

Recommender Algorithm for Supporting Self-Management of CVD Risk Factors in an Adult Population at Home

Tatiana V. Afanasieva

*Department of Informatics,
Plekhanov Russian University of Economics
36, Stremyanny lane, Moscow, Russia*

afanaseva.tv@rea.ru

Pavel V. Platov

*Department of Information Systems,
Ulyanovsk State Technical University
32 Severny Venetz Street, Ulyanovsk, Russia*

Anastasia I. Medvedeva

*Department of Informatics,
Plekhanov Russian University of Economics
36, Stremyanny lane, Moscow, Russia*

Corresponding Author: Tatiana V. Afanasieva

Copyright © 2024 Tatiana V. Afanasieva, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Since CVDs are the leading cause of death worldwide, this article solves the problem of developing a recommendation algorithm for the prevention of cardiovascular diseases (CVDs), as the main component of health recommender systems. To address this issue, a knowledge-based recommendation algorithm was proposed to support self-management of CVD risk factors in adults at home. The proposed algorithm is based on a new user profile and an original multidimensional recommendation model. The user profile includes not only descriptive but also predictive assessments of cardiovascular health based on the SCORE model which are outlined in official guidelines. A multidimensional recommendation model contains target, informational, and explanatory components for higher health lifestyle adherence. We found that ChatGPT provided plausible and understandable explanations of cardiovascular disease risk factors. therefore, these explanations were used in the texts of the recommendations. The main feature of the proposed algorithm is the combination of rule-based logic with the capabilities of a large language model to generate human-like text of multidimensional recommendations. Evaluation of the proposed algorithm showed high user satisfaction with the proposed recommendation algorithm.

Keywords: Knowledge-based recommender, Multidimensional recommendation, Multifactorial user profile, Explanations, ChatGPT.

1. INTRODUCTION

Early detection and prevention of CVD, including management risk factors [1], is an important focus of preventive services in all countries. To solve this problem, scientific research is being carried out [2], programs [3], algorithms [4], models [5], and clinical guidelines for the prevention of CVD [6, 7], intended for specialists of medical organizations. Systematization of the main problems in effective strategies for the prevention of CVD at the individual level is discussed in the article [8]: a) lack of knowledge among the population about signs of diseases, risk factors and individual risk; (b) lack of digital tools for disease prevention; (c) insufficient physician time to provide detailed patient-centered prevention recommendations during clinical visits; (d) targeting prevention at high-risk individuals, whereas most acute cardiovascular events occur in low- or moderate-risk individuals.

There is challenge in self-manage of CV risk factors at home as they are characterized by diversity and heterogeneity, their indicators are presented on different scales, assessment of which depends on gender, age, non-modifiable CV risk factors and clinical conditions [9].

The development of digital technologies has enabled the growth of digital prevention systems among which Health Recommendation Systems (HRS) focused on self-management at home and promoting healthy lifestyles are promising for disease prevention [10]. HRS is a specialized recommendation system that, based on a user's profile (model), generates a list of recommendation items using information sources related to healthcare and health lifestyle. The main component of HRS is recommendation algorithm providing selecting recommended items, in which the selection conditions are determined by user profile, or by attributes of recommended items or based on domain knowledge [11–13]. As noted in the review [10], the most common HRSs are knowledge-based systems, where knowledge is represented in the form of rules. Knowledge-based recommendation algorithms for self-manage of CVD risk factors at home are the most adequate too, as CVD prevention is a high-risk domain of healthcare in which precise and timely recommendations based on knowledge can prevent or delay fatal CV events.

At the same time incorrect recommendations obtained using another type of recommendation algorithm can have irreversible negative consequences for a person [11].

The goal of this study is to create a knowledge-based recommendation algorithm for supporting self-management of CV risk factors in an adult population at home, which will generate individual recommendations based on 2021 ESC Guidelines on CVD prevention in clinical practice [6].

Another motivation for developing knowledge-based recommendation algorithm for supporting self-management of CV risk factors in an adult population at home is the limitations and the gaps in research in this domain, namely:

1. The set of risk factors used is incomplete. Recommender algorithms utilize few CVD risk factors that characterize a person's lifestyle [10, 14–16].
2. Limitations of the user profile model. Recommender algorithms create recommendations based on the user's current profile and do not consider predictive assessments of the CVD development [10].

3. Lack of a formal model of the recommendation. Recommendations are presented as a list of recommendations items of the same dimension of meaning, of what is recommended to be done, for example, “Read an online article about the symptoms of stress and good practices for stress management” [16], or “Forget about sugar” [15].

To overcome these limitations and fill the gaps in research of recommendation algorithms, this article presents knowledge-based recommender algorithm, called CVD-Risk-Prevent, for supporting self-management of multifactorial person’s CV risk at home.

The contribution of the study is defined as follows:

In order to solve the problem of incomplete recommendations regarding the number and type of risk factors the multifactorial CVD risk model is presented based on the set of risk factors described in official guidelines [6, 7]. A user profile is proposed containing descriptive assessments of CVD risk factors and predictive ones of 10-year risk of CVD events.

The formal model of recommendation is described as a structure with 3 dimensions of meaning similar to the content of CVD preventive recommendation given by the physician. Each dimension of meaning is associated with one type of recommendation items, that are targeted, informational, explanatory. Thus, multidimensional recommendations for any person have the same three-dimensional structure but with individual content, determined by the user profile.

Based on above models, CVD-Risk-Prevent algorithm was developed to support self-management of CV risk factors in adults at home. The logic of the algorithm is based on rules of three types: rules for identifying individual CV risk factors, rules for descriptive and predictive assessment of CVD risk, and rules for creating multidimensional recommendations. The peculiarity of the algorithm for creating recommendations is that a large language model, ChatGPT (Generative Pre-trained Transformer), is also used to obtain the explanatory items of the recommendations.

Framework of the study design is presented in the FIGURE 1.

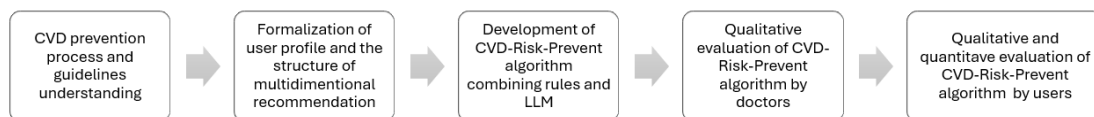


Figure 1: Framework of study design

The rest of the article is organized as follows. The second Section provides a brief overview of the field of health recommendation algorithms and recommender systems for person’ support.. A model of multidimensional recommendations based on a formal representation of multiple CVD risk factors, including 3 types of recommendation items, is presented in Section 3. The fourth Section describes multifactorial user’s CVD risk model. Section 5 presents the proposed recommendation algorithm, which we called CVD-Risk-Prevent, for supporting self-management of CV risk factors in an adult at home. Using ChatGPT for explanatory items creation is described and discussed in Section 6. In the Section 7 evaluation of CVD-Risk-Prevent algorithm is provided.. Discussion and conclusions are presented in the last Sections of the article.

2. LITERATURE REVIEW

The existing research landscape of HRS is provided in survey [10], where authors highlighted that HRS is a specialized recommender system that supplies a user with personalized and highly relevant health information, based on their profile, specific health conditions and needs. HRS, in contrast to the clinic decision support system, are focused on supporting patients in following the principles of a healthy lifestyle and balanced diet, improving medical literacy and have undeniable value for disease prevention and increasing patient adherence, in particular in their own health management. As follows from studies [9–11, 14–16], knowledge-based recommendation algorithms are becoming more and more suitable and widespread among all HRS, aimed at forming a healthy lifestyle [17]. Knowledge-based recommendation algorithms take as input the data of one user X and, using a set of pre-generated rules, associate the user's profile with a set of recommendation items $H \subset M$. Unlike recommender systems that create recommendations for similar recommendation items or similar patients, knowledge-based recommender systems can create recommendations that consider the domain knowledge, limitations, health conditions, and preferences of a particular patient. The sources of recommendation items in such algorithms are expert knowledge, knowledge extracted from official medical guidelines and knowledge about the person's health status. Authors of review [11], emphasize that the fields of application of knowledge-based recommendation are those in which irrelevant recommendations can cause serious negative consequences. In the paper [18], user needs for explanation of recommendation received by recommender systems are studied including knowledge-based recommendation.

A study [15], focuses on creating a HRS for recommending the list of challenges that positively impact the health of individuals. The pool of 30 challengers is created by experts and user profile includes physical, mental, and nutritional aspects. By challenge the authors mean a brief recommendation, for example, "Sleep eight hours a day" for people from the group of sleep and wakefulness disorders. The lack of explanation of why these recommendations should be implemented is a limitation of the proposed recommendation system. Article [16], notes the limitations of knowledge-based HRS of the range of lifestyle risk factors used, namely physical activity and diet. They proposed knowledge-based recommender algorithm expanding multiple lifestyle risk factors from a multidimensional viewpoint. The recommendations received were intended to support identifying appropriate coaching plans for stress management. Proposed system was designed to be used only at the beginning of the health-coaching program, which limits its usefulness as a stand-alone HRS for the long-term support of health behavior change. However, the studies [15, 16], focused on persons without any medical conditions, while patients with CVD need supportive advice as well.

Integration of rule-based and Natural Language Processing (NLP) approaches to create knowledge-based recommendation on personalized educational materials for chronic disease patients in China was proposed in the study [9]. The patient profile was presented using ontology, and then user vector was generated by the 80 SWRL rules, previously defined. The ontology structure includes patient characteristics in 5 areas: chronic diseases, lifestyle and biological risk factors, demographic features and medication use. NLP technology was applied to obtain text vectors based on the keywords of each educational document and Word2Vec embeddings to map the extracted keywords to the ontology vector space. Recommendations generated through this method are more interpretable than in content-based and collaborative HRS. However, in this study patients' opinion of usefulness are not provided and the constructed ontology remains to be further validated on official guidelines for consistency, correctness, and completeness. Recent advancements in natural language

processing have introduced Large Language Models (LLMs) that exhibit remarkable capabilities in understanding and generating human-like text [19]. As noted in [11], the implementation of LLM in knowledge-based recommender systems is beginning to play an important role in the context of various tasks, including the creation of explanations in specific recommendation contexts. These models have the big potential in cardiology improving medical diagnosis and decision support by analyzing vast amounts of medical data and provide educational materials for patients [20]. However, to the best of our knowledge, no studies to date have concentrated on a combining knowledge-based recommendation algorithms with ChatGPT for supporting self-management of the set of CV risk factors outlined in official guidelines [6, 7].

3. FORMAL MODEL OF MULTIDIMENSIONAL RECOMMENDATION

We propose a formal model of multidimensional recommendations for supporting self-management CV risk factors in adults at home in the form of a set of recommendation components of different dimensions, namely: targeted *Es*, informational *Inf*, explanatory *Expl*:

$$\begin{aligned} R &= \{Es, Inf, Expl\}, \{Inf(1), Inf(2), \dots, Inf(N)\}, \\ Expl &= \{Expl(1), Expl(2), \dots, Expl(N)\}, \end{aligned} \quad (1)$$

where N is the number of person's CV risk factors.

The targeted recommendation component *Es* includes goals at the strategic level and tactical level. The informational component *Inf* contains descriptive and predictive assessments the state of CV health. The explanatory component *Expl* answers the question "Why is this recommended?". The key difference between the proposed recommendation model and those used in other knowledge-based recommendation algorithms is its structure and semantic multidimensionality, corresponding to the logic of counseling on the management of CVD risk factors.

4. MULTIFACTORIAL USER'S CVD RISK MODEL

Let's define a multifactorial model of CV risk factors in the form of the following expressions:

$$Z = \{S, NonModF, Biol, Beh\}, NonModF = \{CVD, KD, DM2, GEN\}, \quad (2)$$

$$Biol = \{OBS, CH, NHL, SBP, GL\}, Beh = \{Sm, Fa, Diet\}. \quad (3)$$

The variable designation used in expressions (2-3) is presented in TABLE 1, where the sign of the CV risk factor is coded by value "1", and its absence - by value "0".

In the formal model *Z*, we use 13 CVD risk factors grouped by their types. In accordance with [6], when defining person's CVD risk the class of person and prediction assessment should be determined.

Using the multifactorial model of CV risk factors we presented the model of user profile as following

$$UP = \{Factor, C, Class\}. \quad (4)$$

Table 1: Designation of CV risk factors. If there is a risk factor, its value will be set to “1”, otherwise “0”

Factor No.	CV risk factor designation and description
Type 1: Symptoms	
1	S– a sign of symptoms of angina pectoris with a significant deterioration in health:
Type 2: Non-modifiable factors (NonModF)	
2	CVD - a sign that the person has documented one of the CVD
3	KD– a sign of chronic kidney disease
4	DM2– a sign of type 2 diabetes
5	GEN– sign of family history of early CV diseases
Type 3: Modifiable biological factors (Biol)	
6	OBS– a sign of obesity if Body Mass Index is more than 24 kg/m
7	CH– a sign of increased total cholesterol, if it is more than 5 mmol/l
8	NHL– a sign of high level of non-high-density lipoprotein cholesterol, more than 3 Mmol/l
9	SBP– a sign of increased systolic blood pressure, if it is more than 120/90 mmHg
10	GL– a sign of increased glucose levels in the blood serum, if it is more than 7 mmol/l
Type 4: Modifiable behavioral factors of lifestyle (Beh)	
11	Sm– a sign of tobacco smoking
12	Fa– a sign of insufficient physical activity
13	Diet– a sign of unhealthy food abuse

Here *Factor* is a binary vector of descriptive assessments of individual CVD risk factors based on the model *Z*, *C* denotes the predictive assessment of total CVD risk in 10-year perspective, commonly used in clinic practice by application of SCORE algorithms [4, 6, 7], and *Class* is a binary vector which components describe person belonging to one of the risk factor types presented in TABLE 1. To determine the user profile *UP* we use the set of CVD health indicators in the form

$$X = \{x_j, x_j \in \mathbb{R}, x_j \geq 0, \forall x_j \in [\min_{x_j}, \max_{x_j}], j = 1, 2 \dots 17\}. \tag{5}$$

TABLE 2 shows the set of CV health indicators, grouped by types of CV risk factors. Questionnaires, texts from electronic medical records, and data from digital medical gadgets can be used to gather user CVD health indicators [10].

5. RECOMMENDER ALGORITHM

We propose a knowledge-based recommendation algorithm, namely CVD-Risk-Prevent, for supporting adults in self-management their CVD risk factors at home, based on official guideline [6]. The main user of the recommendation algorithm is a person interested in consciously managing their health to increase the duration of an active and full life. The key feature and contribution of this algorithm is the formal description of the transformation of one data structure in the form of a set of

Table 2: Indicators of a person’s CV health. Here 1 denotes “Yes”, 0 – “No”

X	Indicator	X	Indicator
Basic indicators		Modifiable biological indicators	
X(1)	Gender: male (1)/female (0)	X(10)	Total cholesterol (mmol/l)
X(2)	Age (years)	X(11)	Non-high-density lipoprotein cholesterol (mmol/l)
X(3)	Height (cm)	X(12)	Systolic blood pressure (mmHg)
X(4)	Weight (kg)	X(13)	Glucose level (mmol/l)
Non-modifiable indicators		Modifiable lifestyle indicators	
X(5)	Family history of CVD (1/0)	X(14)	Physical inactivity (1/0)
X(6)	Presence of CVD (1/0)	X(15)	Smoking (1/0)
X(7)	Chronic kidney disease (1/0)	X(16)	Unhealthy diet (1/0))
X(8)	History of CV events (1/0)	Symptoms of CVD or CV event	
X(9)	Diabetes mellitus type 2 (1/0)	X(17)	Symptoms of angina pectoris, with a significant deterioration in health (1/0)

person’s CV health indicators X (see TABLE 2) into another data structure presented in the form of multidimensional recommendations R (see expression 1), using formal model of user profile UP, a set of rules and LLM.

The CVD-Risk-Prevent algorithm for each person includes 4 stages:

Stage 1. Using CV health indicators X and model Z of CV risk factors creation a descriptive component for user profile UP, that is a vector *Factor*. If a person did not specify CV health indicator in X, this indicator is set to 0. Since the components in model Z are given in TABLE 1, by binary signs (1-yes, 0-no), we propose to use a binary vector *Factor* which initial values of each component are set to 0:

$$Factor = (f(1), f(2), \dots, f(13)), f(i) \in \{0, 1\}, i = 1, 2, \dots, 13 \tag{6}$$

here *i* – CV risk factor number, presented in TABLE 1, $f(i) = 1$ if user CV health indicator corresponds to the presence of a CV risk factor in model Z, $f(i) = 0$ otherwise.

Stage 2. Determining a descriptive component Class of user profile UP based on vector Factor (6). Using types of CVD risk factors in multifactorial model Z and vector Factor the CVD-Risk-Prevent algorithm classifies persons into 5 classes:

Class 1. Individuals are generally healthy and do not have CV risk factors or if a person did not specify CV health indicators. We will classify such persons as persons of class 1 without CV risk factors.

Class 2. Persons for whom modifiable behavioral risk factors for CV risk have been identified.

Class 3. Individuals for whom modifiable biological CV risk factors have been identified.

Class 4. Individuals for whom non-modifiable CV risk factors have been identified.

Class 5. Persons who have symptoms of angina associated with a sudden significant deterioration in health.

Considering the situation that a person can have different CV risk classes identified, we will simultaneously determine the person CV risk class in the form of a binary *Class* vector with initial values zero of each component:

$$Class = \{Class(1), Class(2), Class(3), Class(4), Class(5)\} . \tag{7}$$

To evaluate the person’s class, the following logical rules are used, which change the corresponding component of the Class vector to the value 1 analyzing the vector Factor (6):

$$Rule\ 1 : IF \max_{i=1,2,\dots,13} f(i) = 0 THEN Class(1) = 1, \tag{8}$$

$$Rule\ 2 : IF \max_{i=11,12,13} f(i) = 1 THEN Class(2) = 1, \tag{9}$$

$$Rule\ 3 : IF \max_{i=6,7,\dots,10} f(i) = 1 THEN Class(3) = 1, \tag{10}$$

$$Rule\ 4 : IF \max_{i=2,3,4,5} f(i) = 1 THEN Class(4) = 1, \tag{11}$$

$$Rule\ 5 : IF f(1) = 1 THEN Class(5) = 1. \tag{12}$$

According to the above rules and vector *Class*, the person can be assigned to one of 16 sub-classes. One sub-class will include persons without CV risk factors, and the remaining ones will include persons with various combinations of CV risk factors and clinical conditions.

Stage 3. Assessment of total CV risk over a 10-year using the SCORE/SCORE2-OP models [4, 6, 7], and calculation of predictive estimate C of user profile UP.

The probability of the person’s total CV risk in a ten-year perspective is calculated as CVrisk using the Rule 6, in which FUNC denotes the SCORE2/SCORE2-OP algorithm proposed in [4]:

$$Rule\ 6 : IF Sco = 0 THEN CVrisk = FUNC(x(1), x(2), x(11), x(12), x(15)) \tag{13}$$

$$here\ Sco = f(1) \vee f(2) \vee f(4) \vee f(5). \tag{14}$$

The logical expression (14) sets a constraint on the use of the SCORE2/SCORE2-OP algorithm outlined in guideline [6]. We consider SCORE2/SCORE2-OP algorithm as external function FUNC and obtain predictive assessment of person total CVD risk based on Rule 6 following guideline [6]:

$$C = \begin{cases} low, & if\ CVrisk < 2.5\% \\ moderate, & if\ 2.5\% \leq CVrisk < 5\% \\ high, & if\ 5\% \leq CVrisk < 10\% \\ very\ high, & if\ CVrisk \geq 10\% or (Sco = 1). \end{cases} \tag{15}$$

At the end of this stage all components of user profile UP are defined and could be used to generate multidimensional recommendation for supporting self-management of person CV risk factors at home.

Stage 4. Mapping user profile UP with recommendation items R. Please note that the contents of recommendation items $Es \in R, Inf \in R$, were manually extracted from the guidelines [6, 7], while the content of recommendation items. $Expl \in R$ was generated by LLM, namely ChatGPT-4.

To make the received text of the recommendations more diverse without violating their semantics explanatory component *Expl* of the recommendation was created using ChatGPT-4, since this model is able to generate human-like text [19], including the creation of explanations in specific contexts [11], in particular in cardiology [20].

The content of recommendation items of the *Es* type which sets a strategic goal for the CVD prevention is determined by non-zero component of the *Class* vector. Therefore, to generate this type of recommendation items, the following rules are used, ranked in descending order of class number *j*:

$$Rule7j : IF Class(j) = 1 THEN Es(j), j = 1, 2, \dots, 5. \quad (16)$$

Descending ordering of the *Class* number is advisable because the larger the class number, the higher the risk of CVD in a person. For obtaining the informational *Inf* and explanatory *Expl* components of user recommendation, the rules corresponding to non-zero values of the vector *Factor* (14) are used:

$$Rule8i : IF f(i) = 1 THEN (C \cup Inf(i)) \cup Expl(i), i = 1, 2, \dots, 13. \quad (17)$$

In expression (17) the sign " \cup " denotes the operation of concatenation of the items, and *i* designates the risk factor number.

The ranking of recommendation items obtained using *Rule 8i* (17) is carried out in 2 stages. Since each non-zero $f(i)$ corresponds to a single user *Class* at the first stage, the recommendation items in expression (17) are ordered in descending order of the user's *Class* numbers. Thus, at first, the user will be shown recommendation items concerning the most serious CV risk factors. Then, for user of classes 2 and 3, the recommendation items are additionally ordered in ascending order using the *Utility* function, as suggested in [11]:

$$Utility(h) = \sum_{i=6}^{13} f(i) \cdot S_i(h), \quad (18)$$

here *h* is a recommendation item, $f(i)$ denotes *i*-th component of *Factor* vector, $S_i(h)$ designates contribution of recommendation item *h* to *i*-th CVD risk factor. We assumed that the contribution value of the recommendation item *h* corresponds to the rank of the negative impact of CV risk factors on CVD and CV events considered in the Prospective Urban Rural Epidemiology (PURE) study [21]. The PURE study focused on examining the influence of CV risk factors on CVD and CV events in 155 722 people. As a result of this study, risk factors for CVD were ordered by their ranks, where the lower the rank, the greater the degree of negative impact of this risk factor..

Since the CVD-Risk-Prevent algorithm evaluates the predictive estimate of total CV risk using the SCORE2/SCORE2-OP model, to obtain the CV risk probability value, the medical calculator recommended by the EAPC (European Society of Cardiology) for CVD prevention and the SCORE2/SCORE2-OP scale given in [4, 6], were used. To obtain initial data about person CV health, a questionnaire was developed, the questions of which correspond to CVD health indicators represented in TABLE 2. To assess the symptoms of angina pectoris, the Rose questionnaire was used [22].

FIGURE 2 shows a diagram of the proposed algorithm for generating personal recommendation contained targeted *Es*, informational *Inf* and explanatory *Expl* components.

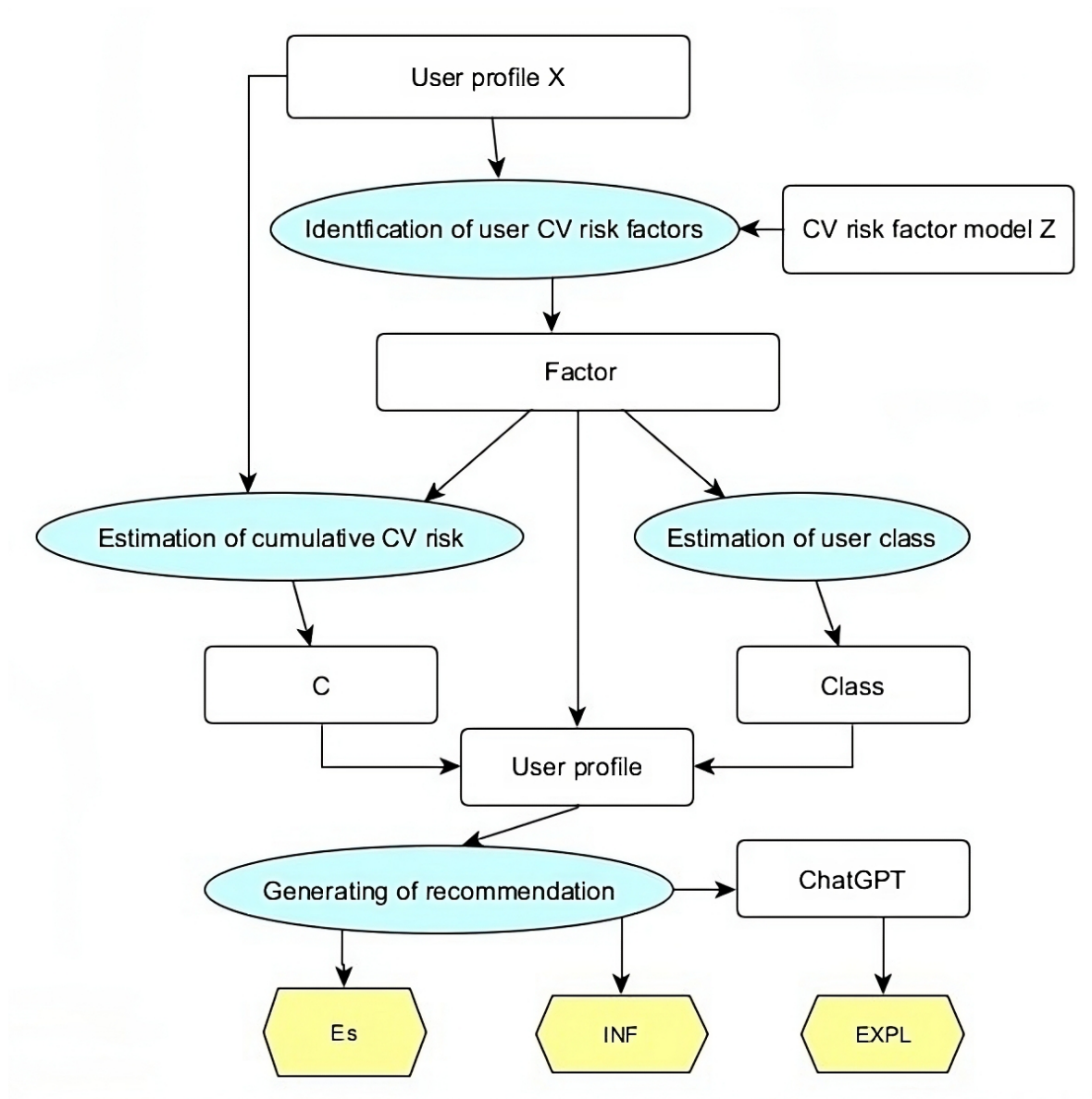


Figure 2: The diagram of the CVD-Risk-Prevent algorithm

6. USING CHATGPT FOR EXPLANATORY ITEMS CREATION

We decide to use ChatGPT-4 for generating only explanatory items *Expl* of the recommendation for two reasons. Firstly, the explanation of the negative impact of the CVD risk factor is not sufficiently presented in the guidelines [6, 7], and secondly, in our opinion, large language model can create unusual metaphors that could draw attention to the main content of the recommendation.

In the review [23], the authors emphasize that modern recommendation systems lack the explicability of recommendations, as well as knowledge about the open world to improve the effectiveness

of recommendations. Prompt to ChatGPT contains three components *task*, *parameters*, *constraints* as shown in FIGURE 3, where the *task* is given in the beginning of a prompt, then the list of CVD risk factors as *parameters* is provided. At the end of the prompt the sentence about *constraints* is presented. Note the blue symbols refer to LLM responses to the prompt.

Please give the explanations why person need to control her/his CV risk factors such as high blood pressure – 160/90 mmHg, physical inactivity, smoking, unhealthy diet. The explanation must be understandable to the person, include only person CV risk factors and contain no more 4 propositions for each CVD risk factors.

Let's go over the person cardiovascular (CV) risk factors:

1. **High Blood Pressure (Systolic blood pressure – 160/90 mmHg)**
High blood pressure is like putting too much air in a balloon—it strains the walls and can cause them to burst. In the body, this can lead to damage in the arteries, heart, and other organs. It's often called the "silent killer" because it doesn't have symptoms but can lead to serious problems like heart attacks and stroke. Keeping blood pressure lower reduces the strain on your heart and the risk of heart disease.

2. **Physical Inactivity**
Being inactive is like leaving a car unused for a long time; eventually, it won't run as well. Regular physical activity keeps your heart and body running smoothly and can help lower high blood pressure, manage weight, and reduce stress on your heart. It can also improve your mood and energy levels.

3. **Smoking**
Smoking damages your blood vessels and can make them narrow, making it harder for blood to get around your body. Think of it as a traffic jam in your arteries. Stopping smoking can help to clear the traffic and reduce your risk of heart disease, as well as improve your overall health.

4. **Unhealthy Diet**
Eating foods high in fat, salt, and sugar is like using low-quality fuel in a car—it won't run properly and will break down sooner. A healthy diet is like high-quality fuel; it makes sure everything runs smoothly and reduces the risk of blockages in your arteries. It's good for your whole body, including your heart.

Addressing these risk factors involves lifestyle changes, such as getting more exercise, eating a healthier diet, and quitting smoking. These changes can significantly lower the risk of developing serious CV conditions, leading to a healthier heart and a longer life. It's like maintaining a car; you do it not just to prevent it from breaking down, but to make sure it runs well for a long time.

Please give the explanations why person need to control her/his CV risk factors such as moderate total CV risk on SCORE model, family history of early CV diseases, high blood pressure. The explanation must be understandable to the person, include only person CV risk factors and contain no more 4 propositions for each CVD risk factors.

That's why now is the time to self-manage your cardiovascular risk factors:

1. **Moderate Total CV risk**
You are at moderate risk for cardiovascular disease according to the SCORE model. This means you have a 1-4% chance of developing serious heart or vascular problems within the next 10 years. This is not the highest level of risk, but it is not the lowest, which means we need to pay attention to your health now to prevent possible problems in the future.

2. **Family history**
Having cardiovascular disease in your immediate family increases your own risk. This means you should be more careful about prevention and get regular medical checkups to ensure any potential problems are caught and treated early.

3. **Blood Pressure**
Please note that elevated blood pressure (BP) may be a sign of incipient hypertension. Systematic elevation of blood pressure is a major risk factor for heart disease, stroke and other serious health problems. As blood pressure rises, the load on the heart and blood vessels increases, and the heart begins to work harder.

Not smoking greatly helps reduce your risk.

Figure 3: Two examples of the prompts and the explanatory item lists received as responses from ChatGPT

The prompt parameters "*params*" are a person's CVD risk factors, which are contained in the information component of his recommendation in the form of a list items identified in formula (17) as $(C \cup Inf(i))$. Based on 30 different ChatGPT prompts and analysis of the generated explanatory texts we found that ChatGPT produces plausible and stylistically correct explanations of CVD risk factors. Therefore, the itemset Expl was formed using explanations generated by ChatGPT after expert validation and then was used in the proposed recommender algorithm.

7. EVALUATION OF THE RECOMMENDATION ALGORITHM

The proposed CVD-Risk-Prevent algorithm refers to the knowledge-based recommendation algorithms, since it is based on explicit knowledge about person, about recommendation items and about their relations. Our approach to evaluating the effectiveness of the proposed CVD-Risk-Prevent algorithm is based on verification methods commonly used in computer science: (1) pilot study of user's satisfaction in recommendation generated by algorithm; (2) quality assessment of the algorithm, recommendation items and texts of recommendations by doctors. This evaluation scenario is appropriate for knowledge-based recommendation algorithms and is often used as noted in [10].

7.1 Study of User Satisfaction in Recommendations Received

To understand the effectiveness of the proposed CVD-Risk-Prevent algorithm from the user’s perspective, additional study was conducted involving 15 independent persons who provided consent for the use of CV health indicators. The age of the participants varied from 34 to 70 years, among them 53% were men, 47% were women. The purpose of the study is to find out the degree of user satisfaction with the content of the recommendations received and the degree of agreement of their opinions. To obtain recommendations in this study a software implementation of the CVD-Risk-Prevent algorithm was developed, a fragment of the user interface for input CV health indicators is depicted in FIGURE 4. To analyze user opinions, a special questionnaire was developed, the questions of which are represented in TABLE 3. Likert scale was used to create a survey of persons’ satisfaction with the content of the recommendations received. Users assessed the individual recommendations according to 3 criteria, using score from 1 to 5, where score 1 corresponds to completely disagree and 5 – to completely agree with the statement given in the questionnaire.

<p>Have any close relatives had strokes, heart attacks, angina pectoris?(Men up to 55 years old, women up to 65 years old).</p> <p><input checked="" type="checkbox"/></p> <p><input type="checkbox"/> Do you take statins?</p> <p><input type="checkbox"/> I've been diagnosed with chronic kidney disease.</p> <p><input type="checkbox"/> I've been diagnosed with arterial hypertension.</p> <p><input checked="" type="checkbox"/> I've been diagnosed with coronary artery disease.</p> <p><input type="checkbox"/> I've been diagnosed with type 2 diabetes.</p> <p><input type="checkbox"/> I've got a history of myocardial infarction</p> <p><input checked="" type="checkbox"/> I've had a stroke</p>	<p>Weight in kg _____</p> <p>Weight 108 kg</p> <p>Height in cm _____</p> <p>Height 171 cm</p> <p>Systolic blood pressure in mmHg _____</p> <p>SBP 140 mmHg ?</p> <p>Total cholesterol in mmol/L _____</p> <p>Total cholesterol 5,2 mmol/L ?</p> <p><input checked="" type="checkbox"/> I'm smoking</p>
--	--

Figure 4: Fragment of the user interface for entering CV health indicators

The results of the user survey regarding satisfaction with the recommendations for CVD risk factors management are presented in the TABLE 3. A Degree of User Satisfaction (DUS) is calculated by dividing the mean by the maximum score, i.e. 5:

$$DUS_l = \frac{1}{5k} \sum_{v=1}^k y_{lv}, \tag{19}$$

here y_{lv} denotes the score of the v-th participant on the l-th question of the questionnaire, and k is the number of participants.

Table 3: The results of the user survey. Here SD designates standard deviations, DUS denotes degree of users' satisfaction with a feature of recommendation received

No.	Feature of the individual recommendations	Mean	SD	DUS
1	Utility. The recommendations received are useful for you	4.8	0.18	0.96
2	Explainability. Recommendations explain the impact of risk factors on cardiovascular health	4.5	0.28	0.9
3	Accuracy and individual focus. The recommendations received are specific to you and presented in understandable language.	4.2	0.4	0.84
	Mean	4.5	0.29	0.9

Data in TABLE 3, show that participants agree that individual recommendations meet all quality criteria. Moreover, some persons noted that they had previously received exactly such recommendations from their attending physicians. An analysis of the average ratings of the features of recommendations shows that the Utility property received the highest rating, the explainability of individual recommendations, for the generation of which ChatGPT was used, received a high score of DUS = 0.9.

7.2 Expert Evaluation

Two independent experts specializing in the prevention and treatment of CVD gave their consent for a qualitative assessment of the proposed CVD-Risk-Prevent algorithm from the point of view of evidence-based medicine. The first expert is a Doctor of Medical Sciences, professor of the medical faculty of the Institute of Medicine, Ecology and Physical Culture of Ulyanovsk State University, Russia, the second one is a cardiologist and therapist working at the Moscow clinic GLOBAL MEDICAL SYSTEMS (GMS LLC), Russia. Both specialists have extensive experience, more than 20 years of practice in the treatment of persons with CVD, as well as their preventive counseling. The description of the algorithm, the models, the rules it uses, information sources, person's profiles as well as the resulting texts of recommendations were independently analyzed by experts. The experts focused on assessing the correctness of recommendations obtained by the CVD-Risk-Prevent algorithm. the correctness of recommendation was considered as their compliance with the person's profiles, their individual CV risk factors, and official guidelines [6, 7]. Both experts expressed the opinion that the proposed algorithm generates recommendations consistent with official guidelines and clinical practice, so it can be used by adults in self-monitoring of CVD risk factors. They also noted the usefulness and safety of the recommendations and gave a positive assessment of the completeness of the set of CVD risk factors and their explanations obtained using LLM. Doctors agree that the proposed algorithm is also promising in clinical practice, since it can reduce the doctor's time when preparing recommendations for the patient. At the same time, experts made some useful comments. In particular, to support adults with cardiovascular disease at home, the content of recommendations should be more specific and could be expanded using electronic health record data. For example, it would be useful for such patients to add recommendations for taking prescribed medications or a description of the dynamics of CVD health indicators.

8. DISCUSSION

The CVD-Risk-Prevent algorithm proposed in this article does not require training to calculate the optimal parameters of models from many examples, as its rules are based on proven medical knowledge, outlined in [6, 7], and pre-training large language model, namely ChatGPT-4.

To highlight the opportunity of CVD-Risk-Prevent algorithm for HRS the comparison with similar knowledge-based recommendation algorithms focused on risk factors management was done and presented in TABLE 4. Understanding that dietary nutrition recommendations are useful in preventing many diseases authors of the article [14], proposed recommender based on knowledge to present individualized diet menus for obese youth, while in our study recommendations associated with management of a set of CV risk factors.

Table 4: The comparison knowledge-based recommendation algorithms focused on risk factors management

Criteria of comparison		Spoladore [17]	Lopez-Barreiro [15]	Wang [9]	Proposed CVD-Risk-Prevent
Perspective of comparison	Indicator of comparison				
User profile corresponding to CVD risk factors	Number of risk factors	2	9	12	14
Recommendation associated with self-management of CVD risk factors	User focus	One class. Older adults with chronic pathologies	One class. Persons without medical conditions	One class. Chronic disease patients, including CVD	Five classes including adults with or without any CVD risk factors as well symptoms of angina
	Knowledge models and algorithms	Ontology, expert knowledge	Expert ranking model	Rules, ontology, natural language processing	Rules, ChatGPT

Extended well-being and healthy lifestyle factors of risk such as unhealthy eating rhythm, insufficient sleep, lack of exercise and dietary habits are considered in HRS with taking an account readiness to behavior change in the study [16].The author focused on real-life health-coaching program creating and recommendations included change objectives, activities and plan. However, to manage CVD risk factors it is necessary to analyze not only behavioral factors, but also biological, family history and symptoms as was provided in our study.

As follows from TABLE 4, the proposed algorithm estimates a larger number of CVD risk factors in user profile to construct recommendations. Moreover, the proposed algorithm not only evaluates a wide range of CVD risk factors listed in official guidelines, but also, for the first time in the HRS field, contains predictive estimates of 10-year CVD risk. As a consequence, this provides a greater coverage of recommendations generated by the proposed algorithm than in the compared ones [9, 14–17].

In the articles [9, 15, 17], the recommendations are intended for one class of users, namely, adults with chronic pathologies or persons without medical conditions, which is narrower than in the proposed algorithm, since in addition to such users, recommendations are intended for five different classes, including adults with or without any CVD risk factors as well symptoms of angina. Regarding the content of the recommendations, in contrast to one-dimensional recommendations presented as a list of dishes [14], exercises [17], challenges [15], or educational materials [9], the proposed recommendations are multifaceted in nature, the content of which includes information and explanation of each CVD risk factor. The explanation component is important because explanations are believed to be a motivating factor for adherence to health care.

Basically, knowledge-based algorithms in HRS use only expert knowledge, presented in the form of rules or ontologies [9–11, 14–18], that guarantee the correctness and safety of recommendations. In the proposed algorithm, this approach is extended by leverage the capabilities of ChatGPT for generating explanations of recommendations. This brings additional value to the generated recommendations and is in line with modern trends in the field of explainable recommender systems and AI in general.

To evaluate recommender algorithm performance, we assess the generated recommendations from two main perspectives: quality and user satisfaction. The proposed recommendation algorithm received a positive qualitative assessment by experienced therapists for safety and compliance with official recommendations used in clinical practice. As was shown the degree of users' satisfaction with personal recommendation generated by CVD-Risk-Prevent algorithm was high, the utility and explainability have received 0.96 and 0.9 score respectively. A limitation of the proposed recommender algorithm is the lack of recommendations for the use of medications, such as statins or aspirin. It is assumed that the appointment of such an intervention is the task of the attending physician. The lack of analysis of tendencies in the person's profile and his CV risk factors is another limitation of the proposed recommendation algorithm. In our opinion, a common limitation of knowledge-based health recommendation algorithms, including proposed in this article, is the use of subjective quality assessments. CVD-Risk-Prevent algorithm does not consider such risk modifiers as psychological stress, ethnicity, imaging data and comorbidity. Another limitation relies at focusing on European adult population and on prevention only CVD. We recognize that larger studies in correctness of LLM explanations of CV risk factors involving medical professionals and different types of users are needed to evaluate its effectiveness in preventing CVD.

9. CONCLUSION

In the article knowledge-based CVD-Risk-Prevent algorithm for supporting self-management CVD risk factors in adults at home was proposed and studied. This recommendation algorithm combines LLM and rules to make multidimensional recommendations with explanations by converting general clinical practice, CVD prevention guidelines [6, 7], and open world knowledge into rule-based logic. Proposed CVD-Risk-Prevent algorithm has several key advantages as knowledge-based recommendation algorithm. The first advantage is that it overcomes the data shortage obstacles that plague collaborative and content filtering recommendation algorithms. The novelty of the proposed algorithm from knowledge-based HRS lies in its focus on supporting self-management of CV risk factors in adults at home, in the original recommendation model, in the completeness of the set of CVD risk factors, and in the method of generating recommendations. We believe that the proposed

algorithm is one of the ways to solve the problems mentioned in article [8], related to effective strategies for the prevention of cardiovascular diseases at the individual level. The recommendation system based on the proposed algorithm, as can be concluded from the results of the survey of future users, will be in demand both among the adult population for managing cardiovascular health at home and among doctors.

Further research will be focused to the developing HRS for supporting adults in self-management of CVD risk factors at home and its experimental studies to obtain numerical performance estimates for CVD-Risk-Prevent algorithm and ChatGPT explanations. To expand the opportunities of the software implementation of the proposed recommendation algorithm, research will be conducted in four directions: 1) developing an interface for integration with digital gadgets, medical information systems and electronic medical records; 2) exploring ways to integrate the proposed algorithm with recommendation algorithms for collaborative and content filtering when a large database of persons is created; 3) solving problems related to ethical issues and confidentiality; 4) the use of LLM to automate the addition of knowledge about new CVD risk factors from new guidelines.

10. AUTHOR CONTRIBUTIONS

Methodology, writing, visualization, T.V.A; formal analysis, data curation, P.V.P.; writing, original draft preparation and editing A.I.M. All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

11. FUNDING

This research was performed in the framework of the state task in the field of scientific activity of the Ministry of Science and Higher Education of the Russian Federation, project "Models, methods, and algorithms of artificial intelligence in the problems of economics for the analysis and style transfer of multidimensional datasets, time series forecasting, and recommendation systems design", grant no. FSSW-2023-0004.

12. CONFLICTS OF INTEREST

The authors declare no conflict of interest.

References

- [1] Kobrinskii BA, Grigoriev OG, Molodchenkov AI, Smirnov IV, Blagosklonov NA. Artificial Intelligence Technologies Application for Personal Health Management. IFAC PapersOnLine. 2019;52:70-74.
- [2] Kaasenbrood L, Boekholdt SM, van der Graaf Y, Ray KK, Peters RJ, Kastelein JJ et al. Distribution of Estimated 10-Year Risk of Recurrent Vascular Events and Residual Risk in

- a Secondary Prevention Population. *Circulation*. 2016;134:1419-1429.
- [3] Kennedy O, Su F. Evaluating the Effectiveness of the NHS Health Check Program in South England: A Quasi-Randomised Controlled Trial. *BMJ Open J*. 2019.
- [4] SCORE2 working group and ESC cardiovascular risk collaboration. SCORE2 Risk Prediction Algorithms: New Models to Estimate 10-Year Risk of Cardiovascular Disease in Europe. *Eur Heart J*. 2021;42:2439-54.
- [5] Marques MD, Pires R, Perdigão M, Sousa L, Fonseca C, et al. Patient-Centered Care for Patients With Cardiometabolic Diseases: An Integrative Review. *J Pers Med*. 2021;11:1289.
- [6] Visseren FL, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M et al. 2021 ESC Guidelines on Cardiovascular Disease Prevention in Clinical Practice. *Eur Heart J*. 2021;42:3227-337.
- [7] https://russjcardiol.elpub.ru/jour/article/view/5452?locale=ru_RU
- [8] Feigin VL, Krishnamurthi R, Merkin A, Nair B, Kravchenko M, et al. Digital Solutions for Primary Stroke and Cardiovascular Disease Prevention: A Mass Individual and Public Health Approach. *Lancet Reg Health West Pac*. 2022;29:100511.
- [9] Wang Z, Huang H, Cui L, Chen J, An J, Duan H et al. Using Natural Language Processing Techniques to Provide Personalized Educational Materials for Chronic Disease Patients in China: Development and Assessment of a Knowledge-Based Health Recommender System. *JMIR Med Inform*. 2020;8:e17642.
- [10] Cai Y, Yu F. Health Recommender Systems Development, Usage, and Evaluation From 2010 to 2022: A Scoping Review. *Int J Environ Res Public Health*. 2023.
- [11] Uta M, Felfernig A, Le VM, Burgstaller T. Knowledge-Based Recommender Systems: Overview and Research Directions. *Big Data J*. 2024;7.
- [12] Burke R. Knowledge-Based Recommender Systems. In: *The Encyclopedia of Library and Information Science*. 2000;69:180-200.
- [13] Aggarwal CC. *Recommender Systems: The Textbook*. Switzerland: Springer International Publishing. 2016.
- [14] Jung H, Chung K. Knowledge-Based Dietary Nutrition Recommendation for Obese Management. *Inf Technol Manag*. 2016;17:29-42.
- [15] Lopez-Barreiro J, Garcia-Soidan JL, Alvarez-Sabucedo L, Santos-Gago JM. Practical Approach to Designing and Implementing a Recommendation System for Healthy Challenges. *Appl Sci J*. 2023;13:9782.
- [16] Honka AM, Nieminen H, Similä H, Kaartinen J, Gils MV. A Comprehensive User Modeling Framework and a Recommender System for Personalizing Well-Being Related Behavior Change Interventions: Development and Evaluation. *IEEE Access*. 2022;10:116766-83.
- [17] Spoladore D, Colombo V, Arlati S, Mahroo A, Trombetta A, et al. An Ontology-Based Framework for a Telehealthcare System to Foster Healthy Nutrition and Active Lifestyle in Older Adults. *Electronics*. 2021;10:2129.

- [18] Tran TNT, Felfernig A, Le VM, Chau TMN, Mai TG. User Needs for Explanations of Recommendations: In-Depth Analyses of the Role of Item Domain and Personal Characteristics. In: Proceedings of the 31st ACM conference on user modeling, adaptation and personalization (UMAP '23). 2023:54-65.
- [19] Li Z, Chen Y, Zhang X, Liang X. BookGPT: A General Framework for Book Recommendation Empowered by Large Language Model. *Electronics*. 2023;12:4654.
- [20] Gala D, Makaryus AN. The Utility of Language Models in Cardiology: A Narrative ReView of the Benefits and Concerns of ChatGPT-4. *Int J Environ Res Public Health*. 2023;20:6438.
- [21] Yusuf S, Joseph P, Rangarajan S, Islam S, Mentz A, et al. Modifiable Risk Factors, Cardiovascular Disease, and Mortality in 155,722 Individuals From 21 High-Income, Middle-Income, and Low-Income Countries (Pure): A Prospective Cohort Study. *Lancet*. 2020:795-808.
- [22] Rose GA. The Diagnosis of Ischemic Heart Pain and Intermittent Claudication in Field Surveys. *Bull World Health Organ*. 1962;27:645-658.
- [23] Lin J, Dai X, Xi Y, Liu W, Chen B, Li X et al. How Can Recommender Systems Benefit From Large Language Models: A Survey. 2023. Arxiv Preprint: <https://arxiv.org/pdf/2306.05817>