Chang, H.-T., Lin, C-.Y., Jheng, W.-B., Chen, S.-H., Wu, H-.H., Tseng, F.-C., & Wang, L.-C. (2023). AI, Please Help Me Choose a Course: Building a Personalized Hybrid Course Recommendation System to Assist Students in Choosing Courses Adaptively. *Educational Technology & Society*, *26*(1), 203-217. https://doi.org/10.30191/ETS.202301_26(1).0015

AI, Please Help Me Choose a Course: Building a Personalized Hybrid Course Recommendation System to Assist Students in Choosing Courses Adaptively

Hui-Tzu Chang[1], Chia-Yu Lin[2*], Wei-Bin Jheng[3], Shih-Hsu Chen[4], Hsien-Hua Wu[5], Fang-Ching Tseng[6] and Li-Chun Wang[4]

[1]Center for Institutional Research and Data Analytics, National Yang Ming Chiao Tung University, Hsinchu // [2]Department of Computer Science and Information Engineering, National Central University, Taoyuan // [3]Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu // [4]Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu // [5]Department of Transportation and Logistics Management, National Yang Ming Chiao Tung University, Hsinchu // [6]Electrical Engineering and Computer Science Undergraduate Honors Program, National Yang Ming Chiao Tung University, Hsinchu // simple@nycu.edu.tw // sallylin0121@gmail.com // sozai97.cs06@nctu.edu.tw // brian880403@gmail.com // simonwu.mg09@nycu.edu.tw // claire.eecs07@nctu.edu.tw // wang@nycu.edu.tw

[*]Corresponding author

ABSTRACT: The objective of this research is based on human-centered AI in education to develop a personalized hybrid course recommendation system (PHCRS) to assist students with course selection decisions from different departments. The system integrates three recommendation methods, item-based, user-based and content-based filtering, and then optimizes the weights of the parameters by using a genetic algorithm to enhance the prediction accuracy. First, we collect the course syllabi and tag each course from twelve departments for the academic years of 2015 to 2020. Next, we use the course tags, student course selection records and grades to train the recommendation model. To evaluate the prediction accuracy, we conduct an experiment on 1490 different courses selected by 5662 students from the twelve departments and then use the root-mean-squared error and the normalized discounted cumulative gain. The results show that the influence of item-based filtering on the course recommendation results is higher than that of user- and content-based filtering, and the genetic algorithm can find the optimal solution and the corresponding parameter settings. We also invite 61 undergraduate students to test our system, complete a questionnaire and provide their grades. Overall, 83.60% of students are more interested in courses at the top of the recommendation lists. The students are more autonomously motivated rather than holding extrinsic informational motivation across the hybrid recommendation method. Finally, we conclude that PHCRS can be applied to all students by tuning the optimal weights for each course selection factor for each department, providing the best course combinations for students' reference.

Keywords: Human-centered AI in education, AI course recommendation system, Learning aids in systems

1. Introduction

In recent years, the number of research works applying Artificial Intelligence (AI) to educational systems have increased rapidly. AI offered a new solution for education as it helped develop an adaptable, inclusive, agile, individualized, and effective learning environment to overcome the disadvantages of traditional education or training. Additionally, it also brought hope and potential of innovation for education (Renz et al., 2020; Renz & Vladova, 2021). In those AI systems, human-centered AI in education enables us to gain a deeper understanding of students' learning behaviors, reaction time, emotion, or needs (Renz et al., 2020; Yang et al., 2021). It also helped students find the potential and problems, then set up study plans for them using information and communications technologies (ICTs; Yang et al., 2021). A system that assists students with planning their courses was extremely important (Lin et al., 2018). Recent course recommendation system research has focused on how to precisely recommend students courses that suit their needs, with many works proposing course selection methods and algorithms to deal with course recommendation, though none of the methods were designed based on human-centered AI in education. The focus of these studies was on raising the grades of the students (Chang et al., 2016), their graduation rate (Kurniadi et al., 2019) or their employment rate (Farzan & Brusilovsky, 2006) rather than the personal factors affecting the recommendation process.

Course options are important for students to fulfill their degree requirements and to determine their future career directions (Farzan & Brusilovsky, 2006; Kurniadi et al., 2019). In response to the trend of higher education, institutions promote interdisciplinary courses and distance learning courses, and AI systems to contribute to the selection of courses with more diversity (Chang & Chen, 2021). When students are faced with information overload as they are selecting courses, students' adaptive development would be secured if their school provided

ISSN 1436-4522 (online) and 1176-3647 (print). DOI 10.30191/ETS. This article of *Educational Technology & Society* is available under Creative Commons CC-BY-NC-ND 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Editors at ets.editors@gmail.com.

a course recommendation system that recommended courses based on their interests, abilities, and career goals (Iatrellis et al., 2017; Sawarkar et al., 2018). Thus, this study proposes the personalized hybrid course recommendation system (PHCRS) that considers students' course selection factors to provide better course selection advice. PHCRS utilizes the course selection data (e.g., courses, grades) and course data (e.g., objectives, knowledge area, skills) accumulated by the school for system development. To ensure that these factors can help generate a better recommendation result, this study uses a genetic algorithm to determine the importance of each indicator and recommendation method, applies weights to the recommendation process, and provides advice to students.

2. Literature review

2.1. Human-centered AI in education

Previous AI technologies focused on how to behave and think like a human, while recent research switched their focus to human-centered AI (HCAI), a technology that approaches AI from a human perspective through human environments (Renz & Vladova, 2021; Yang et al., 2021). Human-centered AI needs explainable computation and decision-making processes, social phenomenona, and mankind characteristics to adjust its algorithms to help enhance human intelligence using machine learning to increase human welfare (Yang et al., 2021). HCAI has been applied to a wide variety of domains, and its effectiveness in education is of great importance. In addition, HCAI can help students learn, adapt, integrate, self-correct, and use data to tackle complicated tasks in the hopes of solving more learning, emotion or career development problems that students may face. AI is superior to humans when it comes to computing and decision making, and it can also educate and train humans to enhance their performances, as well as mine implicit values (Yang et al., 2021). With the development of AI, the trend of education has shifted from the one-size-fits-all approach to the precision approach (Zawacki-Richter et al., 2019), which utilizes AI for analysis. The precision approach identifies students in need and offers real-time assistance, which enhances the teaching quality and learning outcomes for students. It also enables students to develop their skills and knowledge in a more personalized way by providing more precise information, understanding the students' progress of, and what should be done to realize their goals (Yang et al., 2021).

Even though more and more services offer data-driven smart learning solutions for education, only a small portion of them apply AI techniques (Liu et al., 2023). Ahmad et al. (2020) reviewed previous research on applications of AI in education and split the domains into intelligent tutoring systems (ITS), evaluation, adaptive learning, recommendation systems, student performance, sentiment analysis, detention or drop out, and course monitoring. Among those topics, ITS is the most popular and the most important because it allows teachers to provide adaptive learning routes in educational environments and assist students with planning their own learning routes based on their personal interests, abilities, or future career development (Alkhatlan & Kalita, 2019). Even though AI has a lot of potential if applied in education and is increasingly gaining popularity, only few are implementing AI in education tools and even less of them use these tools in their institutions; thus we can conclude that people still doubt the ability or reliability of AI, which limits the development of HCAI. More research has advocated not use AI to replace humans (Xu, 2019), but to support humans based on human's benefit (Schmidt, 2020). Education relatives have come to an understanding that the use of HCAI is to help realize the goals of positive learning outcomes and teaching success instead of replacing traditional education methods, then diminish the fear of AI from students and teachers afterwards (Renz & Vladova, 2021).

2.2. AI recommendation systems in education

ICTs play a huge role in the globalization era and information society, while also providing new opportunities for many domains. In education, ICTs are utilized for the teaching and learning process (Urdaneta-Ponte et al., 2021). However, the development of ICTs poses some challenges, including the increasing complexity and loading of information can make students spend too much time on searching for information and consumes the amount of time they are able to spend studying, which would decrease and their grades would decrease accordingly. If students can get reliable and adequate information easier and quicker, it would be a decisive factor in their learning outcomes. To resolve this problem, the course recommendation system is developed, and the goal of the course recommendation system is to offer choices and recommendations for each student based on their needs, helping students find the courses that truly meet their requirements through information filtering, data mining and predictive algorithms.

The main approaches used in course recommendation system are the collaborative filtering, content-based filtering, and hybrid recommendation methods (Urdaneta-Ponte et al., 2021). (1) Collaborative filtering. There are two main filtering methods, including item-based and user-based filtering. Item-based filtering uses students' grades in other subjects or domains to provide course recommendations (Dwivedi & Roshni VS, 2017). User-based filtering matches the course selection route history of a current student to an alumnus who shared a similar route, then recommends the course list of the alumnus to the current student (Zhang et al., 2015). (2) Content-based filtering. The filtering mechanism is built upon the characteristics of the course syllabi, such as the subject field or the lecture content, thus providing a course list similar to one's interested subjects or domains (Esteban et al., 2020). However, these methods have their respective strengths and weaknesses; to address the disadvantages of the methods mentioned above, researchers have proposed (3) hybrid recommendation methods (Çano & Morisio, 2017). The collaborative filtering and content-based filtering hybrid recommendation method is the most common method since it overcomes the limitations of both filtering methods above, increases predictability, and decreases the degree of sparsity and the loss of information (Esteban et al., 2020).

Several AI technologies are introduced for the construction of the course recommendation system in recent years, including Bayesian techniques, artificial neural networks, machine learning techniques, genetic algorithms, and fuzzy set techniques. These AI techniques prove to be adequate for designing recommendation systems in the big data era (Urdaneta-Ponte et al., 2021), and a genetic algorithm is one of the most often used method. A genetic algorithm, proposed by Holland (1975), was inspired by the encoding and decoding process of DNA and applied to the artificial environment. A genetic algorithm can automatically optimize the weights of each criterion and variable in the recommendation system through the optimization of likelihood function (Esteban et al., 2020) to obtain the final estimation for the system (Esteban et al., 2020). However, even though a genetic algorithm has shown good performance when used in building recommendation systems, only research applies this method (Esteban et al., 2020). Esteban et al. (2020) used hybrid filtering combining collaborative filtering and content-based filtering to train a course recommendation model, then applied a genetic algorithm to optimize the weights of student information, course information, recommendation methods, and system attributes to build a course recommendation system with high accuracy for students. A genetic algorithm has also been applied to estimate the best learning path. Dwivedi and Roshni (2017) matched the learning path of current students with alumni history data and then used a genetic algorithm to find the best learning path for each current student. Huang et al. (2007) applied computerized adaptive testing combined with a genetic algorithm and case-based reasoning to build the best learning path of online courses. In conclusion, a genetic algorithm is a useful tool in learning systems; it provides the best solution for complicated problems that students encounter, and its computation results can also be a reference for students' course selection and learning path.

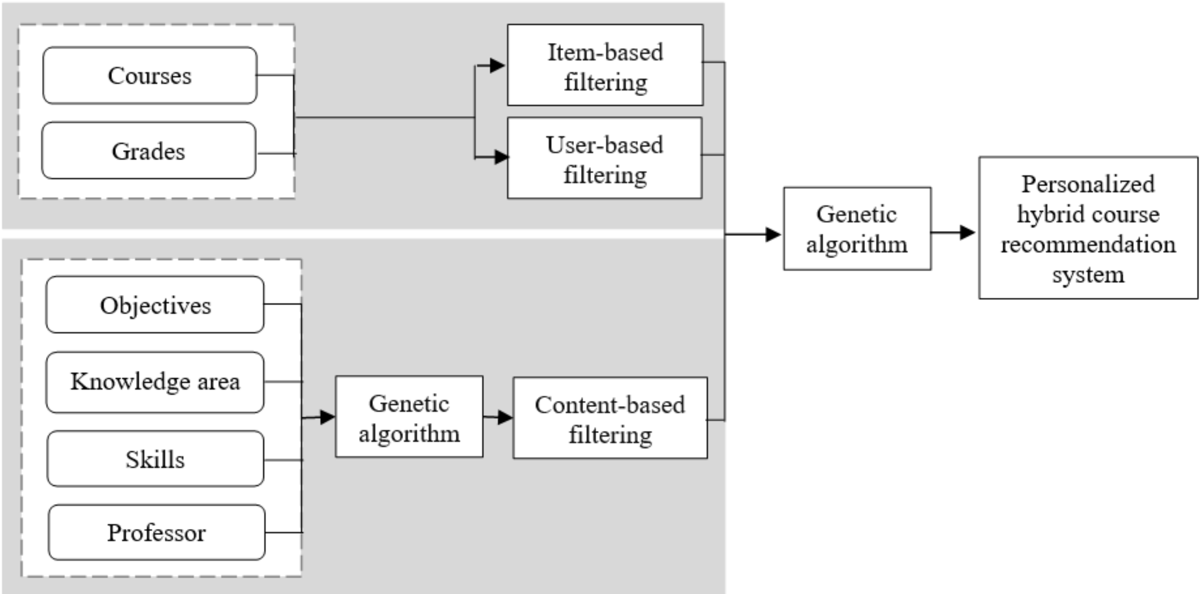
For the reasons mentioned above, we propose PHCRS for formal offline courses to consider the different learning needs of students. The system offers a course recommendation list based on personalized course selection factors, decision sequences and course importance to satisfy the personalized study, capacity building and career exploration needs of students. To achieve this goal, we first filter the factors affecting the students' course selection decisions as the indicators of system development and then use a hybrid multicriteria recommendation method to develop the recommendation system. Last, we use a genetic algorithm to find the weights of student information, course information and system attributes with the goal of determining weights in a standardized manner and optimizing system attributes automatically. This study proposes three hypotheses to verify the effectiveness of PHCRS.

- Hypothesis 1: Students' degree of interest in the courses recommended by the hybrid recommendation method will differ among the course recommendation order.
- Hypothesis 2: The degree of interest in the courses recommended to a student will be affected by the student's internal and external motivations for taking a course.
- Hypothesis 3: Students' degree of academic performance in the courses recommended will differ among those following and not following the recommendation list.

3. Development of a personalized hybrid course recommendation system

The steps of the research design process are shown in Figure 1. We first transform and encode the data used for system development and then apply item-based, user-based and content-based filtering to compute information regarding students and courses to obtain the results of each recommendation method. Then, we use a genetic algorithm to automatically optimize the weights for all filtering methods, and the optimal parameter settings for each student can be found, thus achieving the effect of adaptive recommendation. Finally, we use the root-mean-squared error (RMSE) and normalized discounted cumulative gain (NDCG) to evaluate the effectiveness of the system, thus forming PHCRS. The detailed process is discussed in the following sections.

Figure 1. A Framework for PHCRS



3.1. Data description and preparation

We used course and student data from the Center for Institutional Research and Data Analytics at National Yang Ming Chiao Tung University (NYCU) to train the recommendation system (see Table 1). These data included 12 departments from the Colleges of Electrical and Computer Engineering, Computer Science, Engineering, Management, and Hakka Studies, and a total of 6766 courses were provided from the fall 2015 semester to the fall 2020 semester. For student information, a total of 5662 students from the 12 departments who were enrolled between 2015 and 2020 were selected. To prepare the training data, the researchers collected the course outlines and interviewed the teachers via telephone. The two researchers in each department discussed and agreed upon the labeling rules and then compared the similarities and differences in the labeling results after making the labels. In cases of disagreement, the scorers discussed the issue until a consensus was reached. The interrater reliability was between .7 and .8. The attributes of each course were labeled as follows: (1) Course objectives: This label indicates what the course mainly teaches students, such as signal processing or communication systems. There is a total of 377 possible labels from the 12 departments. (2) Knowledge areas: This label is based on the theories, methods or empirical theories from the field of electrical engineering that are taught to students, such as information and communication, system-on-chip, and 126 other areas from the 12 departments. (3) Skills: This label is based on the relevant technologies, resources or tools used in each course, such as Python or MOSFET. There are a total of 1744 possible labels from the 12 departments. (4) Professors: This label indicates who the course instructor is. After the data preparation, three recommendation methods and a genetic algorithm optimization are implemented in PHCRS for students with different learning needs, as shown below.

Table 1. Student and course information

College/Department	Students	Courses	Total number of courses	Label			
				Course objectives	Knowledge areas	Skills	Professor
College of Electrical and Computer Engineering							
• Department of Electrical and Computer Engineering	1258	332	1890	45	15	373	188
• Department of Photonics	209	94	369	19	3	76	47
College of Computer Science							
• Department of Computer Science	1171	235	1013	54	7	306	114
College of Engineering							
• Department of Civil	473	122	688	70	6	103	51

Engineering							
• Department of Mechanical Engineering	596	119	671	47	11	87	58
• Department of Materials Science and Engineering	299	74	367	24	8	68	37
College of Management							
• Department of Management Science	285	72	235	11	11	204	23
• Department of Transportation & Logistics Management	280	70	336	10	2	81	23
• Department of Industrial Engineering and Management	314	68	256	9	20	149	21
• Department of Information Management and Finance	268	64	283	10	3	163	29
College of Hakka Studies							
• Department of Humanities and Social Sciences	268	132	370	21	5	52	29
• Department of Communication and Technology	241	108	288	57	35	82	26
Total	5662	1490	6766	377	126	1744	646

Note. #in academic years 104 to 109.

3.2. Recommendation model construction

The method we used for recommendation is a multicriteria hybrid recommendation method integrating item-based, user-based and content-based filtering. The formula for predicting the score that *student i* gives to *course j* is as follows: $p_{ij} = \alpha \cdot ICF_{ij} + \beta \cdot UCF_{ij} + \gamma \cdot CBF_{ij}$ (1), where $\alpha + \beta + \gamma = 1$, ICF_{ij} is the score that *student i* gives to *course j* based on item-based filtering, UCF_{ij} is the score that *student i* gives to *course j* based on user-based filtering, and CBF_{ij} is the score that *student i* gives to *course j* based on content-based filtering. The range of the predicted scores of all methods is between 1 and 4.3.

Item-based filtering: Item-based collaborative filtering calculates the similarity score between courses and recommends similar courses (Sarwar et al., 2001). We find the students who have taken the two courses and calculate the difference of their scores in the two courses. The smaller the difference is, the higher the similarity. The similarity is represented as $w_{i,j}$ and is shown in (2), where A is the set of students who have taken *course i* and *course j*. Assuming *student x* has taken *course i*, if PHCRS wants to recommend *course k* to *student x*, the predicted score is calculated by formula (3). The numerator is equal to the product of $w_{i,k}$ and the student's grade in *course i*. The denominator is the summation of the similarity between *course i* and *course k*.

Similarity between *course i* and *course j* ($w_{i,j}$)= $\frac{1}{1+\sqrt{\sum_{A \epsilon M(i) \cap M(j)} (grade(A,i)-grade(A,j))^2}}$ (2)

Prediction score of *course k* for *student x* = $\frac{\sum_{w_{i,k}>0} grade(x,i) * w_{i,k}}{\sum_{w_{i,k}>0} w_{i,k}}$ (3)

User-based filtering: User-based collaborative filtering utilizes students' past course data to calculate the similarity between students and recommend courses taken by similar students (Han et al., 2016). To calculate the similarity between two students, we have to determine the courses the students have both taken. We utilize the scores of two students in the courses to calculate the similarity. The similarity of *student x* and *student y* is represented as a weighted value ($w_{x,y}$) as shown in (4), where *N(x)* are the courses that *student x* has taken and *N(y)* are the courses that *student y* has taken. If the scores are closer, the similarity of the two students is higher. If PHCRS wants to recommend *course k* to *student y*, the similarity of *student x* and *student y* is multiplied by the scores of *student x* on *course k*. The average weighted value is the predicted score, as shown in (5).

Similarity of *student x* and *student y* $(w_{x,y}) = \frac{1}{1+\sqrt{\sum_{i \in N(x) \cap N(y)} (grade(x,i) - grade(y,i))^2}}$ (4)

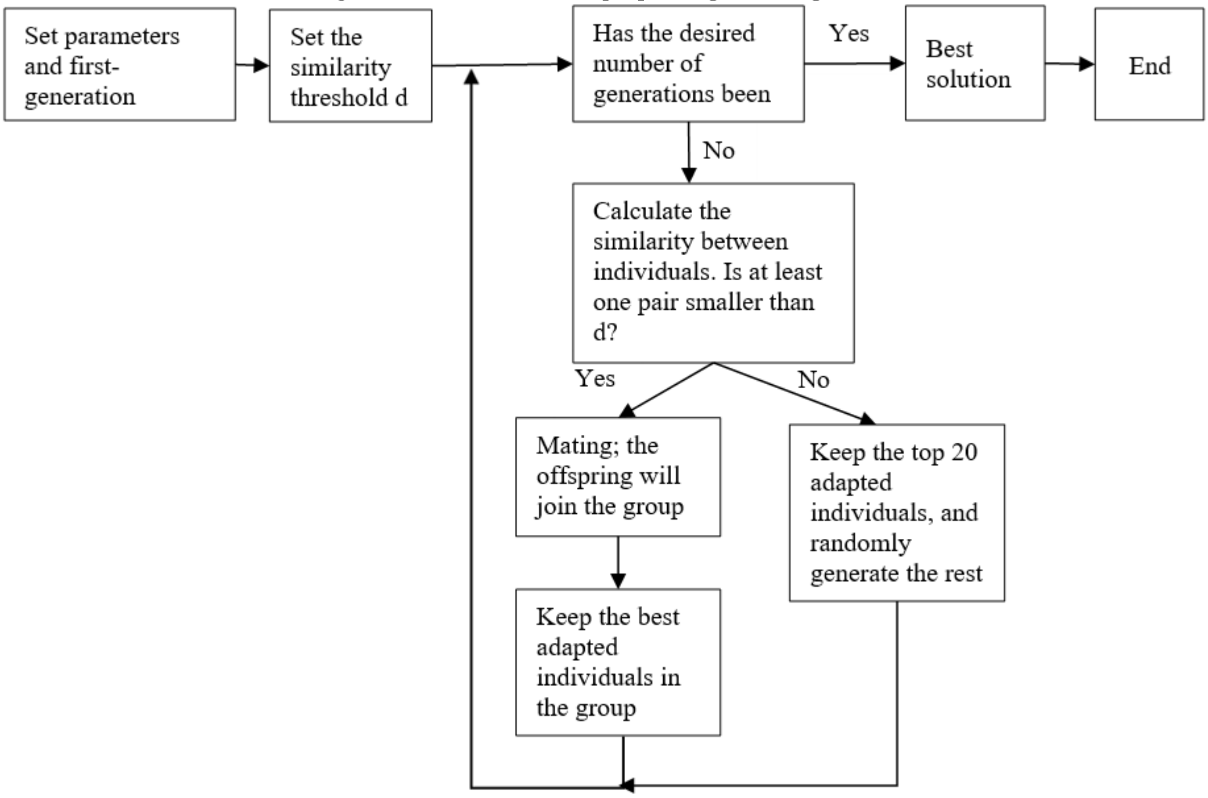
Predicted score for *student y* on *course k* $= \frac{\sum_{w_{x,y}>0.2} grade(y,k) * w_{x,y}}{\sum_{w_{x,y}>0.2} w_{x,y}}$ (5)

Content-based filtering: Content-based filtering recommends similar courses based on the characteristics of students' past courses (Esteban et al., 2020). In the first step, the feature vectors of the courses are extracted. The course feature vector indicates which domains the courses belong to and which objectives the courses contain. To calculate the feature vectors of *student x* for *course i*, the feature vector of *course i* is multiplied by the score of *student x* on *course i*. We add up all the feature vectors of *student x* on each course and define this value as the feature vector of *student x*. To recommend *course j* to *student x*, we use the feature vector of *student x* and the feature vector of *course j* to calculate the cosine value ($\cos\theta = \frac{\vec{i} \cdot \vec{j}}{||\vec{i}|| \cdot ||\vec{j}||}$) as the similarity. If the similarity is close to 1, *student x* is more likely to like *course j*.

3.3. Weight selection

We apply a genetic algorithm to find the optimal solution for course recommendation. The genetic algorithm is a type of machine learning algorithm that finds new and better individuals through crossover or mutation of candidate individuals; this procedure iterates for multiple generations until the ending criteria are satisfied (Holland, 1975). The ending criterion in this study is a fixed number of evolutions. Our algorithm follows the algorithm proposed by Esteban et al. (2020). The flow chart of the genetic algorithm in computing the optimum solution is shown in Figure 2, and the details of each step are explained in the next section.

Figure 2. Flowchart of the proposed genetic algorithm



Each individual has eleven genes and is split into four parts (Figure 3), where z_i represents the i^{th} gene.

Figure 3. Gene paradigm

7	1	76	83	1	60	55	100	1	100	2

The first three genes represent the weights of item-based, user-based and content-based filtering, respectively, when combining their solutions. In other words, $\alpha = z_1/(z_1 + z_2 + z_3)$, $\beta = z_2/(z_1 + z_2 + z_3)$ and $\gamma = z_3/(z_1 + z_2 + z_3)$. For example, if $z_1 = 7$, $z_2 = 1$ and $z_3 = 76$, then $\alpha = \frac{7}{7+1+76} = 0.083$, $\beta = \frac{1}{7+1+76} = 0.011$ and $\gamma = \frac{76}{7+1+76} = 0.904$.

The fourth to seventh genes represent the weights of content-based filtering for each variable, which are used to calculate the similarity between students. The variables include the domain of the course, the overview of the course and the detailed course context and lecturer, represented by α, β, γ and δ. For example, if $z_4 = 83$, $z_5 = 1$, $z_6 = 60$ and $z_7 = 55$, then $\alpha = \frac{83}{83+1+60+55} = 0.417$, $\beta = \frac{1}{83+1+60+55} = 0.005$, $\gamma = \frac{60}{83+1+60+55} = 0.301$ and $\delta = \frac{55}{83+1+60+55} = 0.276$.

The eighth and ninth genes represent the weights of user-based filtering for each variable, which are used to calculate the similarity between students, where z_8 is always 100. For example, $z_9 = 1$ means that the threshold of user-based filtering is 0.1.

The tenth and eleventh genes represent the weights of item-based filtering for each variable, which are used to calculate the similarity between students, where z_{10} is always 100. For example, $z_{11} = 2$ means that the threshold of item-based filtering is 0.2.

3.4. Parameters in a genetic algorithm

The following sections introduce different formulas for the genetic algorithm that were designed.

3.4.1. Distance threshold d

To address the inability of highly similar existing individuals to generate a different child generation and find the optimal solution, the generation process restarts when two genes of a child generation are too similar. The similarity threshold of *distance d* is set to 0.8. If the similarity between every individual pair is higher than *d*, then the process enters the "restart" phase, meaning that the 20 best individuals are kept while the others are generated randomly.

3.4.2. Individual dissimilarity

We use the Hamming distance to calculate the distance between each pair of individuals and then transform the distance into a similarity value, which is the number of genes that are the same divided by the length of the individual (L = 11). For example, when the first, third and fourth genes in a pair of individuals are the same, the similarity is $\frac{3}{11} \approx 0.27$.

3.4.3. Crossover operator

The method of generating a child generation is to cross the same set of genes from two parent generation individuals. For example, the first and third genes of a child-generation individual may be from the father, and the second and fourth genes may be from the mother. The crossover probability of a set of genes is 50%.

3.4.4. Update process

The best individuals of each child generation are kept, maintaining the total number of individuals, and then the next generation is generated.

3.5. Evaluation metrics

This study uses the Root Mean Square Error (RMSE) and the Normalized Discounted Cumulative Gain (NDCG) to evaluate the recommendation results.

RMSE: The RMSE has been used as a standard statistical metric to measure model performance in the recommendation system (Esteban et al., 2020). When there are more samples or the error distribution is expected to be Gaussian, reconstructing the error distribution using RMSEs will be even more reliable (Chai & Draxler, 2014). The purpose of the RMSE is to compare the predicted score of *student i* for *course j*, v_{ij}, and the real score given by the student, v_{ij}. For the testing data set *K*={(*i*,*j*)}, $\mathrm{RMSE} = \sqrt{\frac{\sum_{(i,j)\epsilon K}\sum(p_{ij}-v_{ij})^2}{\#K}}$, and a smaller RMSE value means that the predicted score is closer to the real score given by the student.

NDCG: The NDCG is a family of ranking measures widely used in applications. It has two advantages. First, the NDCG allows each retrieved document has graded relevance while most traditional ranking measures only allow binary relevance. Second, the NDCG involves a discount function over the rank, while many other measures uniformly weight all positions (Wang et al., 2013). For the *k* example courses, we sort the courses by the recommendation scores and calculate the discounted cumulative gain (DCG). The DCG is shown in (5), where *k* represents the number of courses the system recommends and rel_i is the gain for each recommended course. In the evaluation, when the recommended course overlaps the real record, we set the gain rel_i to 1; otherwise, it is set to 0. The ideal course order based on the predicted score is used to calculate the ideal discounted cumulative gain (IDCG), as shown in (6). We can use the DCG and IDCG to calculate the NDCG, as shown in (7).

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i}-1}{log_2(i+1)} \quad (5)$$

$$IDCG_k = \sum_{i=1}^{|rel_k|} \frac{2^{rel_i}-1}{log_2(i+1)} \quad (6)$$

$$NDCG_k = \frac{DCG_k}{IDCG_k} \quad (7)$$

3.6. Experimental work

The experiment is divided into two parts. First, we determine the optimized weight for each index in PHCRS (including item-based filtering, user-based filtering, and content-based filtering) separately. Then, we use the RMSE and NDCG to evaluate the accuracy of the recommendation provided by PHCRS. The system is built in the Python environment, including the recommendation criterion, genetic algorithm, and system performance evaluation. The data source is the course selection records of college students from twelve departments at NYCU from academic years 2015-2020. The unit of the experiment during system development is per department, the training data consist of the course selection data from 2015-2018 and the 2020 academic year, and the testing data are the course selection data from 2019. The results of the experiment are given below.

3.6.1. Criteria weight optimization

The first part of the experiment uses a genetic algorithm to determine the weights of the three recommendation methods of PHCRS, to optimize their relative parameters, and to evaluate the influence of the weights on PHCRS. In PHCRS, there are nine weights that need to be optimized, including the weights of item-based, user-based and content-based filtering, the sizes of the filters of item-based filtering and user-based filtering, and the weights of the objectives, knowledge areas, skills and professors in content-based filtering. The settings of the important parameters of the genetic algorithm are shown in Table 2, which we applied for the experiment.

Table 2. Configuration of the genetic algorithm parameters

Parameter	Value
Number of generations	100
Population size	209-1258
Crossover probability	0.9
Initial value for incest prevention threshold	4
Allowed range for weight genes	[0, 50]
Allowed range for neighborhood gene	[1, 50]
Allowed range for metric genes	[0, 4] or [0, 1]

Table 3 shows the optimized weights of each department obtained by the genetic algorithm. The results showed that there is a large difference between the weights of the four indexes in content-based filtering, with the weight of "Course objectives" lying within .339% ~ 78.723%, the weights of "Knowledge areas" lying within 1.613% ~ 38.525%, the weights of "Skills" lying within .633% ~ 38.672%, and the weights of "Professor" lying within 7.447% ~ 53.714%, indicating that the influence of the indexes differs from department to department. For example, students from the Department of Electrical and Computer Engineering mainly consider "Professor" (53.459%), and students from the Department of Mechanical Engineering mainly consider "Course objectives." We further compare the weights of the three recommendation methods in PHCRS, and the results show that for all departments, the weight of item-based filtering is always the highest, lying within 94.118% ~ 98.039%, while the weights of user-based and content-based filtering are both low in PHCRS; the former lies within .971% ~ 5.208%, and the latter lies within .908% ~ 2.913%. Thus, item-based filtering is the method that mainly influences the results of course recommendation provided by PHCRS.

Table 3. Criteria weights, similarity measures chosen by genetic algorithm, and RS evaluation

College/ Department	Content-based filtering				Hybrid recommendation			Evaluation	
	Course objectives	Knowledge areas	Skills	Professor	Item-based filtering	User-based filtering	Content-based filtering	RMSE	NDCG
College of Electrical and Computer Engineering									
• Department of Electrical and Computer Engineering	.63%	13.84%	32.08%	53.46%	97.47%	1.27%	1.27%	.61	.93
• Department of Photonics	.49%	31.53%	37.93%	30.05%	96.77%	1.08%	2.15%	.37	.94
College of Computer Science									
• Department of Computer Science	17.62%	31.09%	19.69%	31.61%	96.15%	2.56%	1.28%	.90	.90
College of Engineering									
• Department of Civil Engineering	.41%	38.53%	25.00%	36.07%	95.89%	2.74%	1.37%	.80	.93
• Department of Mechanical Engineering	78.72%	6.38%	7.45%	7.45%	95.83%	3.13%	1.04%	.58	.95
• Department of Materials Science and Engineering	.34%	33.22%	33.90%	32.54%	94.12%	4.90%	.98%	.56	.96
College of Management									
• Department of Management Science	49.37%	32.28%	.63%	17.72%	98.04%	.98%	.98%	.42	.96
• Department of Transportation & Logistics Management	10.86%	21.14%	14.29%	53.71%	95.75%	2.13%	2.13%	.59	.95
• Department of Industrial Engineering	69.36%	1.61%	8.07%	20.97%	93.75%	5.21%	1.04%	.54	.93

and Management									
• Department of Information Management and Finance	.39%	28.52%	38.67%	32.42%	96.12%	.97%	2.91%	.39	.97
College of Hakka Studies									
• Department of Humanities and Social Sciences	70.27%	8.11%	5.41%	16.22%	97.00%	1.00%	2.00%	.47	.95
• Department of Communication and Technology	29.31%	22.66%	29.31%	18.73%	97.67%	1.16%	1.16%	.75	.96

3.6.2. RS evaluation

The second part of the experiment uses the RMSE and NDCG to evaluate the accuracy of the course recommendation results provided by PHCRS. The value of RMSE indicates the difference between the predicted score and the score provided by students who finished the course. A larger RMSE value means that the difference between the predicted and real scores is larger. The results showed that the RMSE values of all departments lie within .365 ~ .898, with the departments with fewer courses having lower RMSE values (e.g., the Department of Photonics) and the departments with more courses having higher RMSE values. On the other hand, the value of NDCG indicates the sequence of recommendations, and a larger value of NDCG means that a more highly correlated course could be recommended first (e.g., courses that could yield higher grades). The results showed that the value of NDCG lies within .902 ~ .970 for all departments, meaning that for all departments, the collaborative filtering method applied by PHCRS is able to recommend courses to students based on the importance of the course (Table 3). It is worth noting that even though there is no direct relationship between the performance of RMSE and NDCG, generally, the departments with good RMSE performance also have sufficient NDCG values.

Figure 4 shows the results of the genetic algorithm iterating for 100 generations on each department. From the scree plot of each department, the RMSE values of the first generation lie within .4 ~ 2.8, and as the evolution continues, the RMSE values for every department decrease to .4 ~ .9, indicating that using a genetic algorithm in collaborative filtering can yield the optimal solution. We also find that the convergence for the College of Engineering is more obvious, and the similarity of the College of Management courses is higher, but both are able to minimize the recommendation error as evolution continues.

4. Research design

This study uses a survey method to verify the accuracy of PHCRS. The survey uses nonprobability sampling to invite undergraduates from the Department of Electrical and Computer Engineering, and Computer Science, NYCU, who volunteered as participants. As the freshmen's course selection and grade data were not yet completed, they were excluded to avoid interference in the research results. A total of 61 students were selected (28 sophomores, 15 juniors, and 18 seniors; 44 males and 17 females). In this research, recruitment posters were sent out by online student communities. After the students signed up, the researchers explained the research process and the parameters via phone or mail. To collect the data, students were required to log in to the course recommendation system. After reading the description of the hybrid recommendation method, students were asked to evaluate whether the courses recommended by the method was of interest, and if so, to provide their reasoning. Finally, they were asked to fill in their personal information and offer suggestions for the system.

This study uses a recommendation effect scale defined by our research group. When students browsed the course recommendation list, they were asked to evaluate whether each course was of interest to them and the reasons for their answer. For example, when students answered, "yes", they would select from reasons aligned with “autonomous motivation,” which comes from careful consideration and self-determination (Lee & Sun, 2010) and includes reasons, such as the practicality of the course content, individual learning plans and personal interests. In addition, there were other reasons aligned with passive “external information motivation” (Lee & Sun, 2010), which included reasons, such as making up for missed credits, the course being easy to pass, and seeing good reviews about the teacher. This study also uses the students’ true course selection list and grade data to verify the accuracy of PHCRS.

5. Data analysis and results

5.1. An analysis of the difference among the students’ degree of interest in the courses recommended according to the order of the recommendations

Chi-Square test is used in this section. The data follow a normal distribution (skewness between -.48 and 1.30; kurtosis between -2.28 and 1.08). Table 4 shows that the course recommendation order is related to the students’ interest or not ($\chi^2 = 10.38$; $p < .05$). The results indicated that the students were more interested in the courses at the top of the recommendation lists.

Table 4. A difference analysis between the students’ degree of interest in the courses recommended in the course recommendation order

Recommendation Courses	*n*	Interest		No interest		χ^2	*p*
		Frequency	Percentage	Frequency	Percentage		
First course	61	51	83.60%	10	16.40%	10.38	$.03^{*}$
Second course	61	46	75.40%	15	24.60%		
Third course	61	43	70.50%	18	29.50%		
Fourth course	48	30	62.50%	18	37.50%		
Fifth course	46	27	58.70%	19	41.30%		

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

5.2. The degree of interest in the recommended courses is affected by students’ internal and external motivations for taking a course

The Mann-Whitney U nonparametric test is used in this section. The data follow a normal distribution (skewness between .13 and 1.58; kurtosis between -1.06 and 2.10). Table 5 shows that the proportion of students with autonomous motivation (M = 89.13% ~ 100%) was higher than that of students with extrinsic informational motivation (M = 29.63% ~ 44.19%; $p < .001$) across the five recommendation courses. The results indicated that most students choose courses according to their plans, interests, or needs.

Table 5. A difference analysis of the students’ motivation of course-taking in hybrid recommendation method

Recommendation Courses	*n*	Autonomous motivation		Extrinsic informational motivation		*p*
		M	*SD*	*M*	*SD*	
First course	51	92.16%	27.15%	43.14%	50.02%	$.00^{***}$
Second course	46	89.13%	31.47%	41.30%	49.78%	$.00^{***}$
Third course	43	97.67%	15.25%	44.19%	50.25%	$.00^{***}$
Fourth course	30	96.67%	18.26%	40.00%	49.83%	$.00^{***}$
Fifth course	27	100.00%	0.00%	29.63%	46.53%	$.00^{***}$

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

5.3. The degree of academic performance in the courses recommended is affected by students’ following and not following the recommendation list

The Mann-Whitney U nonparametric test is again used in this section. The data follow a normal distribution (skewness = -.34; kurtosis = .54). Table 6 shows that the students’ degree of academic performance in the courses recommended will not differ among the following (M = 85.90) and not following the recommendation list (M = 84.98; $p > .05$). The results indicated that there is same on their academic performance.

Table 6. A difference analysis between the students' degree of academic performance in the courses recommended among following and not following the recommendation list

Recommendation Courses	*n*	Recommendation and true course selection list overlap proportion		Academic performance				*p*
				Following the recommendation list		Not following the recommendation list		
		Min %	Max %	*M*	*SD*	*M*	*SD*	.45
Given 5 recommended courses	61	20%	100%	85.90	7.37	84.98	5.06	

6. Conclusions

This study proposes PHCRS based on human-centered AI in education combining item-based filtering, user-based filtering, and content-based filtering to recommend courses for students from different departments in universities. Our system used a genetic algorithm to automatically optimize the weights of indexes. In addition to enhance the accuracy of PHCRS, a genetic algorithm also configures the weights of different recommendation methods for each student to suit their needs. The results show that the weights of recommended methods are slightly different between departments. However, the influence of item-based filtering on the course recommendation result is higher than that of user-based and content-based filtering, meaning that students tend to select courses with similar characteristics. This result is in line with that of Chang et al. (2022), who found that the accuracy of item-based filtering is better than that of other recommendation methods through the receiver operating characteristic (ROC) curve. We also found that after the experiment, the weights of the four parameters in content-based filtering (course objectives, knowledge areas, skills, and professors) were not the same, meaning that the focus of students on courses was different.

We use RMSE and NDCG to evaluate the effectiveness of PHCRS, and the results show superior performance compared to previous research (ex: Esteban et al., 2020; Defiebre et al., 2022; Ngaffo et al., 2020). This study also collects data of university students who used PHCRS to evaluate the helpfulness of the system on students in real world situations. The results show that students are more interested in courses that ranked higher in the recommendation list, especially the top 3 ranked courses, which 70 ~ 83% of students are interested in. However, individual differences are also found in the course selection preference of students, with some students only interested in 1 to 2 courses in the recommendation list, while most students are interested in 3 to 5 courses in the list. When students are interested in the course being recommended to them, 90% of them are based on intrinsic motivation reasons, including personal interest or attracted by the syllabi, indicating that most students approved the courses recommended by PHCRS. In contrast, 10% of them select the recommended courses based on extrinsic motivation reasons, including obtaining the necessary credits for graduation or is easier to pass the course. This study further utilizes the actual course selection data of the students to discuss whether they select courses based on the recommendation list, Moreover, the results show a considerable gap in matching between 20% to 100%, meaning that even though some students are interested in the course recommendation list, they may not consider taking those courses. The possible reasons for this may be personal or environmental interference, but there is no substantial difference in learning outcomes whether they follow the recommendation list or not, indicating that the recommended course provided by PHCRS are not necessarily those that are easier to receive good grades.

7. Contributions, limitations, and future work

The PHCRS proposed in this study proves its ability of recommending adequate course lists for students from different departments while taking human factors into consideration and providing recommendations that suit the students' needs. Only few research studies proved the effectiveness of the AI course recommendation system on students' learning outcomes (e.g., Esteban et al., 2020), but these systems focused on specific subjects by collecting additional data for their experiments. The PHCRS proposed in this study eliminates this downside by developing the system directly utilizing the course selection data, then uses AI to find out the potentials and disadvantages of students and recommend adequate courses for students to select. The system is now available for all students in NYCU. Moreover, the PHCRS database can track the learning progress and learning outcomes of students through its own database or concatenate the data from the university database and then provide recommendations by taking these data into consideration. In the future, we can adjust or expand the functionality of the PHCRS through historic data and provide interdisciplinary course recommendations and real-time learning outcome feedback, making the recommendation results more focused on the need of students in different learning stages.

Second, our research proved the potential of the genetic algorithm in finding the optimum weights of the parameters in a recommendation system, especially in chromosome modeling, in which the genetic algorithm can optimize the relative parameters, such as the size of neighbors and similarity metrics. This method can set up the best parameter setting combinations for each student. However, the recommended courses can be affected by personal preferences, course selection regulations, or the environment that the student is in, making the recommendation not 100% accurate (Chang et al., 2022; Esteban et al., 2020). This is a common restriction in human-centered recommendation systems; no state-of-the-art systems can include all algorithms, and no state-of-the-art algorithms can be applied without sacrificing accuracy in some fields (Lee et al., 2023). Although it is a tough task, to make the recommendation more accurate, we will keep using AI techniques to find out the factors affecting students' course selection decisions and their needs. By taking these human or environment factors into the construction of the recommendation model, the recommendation results can be closer to the true personal needs of students and can be more accurate.

Finally, the case study only tracks one semester of use of PHCRS, and the results indicated that those who selected courses based on the recommendations provided by PHCRS did not have higher motivations nor higher grades than those who did not. Based on the records collected by PHCRS, even though this study practiced the value of human-centered AI while developing PHCRS, there are still some issues that can be solved by further studies or system development. With the PHCRS being open to all undergraduate students, what kind of characteristics or student needs made them more intrigued to use PHCRS? Do departments with more students and courses hold higher standards towards PHCRS? Do students change their course selection preferences after using PHCRS for some time? How does PCHRS change its recommendation algorithm accordingly? Future research can concatenate with other databases of interest, adding real-time feedback or learning analysis, offering this information to students and teachers to achieve the goal of learning outcome optimization.

Acknowledgement

The authors would like to thank the National Science and Technology Council of Republic of China for financially supporting this research under Contract No. NSC 110-2222-E-008-008-MY3 and NSC 111-2410-H-A49-030-.

Reference

Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Bendaddou, D., Ayyash, M., & Al-Fuquaha, A. (2020). *Artificial Intelligence in Education: A Panoramic review*. https://doi.org/10.35542/osf.io/zvu2n

Alkhatlan, A., & Kalita, J. (2019). Intelligent tutoring systems: A Comprehensive historical survey with recent developments. *International Journal of Computer Applications, 181*(43), 1-20. https://doi.org/10.5120/ijca2019918451

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A Systematic literature review. *Intelligent Data Analysis, 21*(6), 1487-1524. https://arxiv.org/abs/1901.03888

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*, 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

Chang, H. T., & Chen, H. H. (2021). The Learning outcomes in short-term advanced technology interdisciplinary hands-on courses using multigroup SEM. *Chinese Journal of Science Education, 29*(3), 245-266. https://doi.org/10.6173/CJSE.202109_29(3).0003

Chang, P. C., Lin, C. H., & Chen, M. H. (2016). A Hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. *Algorithms, 9*(3), 47. https://doi.org/10.3390/a9030047

Chang, H. T., Lin, C. Y., Wang, L. C., & Tseng, F. C. (2022). How students can effectively choose the right courses: Building a recommendation system to assist students in choosing courses adaptively. *Educational Technology & Society, 25*(1), 61-74.

Defiebre, D., Sacharidis, D., & Germanakos, P. (2022). A Human-centered decentralized architecture and recommendation engine in SIoT. *User Modeling and User-Adapted Interaction, 32*(3), 297-353. https://doi.org/10.1007/s11257-022-09320-3

Dwivedi, S., & Roshni VS, D. K. (2017, August 3-4). Recommender system for big data in education. In *Proceedings of 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH).* https://doi.org/10.1109/ELELTECH.2017.8074993

Esteban, A., Zafra, A., & Romero, C. (2020). Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. *Knowledge-Based Systems, 194*(22), 1-14. https://doi.org/10.1016/j.knosys.2019.105385

Farzan, R., & Brusilovsky, P. (2006). Social navigation support in a course recommendation system. In *Adaptive Hypermedia and Adaptive Web-Based Systems. AH 2006. Lecture Notes in Computer Science, 4018*, 91-100. https://doi.org/10.1007/11768012_11

Han, J. W., Jo, J. C., Ji, H. S., & Lim, H. S. (2016). A Collaborative recommender system for learning courses considering the relevance of a learner's learning skills. *Cluster Computing, 19*, 2273-2284. https://doi.org/10.1007/s10586-016-0670-x

Holland, J. (1975). *Adaptation in natural and artificial systems: An Introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.

Huang, M., Huang, H., & Chen, M. (2007). Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Expert Systems with Applications, 33*(3), 551-564. https://doi.org/10.1016/j.eswa.2006.05.019

Iatrellis, O., Kameas, A., & Fitsilis, P. (2017). Academic advising systems: A Systematic literature review of empirical evidence. *Education Science, 7*(4), 90. https://doi.org/10.3390/educsci7040090

Kurniadi, D., Abdurachman, E., Warnars, H. L. H. S., & Suparta, W. (2019). A Proposed framework in an intelligent recommender system for the college student. *Journal of Physics: Conference Series, 1402*(6), 1-7. https://doi.org/10.1088/1742-6596/1402/6/066100

Lee, A. V. Y., Luco, A. C., & Tan, S. C. (2023). A Human-centric automated essay scoring and feedback system for the development of ethical reasoning. E*ducational Technology & Society, 26*(1), 147-159.

Lee, Y. M., & Sun, S. H. (2010). The relationship between autonomous motivation of course-taking and learning engagement on college students. *Journal of Research in Education Sciences, 55*(1), 155-182. https://doi.org/10.3966/2073753X2010035501006

Lin, J., Pu, H., Li, Y., & Lian, J. (2018). Intelligent recommendation system for course selection in smart education. *Procedia Computer Science, 129*, 449-453. https://doi.org/10.1016/j.procs.2018.03.023

Liu, Z., Zhang, N., Peng, X., Liu, S., & Yang, Z. (2023). Students' social-cognitive engagement in online discussions: An Integrated analysis perspective. *Educational Technology & Society, 26*(1), 1-15.

Ngaffo, A. N., Ayeb, W. E., & Choukair, Z. (2020). A Bayesian inference based hybrid recommender system. *IEEE Access, 8*, 101682-101701. https://doi.org/10.1109/ACCESS.2020.2998824

Popenici, S. A. D. & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Learning, 12*(22). https://doi.org/10.1186/s41039-017-0062-8

Renz, A., Krishnaraja, S., & Gronau, E. (2020). Demystification of artificial intelligence in education: How much AI is really in the educational technology? *International Journal of Learning Analytics and Artificial Intelligence in Education (ijAI), 2*(1), 14-30. https://doi.org/10.3991/ijai.v2i1.12675

Renz, A., & Vladova, G. (2021). Reinvigorating the discourse on human-centered artificial intelligence in educational technologies. *Technology Innovation Management Review, 11*(5), 5-16.

Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation systems for education: Systematic review. *Electronics, 10*(14), 1-21. https://doi.org/10.3390/electronics10141611

Sawarkar, N., Raghuwanshi, M. M., & Singh, K. R. (2018). Intelligent recommendation system for higher education. *International Journal on Future Revolution in Computer Science and Communication Engineering, 4*(4), 311-320.

Schmidt, A. (2020). Interactive human-centered artificial intelligence: A Definition and research challenges. In *Proceedings of the International Conference on Advanced Visual Interfaces, 3*, 1-4. https://doi.org/10.1145/3399715.3400873

Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013). A Theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory (COLT 2013), 8*, 6.

Xu, W. (2019). Toward human-centered AI: A Perspective from human-computer interaction. *Interactions, 26*(4), 42-46. https://doi.org/10.1145/3328485

Yang, S. J. H., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence, 2*, 100008. https://doi.org/10.1016/j.caeai.2021.100008

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(39). https://doi.org/10.1186/s41239-019-0171-0

Zhang, H. R., Min, F., He, X., & Xu, Y. Y. (2015). A Hybrid recommender system based on user-recommender interaction. *Mathematical Problems in Engineering, 2015*, 1-12. https://doi.org/10.1155/2015/145636