus and Expectations
Phase 1 **EXPLORATION** *Weeks 1–3*	Primary focus: Establishing the daily logging habit and building awareness of current intake and activity patterns. Participants focus on completeness and honesty of reporting rather than perfection of adherence. All food items and activities should be recorded, including deviations from the plan. The log itself is the goal during this phase. ChatGPT (NEX) provides educational feedback, clarifies portions, and reinforces the logging habit.
Phase 2 **CONSOLIDATION** *Weeks 4–8*	Primary focus: Strengthening adherence patterns and refining dietary and exercise behaviors. Participants are expected to log consistently and with increasing specificity (e.g., more precise portion estimates, noting cooking methods). Self-rating scores (1–10) become a focus for weekly improvement. ChatGPT (NEX) identifies patterns from accumulated logs and provides targeted behavioral coaching.
Phase 3 **REINFORCEMENT** *Weeks 9–12*	Primary focus: Sustaining habits autonomously and preparing for post-intervention maintenance. Participants demonstrate internalized self-monitoring skills. Logging remains daily but self-reflection components (challenges, successes, weekly improvements) become more prominent. ChatGPT (NEX) reinforces long-term habit sustainability, reviews progress across all weeks, and prepares participants for the follow-up period without active AI support.

SECTION F: LOG COMPLIANCE TRACKING AND SIGN-OFF

F1. PARTICIPANT-LEVEL LOG COMPLIANCE RECORD

Metric	Weeks 1–3	Weeks 4–8	Weeks 9–12	Total	Rate	Notes
Total possible logging days	21	35	28	**84**	—	
Days with completed daily logs					____%	
Days with dietary section completed					____%	
Days with PA section completed					____%	
Weekly summaries completed	___ / 3	___ / 5	___ / 4	___ / 12	____%	
Overall log compliance classification	☐ High (≥80%) ☐ Moderate (50–79%) ☐ Low (<50%)					

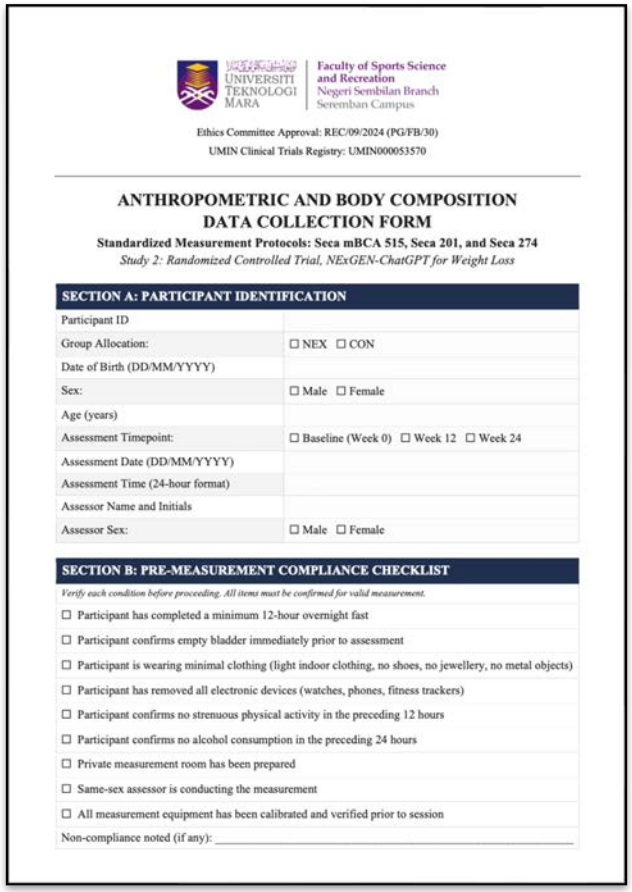
F2. VERIFICATION AND SIGN-OFF

I confirm that this participant's ChatGPT conversation history was reviewed for daily adherence log entries, that the compliance counts above are accurate based on the conversation record, and that the log template was provided to the participant in its standardized format at the baseline visit.

Research Personnel Name | *Signature* | *Date*

APPENDIX 18

Anthropometric and Body Composition Assesement



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

ANTHROPOMETRIC AND BODY COMPOSITION DATA COLLECTION FORM

Standardized Measurement Protocols: Seca mBCA 515, Seca 201, and Seca 274

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: PARTICIPANT IDENTIFICATION

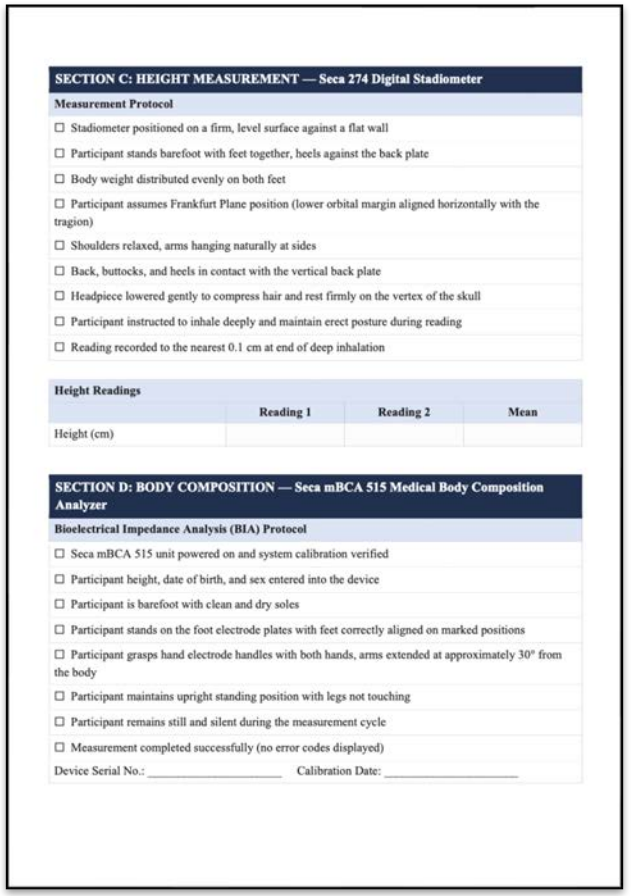
Participant ID	
Group Allocation:	☐ NEX ☐ CON
Date of Birth (DD/MM/YYYY)	
Sex:	☐ Male ☐ Female
Age (years)	
Assessment Timepoint:	☐ Baseline (Week 0) ☐ Week 12 ☐ Week 24
Assessment Date (DD/MM/YYYY)	
Assessment Time (24-hour format)	
Assessor Name and Initials	
Assessor Sex:	☐ Male ☐ Female

SECTION B: PRE-MEASUREMENT COMPLIANCE CHECKLIST

Verify each condition before proceeding. All items must be confirmed for valid measurement.

- ☐ Participant has completed a minimum 12-hour overnight fast
- ☐ Participant confirms empty bladder immediately prior to assessment
- ☐ Participant is wearing minimal clothing (light indoor clothing, no shoes, no jewellery, no metal objects)
- ☐ Participant has removed all electronic devices (watches, phones, fitness trackers)
- ☐ Participant confirms no strenuous physical activity in the preceding 12 hours
- ☐ Participant confirms no alcohol consumption in the preceding 24 hours
- ☐ Private measurement room has been prepared
- ☐ Same-sex assessor is conducting the measurement
- ☐ All measurement equipment has been calibrated and verified prior to session

Non-compliance noted (if any): _____



SECTION C: HEIGHT MEASUREMENT — Seca 274 Digital Stadiometer

Measurement Protocol

- ☐ Stadiometer positioned on a firm, level surface against a flat wall
- ☐ Participant stands barefoot with feet together, heels against the back plate
- ☐ Body weight distributed evenly on both feet
- ☐ Participant assumes Frankfurt Plane position (lower orbital margin aligned horizontally with the tragion)
- ☐ Shoulders relaxed, arms hanging naturally at sides
- ☐ Back, buttocks, and heels in contact with the vertical back plate
- ☐ Headpiece lowered gently to compress hair and rest firmly on the vertex of the skull
- ☐ Participant instructed to inhale deeply and maintain erect posture during reading
- ☐ Reading recorded to the nearest 0.1 cm at end of deep inhalation

Height Readings

	Reading 1	Reading 2	Mean
Height (cm)			

SECTION D: BODY COMPOSITION — Seca mBCA 515 Medical Body Composition Analyzer

Bioelectrical Impedance Analysis (BIA) Protocol

- ☐ Seca mBCA 515 unit powered on and system calibration verified
- ☐ Participant height, date of birth, and sex entered into the device
- ☐ Participant is barefoot with clean and dry soles
- ☐ Participant stands on the foot electrode plates with feet correctly aligned on marked positions
- ☐ Participant grasps hand electrode handles with both hands, arms extended at approximately 30° from the body
- ☐ Participant maintains upright standing position with legs not touching
- ☐ Participant remains still and silent during the measurement cycle
- ☐ Measurement completed successfully (no error codes displayed)

Device Serial No.: _____ Calibration Date: _____

Body Composition Readings			
Parameter	**Baseline**	**Week 12**	**Week 24**
Body Weight (kg)			
Body Mass Index (kg/m²)			
Fat Mass (kg)			
Fat Mass Index (kg/m²)			
Body Fat Percentage (%)			
Fat-Free Mass (kg)			
Fat-Free Mass Index (kg/m²)			
Skeletal Muscle Mass (kg)			
Total Body Water (L)			
Visceral Adipose Tissue (L)			

SECTION E: WAIST CIRCUMFERENCE — Seca 201 Ergonomic Measuring Tape

Measurement Protocol (WHO Procedure)

- ☐ Participant standing upright with arms relaxed at sides and feet positioned together
- ☐ Measurement area exposed (shirt lifted/removed) and abdomen is relaxed
- ☐ Assessor palpates and identifies the lower border of the last rib and the top of the iliac crest on each side
- ☐ Measurement landmark identified at the midpoint between the lower rib margin and the iliac crest
- ☐ Seca 201 tape placed horizontally around the abdomen at the identified midpoint
- ☐ Tape is snug against the skin without compressing underlying tissue
- ☐ Tape alignment verified: horizontal and parallel to the floor at all points
- ☐ Participant instructed to breathe normally; reading taken at the end of a gentle exhalation
- ☐ Reading recorded to the nearest 0.1 cm
- ☐ Measurement repeated; third measurement taken if first two differ by more than 1.0 cm

Waist Circumference Readings				
	Reading 1	**Reading 2**	**Reading 3***	**Mean**
Waist Circumference (cm)				

**Reading 3 required only if Readings 1 and 2 differ by > 1.0 cm. Mean calculated from the two closest readings.*

SECTION F: DERIVED CALCULATIONS (ASSESSOR USE)

Parameter	Formula	Value
BMI (kg/m²)	*Weight (kg) / Height (m)²*	
BMI Classification	*WHO Asian BMI Cut-offs*	

Eligibility Confirmed (BMI 27.5 – 34.9 kg/m²): ☐ Yes ☐ No

SECTION G: QUALITY ASSURANCE AND ASSESSOR SIGN-OFF

Measurement Environment Conditions	
Room Temperature (°C)	
Measurement Room Number / Location	

Data Quality Verification

- ☐ All required measurements have been completed and recorded
- ☐ All readings fall within plausible physiological ranges
- ☐ Any anomalous readings have been repeated and verified
- ☐ Participant was comfortable throughout the assessment procedure

Deviations from Standard Protocol (describe): _____

_____

SECTION H: VERIFICATION AND SIGN-OFF

I confirm that this form accurately records all anthropometric and body composition measurements obtained during this assessment session, that all measurements were performed according to the standardized protocols specified above, and that this documentation is complete and consistent with the participant's Case Report Form.

Assessor

_____
Name

_____
Signature

_____
Date

Data verified by (independent reviewer):

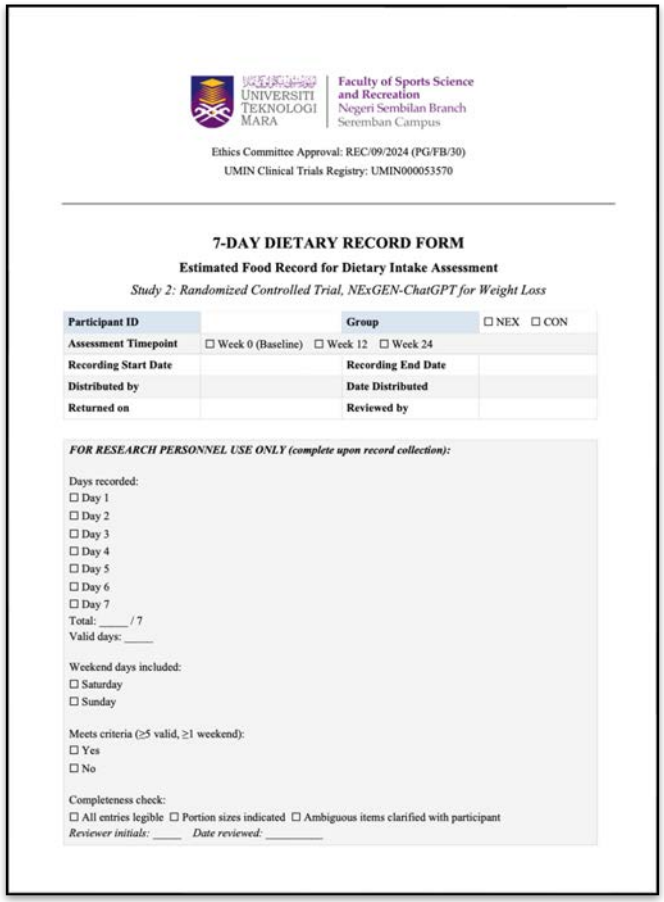
_____
Name

_____
Signature

_____
Date

APPENDIX 19

7-Day Dietary Record Form (Study 2)



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

7-DAY DIETARY RECORD FORM

Estimated Food Record for Dietary Intake Assessment

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

Participant ID		**Group**	☐ NEX ☐ CON
Assessment Timepoint	☐ Week 0 (Baseline) ☐ Week 12 ☐ Week 24		
Recording Start Date		**Recording End Date**	
Distributed by		**Date Distributed**	
Returned on		**Reviewed by**	

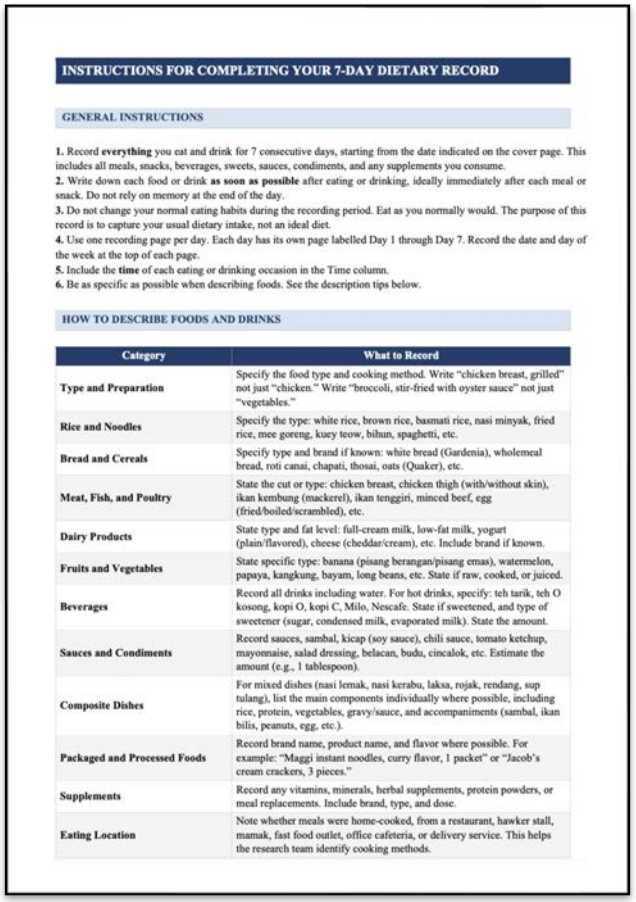
FOR RESEARCH PERSONNEL USE ONLY (complete upon record collection):

Days recorded:
☐ Day 1
☐ Day 2
☐ Day 3
☐ Day 4
☐ Day 5
☐ Day 6
☐ Day 7
Total: ____ / 7
Valid days: ____

Weekend days included:
☐ Saturday
☐ Sunday

Meets criteria (≥5 valid, ≥1 weekend):
☐ Yes
☐ No

Completeness check:
☐ All entries legible ☐ Portion sizes indicated ☐ Ambiguous items clarified with participant
Reviewer initials: ____ *Date reviewed:* _____



INSTRUCTIONS FOR COMPLETING YOUR 7-DAY DIETARY RECORD

GENERAL INSTRUCTIONS

1. Record **everything** you eat and drink for 7 consecutive days, starting from the date indicated on the cover page. This includes all meals, snacks, beverages, sweets, sauces, condiments, and any supplements you consume.
2. Write down each food or drink **as soon as possible** after eating or drinking, ideally immediately after each meal or snack. Do not rely on memory at the end of the day.
3. Do not change your normal eating habits during the recording period. Eat as you normally would. The purpose of this record is to capture your usual dietary intake, not an ideal diet.
4. Use one recording page per day. Each day has its own page labelled Day 1 through Day 7. Record the date and day of the week at the top of each page.
5. Include the **time** of each eating or drinking occasion in the Time column.
6. Be as specific as possible when describing foods. See the description tips below.

HOW TO DESCRIBE FOODS AND DRINKS

Category	What to Record
Type and Preparation	Specify the food type and cooking method. Write "chicken breast, grilled" not just "chicken." Write "broccoli, stir-fried with oyster sauce" not just "vegetables."
Rice and Noodles	Specify the type: white rice, brown rice, basmati rice, nasi minyak, fried rice, mee goreng, kuey teow, bihun, spaghetti, etc.
Bread and Cereals	Specify type and brand if known: white bread (Gardenia), wholemeal bread, roti canai, chapati, thosai, oats (Quaker), etc.
Meat, Fish, and Poultry	State the cut or type: chicken breast, chicken thigh (with/without skin), ikan kembung (mackerel), ikan tenggiri, minced beef, egg (fried/boiled/scrambled), etc.
Dairy Products	State type and fat level: full-cream milk, low-fat milk, yogurt (plain/flavored), cheese (cheddar/cream), etc. Include brand if known.
Fruits and Vegetables	State specific type: banana (pisang berangan/pisang emas), watermelon, papaya, kangkung, bayam, long beans, etc. State if raw, cooked, or juiced.
Beverages	Record all drinks including water. For hot drinks, specify: teh tarik, teh O kosong, kopi O, kopi C, Milo, Nescafe. State if sweetened, and type of sweetener (sugar, condensed milk, evaporated milk). State the amount.
Sauces and Condiments	Record sauces, sambal, kicap (soy sauce), chili sauce, tomato ketchup, mayonnaise, salad dressing, belacan, budu, cincalok, etc. Estimate the amount (e.g., 1 tablespoon).
Composite Dishes	For mixed dishes (nasi lemak, nasi kerabu, laksa, rojak, rendang, sup tulang), list the main components individually where possible, including rice, protein, vegetables, gravy/sauce, and accompaniments (sambal, ikan bilis, peanuts, egg, etc.).
Packaged and Processed Foods	Record brand name, product name, and flavor where possible. For example: "Maggi instant noodles, curry flavor, 1 packet" or "Jacob's cream crackers, 3 pieces."
Supplements	Record any vitamins, minerals, herbal supplements, protein powders, or meal replacements. Include brand, type, and dose.
Eating Location	Note whether meals were home-cooked, from a restaurant, hawker stall, mamak, fast food outlet, office cafeteria, or delivery service. This helps the research team identify cooking methods.

PORTION SIZE ESTIMATION GUIDE

Estimate the amount of each food or drink you consumed using any of the following references. You do not need to weigh your food.

Reference	How to Use
Household Measures	Use cups, tablespoons (sudu besar), teaspoons (sudu kecil), glasses, bowls, and plates as references. For example: "1 cup rice," "2 tablespoons sambal," "1 glass (250 mL) water."
Hand and Finger Guides	Palm (no fingers) = 1 serving of meat, chicken, or fish (approximately 90g cooked) Fist = 1 cup (approximately 1 serving of rice, noodles, or cut fruit) Cupped hand = 1 serving of cereal, nuts, or small snacks Thumb tip = 1 teaspoon (approximately 5g of butter, sugar, or oil) Thumb (full) = 1 tablespoon (approximately 15g)
Standard Portions	1 plate of rice = approximately 2 cups (common Malaysian plate serving) 1 senduk (ladle) of rice = approximately 1/2 cup 1 slice of bread = 1 serving 1 piece of roti canai = 1 piece (specify if plain, egg, or tissue) 1 glass of water or beverage = approximately 250 mL 1 small bowl of soup = approximately 250 mL
Comparative Objects	Deck of cards = 1 serving of meat (approximately 90g) Tennis ball = 1 medium fruit (apple, orange) Matchbox = 1 serving of cheese (approximately 30g) Golf ball = 2 tablespoons (peanut butter, jam, dressing)
Fractions and Multiples	You may describe amounts as fractions (½ plate, ¼ cup, ¾ glass) or multiples (2 cups, 3 tablespoons). Be as precise as you can, but estimates are acceptable.

DAY 1

Date:	____ / ____ / _____	**Day:**	Monday	☐ *Weekday*

Time	BREAKFAST — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	MORNING SNACK — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	LUNCH — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	AFTERNOON SNACK — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	DINNER — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	EVENING SNACK / SUPPER — Food / Drink Description	Amount / Portion	Cooking Method / Notes

Time	BEVERAGES (THROUGHOUT THE DAY) — Food / Drink Description	Amount / Portion	Cooking Method / Notes

DAILY NOTES

Meals eaten at:

☐ Home
☐ Restaurant / Hawker
☐ Office / School Cafeteria
☐ Fast Food
☐ Delivery
☐ Other: _____

Was today a typical day for your eating? ☐ Yes ☐ No (If No, explain: _____)

Did you take any supplements today? ☐ No ☐ Yes (Name and dose: _____)

Additional notes: _____

_____

_____

_____

_____

RECORD COMPLETION AND SUBMISSION

PARTICIPANT CHECKLIST (complete before returning this form)

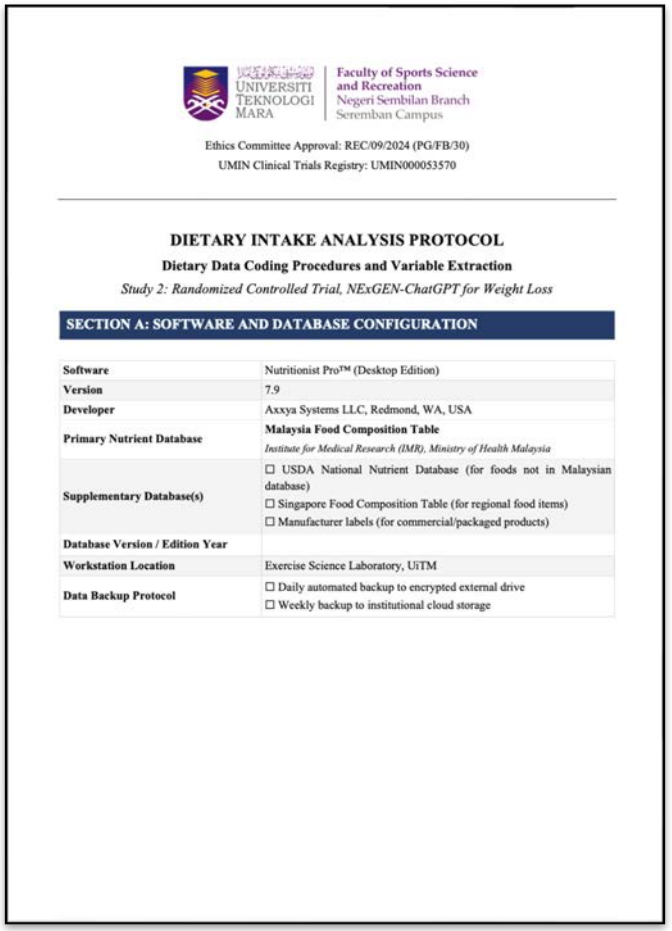
No.	Item	✓
1	I have recorded my food and drink intake for all 7 days.	☐
2	I have included at least 1 weekend day (Saturday or Sunday) in my records.	☐
3	I have recorded all meals, snacks, and beverages including water.	☐
4	I have included portion sizes or amounts for each food and drink item.	☐
5	I have described foods specifically (e.g., "brown rice" not just "rice").	☐
6	I have noted cooking methods where applicable (fried, grilled, steamed, etc.).	☐
7	I have recorded sauces, condiments, and additions (sugar in drinks, sambal, etc.).	☐
8	I have recorded the time of each eating or drinking occasion.	☐
9	I have noted whether meals were home-cooked or eaten out.	☐
10	I have recorded any supplements taken.	☐

I confirm that I have completed this dietary record to the best of my ability, recording all food and drink consumed over the 7-day period as accurately and honestly as possible.

Participant Name _____ *Signature* _____ *Date* _____

APPENDIX 20

Dietary Intake Analysis Protocol (Study 2)



UNIVERSITI TEKNOLOGI MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

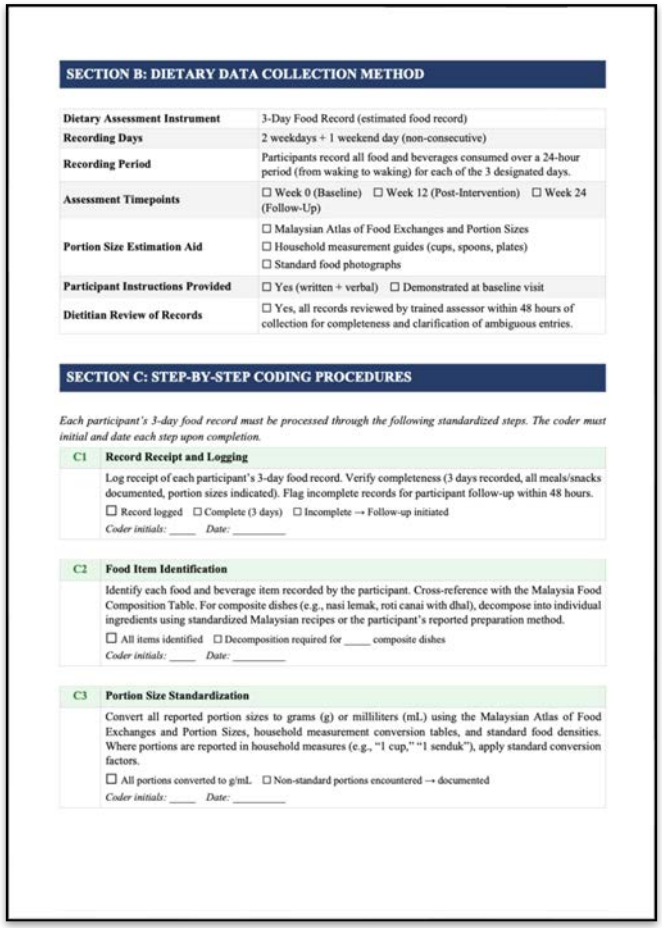
DIETARY INTAKE ANALYSIS PROTOCOL

Dietary Data Coding Procedures and Variable Extraction

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: SOFTWARE AND DATABASE CONFIGURATION

Software	Nutritionist Pro™ (Desktop Edition)
Version	7.9
Developer	Axxya Systems LLC, Redmond, WA, USA
Primary Nutrient Database	**Malaysia Food Composition Table** *Institute for Medical Research (IMR), Ministry of Health Malaysia*
Supplementary Database(s)	☐ USDA National Nutrient Database (for foods not in Malaysian database) ☐ Singapore Food Composition Table (for regional food items) ☐ Manufacturer labels (for commercial/packaged products)
Database Version / Edition Year	
Workstation Location	Exercise Science Laboratory, UiTM
Data Backup Protocol	☐ Daily automated backup to encrypted external drive ☐ Weekly backup to institutional cloud storage



SECTION B: DIETARY DATA COLLECTION METHOD

Dietary Assessment Instrument	3-Day Food Record (estimated food record)
Recording Days	2 weekdays + 1 weekend day (non-consecutive)
Recording Period	Participants record all food and beverages consumed over a 24-hour period (from waking to waking) for each of the 3 designated days.
Assessment Timepoints	☐ Week 0 (Baseline) ☐ Week 12 (Post-Intervention) ☐ Week 24 (Follow-Up)
Portion Size Estimation Aid	☐ Malaysian Atlas of Food Exchanges and Portion Sizes ☐ Household measurement guides (cups, spoons, plates) ☐ Standard food photographs
Participant Instructions Provided	☐ Yes (written + verbal) ☐ Demonstrated at baseline visit
Dietitian Review of Records	☐ Yes, all records reviewed by trained assessor within 48 hours of collection for completeness and clarification of ambiguous entries.

SECTION C: STEP-BY-STEP CODING PROCEDURES

Each participant's 3-day food record must be processed through the following standardized steps. The coder must initial and date each step upon completion.

C1 Record Receipt and Logging

Log receipt of each participant's 3-day food record. Verify completeness (3 days recorded, all meals/snacks documented, portion sizes indicated). Flag incomplete records for participant follow-up within 48 hours.

☐ Record logged ☐ Complete (3 days) ☐ Incomplete → Follow-up initiated

Coder initials: ____ *Date:* _____

C2 Food Item Identification

Identify each food and beverage item recorded by the participant. Cross-reference with the Malaysia Food Composition Table. For composite dishes (e.g., nasi lemak, roti canai with dhal), decompose into individual ingredients using standardized Malaysian recipes or the participant's reported preparation method.

☐ All items identified ☐ Decomposition required for ____ composite dishes

Coder initials: ____ *Date:* _____

C3 Portion Size Standardization

Convert all reported portion sizes to grams (g) or milliliters (mL) using the Malaysian Atlas of Food Exchanges and Portion Sizes, household measurement conversion tables, and standard food densities. Where portions are reported in household measures (e.g., "1 cup," "1 senduk"), apply standard conversion factors.

☐ All portions converted to g/mL ☐ Non-standard portions encountered → documented

Coder initials: ____ *Date:* _____

C4 **Database Matching**

Match each food item to the corresponding entry in the Nutritionist Pro™ Malaysia Food Composition Table. Use the closest match available. If an item is not found in the Malaysian database, search in this priority order: (1) Singapore Food Composition Table, (2) USDA National Nutrient Database, (3) manufacturer nutrition label data entered manually.

☐ Primary database match: ____% ☐ Supplementary database used for ____ items

Coder initials: ____ *Date:* _____

C5 **Data Entry into Nutritionist Pro™**

Enter each day's dietary intake as a separate food record within the participant's profile in Nutritionist Pro™. Organize entries by meal occasion: Breakfast, Morning Snack, Lunch, Afternoon Snack, Dinner, Evening Snack, Beverages. Enter quantities in grams/milliliters as standardized in Step C3.

☐ Day 1 entered ☐ Day 2 entered ☐ Day 3 entered ☐ All meals coded

Coder initials: ____ *Date:* _____

C6 **Quality Control — First Verification**

After data entry for each participant, run the Nutritionist Pro™ nutrient analysis report. Review output for implausible values: total energy <500 kcal/day or >5000 kcal/day, protein <10 g/day or >300 g/day, fat <10 g/day or >250 g/day. Flag and re-check any values outside plausible ranges against the original food record.

☐ Output reviewed ☐ All values plausible ☐ Re-checked ____ items

Coder initials: ____ *Date:* _____

C7 **Quality Control — Independent Double Entry (10% Sample)**

Randomly select 10% of participant records (n = 16) for independent double entry by a second trained coder. Calculate inter-rater reliability (intraclass correlation coefficient). Acceptable threshold: ICC ≥ 0.90 for total energy intake.

☐ Double entry completed for ____ participants ☐ ICC = ____ ☐ Meets threshold (≥0.90)

Coder initials: ____ *Date:* _____

C8 **3-Day Average Computation**

After all 3 days are entered and verified, use the Nutritionist Pro™ averaging function to compute the 3-day mean for each nutrient variable. This 3-day average represents the participant's habitual dietary intake for that assessment timepoint.

☐ 3-day average computed ☐ Exported to data extraction template

Coder initials: ____ *Date:* _____

SECTION D: VARIABLES EXTRACTED FROM NUTRITIONIST PRO™

The following variables are extracted for each participant at each assessment timepoint (Week 0, 12, 24). All values represent the 3-day mean unless otherwise specified.

D1. TOTAL ENERGY AND MACRONUTRIENTS

Variable	Unit	Computation / Notes
Total Energy Intake	kcal/day	3-day mean
Total Protein	g/day	3-day mean
Total Protein (% of energy)	% kcal	(protein g × 4 / total kcal) × 100
Total Fat	g/day	3-day mean
Total Fat (% of energy)	% kcal	(fat g × 9 / total kcal) × 100
Total Carbohydrate	g/day	3-day mean
Total Carbohydrate (% of energy)	% kcal	(CHO g × 4 / total kcal) × 100
Total Dietary Fiber	g/day	3-day mean
Total Sugar	g/day	3-day mean
Saturated Fatty Acids (SFA)	g/day	3-day mean
Monounsaturated Fatty Acids (MUFA)	g/day	3-day mean
Polyunsaturated Fatty Acids (PUFA)	g/day	3-day mean
Trans Fatty Acids	g/day	*If available in database*
Cholesterol	mg/day	3-day mean

D2. MICRONUTRIENTS

Variable	Unit	Computation / Notes
Sodium	mg/day	3-day mean
Potassium	mg/day	3-day mean
Calcium	mg/day	3-day mean
Iron	mg/day	3-day mean
Zinc	mg/day	3-day mean
Vitamin A	RE/day or µg RAE/day	*As available in database*
Vitamin C	mg/day	3-day mean
Vitamin D	IU/day or µg/day	*If available in database*
Vitamin E	mg α-TE/day	*If available in database*
Thiamin (B1)	mg/day	3-day mean
Riboflavin (B2)	mg/day	3-day mean
Niacin (B3)	mg NE/day	3-day mean
Folate	µg DFE/day	3-day mean

D3. ENERGY DENSITY (EXCLUDING BEVERAGES)

Variable	Unit	Computation / Notes
Total Energy Density (ED)	kcal/g	Total energy (kcal) / Total food weight (g), excluding all beverages
Total Food Weight (excl. beverages)	g/day	Sum of all solid food items, excluding water, tea, coffee, juices, milk as beverage, and other drinks
Beverage Energy	kcal/day	Energy from all beverages (reported separately)
Beverage Weight	g/day or mL/day	Total weight of all beverages consumed

ENERGY DENSITY CALCULATION NOTE

Energy density is calculated by dividing total energy from food (kcal) by total food weight (g), **excluding all beverages**. This approach follows the method recommended by Ledikwe et al. (2005) and Rolls et al. (2005), as beverages disproportionately increase food weight without proportional energy contribution, thereby artificially lowering energy density values.

Beverages excluded: Water, tea, coffee (black or with additions), fruit juices, carbonated drinks, milk consumed as a beverage, sports drinks, energy drinks, alcoholic beverages, and any other liquid consumed as a drink rather than as a component of a food dish.

Beverages included in food weight: Soups, gravies, sauces, and liquids that are integral components of a food dish (e.g., curry gravy consumed with rice).

D4. FOOD GROUP PATTERNS (Malaysian Food Pyramid Classification)

Food Group	Unit	Examples
Cereals and Grains	servings/day	Rice, noodles, bread, cereal products
Meat, Poultry, and Eggs	servings/day	All animal protein sources excluding fish
Fish and Seafood	servings/day	All fish and seafood items
Legumes and Nuts	servings/day	Beans, lentils, nuts, seeds, soy products
Milk and Dairy Products	servings/day	Milk, yogurt, cheese
Vegetables	servings/day	All vegetable items
Fruits	servings/day	All fruit items
Fats and Oils	servings/day	Cooking oils, butter, margarine, coconut milk
Sugar and Sweetened Beverages	servings/day	Sugar, syrup, sweetened drinks, confectionery
Condiments and Sauces	servings/day	Soy sauce, chili sauce, sambal, belacan

Food group servings are classified according to the Malaysian Dietary Guidelines (MDG) 2020 and the Malaysian Food Pyramid. Serving sizes follow the Malaysian Atlas of Food Exchanges and Portion Sizes.

SECTION E: PARTICIPANT-LEVEL DATA EXTRACTION RECORD

Complete one record per participant per assessment timepoint. Attach the Nutritionist Pro™ nutrient analysis printout to this form.

Participant ID		**Group**	☐ NEX ☐ CON
Assessment Timepoint	☐ Wk 0 ☐ Wk 12 ☐ Wk 24	**Date of Record Collection**	
Days Recorded	Day 1: _____ Day 2: _____ Day 3: _____		
Record Completeness	☐ Complete (3 days, all meals) ☐ Incomplete (specify): _____		

Variable	Day 1	Day 2	Day 3	3-Day Mean
Total Energy (kcal)				
Protein (g)				
Fat (g)				
Carbohydrate (g)				
Dietary Fiber (g)				
Total Sugar (g)				
SFA (g)				
MUFA (g)				
PUFA (g)				
Cholesterol (mg)				
Sodium (mg)				
Calcium (mg)				
Iron (mg)				
Vitamin C (mg)				
Food Weight excl. beverages (g)				
Energy Density (kcal/g)				

Coded by: _____ Date: _____ Verified by: _____ Date: _____

Nutritionist Pro™ printout attached: ☐ Yes ☐ No

SECTION F: QUALITY ASSURANCE AND SIGN-OFF

F1. QUALITY ASSURANCE CHECKLIST

No.	Quality Assurance Item	Yes	No	N/A
1	All food records were reviewed by a trained assessor within 48 hours of collection.	☐	☐	☐
2	Ambiguous food descriptions were clarified with the participant before coding.	☐	☐	☐
3	Composite Malaysian dishes were decomposed into individual ingredients using standardized recipes.	☐	☐	☐
4	All portion sizes were converted to grams/milliliters using the Malaysian Atlas of Food Exchanges.	☐	☐	☐
5	Food items not in the Malaysian database were matched from supplementary databases following the priority hierarchy (Singapore > USDA > manufacturer labels).	☐	☐	☐
6	Any manually entered food items were documented and verified against original labels or references.	☐	☐	☐
7	Implausible nutrient values were flagged and re-checked against the original food record.	☐	☐	☐
8	A 10% random sample was independently double-entered with ICC ≥ 0.90 for total energy.	☐	☐	☐
9	Energy density was calculated excluding all beverages as per the defined protocol.	☐	☐	☐
10	All participant records have Nutritionist Pro™ analysis printouts attached.	☐	☐	☐
11	Data have been exported to the master dataset (Excel/SPSS) and verified for transcription accuracy.	☐	☐	☐

F2. INTER-RATER RELIABILITY RECORD

Parameter	Value
Total participant records coded	
Number selected for double entry (10%)	
Coder 1 (Primary)	Name: _____
Coder 2 (Independent verifier)	Name: _____
ICC for Total Energy Intake (kcal/day)	ICC = _____ ☐ ≥0.90 ☐ <0.90 (action taken: _____)
ICC for Total Protein (g/day)	ICC = _____
ICC for Total Fat (g/day)	ICC = _____
ICC for Total Carbohydrate (g/day)	ICC = _____
Date reliability check completed	

F3. PROTOCOL COMPLIANCE SIGN-OFF

I confirm that I have coded the dietary data for all participants in accordance with this standardized analysis protocol, that the Malaysia Food Composition Table was used as the primary nutrient database, that energy density was computed excluding beverages, and that all quality assurance procedures documented above were completed satisfactorily.

_____ _____ _____

Dietary Data Coder Name *Signature* *Date*

Designation: _____

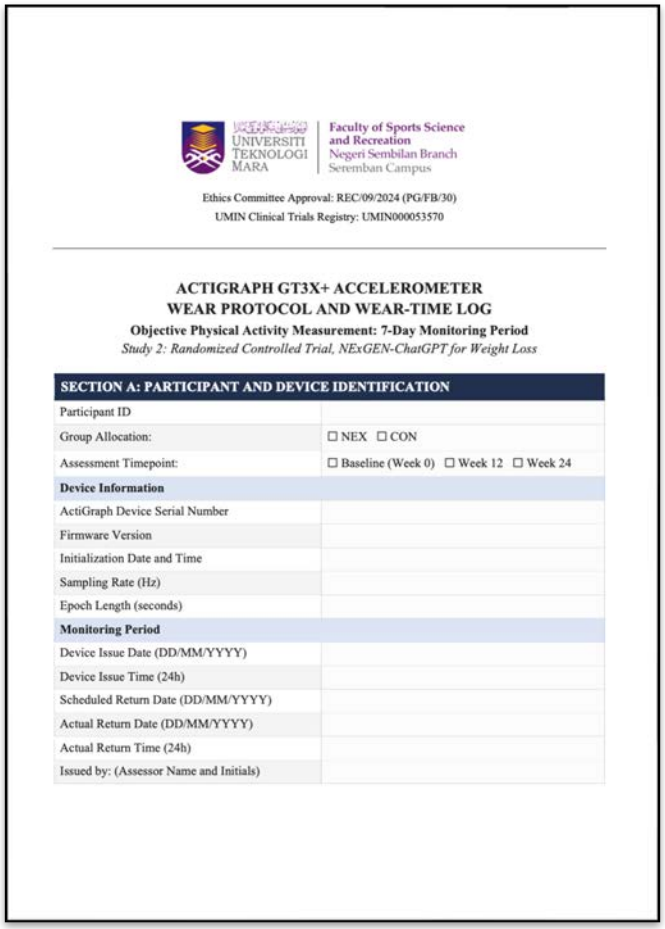
Reviewed and Approved by (Principal Investigator):

_____ _____ _____

Principal Investigator Name *Signature* *Date*

APPENDIX 21

ActiGraph GT3x+ Accelerometer Wear Protocol And Wear-Time Log (Study 2)

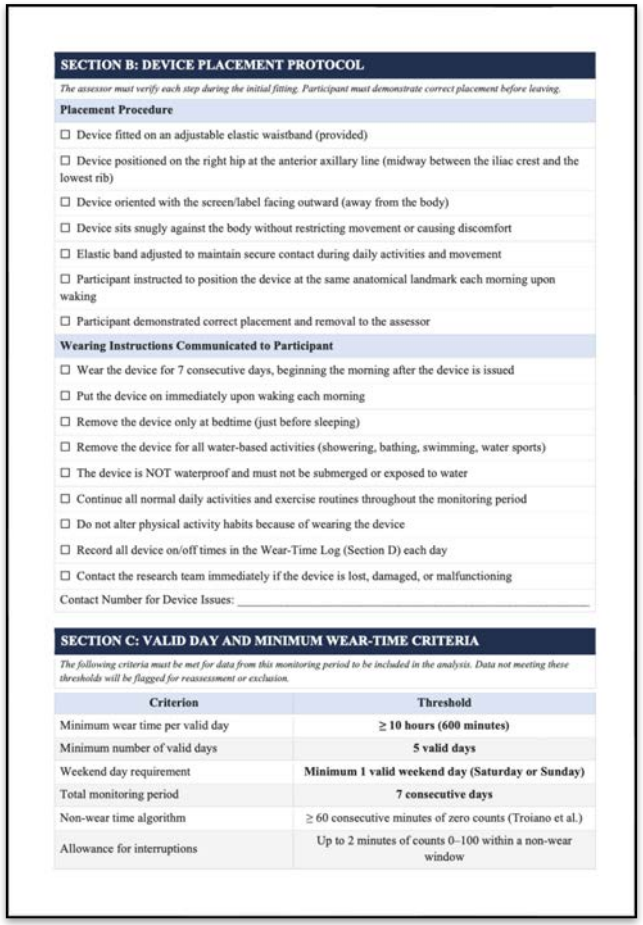


Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)
UMIN Clinical Trials Registry: UMIN000053570

ACTIGRAPH GT3X+ ACCELEROMETER
WEAR PROTOCOL AND WEAR-TIME LOG
Objective Physical Activity Measurement: 7-Day Monitoring Period
Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: PARTICIPANT AND DEVICE IDENTIFICATION	
Participant ID	
Group Allocation:	☐ NEX ☐ CON
Assessment Timepoint:	☐ Baseline (Week 0) ☐ Week 12 ☐ Week 24
Device Information	
ActiGraph Device Serial Number	
Firmware Version	
Initialization Date and Time	
Sampling Rate (Hz)	
Epoch Length (seconds)	
Monitoring Period	
Device Issue Date (DD/MM/YYYY)	
Device Issue Time (24h)	
Scheduled Return Date (DD/MM/YYYY)	
Actual Return Date (DD/MM/YYYY)	
Actual Return Time (24h)	
Issued by: (Assessor Name and Initials)	



SECTION B: DEVICE PLACEMENT PROTOCOL

The assessor must verify each step during the initial fitting. Participant must demonstrate correct placement before leaving.

Placement Procedure

- ☐ Device fitted on an adjustable elastic waistband (provided)
- ☐ Device positioned on the right hip at the anterior axillary line (midway between the iliac crest and the lowest rib)
- ☐ Device oriented with the screen/label facing outward (away from the body)
- ☐ Device sits snugly against the body without restricting movement or causing discomfort
- ☐ Elastic band adjusted to maintain secure contact during daily activities and movement
- ☐ Participant instructed to position the device at the same anatomical landmark each morning upon waking
- ☐ Participant demonstrated correct placement and removal to the assessor

Wearing Instructions Communicated to Participant

- ☐ Wear the device for 7 consecutive days, beginning the morning after the device is issued
- ☐ Put the device on immediately upon waking each morning
- ☐ Remove the device only at bedtime (just before sleeping)
- ☐ Remove the device for all water-based activities (showering, bathing, swimming, water sports)
- ☐ The device is NOT waterproof and must not be submerged or exposed to water
- ☐ Continue all normal daily activities and exercise routines throughout the monitoring period
- ☐ Do not alter physical activity habits because of wearing the device
- ☐ Record all device on/off times in the Wear-Time Log (Section D) each day
- ☐ Contact the research team immediately if the device is lost, damaged, or malfunctioning

Contact Number for Device Issues: _____

SECTION C: VALID DAY AND MINIMUM WEAR-TIME CRITERIA

The following criteria must be met for data from this monitoring period to be included in the analysis. Data not meeting these thresholds will be flagged for reassessment or exclusion.

Criterion	Threshold
Minimum wear time per valid day	**≥ 10 hours (600 minutes)**
Minimum number of valid days	**5 valid days**
Weekend day requirement	**Minimum 1 valid weekend day (Saturday or Sunday)**
Total monitoring period	**7 consecutive days**
Non-wear time algorithm	≥ 60 consecutive minutes of zero counts (Troiano et al.)
Allowance for interruptions	Up to 2 minutes of counts 0–100 within a non-wear window

SECTION D: DAILY WEAR-TIME LOG (PARTICIPANT SELF-RECORD)

Instructions: Record the exact times you put on and take off the device each day. If you remove the device during the day (e.g., for showering), record the removal time, reason, and the time you put it back on. Use 24-hour format (e.g., 07:30, 22:15).

Day	Date	Time On (waking)	Time Off (sleep)	Removed (time)	Reason for Removal	Back On (time)	Est. Total Wear (hrs)
Day 1 *(circle day)*	__/__/____						
(2nd removal)							
Day 2 *(circle day)*	__/__/____						
(2nd removal)							
Day 3 *(circle day)*	__/__/____						
(2nd removal)							
Day 4 *(circle day)*	__/__/____						
(2nd removal)							
Day 5 *(circle day)*	__/__/____						
(2nd removal)							
Day 6 *(circle day)*	__/__/____						
(2nd removal)							
Day 7 *(circle day)*	__/__/____						
(2nd removal)							

Common Removal Reasons: S = Showering/Bathing | W = Water activity/Swimming | Sp = Sport (contact/water) | M = Medical procedure | D = Discomfort | O = Other (specify)

SECTION E: DATA VALIDATION SUMMARY (ASSESSOR USE ONLY)

Complete upon data download using ActiLife software. Cross-reference with participant self-reported wear-time log.

Day	Date	Wear Time (minutes)	Valid Day? (≥10 hrs)	MVPA (minutes)	Steps	Notes
Day 1			☐ Yes ☐ No			
Day 2			☐ Yes ☐ No			
Day 3			☐ Yes ☐ No			
Day 4			☐ Yes ☐ No			
Day 5			☐ Yes ☐ No			
Day 6			☐ Yes ☐ No			
Day 7			☐ Yes ☐ No			

Compliance Summary	
Total valid days	
Valid weekdays (Mon–Fri)	
Valid weekend days (Sat–Sun)	
Average daily wear time (minutes)	
Average daily MVPA (minutes)	
Average daily steps	
Meets minimum criteria (≥5 valid days including ≥1 weekend day): ☐ Yes ☐ No	
If No, action taken: ☐ Extended monitoring period ☐ Data excluded ☐ Other: _____	

SECTION F: DEVICE RETURN AND CONDITION CHECK

Device returned on (DD/MM/YYYY)	
Device condition:	☐ Good ☐ Damaged ☐ Lost
Damage description (if applicable)	
Data download completed:	☐ Yes ☐ No
Data download date and time	
ActiLife software version used	
Data file name / path	
Data quality check completed:	☐ Yes ☐ No
Issues identified during download: _____	

SECTION G: VERIFICATION AND SIGN-OFF

I confirm that this form accurately documents the ActiGraph GT3X+ accelerometer wear protocol delivery, participant wear-time self-record, data validation results, and device return condition. All data processing was completed according to the pre-specified analysis protocol, and this documentation is complete and consistent with the participant's Case Report Form.

Assessor

_____
Name

_____
Signature

_____
Date

Data verified by (independent reviewer):

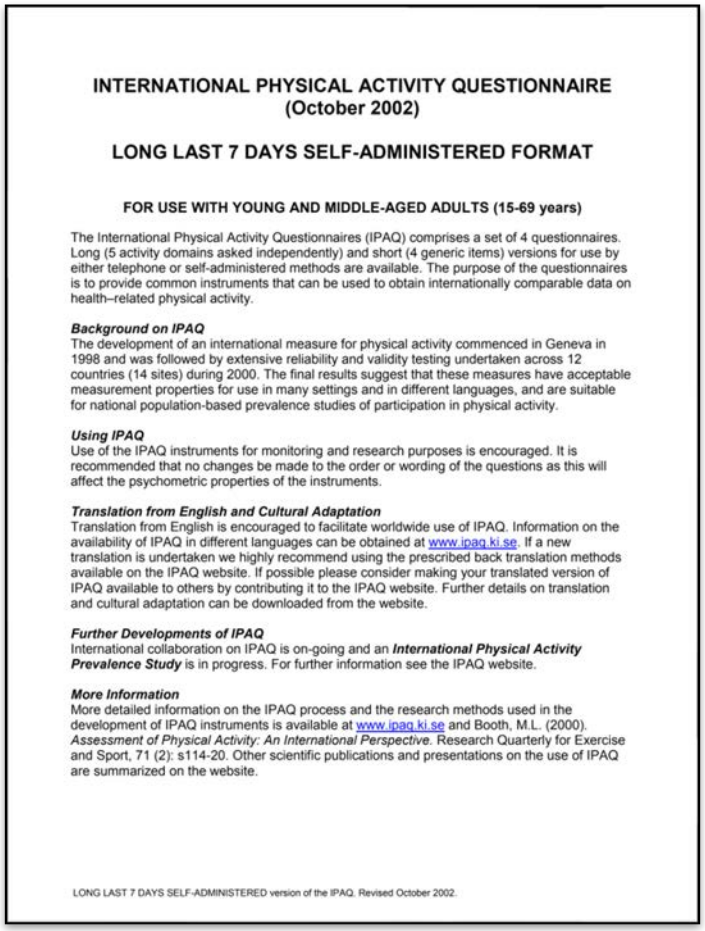
_____
Name

_____
Signature

_____
Date

APPENDIX 22

International Physical Actvity - Long Form Version (Study 2)



INTERNATIONAL PHYSICAL ACTIVITY QUESTIONNAIRE (October 2002)

LONG LAST 7 DAYS SELF-ADMINISTERED FORMAT

FOR USE WITH YOUNG AND MIDDLE-AGED ADULTS (15-69 years)

The International Physical Activity Questionnaires (IPAQ) comprises a set of 4 questionnaires. Long (5 activity domains asked independently) and short (4 generic items) versions for use by either telephone or self-administered methods are available. The purpose of the questionnaires is to provide common instruments that can be used to obtain internationally comparable data on health–related physical activity.

Background on IPAQ
The development of an international measure for physical activity commenced in Geneva in 1998 and was followed by extensive reliability and validity testing undertaken across 12 countries (14 sites) during 2000. The final results suggest that these measures have acceptable measurement properties for use in many settings and in different languages, and are suitable for national population-based prevalence studies of participation in physical activity.

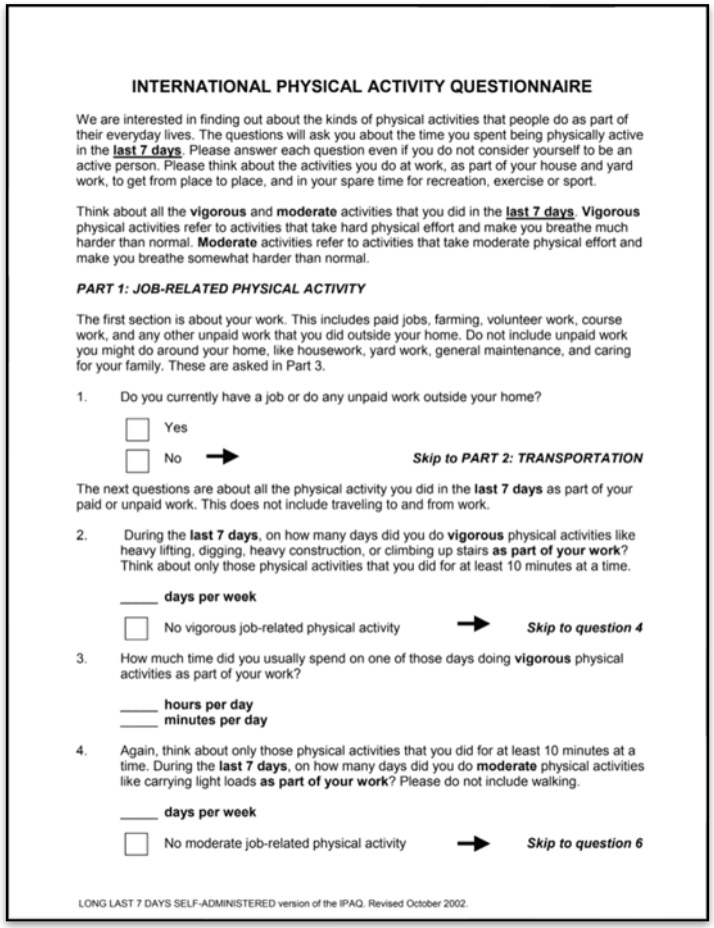
Using IPAQ
Use of the IPAQ instruments for monitoring and research purposes is encouraged. It is recommended that no changes be made to the order or wording of the questions as this will affect the psychometric properties of the instruments.

Translation from English and Cultural Adaptation
Translation from English is encouraged to facilitate worldwide use of IPAQ. Information on the availability of IPAQ in different languages can be obtained at www.ipaq.ki.se. If a new translation is undertaken we highly recommend using the prescribed back translation methods available on the IPAQ website. If possible please consider making your translated version of IPAQ available to others by contributing it to the IPAQ website. Further details on translation and cultural adaptation can be downloaded from the website.

Further Developments of IPAQ
International collaboration on IPAQ is on-going and an ***International Physical Activity Prevalence Study*** is in progress. For further information see the IPAQ website.

More Information
More detailed information on the IPAQ process and the research methods used in the development of IPAQ instruments is available at www.ipaq.ki.se and Booth, M.L. (2000). *Assessment of Physical Activity: An International Perspective*. Research Quarterly for Exercise and Sport, 71 (2): s114-20. Other scientific publications and presentations on the use of IPAQ are summarized on the website.

LONG LAST 7 DAYS SELF-ADMINISTERED version of the IPAQ. Revised October 2002.



INTERNATIONAL PHYSICAL ACTIVITY QUESTIONNAIRE

We are interested in finding out about the kinds of physical activities that people do as part of their everyday lives. The questions will ask you about the time you spent being physically active in the **last 7 days**. Please answer each question even if you do not consider yourself to be an active person. Please think about the activities you do at work, as part of your house and yard work, to get from place to place, and in your spare time for recreation, exercise or sport.

Think about all the **vigorous** and **moderate** activities that you did in the **last 7 days**. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal.

PART 1: JOB-RELATED PHYSICAL ACTIVITY

The first section is about your work. This includes paid jobs, farming, volunteer work, course work, and any other unpaid work that you did outside your home. Do not include unpaid work you might do around your home, like housework, yard work, general maintenance, and caring for your family. These are asked in Part 3.

1. Do you currently have a job or do any unpaid work outside your home?

   ☐ Yes
   ☐ No → ***Skip to PART 2: TRANSPORTATION***

The next questions are about all the physical activity you did in the **last 7 days** as part of your paid or unpaid work. This does not include traveling to and from work.

2. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, heavy construction, or climbing up stairs **as part of your work**? Think about only those physical activities that you did for at least 10 minutes at a time.

   ____ **days per week**

   ☐ No vigorous job-related physical activity → ***Skip to question 4***

3. How much time did you usually spend on one of those days doing **vigorous** physical activities as part of your work?

   ____ **hours per day**
   ____ **minutes per day**

4. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads **as part of your work**? Please do not include walking.

   ____ **days per week**

   ☐ No moderate job-related physical activity → ***Skip to question 6***

LONG LAST 7 DAYS SELF-ADMINISTERED version of the IPAQ. Revised October 2002.

5. How much time did you usually spend on one of those days doing **moderate** physical activities as part of your work?

_____ **hours per day**
_____ **minutes per day**

6. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time **as part of your work**? Please do not count any walking you did to travel to or from work.

_____ **days per week**

☐ No job-related walking → ***Skip to PART 2: TRANSPORTATION***

7. How much time did you usually spend on one of those days **walking** as part of your work?

_____ **hours per day**
_____ **minutes per day**

PART 2: TRANSPORTATION PHYSICAL ACTIVITY

These questions are about how you traveled from place to place, including to places like work, stores, movies, and so on.

8. During the **last 7 days**, on how many days did you **travel in a motor vehicle** like a train, bus, car, or tram?

_____ **days per week**

☐ No traveling in a motor vehicle → ***Skip to question 10***

9. How much time did you usually spend on one of those days **traveling** in a train, bus, car, tram, or other kind of motor vehicle?

_____ **hours per day**
_____ **minutes per day**

Now think only about the **bicycling** and **walking** you might have done to travel to and from work, to do errands, or to go from place to place.

10. During the **last 7 days**, on how many days did you **bicycle** for at least 10 minutes at a time to go **from place to place**?

_____ **days per week**

☐ No bicycling from place to place → ***Skip to question 12***

11. How much time did you usually spend on one of those days to **bicycle** from place to place?

_____ **hours per day**
_____ **minutes per day**

12. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time to go **from place to place**?

_____ **days per week**

☐ No walking from place to place → ***Skip to PART 3: HOUSEWORK, HOUSE MAINTENANCE, AND CARING FOR FAMILY***

13. How much time did you usually spend on one of those days **walking** from place to place?

_____ **hours per day**
_____ **minutes per day**

PART 3: HOUSEWORK, HOUSE MAINTENANCE, AND CARING FOR FAMILY

This section is about some of the physical activities you might have done in the **last 7 days** in and around your home, like housework, gardening, yard work, general maintenance work, and caring for your family.

14. Think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, chopping wood, shoveling snow, or digging **in the garden or yard**?

_____ **days per week**

☐ No vigorous activity in garden or yard → ***Skip to question 16***

15. How much time did you usually spend on one of those days doing **vigorous** physical activities in the garden or yard?

_____ **hours per day**
_____ **minutes per day**

16. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **moderate** activities like carrying light loads, sweeping, washing windows, and raking **in the garden or yard**?

_____ **days per week**

☐ No moderate activity in garden or yard → ***Skip to question 18***

17. How much time did you usually spend on one of those days doing **moderate** physical activities in the garden or yard?

_____ **hours per day**
_____ **minutes per day**

18. Once again, think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **moderate** activities like carrying light loads, washing windows, scrubbing floors and sweeping **inside your home**?

_____ **days per week**

☐ No moderate activity inside home → ***Skip to PART 4: RECREATION, SPORT AND LEISURE-TIME PHYSICAL ACTIVITY***

19. How much time did you usually spend on one of those days doing **moderate** physical activities inside your home?

_____ **hours per day**
_____ **minutes per day**

PART 4: RECREATION, SPORT, AND LEISURE-TIME PHYSICAL ACTIVITY

This section is about all the physical activities that you did in the **last 7 days** solely for recreation, sport, exercise or leisure. Please do not include any activities you have already mentioned.

20. Not counting any walking you have already mentioned, during the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time **in your leisure time**?

_____ **days per week**

☐ No walking in leisure time → ***Skip to question 22***

21. How much time did you usually spend on one of those days **walking** in your leisure time?

_____ **hours per day**
_____ **minutes per day**

22. Think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **vigorous** physical activities like aerobics, running, fast bicycling, or fast swimming **in your leisure time**?

_____ **days per week**

☐ No vigorous activity in leisure time → ***Skip to question 24***

23. How much time did you usually spend on one of those days doing **vigorous** physical activities in your leisure time?

_____ **hours per day**
_____ **minutes per day**

24. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the **last 7 days**, on how many days did you do **moderate** physical activities like bicycling at a regular pace, swimming at a regular pace, and doubles tennis **in your leisure time**?

_____ **days per week**

☐ No moderate activity in leisure time → ***Skip to PART 5: TIME SPENT SITTING***

25. How much time did you usually spend on one of those days doing **moderate** physical activities in your leisure time?

_____ **hours per day**
_____ **minutes per day**

PART 5: TIME SPENT SITTING

The last questions are about the time you spend sitting while at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading or sitting or lying down to watch television. Do not include any time spent sitting in a motor vehicle that you have already told me about.

26. During the **last 7 days**, how much time did you usually spend **sitting** on a **weekday**?

_____ **hours per day**
_____ **minutes per day**

27. During the **last 7 days**, how much time did you usually spend **sitting** on a **weekend day**?

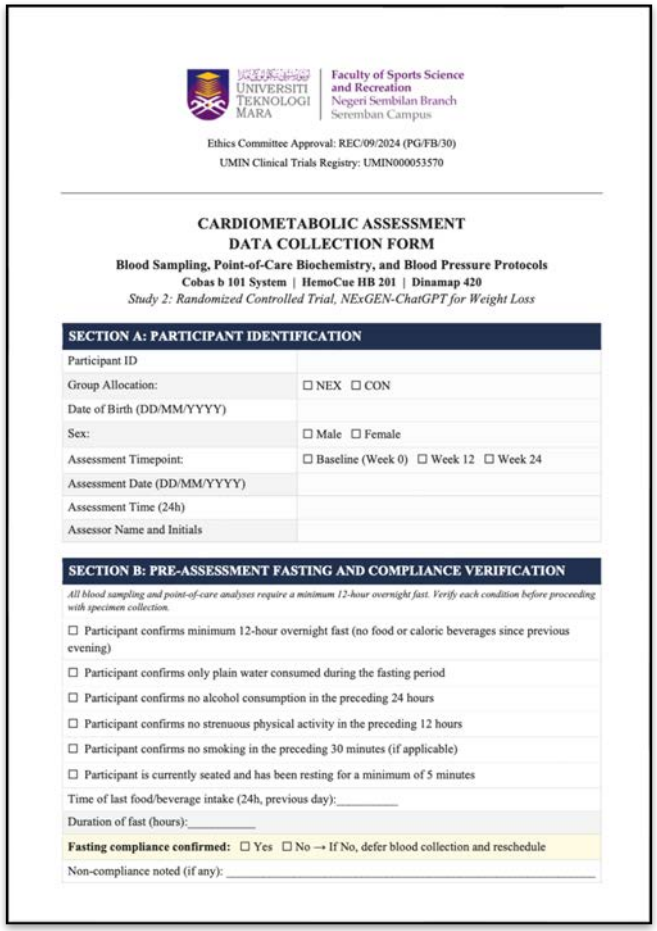
_____ **hours per day**
_____ **minutes per day**

This is the end of the questionnaire, thank you for participating.

LONG LAST 7 DAYS SELF-ADMINISTERED version of the IPAQ. Revised October 2002.

APPENDIX 23

Cardiometabolic Assessment (Study 2)



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)

UMIN Clinical Trials Registry: UMIN000053570

CARDIOMETABOLIC ASSESSMENT
DATA COLLECTION FORM

Blood Sampling, Point-of-Care Biochemistry, and Blood Pressure Protocols
Cobas b 101 System | HemoCue HB 201 | Dinamap 420
Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

SECTION A: PARTICIPANT IDENTIFICATION	
Participant ID	
Group Allocation:	☐ NEX ☐ CON
Date of Birth (DD/MM/YYYY)	
Sex:	☐ Male ☐ Female
Assessment Timepoint:	☐ Baseline (Week 0) ☐ Week 12 ☐ Week 24
Assessment Date (DD/MM/YYYY)	
Assessment Time (24h)	
Assessor Name and Initials	

SECTION B: PRE-ASSESSMENT FASTING AND COMPLIANCE VERIFICATION

All blood sampling and point-of-care analyses require a minimum 12-hour overnight fast. Verify each condition before proceeding with specimen collection.

☐ Participant confirms minimum 12-hour overnight fast (no food or caloric beverages since previous evening)

☐ Participant confirms only plain water consumed during the fasting period

☐ Participant confirms no alcohol consumption in the preceding 24 hours

☐ Participant confirms no strenuous physical activity in the preceding 12 hours

☐ Participant confirms no smoking in the preceding 30 minutes (if applicable)

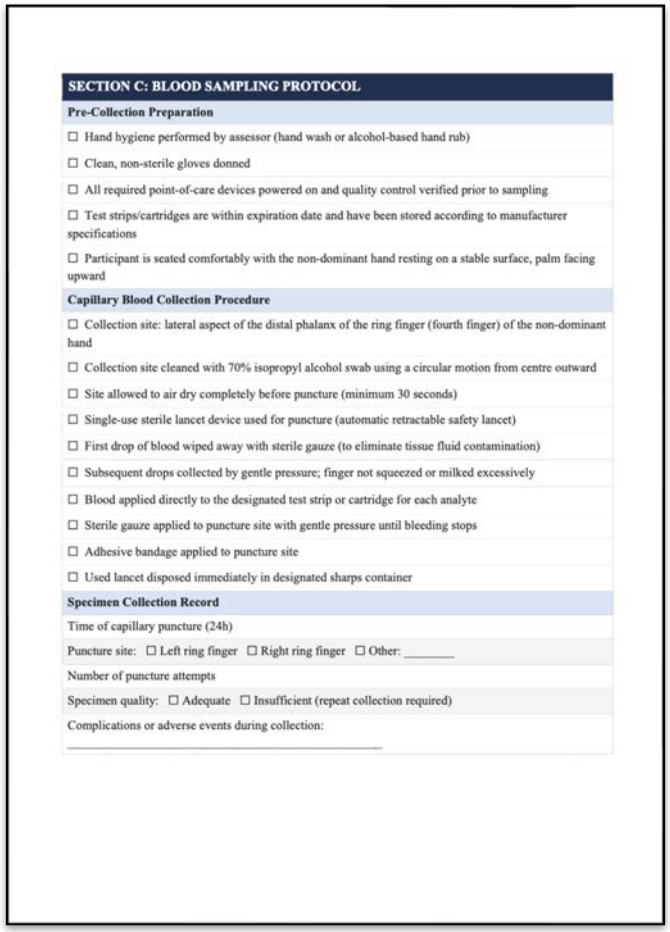
☐ Participant is currently seated and has been resting for a minimum of 5 minutes

Time of last food/beverage intake (24h, previous day):_____

Duration of fast (hours):_____

Fasting compliance confirmed: ☐ Yes ☐ No → If No, defer blood collection and reschedule

Non-compliance noted (if any): _____



SECTION C: BLOOD SAMPLING PROTOCOL

Pre-Collection Preparation

☐ Hand hygiene performed by assessor (hand wash or alcohol-based hand rub)

☐ Clean, non-sterile gloves donned

☐ All required point-of-care devices powered on and quality control verified prior to sampling

☐ Test strips/cartridges are within expiration date and have been stored according to manufacturer specifications

☐ Participant is seated comfortably with the non-dominant hand resting on a stable surface, palm facing upward

Capillary Blood Collection Procedure

☐ Collection site: lateral aspect of the distal phalanx of the ring finger (fourth finger) of the non-dominant hand

☐ Collection site cleaned with 70% isopropyl alcohol swab using a circular motion from centre outward

☐ Site allowed to air dry completely before puncture (minimum 30 seconds)

☐ Single-use sterile lancet device used for puncture (automatic retractable safety lancet)

☐ First drop of blood wiped away with sterile gauze (to eliminate tissue fluid contamination)

☐ Subsequent drops collected by gentle pressure; finger not squeezed or milked excessively

☐ Blood applied directly to the designated test strip or cartridge for each analyte

☐ Sterile gauze applied to puncture site with gentle pressure until bleeding stops

☐ Adhesive bandage applied to puncture site

☐ Used lancet disposed immediately in designated sharps container

Specimen Collection Record

Time of capillary puncture (24h)

Puncture site: ☐ Left ring finger ☐ Right ring finger ☐ Other: _____

Number of puncture attempts

Specimen quality: ☐ Adequate ☐ Insufficient (repeat collection required)

Complications or adverse events during collection:
_____

SECTION D: FASTING GLUCOSE — HemoCue Glucose 201 Analyzer

Device Protocol

- ☐ HemoCue Glucose 201 analyzer powered on and self-test completed successfully
- ☐ Quality control (QC) check performed with control cuvette within acceptable range
- ☐ Microcuvette removed from the vial and used within the manufacturer-specified timeframe
- ☐ Microcuvette filled in a single continuous action by capillary attraction (no air bubbles visible)
- ☐ Excess blood wiped from the outside of the microcuvette tip
- ☐ Microcuvette placed into the holder and inserted into the analyzer within 60 seconds of filling
- ☐ Result displayed and recorded after the measurement cycle (approximately 40–60 seconds)
- ☐ Used microcuvette disposed of in biohazard waste

Fasting Glucose Readings

Parameter	Baseline	Week 12	Week 24
Fasting Glucose (mmol/L)			
Device Serial No.			
QC Check Passed	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
Lot No. / Expiry of Microcuvettes			

SECTION E: HbA1c AND LIPID PROFILE — Cobas b 101 System (Roche Diagnostics)

Device Protocol

- ☐ Cobas b 101 system powered on and initialization self-test completed successfully
- ☐ Quality control (QC) check performed and results within acceptable range
- ☐ Ambient temperature within operating range (15–32°C) and humidity within specifications

HbA1c Test Procedure

- ☐ HbA1c test disc removed from sealed pouch and used within 10 minutes of opening
- ☐ Test disc inserted into the analyzer and disc information read by the system
- ☐ Capillary blood sample applied to the designated application zone on the test disc
- ☐ Analyzer door closed; measurement cycle initiated automatically
- ☐ Result displayed after approximately 5 minutes and recorded
- ☐ Used test disc ejected and disposed of in biohazard waste

Lipid Profile Test Procedure

- ☐ Lipid panel test disc removed from sealed pouch and used within 10 minutes of opening
- ☐ Test disc inserted into the analyzer and disc information read by the system
- ☐ Capillary blood sample applied to the designated application zone on the test disc
- ☐ Analyzer door closed; measurement cycle initiated automatically
- ☐ Result displayed after approximately 6 minutes and recorded
- ☐ Used test disc ejected and disposed of in biohazard waste

HbA1c and Lipid Profile Readings

Parameter	Baseline	Week 12	Week 24
HbA1c (%)			
HbA1c (mmol/mol)			
Total Cholesterol (mmol/L)			
HDL Cholesterol (mmol/L)			
LDL Cholesterol (mmol/L)			
Triglycerides (mmol/L)			
Total Cholesterol / HDL Ratio			

Device Traceability (Cobas b 101)

	Baseline	Week 12	Week 24
Device Serial No.			
QC Check Passed	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
HbA1c Disc Lot No. / Expiry			
Lipid Disc Lot No. / Expiry			

SECTION F: BLOOD PRESSURE — Dinamap 420 Oscillometric Device

Pre-Measurement Protocol

- ☐ Participant has been seated quietly for a minimum of 5 minutes with back supported and feet flat on the floor
- ☐ Legs uncrossed throughout the rest period and measurement
- ☐ Participant has not consumed caffeine in the preceding 30 minutes
- ☐ Participant has emptied bladder prior to measurement (if needed)
- ☐ Participant is not talking and remains silent throughout all three measurements

Cuff Selection and Placement

Mid-arm circumference (cm)

Cuff size selected: ☐ Small ☐ Regular/Adult ☐ Large ☐ Extra-Large

☐ Appropriate cuff size selected (bladder width approximately 40% and length approximately 80% of arm circumference)

☐ Cuff applied to the bare upper left arm (unless contraindicated) at heart level

☐ Cuff positioned with the arterial indicator marker aligned over the brachial artery at the antecubital fossa

☐ Lower edge of the cuff positioned approximately 2–3 cm above the antecubital fossa

☐ Arm resting on a flat surface (desk or armrest) with the midpoint of the upper arm at the level of the heart

Measurement Procedure

☐ Dinamap 420 powered on and self-calibration check completed

☐ Three consecutive measurements taken at 1–2 minute intervals

☐ First reading discarded per protocol; mean of the final two readings calculated and recorded

☐ All readings fall within plausible physiological range (flagged if SBP > 180 or DBP > 110 mmHg)

Blood Pressure Readings (Current Assessment Timepoint)

Reading	SBP (mmHg)	DBP (mmHg)	MAP (mmHg)	Notes
Reading 1 (discarded)				
Reading 2				
Reading 3				
Mean of Readings 2 and 3				
Recorded Mean				*(R2+R3)/2*

Blood Pressure Longitudinal Summary (Mean of Final Two Readings)

Parameter	Baseline	Week 12	Week 24
Systolic Blood Pressure (mmHg)			
Diastolic Blood Pressure (mmHg)			
Mean Arterial Pressure (mmHg)			
Resting Heart Rate (bpm)			

Device Traceability (Dinamap 420)

	Baseline	Week 12	Week 24
Device Serial No.			
Last Calibration Date			
Arm Used: L / R			

MAP Calculation: *MAP = DBP + 1/3 (SBP – DBP). Automatically calculated by the Dinamap 420 device. Verify against manual calculation if reading appears anomalous.*

SECTION G: QUALITY ASSURANCE AND ASSESSOR SIGN-OFF

Specimen and Waste Disposal Verification

☐ All used lancets disposed in designated sharps container

☐ All used test discs and microcuvettes disposed in biohazard waste

☐ Work surfaces cleaned and disinfected after specimen collection

☐ Gloves removed and hand hygiene performed

Data Quality Verification

☐ All required cardiometabolic measurements have been completed and recorded

☐ All readings fall within plausible physiological ranges

☐ Any anomalous readings have been repeated and verified (re-test results documented)

☐ Blood pressure triplicate protocol followed correctly (mean of Readings 2 and 3 recorded)

☐ Device lot numbers, serial numbers, and QC status documented for all instruments

☐ Participant was comfortable throughout the assessment procedure and experienced no adverse events

Deviations from Standard Protocol (describe): _____

_____

Adverse events during assessment (describe): _____

_____

SECTION H: VERIFICATION AND SIGN-OFF

I confirm that this form accurately records all cardiometabolic assessments obtained during this session, that all blood sampling, point-of-care analyses, and blood pressure measurements were performed according to the standardized protocols specified above, and that this documentation is complete and consistent with the participant's Case Report Form.

Assessor

_____
Name

_____
Signature

_____
Date

Data verified by (independent reviewer):

_____
Name

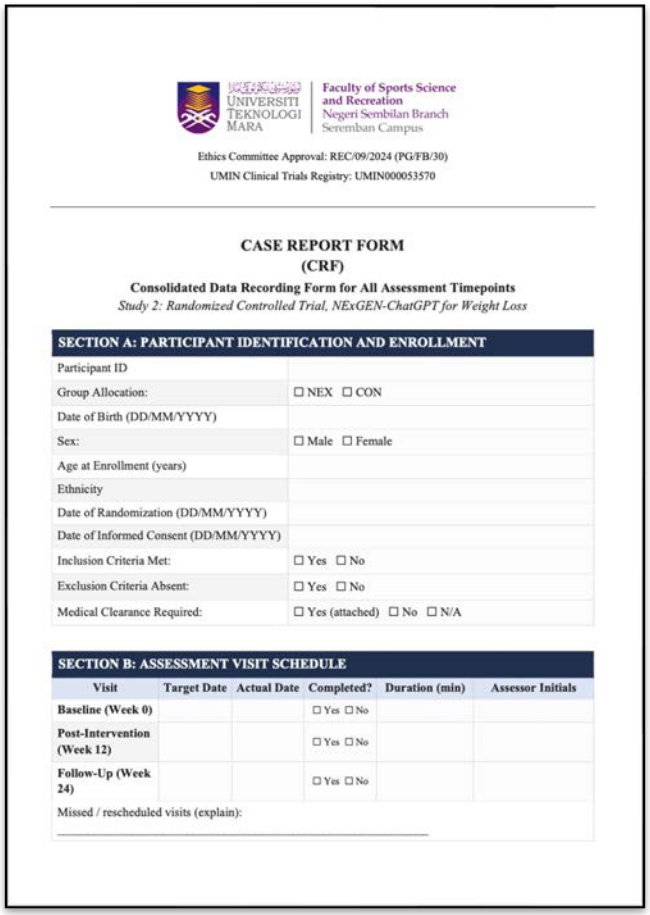
_____
Signature

_____
Date

CONFIDENTIAL: This form contains protected health information. Handle in accordance with the approved data management protocol (Ethics REC/09/2024 (PG/FB/30)). Store securely and restrict access to authorized research personnel only.

APPENDIX 24

Consolidata Data Record Form



Universiti Teknologi MARA | Faculty of Sports Science and Recreation, Negeri Sembilan Branch, Seremban Campus

Ethics Committee Approval: REC/09/2024 (PG/FB/30)

UMIN Clinical Trials Registry: UMIN000053570

CASE REPORT FORM
(CRF)

Consolidated Data Recording Form for All Assessment Timepoints

Study 2: Randomized Controlled Trial, NExGEN-ChatGPT for Weight Loss

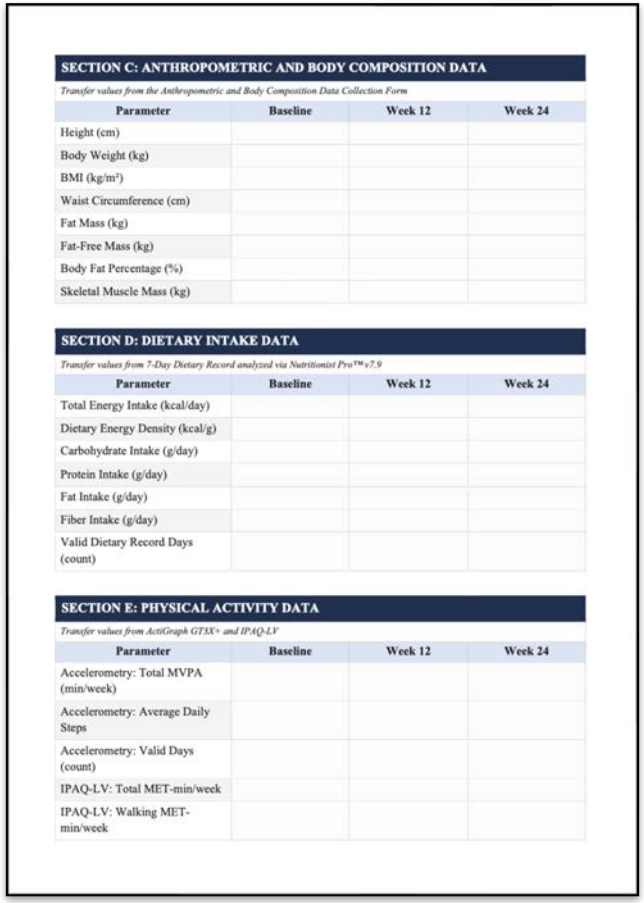
SECTION A: PARTICIPANT IDENTIFICATION AND ENROLLMENT

Participant ID	
Group Allocation:	☐ NEX ☐ CON
Date of Birth (DD/MM/YYYY)	
Sex:	☐ Male ☐ Female
Age at Enrollment (years)	
Ethnicity	
Date of Randomization (DD/MM/YYYY)	
Date of Informed Consent (DD/MM/YYYY)	
Inclusion Criteria Met:	☐ Yes ☐ No
Exclusion Criteria Absent:	☐ Yes ☐ No
Medical Clearance Required:	☐ Yes (attached) ☐ No ☐ N/A

SECTION B: ASSESSMENT VISIT SCHEDULE

Visit	Target Date	Actual Date	Completed?	Duration (min)	Assessor Initials
Baseline (Week 0)			☐ Yes ☐ No		
Post-Intervention (Week 12)			☐ Yes ☐ No		
Follow-Up (Week 24)			☐ Yes ☐ No		

Missed / rescheduled visits (explain): _____



SECTION C: ANTHROPOMETRIC AND BODY COMPOSITION DATA

Transfer values from the Anthropometric and Body Composition Data Collection Form

Parameter	Baseline	Week 12	Week 24
Height (cm)			
Body Weight (kg)			
BMI (kg/m²)			
Waist Circumference (cm)			
Fat Mass (kg)			
Fat-Free Mass (kg)			
Body Fat Percentage (%)			
Skeletal Muscle Mass (kg)			

SECTION D: DIETARY INTAKE DATA

Transfer values from 7-Day Dietary Record analyzed via Nutritionist Pro™ v7.9

Parameter	Baseline	Week 12	Week 24
Total Energy Intake (kcal/day)			
Dietary Energy Density (kcal/g)			
Carbohydrate Intake (g/day)			
Protein Intake (g/day)			
Fat Intake (g/day)			
Fiber Intake (g/day)			
Valid Dietary Record Days (count)			

SECTION E: PHYSICAL ACTIVITY DATA

Transfer values from ActiGraph GT3X+ and IPAQ-LV

Parameter	Baseline	Week 12	Week 24
Accelerometry: Total MVPA (min/week)			
Accelerometry: Average Daily Steps			
Accelerometry: Valid Days (count)			
IPAQ-LV: Total MET-min/week			
IPAQ-LV: Walking MET-min/week			

IPAQ-LV: Moderate MET-min/week			
IPAQ-LV: Vigorous MET-min/week			

SECTION F: CARDIOMETABOLIC MARKERS

Transfer values from Cardiometabolic Assessment Data Collection Form

Parameter	Baseline	Week 12	Week 24
Fasting Glucose (mmol/L)			
HbA1c (%)			
Total Cholesterol (mmol/L)			
HDL Cholesterol (mmol/L)			
LDL Cholesterol (mmol/L)			
Triglycerides (mmol/L)			
Systolic Blood Pressure (mmHg)			
Diastolic Blood Pressure (mmHg)			
Mean Arterial Pressure (mmHg)			
Resting Heart Rate (bpm)			

SECTION G: PARTICIPANT STATUS AND COMPLETION

Study Completion Status:	☐ Completed all 3 visits ☐ Dropout ☐ Excluded (PP)
If dropout: Date of dropout (DD/MM/YYYY)	
Dropout reason code	
Protocol deviations recorded:	☐ None ☐ Yes
Adverse events reported:	☐ None ☐ Yes (describe below)
Adverse event description: _____	
_____	
Included in ITT analysis:	☐ Yes ☐ No
Included in PP analysis:	☐ Yes ☐ No (reason: _____)

SECTION H: CRF VERIFICATION AND SIGN-OFF

I confirm that this Case Report Form accurately consolidates all outcome data for the above participant across all completed assessment timepoints, that all values have been verified against the original source data collection forms, and that this record is complete and consistent with the trial database.

Principal Investigator / Assessor

_____
Name

_____
Signature

_____
Date

Independent Reviewer / Supervisor

_____
Name

_____
Signature

_____
Date

CONFIDENTIAL: This form contains protected health information. Handle in accordance with the approved data management protocol (Ethics REC/09/2024 (PG/FB/30)). Store securely and restrict access to authorized research personnel only.

APPENDIX 25

Table of Normality Test

Normality Test for NEX Condition

		Shapiro Wilk				
Variables	**Time**	**W**	**Sig.**	**Mean**	**Skewness**	**Kurtosis**
Body Weight (kg)	Baseline	0.9514	0.004	87.60	0.074	-1.180
	12-week	0.9727	0.087	84.98	0.171	-0.790
	Follow-up	0.9583	0.011	86.46	-0.110	-1.060
Waist Circumference (cm)	Baseline	0.9550	0.006	99.68	-0.027	-1.174
	12-week	0.9600	0.013	97.54	-0.010	-1.115
	Follow-up	0.9379	<0.001	98.39	-0.136	-1.311
Fat Mass (kg)	Baseline	0.9596	0.014	35.14	0.008	-1.071
	12-week	0.9407	0.001	33.13	0.101	-1.250
	Follow-up	0.9608	0.016	33.72	0.175	-1.078
Fat-free Mass (kg)	Baseline	0.9814	0.305	52.46	0.195	-0.242
	12-week	0.9803	0.261	51.85	0.038	-0.734
	Follow-up	0.9761	0.144	52.74	-0.134	-0.876
Energy Density (kcal/g)	Baseline	0.9867	0.590	4.844	0.310	0.484
	12-week	0.9890	0.741	4.780	0.025	-0.108

	Follow-up	0.9868	0.591	4.844	0.207	0.316
Energy Intake (kcal/day)	Baseline	0.9152	<0.001	1830.0	-0.016	-1.484
	12-week	0.8760	<0.001	1737.0	-0.045	-1.674
	Follow-up	0.9130	<0.001	1807.0	0.070	-1.465
Carbohydrate Intake (g/day)	Baseline	0.9339	<0.001	202.7	-0.027	-1.320
	12-week	0.9047	<0.001	194.4	-0.022	-1.548
	Follow-up	0.9844	0.448	200.2	0.021	-0.599
Fat Intake (g/day)	Baseline	0.9592	0.013	80.51	0.061	-1.084
	12-week	0.9694	0.054	74.31	-0.106	-0.973
	Follow-up	0.9488	0.003	79.76	0.128	-1.197
Protein Intake (g/day)	Baseline	0.9385	<0.001	73.66	0.144	-1.214
	12-week	0.9266	<0.001	72.62	0.025	-1.407
	Follow-up	0.9673	0.040	72.07	0.123	-0.925
Fibre Intake (g/day)	Baseline	0.9638	0.024	21.08	0.009	-1.094
	12-week	0.9401	0.001	23.01	-0.115	-1.267
	Follow-up	0.9271	<0.001	21.19	-0.213	-1.262
PA Accelerometer (MET-min/wk)	Baseline	0.9541	0.006	1708.0	-0.071	-1.163
	12-week	0.9488	0.003	1858.0	-0.028	-1.272
	Follow-up	0.9463	0.002	1761.0	0.068	-1.208

PA IPAQ-LV (min/week)	Baseline	0.9472	0.003	944.5	0.110	-1.034
	12-week	0.9687	0.049	1044.0	-0.236	-0.786
	Follow-up	0.9659	0.033	1043.0	0.073	-0.982
HbA1c (%)	Baseline	0.9738	0.103	5.355	0.112	-0.397
	12-week	0.9469	0.003	5.336	0.066	-1.262
	Follow-up	0.9490	0.003	5.341	-0.073	-1.131
Total Cholesterol (mg/dL)	Baseline	0.9730	0.092	210.3	0.009	-0.895
	12-week	0.9550	0.007	201.4	-0.324	-0.898
	Follow-up	0.9518	0.005	201.1	0.068	-1.183
LDL Cholesterol (mg/dL)	Baseline	0.9435	0.002	137.7	-0.274	-1.156
	12-week	0.9404	0.001	132.1	-0.004	-1.311
	Follow-up	0.9810	0.288	132.4	0.274	-0.563
HDL Cholesterol (mg/dL)	Baseline	0.9817	0.317	48.82	0.395	-0.142
	12-week	0.9870	0.609	46.41	0.159	-0.480
	Follow-up	0.9862	0.555	45.87	-0.035	-0.684
Triglycerides (mg/dL)	Baseline	0.9065	<0.001	118.7	-0.133	-1.488
	12-week	0.9482	0.003	114.4	0.007	-1.251
	Follow-up	0.9595	0.013	114.0	-0.056	-1.056
Systolic BP (mmHg)	Baseline	0.9369	<0.001	128.5	-0.005	-1.336
	12-week	0.9553	0.007	128.7	-0.009	-1.167

	Follow-up	0.9494	0.003	129.7	0.159	-1.118
Diastolic BP (mmHg)	Baseline	0.9427	0.002	87.90	-0.053	-1.283
	12-week	0.9366	<0.001	87.18	-0.061	-1.286
	Follow-up	0.9542	0.006	88.18	-0.035	-1.193
Mean Arterial Pressure (mmHg)	Baseline	0.9175	<0.001	101.4	-0.052	-1.477
	12-week	0.8892	<0.001	101.1	-0.020	-1.622
	Follow-up	0.9238	<0.001	102.0	0.002	-1.434

Note. n = 81 for all variables. W = Shapiro-Wilk test statistic. PA = physical activity; IPAQ-LV = International Physical Activity Questionnaire, Long Version; HbA1c = glycated haemoglobin; LDL = low-density lipoprotein; HDL = high-density lipoprotein; BP = blood pressure. 5% trimmed mean was not available from the analysis software (GraphPad Prism) and is therefore not reported

Normality Test for NEX Condition

Note. n = 81 for all variables. W = Shapiro-Wilk test statistic. PA = physical activity; IPAQ-LV = International Physical Activity Questionnaire, Long Version; HbA1c = glycated haemoglobin; LDL = low-density lipoprotein; HDL = high-density lipoprotein; BP = blood pressure. 5% trimmed mean was not available from the analysis software (GraphPad Prism) and is therefore not reported.

Normality Test for CON Condition

		Shapiro Wilk				
Variables	**Time**	**W**	**Sig.**	**Mean**	**Skewness**	**Kurtosis**
Body Weight (kg)	Baseline	0.9514	0.004	87.60	0.074	-1.180
	12-week	0.9727	0.087	84.98	0.171	-0.790
	Follow-up	0.9583	0.011	86.46	-0.110	-1.060
Waist Circumference (cm)	Baseline	0.9550	0.006	99.68	-0.027	-1.174
	12-week	0.9600	0.013	97.54	-0.010	-1.115
	Follow-up	0.9379	<0.001	98.39	-0.136	-1.311
Fat Mass (kg)	Baseline	0.9596	0.014	35.14	0.008	-1.071
	12-week	0.9407	0.001	33.13	0.101	-1.250
	Follow-up	0.9608	0.016	33.72	0.175	-1.078
Fat-free Mass (kg)	Baseline	0.9814	0.305	52.46	0.195	-0.242
	12-week	0.9803	0.261	51.85	0.038	-0.734
	Follow-up	0.9761	0.144	52.74	-0.134	-0.876
Energy Density (kcal/g)	Baseline	0.9867	0.590	4.844	0.310	0.484
	12-week	0.9890	0.741	4.780	0.025	-0.108
	Follow-up	0.9868	0.591	4.844	0.207	0.316
Energy Intake (kcal/day)	Baseline	0.9152	<0.001	1830.0	-0.016	-1.484
	12-week	0.8760	<0.001	1737.0	-0.045	-1.674

	Follow-up	0.9130	<0.001	1807.0	0.070	-1.465
Carbohydrate Intake (g/day)	Baseline	0.9339	<0.001	202.7	-0.027	-1.320
	12-week	0.9047	<0.001	194.4	-0.022	-1.548
	Follow-up	0.9844	0.448	200.2	0.021	-0.599
Fat Intake (g/day)	Baseline	0.9592	0.013	80.51	0.061	-1.084
	12-week	0.9694	0.054	74.31	-0.106	-0.973
	Follow-up	0.9488	0.003	79.76	0.128	-1.197
Protein Intake (g/day)	Baseline	0.9385	<0.001	73.66	0.144	-1.214
	12-week	0.9266	<0.001	72.62	0.025	-1.407
	Follow-up	0.9673	0.040	72.07	0.123	-0.925
Fibre Intake (g/day)	Baseline	0.9638	0.024	21.08	0.009	-1.094
	12-week	0.9401	0.001	23.01	-0.115	-1.267
	Follow-up	0.9271	<0.001	21.19	-0.213	-1.262
PA Accelerometer (MET-min/wk)	Baseline	0.9541	0.006	1708.0	-0.071	-1.163
	12-week	0.9488	0.003	1858.0	-0.028	-1.272
	Follow-up	0.9463	0.002	1761.0	0.068	-1.208
PA IPAQ-LV (min/week)	Baseline	0.9472	0.003	944.5	0.110	-1.034
	12-week	0.9687	0.049	1044.0	-0.236	-0.786
	Follow-up	0.9659	0.033	1043.0	0.073	-0.982

HbA1c (%)	Baseline	0.9738	0.103	5.355	0.112	-0.397
	12-week	0.9469	0.003	5.336	0.066	-1.262
	Follow-up	0.9490	0.003	5.341	-0.073	-1.131
Total Cholesterol (mg/dL)	Baseline	0.9730	0.092	210.3	0.009	-0.895
	12-week	0.9550	0.007	201.4	-0.324	-0.898
	Follow-up	0.9518	0.005	201.1	0.068	-1.183
LDL Cholesterol (mg/dL)	Baseline	0.9435	0.002	137.7	-0.274	-1.156
	12-week	0.9404	0.001	132.1	-0.004	-1.311
	Follow-up	0.9810	0.288	132.4	0.274	-0.563
HDL Cholesterol (mg/dL)	Baseline	0.9817	0.317	48.82	0.395	-0.142
	12-week	0.9870	0.609	46.41	0.159	-0.480
	Follow-up	0.9862	0.555	45.87	-0.035	-0.684
Triglycerides (mg/dL)	Baseline	0.9065	<0.001	118.7	-0.133	-1.488
	12-week	0.9482	0.003	114.4	0.007	-1.251
	Follow-up	0.9595	0.013	114.0	-0.056	-1.056
Systolic BP (mmHg)	Baseline	0.9369	<0.001	128.5	-0.005	-1.336
	12-week	0.9553	0.007	128.7	-0.009	-1.167
	Follow-up	0.9494	0.003	129.7	0.159	-1.118
Diastolic BP (mmHg)	Baseline	0.9427	0.002	87.90	-0.053	-1.283
	12-week	0.9366	<0.001	87.18	-0.061	-1.286

	Follow-up	0.9542	0.006	88.18	-0.035	-1.193
Mean Arterial Pressure (mmHg)	Baseline	0.9175	<0.001	101.4	-0.052	-1.477
	12-week	0.8892	<0.001	101.1	-0.020	-1.622
	Follow-up	0.9238	<0.001	102.0	0.002	-1.434

Note. n = 79 for body weight, fat mass, fat-free mass, energy density, energy intake, carbohydrate intake, fat intake, protein intake, fibre intake, PA accelerometer, PA IPAQ-LV, HbA1c, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, systolic BP, diastolic BP, and mean arterial pressure; n = 81 for waist circumference. W = Shapiro-Wilk test statistic. PA = physical activity; IPAQ-LV = International Physical Activity Questionnaire, Long Version; HbA1c = glycated haemoglobin; LDL = low-density lipoprotein; HDL = high-density lipoprotein; BP = blood pressure. 5% trimmed mean was not available from the analysis software (GraphPad Prism) and is therefore not reported.

APPENDIX 26

Q-Q Plot for All Dataset

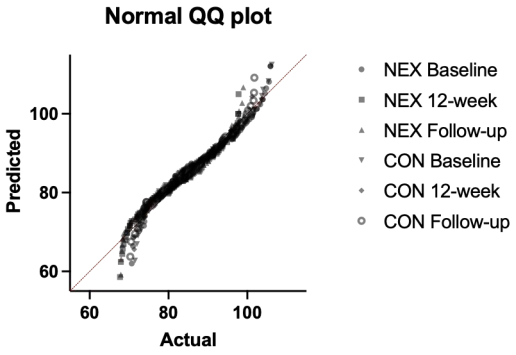


Figure: Q-Q Plot for Body Weight Dataset

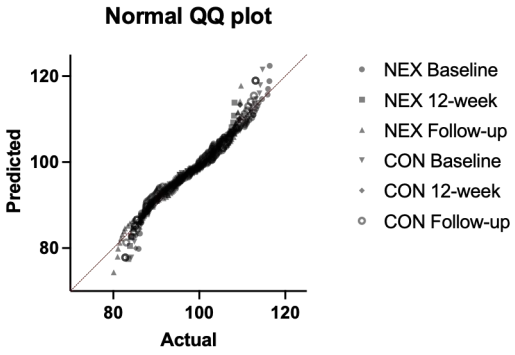


Figure: Q-Q Plot for Waist Circumference Dataset

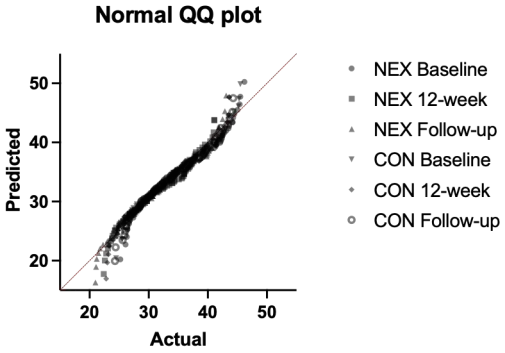


Figure: Q-Q Plot for Fat Mass Dataset

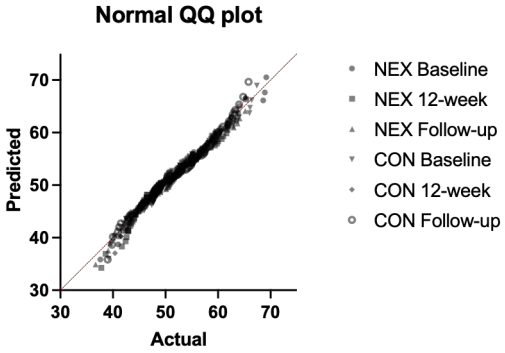


Figure: Q-Q Plot for Fat-free Mass Dataset

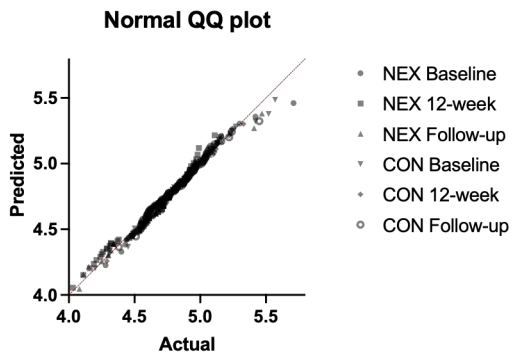


Figure: Q-Q Plot for Energy Density Dataset

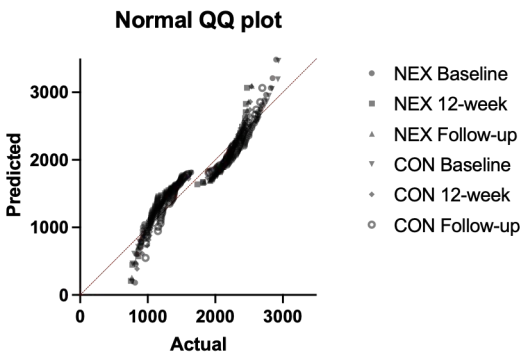


Figure: Q-Q Plot for Energy Intake Dataset

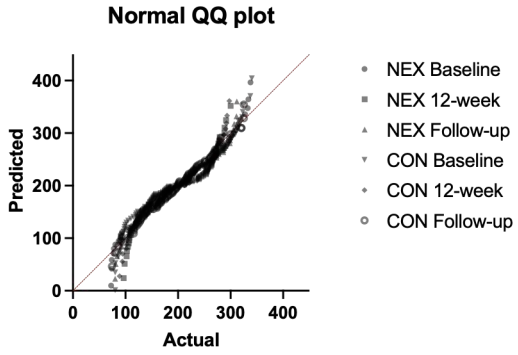


Figure: Q-Q Plot for Carbohydrate Intake Dataset

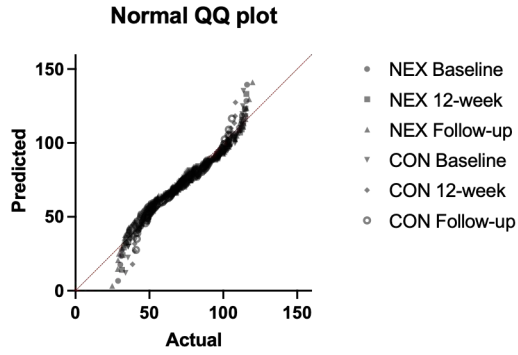


Figure: Q-Q Plot for Protein Intake Dataset

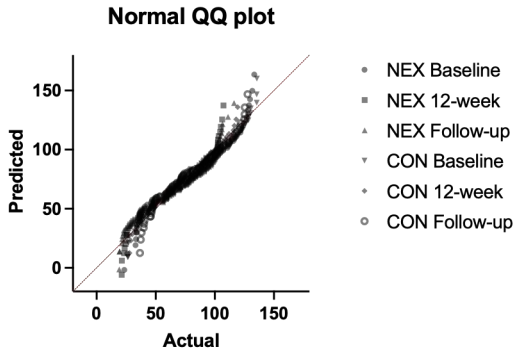


Figure: Q-Q Plot for Fat Intake Dataset

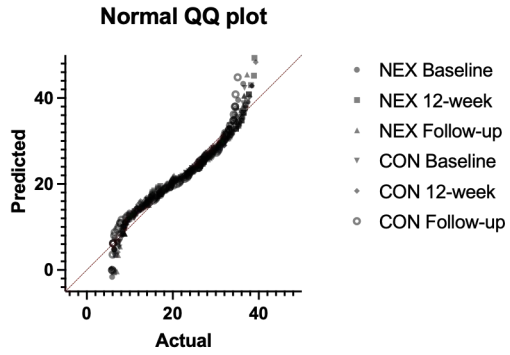


Figure: Q-Q Plot for Fiber Intake Dataset

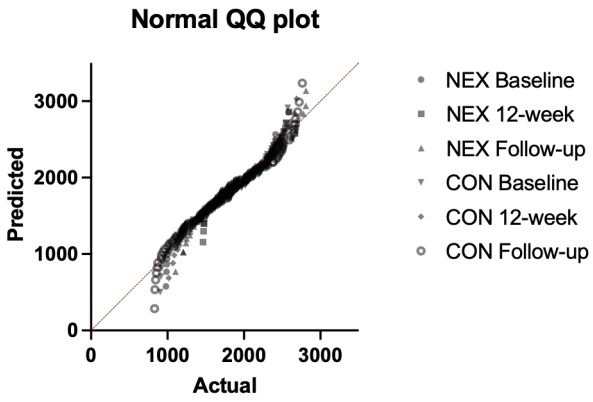


Figure: Q-Q Plot for PA Accelerometer Dataset

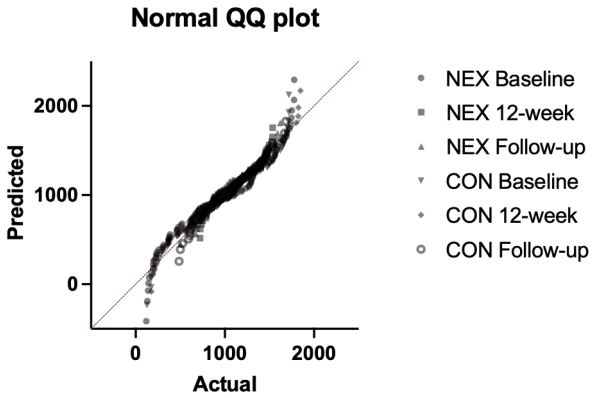


Figure: Q-Q Plot for PA IPAQ-LV Dataset

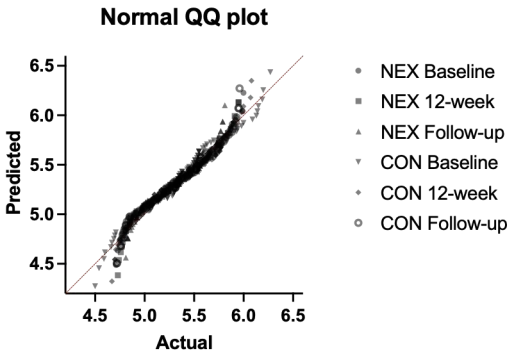


Figure: Q-Q Plot for XXX Dataset

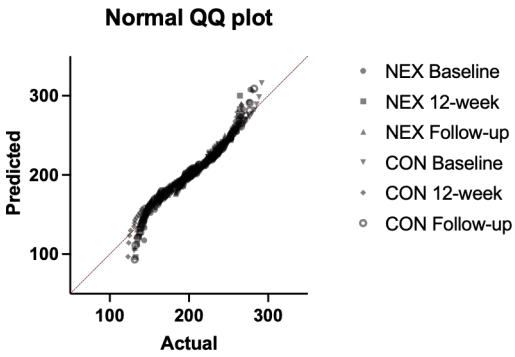


Figure: Q-Q Plot for Total Cholesterol Dataset

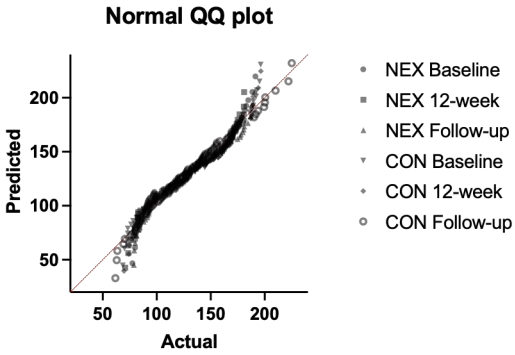


Figure: Q-Q Plot for LDL Cholesterol Dataset

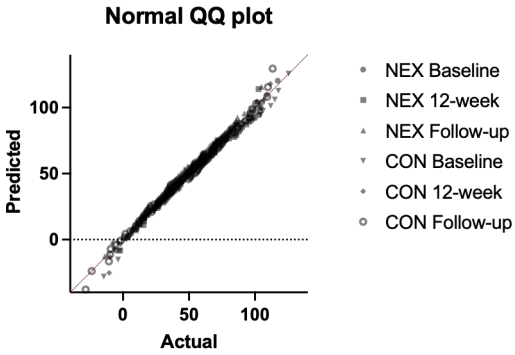


Figure: Q-Q Plot for HDL Cholesterol Dataset

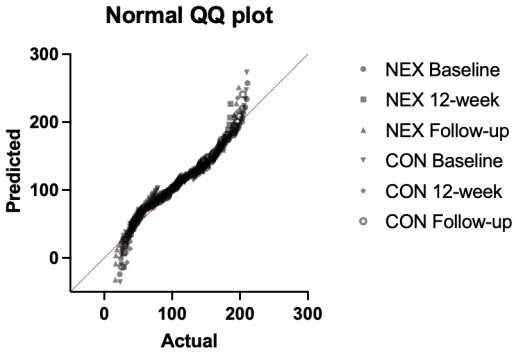


Figure: Q-Q Plot for Triglycerides Dataset

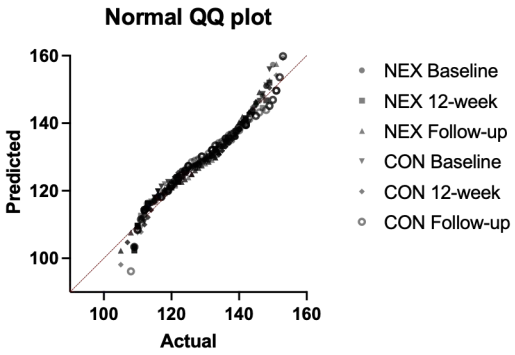


Figure: Q-Q Plot for Systolic BP Dataset

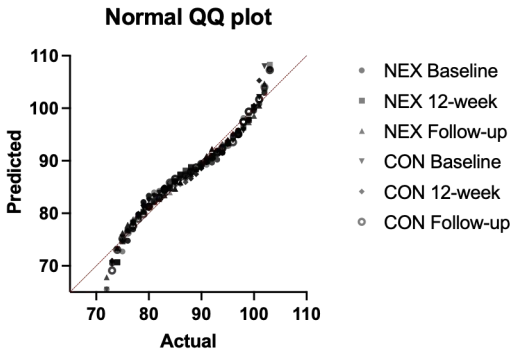


Figure: Q-Q Plot for Diastolic BP Dataset

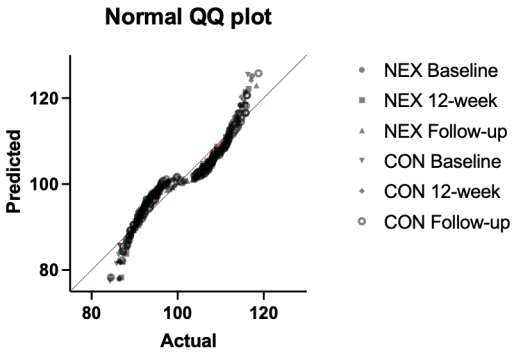


Figure: Q-Q Plot for Mean Arterial Pressure Dataset

APPENDIX 27

Table of Homogeneity Test

Homogeneity Test for the Weight and Body Composition Parameters

Parameters	**Levene's test for Equality of Variance**	
	f	Sig.
Body Weight, Baseline	0.26	0.85
Body Weight, 12-week	2.12	0.10
Body Weight, Follow-up	2.00	0.12
Waist Circumference, Baseline	0.62	0.60
Waist Circumference, 12-week	0.51	0.68
Waist Circumference, Follow-up	1.76	0.16
Fat Mass, Baseline	2.01	0.12
Fat Mass, 12-week	2.72	0.05
Fat Mass, Follow-up	3.35	0.02
Fat-free Mass, Baseline	0.18	0.91
Fat-free Mass, 12-week	3.77	0.01
Fat-free Mass, Follow-up	0.77	0.51

Homogeneity Test for the Dietary Changes Parameters

Parameters	**Levene's test for Equality of Variance**	
	f	Sig.
Energy Density, Baseline	8.40	0.00
Energy Density, 12-week	13.60	0.00
Energy Density, Follow-up	15.55	0.00
Energy Intake, Baseline	0.39	0.76
Energy Intake, 12-week	1.23	0.30
Energy Intake, Follow-up	0.27	0.85
Carbohydrate Intake, Baseline	0.47	0.71
Carbohydrate Intake, 12-week	2.51	0.06
Carbohydrate Intake, Follow-up	0.86	0.47
Protein Intake, Baseline	1.66	0.18
Protein Intake, 12-week	2.86	0.04
Protein Intake, Follow-up	3.06	0.03
Fat Intake, Baseline	0.61	0.61
Fat Intake, 12-week	1.31	0.27
Fat Intake, Follow-up	2.35	0.07
Fiber Intake, Baseline	0.71	0.55
Fiber Intake, 12-week	3.50	0.02
Fiber Intake, Follow-up	3.81	0.01

Homogeneity Test for the Physical Activity Parameters

Parameters	**Levene's test for Equality of Variance**	
	f	Sig.
Accelerometer, Baseline	1.04	0.38
Accelerometer, 12-week	3.45	0.02
Accelerometer, Follow-up	3.07	0.03
IPAQ-LV, Baseline	2.43	0.07
IPAQ-LV, 12-week	14.20	0.00
IPAQ-LV, Follow-up	1.09	0.36

Homogeneity Test for the Metabolic Changes Parameters

Parameters	**Levene's test for Equality of Variance**	
	f	Sig.
HbA1c, Baseline	2.90	0.04
HbA1c, 12-week	0.29	0.83
HbA1c, Follow-up	5.16	0.00
Total Cholesterol, Baseline	3.58	0.02
Total Cholesterol, 12-week	5.35	0.00
Total Cholesterol, Follow-up	3.21	0.02
LDL Cholesterol, Baseline	4.09	0.01
LDL Cholesterol, 12-week	0.61	0.61
LDL Cholesterol, Follow-up	4.07	0.01
HDL Cholesterol, Baseline	2.27	0.08
HDL Cholesterol, 12-week	4.30	0.01
HDL Cholesterol, Follow-up	3.34	0.02
Triglycerides, Baseline	1.28	0.28
Triglycerides, 12-week	0.15	0.93
Triglycerides, Follow-up	0.98	0.40

Homogeneity Test for the Metabolic Changes Parameters

Parameters	**Levene's test for Equality of Variance**	
	f	Sig.
Systolic Blood Pressure, Baseline	0.42	0.74
Systolic Blood Pressure, 12-week	0.87	0.46
Systolic Blood Pressure, Follow-up	4.83	0.00
Diastolic Blood Pressure, Baseline	0.83	0.48
Diastolic Blood Pressure, 12-week	0.63	0.60
Diastolic Blood Pressure, Follow-up	0.32	0.81
Mean Arterial Pressure, Baseline	1.71	0.17
Mean Arterial Pressure, 12-week	0.78	0.51
Mean Arterial Pressure, Follow-up	0.14	0.94

AUTHOR'S PROFILE



Azwa Suraya Binti Mohd Dan graduated with a Bachelor of Sports Science (Hons.) from Universiti Teknologi MARA (UiTM), Negeri Sembilan, in 2019. Her undergraduate research focused on exercise and obesity management through bibliometric analysis and a review of systematic reviews and meta-analyses. She pursued a Master's degree at UiTM, integrating exercise science, nutrition, and artificial intelligence. Her thesis, titled "Development, Validation, and Effectiveness of a Personalized AI Prompt Generator (NExGEN-ChatGPT) for Weight Management in Adults with Overweight and Obesity," produced two indexed and high impact publications, eight non-indexed publications. The first validated the NExGEN-ChatGPT framework using the Fuzzy Delphi Method, published in Biology Methods and Protocols (Oxford University Press, 2025). The second tested its effectiveness through a randomised controlled trial, accepted in the Journal of Technology in Behavioral Science (Springer Nature, 2026). Her research interests include AI-driven health interventions, personalised digital weight management, and prompt engineering for clinical applications.

LIST OF MAJOR PUBLICATION

Dan, A. S. M., Linoby, A., Shahlan Kasim, S., Zaki, S., Sazali, R., Yusoff, Y., Nasir, Z., & Haziq Abidin, A. (2025). Validation of a personalized AI prompt generator (NExGEN-ChatGPT) for obesity management using fuzzy Delphi method. Biology Methods and Protocols, 10 (1) (Web of Science and Scopus-indexed,

SJR Quartile 1, Impact Factor: 1.3). https://doi.org/10.1093/biomethods/bpaf085.

Dan, A. S. M., Linoby, A., Shahlan Kasim, S., Zaki, S., Sazali, R., Yusoff, Y., Nasir, Z., & Haziq Abidin, A. (2026). Personalized AI Prompt Generator and ChatGPT for Weight Loss: Randomized Controlled Trial in Adults with Overweight and Obesity. Journal of Technology in Behavioral Science, Springer Nature (Scopus-indexed, SJR Quartile 1, H-index: 24). (Accepted – ongoing publication; 2026).

LIST OF OTHER PUBLICATION

Dan, A. S. M., Linoby, A., Kasim, S. S., Lamat, S. A., Zaki, S., Sazali, R., Felder, H. Assessing the Quality of a Personalized Prompt Generator and AI-Chatbot (ChatGPT) for Dietary and Exercise Planning in Obese Adults Using the Fuzzy Delphi Method. https://www.medrxiv.org/content/10.1101/2025.08.27.25334528v1.

Dan, A. S. M., Linoby, A., Kasim, S. S., Lamat, S. A., Zaki, S., Sazali, R., Felder, H. Evaluating the Quality of AI-Generated Diet and Exercise Plans: A Preliminary Study Using the NExGEN Prompt Generator System (Extended Abstract). International Online Conference on Exercise Science, Sports Management, Outdoor Recreation, and Physical Education (ExSPORT) 2024 – Online, Malaysia.

Musyid, S., Linoby, A., Hussain, R. N. J. R., Dan, A. S. M., Nugraha R., and Kamalden, N. T. T. Effect of Single Bout Exercise Modalities on Multi-domain Cognitive Functions in Recreationally Active Older Adults. Retos, 66, 313–322. https://doi.org/10.47197/retos.v66.110457.

Khairolnezam,M.Q., Linoby, A., M. Z., Razif.S., Amrun.H., Dan, A. S. M., and Yusandra.M.Y. Personalized prompt Generator and AI Chatbot for Dietary and Exercise Planning: A Controlled Investigation of Adherence and Physical Fitness Over 12 Weeks in Older Malaysian Adults. International Graduate Colloquium: Sports and Physical Exercise Assembly of Knowledge Sharing, i-SPEAK 2025– Online, Malaysia.

Manan, S., Linoby, A., M. Z., Razif.S ., Amrun.H., Dan, A. S. M., and Yusandra.M.Y. AI-Driven Personalized Athletic Interventions: Impact of the NExGEN Prompt

Generator-ChatGPT Dietary and Physical Training Adherence and Endurance performance in Team Sport Athletes. International Graduate Colloquium: Sports and Physical Exercise Assembly of Knowledge Sharing, i-SPEAK 2025– Online, Malaysia.

Rosli, A., Linoby, A., M. Z., Razif.S., Amrun.H., Dan, A. S. M., and Yusandra.M.Y. Enhancing AI-Driven Personalized Exercise and Nutrition Planning for Elderly Through Prompt Refinement and Expert Consensus. International Graduate Colloquium: Sports and Physical Exercise Assembly of Knowledge Sharing, i-SPEAK 2025– Online, Malaysia.

Roslin, A., Linoby, A., M.Z., Razif.S., Amrun.H., Dan, A. S. M., and Yusandra.M.Y. Expert Consensus-Driven Refinement of NExGEN Prompt Generator and AI Chatbot Frameworks for Personalized Athletic Planning. International Graduate Colloquium: Sports and Physical Exercise Assembly of Knowledge Sharing, i-SPEAK 2025– Online, Malaysia.

Musyid, Musyid, S., Linoby, A., Hussain, R. N. J. R., Dan, A. S. M., Nugraha R., and Kamalden, N. T. T. Effect of Single Bout Exercise Modalities on Multi-domain Cognitive Functions in Recreationally Active Older Adults. International Online Conference on Exercise Science, Sports Management, Outdoor Recreation, and Physical Education (ExSPORT) 2024 – Online, Malaysia..

LIST OF CONFERENCE PARTICIPATION

Dan, A. S. M., Linoby, A., Akmal, R. A., Shahlan Kasim, S., Zaki, S.,. Presenting research paper titled ‘Development and Clinical Validation of a Personalized AI Prompt Generator for Obesity Management: From Expert Consensus to Randomized Controll Trial Evidence’ in ASEAN Symposium on Sports Innovation 2025 in conjunction with 4th International Conference on Innovation and Technology in Sports 2025 (30 September - 2 October 2025)