

Article

Fusing Personal Profile and Group Experience: A Major Recommendation Framework Built on LLM-Generated Data

Shiyuan Huang ¹  and Ziwei Chen ^{2,*}¹ College of Intelligent Robotics and Advanced Manufacturing, Fudan University, Shanghai 200433, China² Department of Electronics, Beijing Jiaotong University, Beijing 100044, China

* Correspondence: zwchen@bjtu.edu.cn

Abstract

Major recommendation systems have considered multifaceted user profiles. However, their heavy reliance on questionnaire surveys makes it difficult to extend their application, and also renders recommendation outcomes susceptible to interference from biases. Addressing these issues, this paper proposes a multidimensional decision-fusion major recommendation framework for university applicants. A large language model was innovatively utilized to generate participant responses, constructing a simulated dataset (SimuCareer-25) to overcome the low efficiency and high bias of traditional survey methods. Building upon this, the framework integrates three key decision dimensions: preference, capability, and experience. It then employs fuzzy theory and XGBoost to process multidimensional user features encompassing personality, interests, subject literacy, and skills. Simultaneously, it leverages collaborative filtering to mine group experience. An Analytic Hierarchy Process module is used to perform adjustable weighted fusion of these three dimensions. Through experiments on SimuCareer-25 and real-world data, it was indicated that integrating preference, capability, and historical experience is conducive to generating robust and reliable major recommendations, and that despite structural deviations from real-world samples, the simulated data provides effective feature transferability to support the framework's practical utility.

Keywords: analytic hierarchy process; collaborative filtering; large language model; major recommendation; XGBoost

MSC: 68T05



Academic Editor: Jonathan Blackledge

Received: 6 July 2026

Revised: 18 August 2026

Accepted: 19 August 2026

Published: 30 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Due to the detailed specialization in university admissions and significant information asymmetry in career planning education, most students face considerable confusion regarding major selection. High school graduates are generally required to make a one-time, decisive choice of major within a limited application period, without the opportunity to explore and experience before deciding on their developmental path. Furthermore, a considerable number of students lack a rational understanding of the fit between a major and themselves, leading to dissatisfaction with their chosen major and enrollment decisions after entering university [1]. However, currently, many students still base their major choices on relatively subjective perceptions within small social circles, such as family, friends, teachers, and cluttered online information, lacking reliable and clear information and advice. Higher education urgently needs the assistance of major recommendation systems.

Recommendation systems are currently widely used in scenarios such as e-commerce, media, and education. Career recommendation is an important application direction in the education field, with major recommendation for university application being a more specific branch. Across academic disciplines, outstanding students often share certain characteristics, reflected in factors such as subject ability, interests, and personality traits. Therefore, exploring the use of these traits for major recommendation is feasible [2]. Existing systems for engineering majors [3,4] have achieved good results but lack generalizability and public datasets, and their core fuzzy logic method faces challenges in cross-disciplinary scenarios. Recent systems covering broader disciplines [5,6] remain limited by their feature designs tied to specific user populations. To achieve broad applicability, feature construction needs to move beyond specific evaluation systems toward universal ability and interest assessment. Extending the view to career recommendation, classical methods such as content-based filtering (CBF) [7] and collaborative filtering (CF) [8] are not well-suited for major selection, as CBF relies on deep item understanding, while CF lacks control over personal will. Therefore, these methods cannot be simply transferred to major recommendation. Simultaneously, there is a lack of publicly available datasets for modeling and performance testing for related research. Existing studies all need to collect data by accessing databases or conducting large-scale surveys, further increasing the difficulty of related research.

To address the aforementioned problems in the field of major recommendation, this paper first conducts a systematic review of typical research, investigating from two levels: data construction and model design. In response to the lack of publicly available datasets in this field, drawing on an extensive review of the literature with a specific focus on data collection methods, as well as psychological theories, this paper designs a comprehensive major selection questionnaire in detail. It innovatively utilizes a large language model (LLM) to simulate human participants, forming a simulated major selection dataset, SimuCareer-25. Compared to relying solely on traditional questionnaires, this method offers greater efficiency in data collection and helps reduce biases. Building on prior work, this paper proposes a novel recommendation framework that integrates individual assessment with group behavioral insights. Unlike conventional approaches that rely on isolated information resources or manually engineered feature-item associations, our framework separately models user-specific attributes and collective peer patterns, then fuses these complementary signals through a weighted integration mechanism. To further accommodate the inherent uncertainty of user self-assessment data, a fuzzy processing strategy is introduced, which softens the impact of ambiguous or imprecise responses. The overall framework offers a solution without imposing strong structural priors that limit generalizability. Combined with the proposed data construction method, it enables rapid data collection and model re-training when scaling to new disciplines or features, making it readily adaptable to broader application scenarios. Finally, the intrinsic properties of the LLM-simulated data and the practical utility of the proposed framework are rigorously evaluated through data fidelity assessments, comparative model analysis, weight optimization, and downstream validation on a real-world dataset.

The structure of this paper is organized as follows. Section 2 reviews related work on major and career recommendation; Section 3 elaborates on the LLM-based dataset construction process; Section 4 details the proposed framework; Section 5 presents the experimental design and analyzes the results; Section 6 concludes the paper.

2. Related Work

This section provides a comprehensive review of recommendation systems, particularly related work on major and career recommendation. Key related work primarily encompasses two directions: the selection of multidimensional influencing factors and the design of recommendation frameworks.

2.1. Multidimensional Influencing Factors

Research in social sciences and recommendation systems based on them explores numerous factors influencing individuals in career choice and development [2,9–12]. Among them, key influencing factors involve interests, skills, academic performance, experience, personality, and more.

Based on Social Cognitive Career Theory [9], Oh et al. [10] developed the Science–Technology–Mathematics Educational-Career Interest Scale for high school students. Experiments confirmed the scale’s significant predictive power for Science, Technology, Engineering, and Mathematics (STEM) course selection and career intentions. The practical importance of interest alignment is further evidenced by the finding that congruence between individuals’ interests and their chosen paths significantly predicts long-term satisfaction [11], reinforcing the value of prioritizing interest-based matching in guidance systems. Often considered alongside interest is a student’s own capability, such as skills and academic performance. Verma et al. [12] researched and confirmed that when students map verifiable skills and subject performance to job requirement profiles, it can significantly enhance the match and feasibility of career choices. The article emphasizes that clearly and quantitatively presenting these capability advantages is a core lever for data-driven career counseling. Correspondingly, Qamhie et al. [3] adopted interests, skills, and academic performance as important bases for personalized recommendation. More recently, Roy and Sharma [13] developed a career guidance model that leverages Holland’s vocational interest theory alongside machine learning techniques to generate recommendations, demonstrating the continued relevance and adaptability of interest-based approaches in contemporary recommendation systems.

On the other hand, complex group experience helps students understand career development and influences their choices. Nie et al. [14] distilled the career trajectories of previous students into learnable college prototypes, demonstrating that, even after controlling for personal capability and other factors, this experiential path still significantly influences students’ final choices. This quantified the effect of the experience factor on major and career decisions at the level of behavioral big data for the first time. Other research, such as that by Verma et al. [4], used trust relationships including teachers and parents to assist recommendations based on students’ own indicators.

Furthermore, personality-aware recommendation provides a novel perspective for feature selection in educational recommendation systems. Dhelim et al. [15] studied and compared four personality-aware recommendation systems based on different personality models, namely the Big-Five, Eysenck, and HEXACO models from trait theory, and the Myers–Briggs Type Indicator (MBTI) model from type theory. They further proposed a hybrid personality model capable of effectively addressing cold-start problems. Qamhie et al. [3] used MBTI test results as a personality factor, designing a recommendation system for engineering majors for high school graduates.

2.2. Recommendation Frameworks

Recommendation systems have undergone several paradigm evolutions, with content-based filtering, collaborative filtering, and hybrid methods constituting their foundational methodology [16]. Among them, content-based filtering exhibits unique advantages in

scenarios requiring explicit domain knowledge representation, particularly relevant in major and career recommendation. For example, Yadalam et al. [7] proposed a career recommendation system based on CBF, recommending positions for engineering students based on user profiles like skills. Meanwhile, collaborative filtering, due to its ability to fully utilize social networks and user experience, provides another reliable solution for related recommendation. Wang et al. [8] introduced weighted features influencing college students' occupational values and proposed a CF algorithm based on these features. The CaPaR system built by Patel et al. [17] utilizes text mining, CBF, and CF techniques, analyzing users' personal profiles and resumes to personalize recommendations for jobs, required additional skills, and corresponding learning resources. Such approaches are rarely adopted in isolation for major recommendation, as they rely on a single information source and cannot adequately capture the person–major matching mechanism.

In research related to major and career recommendation, fuzzy logic is a frequently used classical method because it best simulates the working style of school counselors: providing personalized development recommendations to students based on their academic and personal profiles, aided by a set of rules derived from experience. Fuzzy logic, proposed by Zadeh [18], is a computational method based on degrees of truth rather than Boolean logic. It is used for systems with qualitative input observations and imprecise outputs with partial truth values, capable of handling the uncertainty present in most practical problems [19]. A fuzzy system consists of three parts: (1) a fuzzifier, which converts inputs into linguistic variables using membership functions; (2) an inference engine, which uses if–then rules to convert fuzzy inputs into fuzzy outputs; and (3) a defuzzifier, which converts the fuzzy outputs from the inference engine into crisp outputs. Qamhie et al.'s [3] study is a representative one focusing on fuzzy logic. It designed a fuzzy system based on Mamdani inference and constructed 15 if–then rules from two dimensions, namely academic performance and personal profile, to recommend engineering major directions to students. Systems based on this method, however, are constrained by manually crafted explicit rules, resulting in poor scalability and limited capability in processing high-dimensional features.

Some other studies focusing on multidimensional factors have introduced the multi-criteria decision-making technique, Analytic Hierarchy Process (AHP), into recommendation frameworks. AHP, proposed by Saaty [20,21], is a structured technique for organizing and analyzing complex decisions based on mathematics and psychology. Aiming to determine user recommendation criteria, it requires users to evaluate the importance of multiple dimensions to obtain precise multidimensional weights. The study by Wongvilaisakul et al. [22] targeted graduate program recommendation. As each user considers factors like program length, resources, and conditions differently, it used the AHP method to parse user recommendation needs and obtain exact weights for multiple criteria. Verma et al. [4], considering that each student might have different decision criteria and weights regarding preferences for major options, aimed to incorporate students' psychological expectations for recommendations into the model, thus choosing AHP, a multi-criteria decision-making technique suitable for handling multiple domains. The study observed that student groups of different majors and genders had different weighting values and verified the performance difference of the dimensional models with and without them. However, it did not verify the impact of AHP weight changes on recommendations, nor did it form a general weight configuration suggestion.

In recent years, many studies have begun exploring data-driven machine learning methods to automatically learn complex patterns from large-scale behavioral data or student profiles for prediction or recommendation. These methods do not rely on manually formulated rules but instead mine key factors influencing major and career choices and their nonlinear relationships from historical data through algorithms, possessing stronger adapt-

ability and scalability. Nie et al. [14] proposed a hybrid model named ACCBOX, whose core combines XGBoost with prototype cluster center regularization. XGBoost [23], as an efficient gradient-boosted decision tree model, was used for multi-class career choice prediction tasks. The study further introduced prototype cluster centers for each college as a regularization term, enabling the model to balance individual data fitting with group commonalities. Wang et al. [5] approached the problem from a fairness perspective, employing a neural collaborative filtering (NCF) model combined with debiasing embedding techniques. The study trained an NCF model to learn user interest representations from Facebook content, then used linear projection to eliminate gender bias in user embedding vectors, and finally performed major recommendation via logistic regression. Targeting the high school student group, Dang et al. [6] employed deep neural networks (DNN) to process multidimensional academic and behavioral features. The model performed high-order feature extraction through fully connected layers and Dropout layers, ultimately outputting a multi-major probability distribution. It demonstrated classification performance superior to the traditional machine learning method K-Nearest Neighbors (KNN) in testing, showcasing the effectiveness of structured feature engineering and high-order feature interaction in major recommendation. Despite the promising performance demonstrated by these approaches, the integration of multiple information channels or functional modules within a machine learning framework for major recommendation remains largely unexplored.

Furthermore, large language model (LLM)-based recommendation systems have emerged as a new research direction. These works differ in how LLMs are utilized. Some employ LLMs as direct inference engines that reason over user features to generate recommendations [24,25]. Others leverage LLMs as auxiliary tools to generate or augment user behavior data for training recommendation systems [26,27]. To the best of our knowledge, this latter usage has primarily focused on modeling user interactions such as browsing histories, and has not yet been explored for simulating self-assessment responses in major or career recommendation contexts. Nevertheless, it offers a valuable methodological reference for our survey-based setting.

To more clearly understand different types of major and career recommendation systems, Table 1 summarizes representative conventional systems, showing their methods, recommendation targets, considered factors, data sources, and evaluation metrics. Through this review, two notable gaps in the current research landscape emerge. First, existing studies have considered diverse personal factors, but the reliance on survey-collected data makes it difficult to integrate them into a unified modeling framework. Second, while various methods have proven effective individually, few have integrated multiple information channels, particularly group behavioral experience, into a single extensible framework for major recommendation.

Table 1. Summary of relevant studies in major and career recommendation.

Method	Recommended Item	Consideration Factors	Data Source	Evaluation Metrics
Fuzzy logic [3]	Engineering stream	Personality (MBTI); Academic performance; Interests; Skills	Questionnaire	Cohen's κ : 0.23; Satisfaction: 3.0/5.0
Fuzzy logic; AHP [4]	Engineering stream	Personality; Interests; Skills; Trust relationships	Questionnaire	Accuracy: 0.82; Precision: 0.85; Recall: 0.86; F1-score: 0.86

Table 1. Cont.

Method	Recommended Item	Consideration Factors	Data Source	Evaluation Metrics
NCF [5]	College major	Facebook interests; Gender	Public dataset	UPAR: 1.1888; NDCG@3: 0.0024; NDCG@10: 0.0057; NDCG@20: 0.0089;
DNN [6]	College major (for AP course system)	Academic performance; Extracurricular activities; Personal capabilities; Interests	Questionnaire	Accuracy: 0.97; Precision: 0.97; Recall: 0.98; F1-score: 0.89
CBF; NLP [7]	Job	Skills; Job intentions	Multiple databases	User reviews
CF; Clustering [8]	Job	Occupational values	Questionnaire	System diversity: 0.39
XGBoost; Clustering [14]	Post-graduation destination	Academic performance; Behavioral patterns; Reading preferences; Consumption patterns	Campus information system	Accuracy: 0.638; F1-score: 0.647; Precision: 0.629; Recall: 0.666
NLP; CBF; CF [17]	Job; Skill	Job intentions; Resume content	University career center	Precision: 0.799; RMSE: 0.938
Fuzzy AHP [22]	Graduate program	Institution profile; Financial factors; Academic requirements; Support resources	Questionnaire	Top-1 accuracy: 0.902; Top-2 accuracy: 1.0; Precision: 0.319; Recall: 0.103; F1-score: 0.155
Feature selection; Fuzzy logic [28]	Job	NCAE scores; Cutoff scores; Grades; Interests	Questionnaire	MAE: 0.9857; RMSE: 3.861

3. LLM-Based Dataset Construction

3.1. LLM-Based Data Generation

Through a review of related work, this study observes that existing datasets are predominantly sourced from questionnaire surveys or institutional databases. While this ensures high data authenticity for model training and testing, it nevertheless introduces several inherent challenges:

- The collection of survey data is resource-intensive, requiring significant effort and cost. Moreover, the number and selection scope of survey subjects are often limited, making their representativeness difficult to measure.
- Survey data often suffers from various biases such as subjectivity bias and memory bias, which are difficult to correct artificially and will cause unpredictable interference to experiments.
- Existing studies consider various features, all helpful for career recommendation. However, it is infeasible to construct a comprehensive dataset integrating multiple survey dimensions by merging resources from various sources.

Given the advantages of large language models in efficiency and consistency, recent research [29,30] has validated the consistency between LLM-simulated groups and aggregated results from human experiments, proposing the use of LLMs as an alternative to traditional survey-based data collection methods. Therefore, this paper attempts to use LLMs instead of real surveys to construct a dataset, aiming to reduce data collection difficulty and various biases, addressing the three major problems mentioned above. The dataset used during the modeling phase in this paper is generated by the large language model DeepSeek-V3 simulating participants; only during the testing phase are real survey data used to evaluate the system trained on LLM-simulated samples.

Integrating multidimensional influencing factors from related work [3,4,6], the personal features selected for the dataset include: personality, interests, subject literacy, and skills. Personality and interests are categorized as the “preference” dimension, while subject literacy and skills are categorized as the “capability” dimension. This paper designs a questionnaire suitable for both LLM interaction scenarios and real surveys, covering the above personal features. Respondents are guided to provide authentic evaluations of their traits in these aspects. Their responses are expressed via a five-point Likert scale, i.e., “0: Very Disagree/Very Low, 1: Somewhat Disagree/Somewhat Low, 2: Neutral, 3: Somewhat Agree/Somewhat High, 4: Very Agree/Very High”. Table 2 shows some example questions from the dataset.

Table 2. Questionnaire design example.

Section *	Question	Scale
Personality (1)	Your MBTI?	16-type
Interests (9)	Do you enjoy exploring the principles behind how things develop?	0–4
	Are you interested in the mysteries behind life?	0–4
	Are you willing to delve into laws and institutions?	0–4
	... *	...
Subject literacy (16)	Rate your mathematics performance.	0–4
	How good are you at deriving a mathematical theorem?	0–4
	How good are you at understanding the underlying mechanism of a phenomenon?	0–4
	...	...
Skills (5)	Are you skilled at public speaking and expression?	0–4
	Do you have talent in drawing, painting, or handcrafting?	0–4
	Are you knowledgeable about programming or using various software tools?	0–4
	...	...

* Numbers in parentheses indicate total items in each section. The ellipses (...) indicate that only representative examples are shown for each section.

DeepSeek-V3 is used to simulate university student participants from 25 classic majors, completing the questionnaire survey based on their respective identities. The generation is structured in batches, specifying a target quantity of samples for a designated major per batch (e.g., 20 Psychology students). We utilize the model’s default sampling parameters (temperature = 0.7), leveraging the inherent sampling stochasticity of the LLM to generate diverse individual profiles. For a given batch, a shared persona constraint is provided, e.g., “You are a university student majoring in Psychology and confident that you are suited for this major. Please complete a questionnaire survey based on this role”. As the simulation transitions between different sections of the questionnaire, a guiding prompt is inserted: “This section aims to understand your interests. Please continue answering the following questions.” DeepSeek-V3 evaluates the extent to which each description conforms to the

simulated persona using a five-point Likert scale, ultimately reporting a complete set of individualized samples.

3.2. Data Preprocessing

Since the synthetic data originates from LLMs simulating human responses, its quality is susceptible to the model's inherent biases [31]. To ensure the reliability of subsequent analyses, a multi-stage preprocessing pipeline is implemented to filter for high-quality, diverse, and statistically realistic samples. The specific steps include:

1. Removal of duplicate entries. Identical responses may occasionally arise during batch generation under fixed prompts. A “duplicate” is strictly defined as a sample exhibiting an exact match across all features. These redundant entries are removed to ensure data uniqueness and independence.
2. Balancing categorical distributions. To avoid severe group bias, the proportion of each major was controlled within 3–5% ($4 \pm 1\%$), and each MBTI type within 4.75–7.75% ($6.25 \pm 1.5\%$). For any underrepresented categories falling outside these ranges, supplementary sampling was required.
3. Filtering invalid responses. Drawing on survey methodology for identifying careless responding [32,33], unrealistic samples were filtered based on conservative thresholds for the mean and standard deviation of numerical features (on a 0–4 scale). These thresholds were set to capture extreme outliers without artificially clipping the natural variance:
 - Reasonable central tendency: Because the study measures positive traits (e.g., interests and skills), a realistic respondent's overall score should fall within a moderately positive range. Samples with a mean below 1.0 or above 3.5 were removed, as they represent extreme negativity (mostly scoring 0) or extreme positivity (mostly scoring 4) that deviates from normative human psychological profiles.
 - Response variation: Conscientious respondents differentiate their answers based on item content. Samples with a standard deviation ≤ 0.5 were discarded. Mathematically on a 0–4 scale, a standard deviation this low indicates severe straight-lining (i.e., uniform responses across almost all items), which lacks informational value and deflates sample diversity.

The retention rates were tracked throughout the data preprocessing pipeline. Out of the initial pool of generated samples, approximately 8% were removed as exact duplicates, and around 2% were filtered out due to statistical anomalies based on the aforementioned thresholds. This pipeline ultimately retained approximately 90% of the raw data, yielding the dataset of valid samples used for subsequent modeling.

3.3. Data Fidelity Assessment

To further verify the reliability of the simulated dataset, a distributional comparison against the real-world sample ($n = 24$) was conducted, following the synthetic data evaluation framework proposed by Liu et al. [34]. Given the small real sample size, which compromises the reliability of traditional parametric tests, a set of robust analytical metrics was adopted to assess data fidelity from three perspectives: marginal distributions, response variability, and multivariate structure. The results are summarized in Table 3.

For categorical variables, the Jensen–Shannon divergences are 0.410 and 0.347, respectively. These differences are largely due to the simulated data being balanced across categories by design, whereas the real sample is uneven due to its limited size. For numerical features (on a 0–4 scale), the average Wasserstein distance between the simulated and real data is 0.420 (per-dimension range: 0.341–0.463), indicating that the simulated marginal distributions remain within a reasonable range of real-world responses. The

simulated-to-real ratios of mean item-level standard deviations across dimensions range from 0.812 to 0.963 (overall 0.836), confirming that the generation process did not collapse to a single mean response. The ratios below 1.0 suggest that the simulated data are slightly more conservative in their dispersion compared to the real responses.

Table 3. Distributional fidelity assessment of the simulated dataset against real-world sample.

Metric	Interests	Subject Literacy	Skills	Overall
JS divergence (Major)		–		0.410
JS divergence (MBTI)				0.347
Wasserstein distance	0.463	0.420	0.341	0.420
Spearman correlation MAE	0.220 *	0.293 *	0.256 *	0.283 *
Std deviation ratio (Sim/Real)	0.812	0.815	0.963	0.836

* $p < 0.01$ (Significantly outside the 99% CI of the Monte Carlo random baseline).

To examine whether the simulated data preserves feature interdependencies, the mean absolute error (MAE) of the Spearman correlation matrices between the simulated and real data was computed. The overall MAE is 0.283, with per-dimension values ranging from 0.220 to 0.293. To verify that the observed correlations are not random artifacts, a 100-run Monte Carlo random baseline was constructed. At the overall level, the simulated MAE (0.283) falls outside the 99% confidence interval of [0.274, 0.281] for the null baseline. Separate tests conducted within each feature dimension also independently reject the null hypothesis of random generation ($p < 0.01$).

Dimension-wise analyses reveal two distinct patterns. For interests and skills, the simulated MAEs fall below the lower bounds of their respective null baselines, suggesting a reasonable approximation of the psychological and logical patterns observed in the real sample. For subject literacy, the MAE exceeds the null upper bound, indicating stronger structural priors than those observed in the small real-world sample. This is likely due to the LLM injecting population-level associative rules acquired during pre-training, which diverge from the incidental noise inherent in the limited real sample.

These results demonstrate that, despite observable deviations from the small real-world sample caused by strong pre-training priors, the simulated data exhibit highly systematic and structured features rather than random noise. The practical utility of this dataset is further corroborated by the downstream evaluation reported in Section 5.

4. Framework and Methods

The proposed framework, as shown in Figure 1, is divided into three dimensions: preference, capability, and user experience, which are aggregated via an adjustable AHP module. Preference is composed of MBTI personality and interests; capability is composed of subject literacy and skills. They are jointly processed by a fuzzified XGBoost module. The user experience part uses the above features as user profiles, processed by a user-based collaborative filtering module. The two modules produce scores respectively, which are then aggregated into the final recommendation utility scores. This design enables the framework to synthesize individual-level user profiling and group-level behavioral experience within a single architecture, while remaining free from predefined rules or structures that limit the flexibility and extensibility. A summary comparison of these characteristics with the related methods is summarized in Table 4.

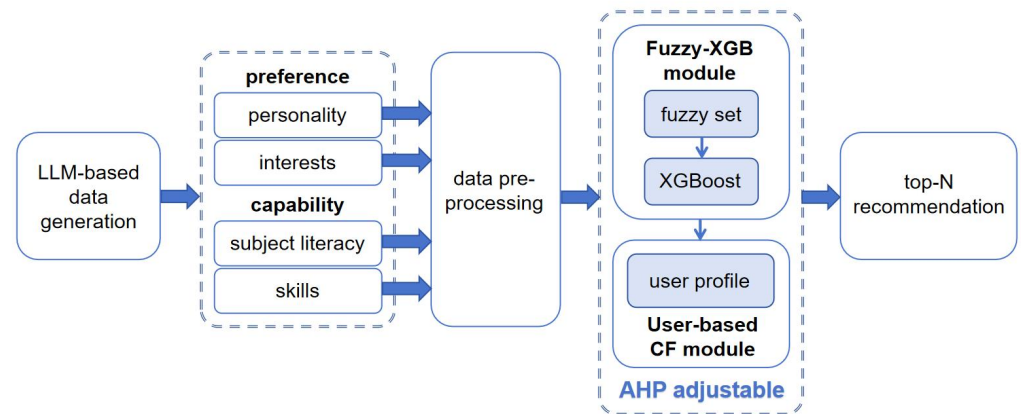


Figure 1. Framework overview.

Table 4. Comparison of key design characteristics with representative existing methods.

Design Dimension	CF	Fuzzy Logic	AHP	XGBoost	Proposed
User attributes modeling	× *	✓ *	×	✓	✓
Group behavior exploitation	✓	×	×	×	✓
No predefined structures	✓	× (requires rules)	× (requires hierarchy)	✓	✓
Handling uncertainty	×	✓	×	×	✓
Extension difficulty	High (cold-start)	High (rule redesign)	Medium (hierarchy rebuild)	Low (retraining)	Low (retraining)

* ✓ indicates that the corresponding dimension is considered in the method; × indicates that it is not.

4.1. Fuzzy XGBoost Module

To capture the nonlinear interaction between user characteristics in major suitability, this paper proposes a joint modeling module based on fuzzy sets and XGBoost (Fuzzy XGBoost, fXGB). This module performs soft partitioning of discrete variables using trapezoidal membership functions, then trains XGBoost on the fuzzified feature space, ultimately outputting a probability utility matrix for the 25 majors for a new user: $\mathbf{S}_{\text{fXGB}} \in [0, 1]^{25}$.

For categorical variables, namely the major and MBTI, integer encoding is applied, mapping the 25 majors to 0–24 and the 16 MBTI personalities to 0–15 respectively.

For numerical discrete variables, ratings for interests, subject ability, and skills undergo trapezoidal fuzzification. The fuzzification process converts input variables into appropriate values in the universe of discourse, using fuzzy qualitative variables to describe measured physical quantities [4].

For each numerical feature $x_k \in \{0, 1, 2, 3, 4\}$ ($k = 0, 1, \dots, K - 1$, where K is the total number of numerical items), define the trapezoidal membership function (as shown in Figure 2):

$$\mu(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & b < x \leq c, \\ \frac{d-x}{d-c}, & c < x < d, \\ 0, & x \geq d. \end{cases} \quad (1)$$

where $x = x_k$, and a, b, c, d are membership function parameters.

For the questionnaire designed in this paper, three linguistic values (low, medium, and high) are used to represent each numerical item. Accordingly, each numerical response is transformed into a three-dimensional membership degree vector:

$$\tilde{x}_k = [\mu_{\text{low}}^{(k)}(x) \quad \mu_{\text{medium}}^{(k)}(x) \quad \mu_{\text{high}}^{(k)}(x)] \in [0, 1]^3. \quad (2)$$

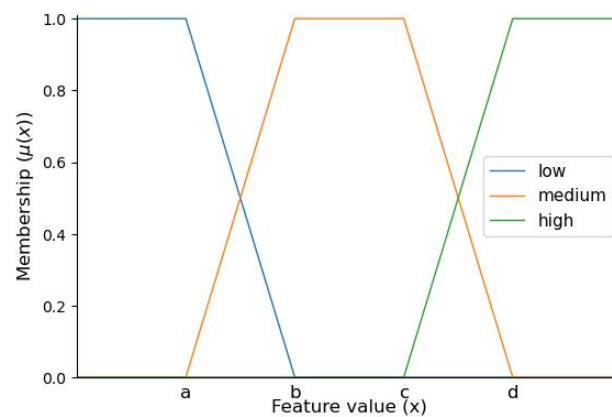


Figure 2. Schematic diagram of trapezoidal membership function.

Let the training set be:

$$\mathcal{D}_{\text{tr}} = \{(\mathbf{z}_i, y_i)\}_{i=1}^N, \mathbf{z}_i = [x_{\text{mbti}}^{(i)}, \tilde{\mathbf{x}}_0^{(i)}, \tilde{\mathbf{x}}_1^{(i)}, \dots, \tilde{\mathbf{x}}_{K-1}^{(i)}], \quad (3)$$

where y_i represents the label for major, and N is the number of samples.

It is noteworthy that, to avoid erroneous assumptions, each user only forms a positive sample pair with one truly chosen major; the remaining 24 majors do not explicitly appear in the current sample. This better aligns with the real scenario of college entrance exam major application. The training objective is to predict, given user features, the utility value for each of the 25 majors. This formulation formalizes major recommendation as a one-to-many multi-class probability prediction task.

XGBoost's multi:softprob objective [23] aligns with this training goal. It can output a multidimensional softmax probability vector in a single training run, eliminating the need to train binary classifiers for each class separately, thereby significantly reducing training complexity. Furthermore, softprob yields normalized probabilities, avoiding issues like class imbalance and false negatives associated with hard classification.

The multi-class softprob objective:

$$\mathcal{L}(\Theta) = \sum_{i=1}^N \ell(y_i, \hat{\mathbf{p}}_i) + \sum_t \Omega(f_t), \quad (4)$$

where $\hat{\mathbf{p}}_i \in [0, 1]^{25}$ is the probability vector, ℓ is the differentiable loss function, f_t represents the t -th independent regression tree, and $\Omega(f_t)$ is the regularization term penalizing tree complexity. For a new user u_{new} , the model outputs:

$$\mathbf{S}_{\text{fXGB}}(u_{\text{new}}) = [\hat{p}_0 \quad \hat{p}_1 \quad \cdots \quad \hat{p}_{24}], \quad \sum_{j=0}^{24} \hat{p}_j = 1. \quad (5)$$

This vector is the probabilistic score from the fXGB module for the 25 majors, used for subsequent fusion and ranking. Theoretically, the inference complexity for a new user is $O(K + d \cdot T)$, where d is the maximum tree depth and T is the number of trees, contributing a negligible constant-time operation.

4.2. Collaborative Filtering Module

To further capture potential associations between user profiles and suitable majors, this paper incorporates user experience by constructing a user-based collaborative filtering (CF) module [35] based on the cosine similarity of user profiles. Its core idea is that users with similar personal characteristics will have similar suitability for majors. This module centers on users' real major choice behaviors, combined with their various features, to achieve personalized ranking of candidate majors.

The user profile concatenates all preference and capability features. MBTI is encoded as a four-dimensional binary vector $\mathbf{x}_{\text{mbti}} \in \{0, 1\}^4$, where each bit represents one dichotomy. All other features retain their original integer values x_k .

To enhance the recommendation gain of the CF module, the top 5 items output by the fXGB module are introduced to construct the feature "major tendency". To reflect the similarity relationships between candidate majors, a major tendency mapping based on circular vector synthesis is employed. Specifically, the indices assigned to the 25 majors are structured as a continuous spectrum based on disciplinary proximity. The sequence progresses from engineering and formal sciences, through natural and social sciences, to the humanities, and then bridges back to technical domains via interdisciplinary applied fields, forming a closed loop. Under this semantic mapping, adjacent indices largely represent related fields. Consequently, circular distance and Gaussian smoothing are applied to appropriately capture this proximity without introducing boundary effects. The calculation process is shown in Algorithm 1.

Algorithm 1 Major Tendency Feature Calculation

Require: User's score vector $\mathbf{S}$ from fXGB module; Number of top majors n ; Gaussian bandwidth σ

Ensure: Major tendency feature x_{td}

```

1:  $\mathcal{C} \leftarrow \text{argsort}(\mathbf{S})_{\text{desc}}[1 : n]$  ▷ Top- $n$  majors
2:  $\mathbf{V} \leftarrow \mathbf{0} \in \mathbb{R}^{25}$ 
3: for each  $c \in \mathcal{C}$  do
4:   for each  $k \in \{0, \dots, 24\}$  do
5:      $d_{k,c} \leftarrow \min(|k - c|, 25 - |k - c|)$  ▷ Circular distance
6:      $V_k \leftarrow V_k + \exp(-d_{k,c}^2 / 2\sigma^2)$  ▷ Gaussian smoothing
7:   end for
8: end for
9: Compute the mean directional angle  $\theta$  of the circular centroid weighted by  $\mathbf{V}$ 
10:  $x_{\text{td}} \leftarrow \text{Map } \theta \text{ to construct the major tendency feature}$ 
11: return  $x_{\text{td}}$ 
```

The complete profile of user u is represented as the concatenated vector:

$$\mathbf{f}_u = [\mathbf{f}_{\text{pref}}^{(u)}, \mathbf{f}_{\text{cap}}^{(u)}, x_{\text{td}}^{(u)}] = [x_{\text{mbti}}^{(u)}, x_0^{(u)}, x_1^{(u)}, \dots, x_{K-1}^{(u)}, x_{\text{td}}^{(u)}], \quad (6)$$

where $\mathbf{f}_{\text{pref}}$ and $\mathbf{f}_{\text{cap}}$ separately represent features of user preference and capability (encompassing the processed MBTI and K initial numeric features), and $x_{\text{td}}^{(u)}$ represents the derived major tendency feature.

Similarity between users is calculated via cosine similarity, incorporating weight parameters w_{pref} and w_{cap} (see Equation (12)). The weighted user profile is:

$$\mathbf{f}'_u = [w_{\text{pref}} \cdot \mathbf{f}_{\text{pref}}^{(u)}, w_{\text{cap}} \cdot \mathbf{f}_{\text{cap}}^{(u)}, x_{\text{td}}^{(u)}], \quad (7)$$

For any two users u and v , their similarity is:

$$\text{sim}(u, v) = \frac{\mathbf{f}'_u \cdot \mathbf{f}'_v}{\|\mathbf{f}'_u\| \|\mathbf{f}'_v\|}. \quad (8)$$

Define a user-major rating matrix $\mathbf{R} \in \{0, 1\}^{|U_{\text{old}}| \times |C|}$, where U_{old} is the set of referenced historical users, C is the set of candidate majors, and $R_{u,c} = 1$ indicates user $u \in U_{\text{old}}$ has truly chosen and is satisfied with major c . For a new user u_{new} , the collaborative filtering utility value for a candidate major c_i is calculated via weighted aggregation:

$$\text{CF}(u_{\text{new}}, c_i) = \frac{1}{|U_{\text{old}}|} \sum_{u \in U_{\text{old}}} \text{sim}(u_{\text{new}}, u) \cdot R_{u,c_i}. \quad (9)$$

To enhance the discriminability of utility values and transform them into a probability distribution, inspired by research [36,37], this paper adopts the idea of temperature-based exponential scaling. An exponential scaling parameter α_{exp} (generally ≥ 1) is introduced to process the utility values:

$$s_{\text{cf}}(u_{\text{new}}, c_i) = \frac{\exp(\alpha_{\text{exp}} \cdot \text{CF}(u_{\text{new}}, c_i))}{\sum_{j=0}^{|C|-1} \exp(\alpha_{\text{exp}} \cdot \text{CF}(u_{\text{new}}, c_j))}. \quad (10)$$

Finally, the scoring vector from the collaborative filtering module for the 25 majors is output:

$$\mathbf{S}_{\text{cf}}(u_{\text{new}}) = [s_{\text{cf}}(u_{\text{new}}, c_0) \quad s_{\text{cf}}(u_{\text{new}}, c_1) \quad \cdots \quad s_{\text{cf}}(u_{\text{new}}, c_{24})]. \quad (11)$$

The inference complexity of this module is $O(|U_{\text{old}}| \cdot K)$.

4.3. Analytic Hierarchy Process Weight Aggregation

To scientifically determine the relative contribution of key dimensions and aggregate the fXGB and CF modules, this paper employs the Analytic Hierarchy Process (AHP) [20,21]. In this framework, AHP provides weight configuration for two core information aggregation steps:

- In collaborative filtering, determining the relative weight of user “preference” and “capability” when calculating user similarity;
- In the model fusion stage, determining the relative weight of the fXGB module and the CF module when generating the final recommendation list.

In the analytic hierarchy process, the core of decision-making lies in constructing the judgment matrix $\mathbf{A} = (a_{ij})_{m \times m}$, where $a_{ji} = 1/a_{ij}$ and m is the number of criteria

(here, $m = 3$). This matrix quantifies the relative importance of each criterion through pairwise comparisons. By calculating the maximum eigenvalue $\lambda_{\max}$ of matrix $\mathbf{A}$ and its corresponding eigenvector, the normalized weight vector $\mathbf{W} = [W_1, W_2, W_3]^T$ is obtained. To ensure consistency in the judgment logic, the consistency index $CI = \frac{\lambda_{\max} - m}{m - 1}$ is calculated. Combined with the random consistency index RI, the consistency ratio $CR = CI/RI$ is determined. When $CR < 0.10$, the consistency of the judgment matrix is considered acceptable.

To ensure that the feature weights have clear comparative meaning, truncation processing is applied to the calculated weight ratios. The relative weights for preference and capability are defined as:

$$w_{\text{pref}} = \max\left(1, \frac{W_1}{W_2}\right), w_{\text{cap}} = \max\left(1, \frac{W_2}{W_1}\right), \quad (12)$$

where the lower bound of 1 prevents any single dimension from being excessively diluted. The resulting weights are then passed to the CF module for user rating matrix computation. These parameters are used in Equation (7) to compute the weighted user profile.

For the aggregation of utility values produced by the fXGB module and the CF module, the relative weights of the two modules are calculated as:

$$w_{\text{fXGB}} = \max\left(1, \frac{W_1 + W_2}{2W_3}\right), w_{\text{cf}} = \max\left(1, \frac{2W_3}{W_1 + W_2}\right), \quad (13)$$

with the lower bound of 1. The aggregated score matrix is represented as:

$$\mathbf{S}_{\text{agg}}(u_{\text{new}}) = w_{\text{fXGB}} \cdot \mathbf{S}_{\text{fXGB}}(u_{\text{new}}) + w_{\text{cf}} \cdot \mathbf{S}_{\text{cf}}(u_{\text{new}}). \quad (14)$$

Accordingly, the system generates a Top- N recommendation list for the new user by selecting the N items with the highest scores.

4.4. Implementation Steps

1. Generate a specified number of simulated questionnaire responses using an LLM with the role-prompting strategy.
2. Perform data preprocessing including duplicate removal, category balancing, and statistical filtering following Section 3.2.
3. Input: User features from the processed dataset. Processing: Train the Fuzzy XGBoost module to learn a mapping from preference and capability features to major predictions (see Section 4.1). Compute the probability score for each candidate major for a given user. Output: Individual-level score matrix $\mathbf{S}_{\text{fXGB}}$.
4. Input: User profiles and AHP-derived weights. Processing: Apply user-based collaborative filtering to capture group-level behavioral experience (see Section 4.2). Compute the utility score for each candidate major. Output: Group-level score matrix $\mathbf{S}_{\text{cf}}$.
5. Input: Score matrices $\mathbf{S}_{\text{fXGB}}$, $\mathbf{S}_{\text{cf}}$, and AHP-derived weights. Processing: Aggregate the two score matrices using AHP fusion weights (see Section 4.3). Generate a ranked recommendation list for each user. Output: Top- N recommendation list.

5. Experiments

To evaluate the framework's performance, this paper used the method described in Section 3 to generate the SimuCareer-25 dataset containing approximately 600 samples (this number can be further expanded). To ensure the robustness of the evaluation, the dataset was repeatedly split into training and test sets in an 8:2 ratio across five independent random splits to simulate historical and new users.

Following Herlocker et al. [38] and the related work summarized in Table 1, this paper adopted Top- N accuracy and its AUC to evaluate the quality of the recommended lists. Top- N accuracy reflects the percentage of correct recommendations appearing within the top N positions of the list, where a recommendation is considered correct if the student's actual major is included in the recommended list. It is formally defined as:

$$\text{Top-}N \text{ Accuracy} = \frac{\left| \left\{ u \in U_{\text{new}} \mid x_{\text{career}}^{(u)} \in \text{argsort}(\mathbf{S}_{\text{agg}}(u))_{1:N} \right\} \right|}{|U_{\text{new}}|}. \quad (15)$$

Considering the practicality of recommendations in real-world scenarios, accuracy at larger N values measures whether a system reaches a usable level, while accuracy at smaller N values reflects whether it can be considered outstanding, as it determines the system's ability to make precise and high-confidence recommendations. In particular, Top-1 accuracy is the strictest metric, as it requires the system to identify the single most suitable item for the user.

The Top- N accuracy values for different N values were concatenated into a curve, and the area under the curve (AUC) was then calculated. This metric summarizes the model's ranking effectiveness across all Top- N positions.

In addition to the metrics used for the simulated dataset, the proposed framework was also evaluated on real-world data where both positive and negative samples were available, allowing for a binary classification assessment. This ensures that the system can not only identify recommendable majors but also rank unsuitable ones lower. Accordingly, Precision, Recall, and F1-score [38] were adopted as supplementary metrics to evaluate the model's discriminative ability. These metrics complement the ranking-oriented evaluation by providing a more fine-grained view of classification performance, and their calculation follows the evaluation practices of studies listed in Table 1.

5.1. Model Comparison

First, to validate the effectiveness of the proposed framework, this paper selected several classical methods as baselines, collectively demonstrating the efficacy of aggregating different modules.

- **Naïve Bayes Classifier (Bayes):** Outputs a preference probability distribution by calculating the posterior probability of an item given user features. This method has been proven simple yet effective in personalized ranking tasks [39]. To ensure a fair comparison, this baseline utilizes the complete set of user features in their raw, unfuzzified form.
- **Collaborative Filtering Model (CF):** Utilizes only the scores generated by the user-based CF module. As a classic method in general recommendation systems, collaborative filtering offers a valuable perspective by discovering collective wisdom solely through choices made by past user groups.
- **XGBoost Model (XGB):** Directly applies the XGBoost method without fuzzification. As a representative state-of-the-art method that has shown strong performance in educational prediction tasks [14], XGBoost serves as a competitive baseline for handling the multidimensional non-linear interactions inherent in user-major compatibility.
- **Fuzzy XGBoost Model (Fuzzy-XGB):** Utilizes only fuzzified scores from the fXGB module to investigate whether uncertainty handling enhances system-user cognitive alignment. This approach fuzzifies numerical features using three linguistic values: low, medium, and high. Based on scale range and semantic continuity requirements, the trapezoidal membership function parameters in Equation (1) are set as $a = 1$, $b = 2$, $c = 2$, $d = 4$. This ensures continuous coverage within the value range and aligns with the semantic hierarchy of the questionnaire design.

As shown in Figure 3 and Table 5, the evaluation on the SimuCareer-25 dataset reveals a performance progression among the models. The Bayes and CF models yield relatively lower comprehensive metrics, with mean AUC values of 0.473 ± 0.018 and 0.490 ± 0.028 , respectively. While CF performs slightly better than Bayes in the top one to five recommendations, it falls behind at Top-7 and Top-10. This highlights the limitations of relying solely on collaborative filtering or probabilistic baselines in major selection scenarios without sufficient user–item interaction density.

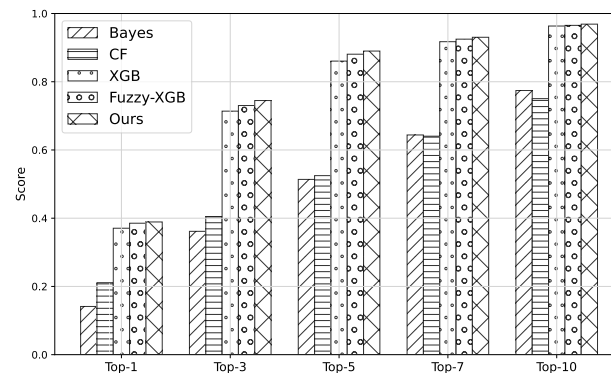


Figure 3. Top- N accuracy comparison on SimuCareer-25 (mean values across 5 random splits).

Table 5. AUC comparison on SimuCareer-25.

Model	Bayes	CF	XGB	Fuzzy-XGB	Ours
AUC *	0.473 ± 0.018	0.490 ± 0.028	0.744 ± 0.008	0.756 ± 0.034	0.763 ± 0.033

* AUC is defined as the area under the Top- N accuracy curve. Mean $\pm$ standard deviation over 5 random splits.

The XGB model provides a powerful solution for high-dimensional user features, significantly boosting the mean AUC to 0.744. Building upon this, the Fuzzy-XGB model demonstrates the value of fuzzification in addressing the inherent uncertainty of user self-assessment. It yields a modest but consistent improvement and raises the mean AUC to 0.756.

Ultimately, the proposed integrated framework (Ours) achieves the most favorable overall ranking quality. By fusing individual-level fuzzy inference with group-level collaborative filtering, it maintains a slight advantage over Fuzzy-XGB at every list length and attains the highest mean AUC (0.763 ± 0.033). Given that samples only have a single positive label, achieving a mean Top-3 accuracy of approximately 0.74 is an encouraging outcome for this recommendation task. The computational cost of the proposed framework is dominated by the CF module ($O(|U_{old}| \cdot K)$). At the current data scale, the average inference time per sample is within one second.

5.2. Exponential Scaling Parameter Optimization

To observe the effect of exponential scaling, the parameter α_{exp} for the CF module (see Equation (10)) was traversed with a step size of 1 (0 to 20), using a trained instance of the framework. Figure 4 plots the change curves of Top- N accuracy as α_{exp} increases.

Overall, α_{exp} does not significantly affect accuracy. This is because the fXGB module can independently provide a recommendation list with reasonably good accuracy. On this basis, α_{exp} determines the degree to which the CF module's score distribution is sharpened or smoothed, thereby affecting the aggregated score ranking. When $\alpha_{exp} = 0$, CF scores are pulled to the same level, and the Top- N list is completely dominated by the fXGB module. As α_{exp} increases, the distribution of CF scores becomes increasingly steep.

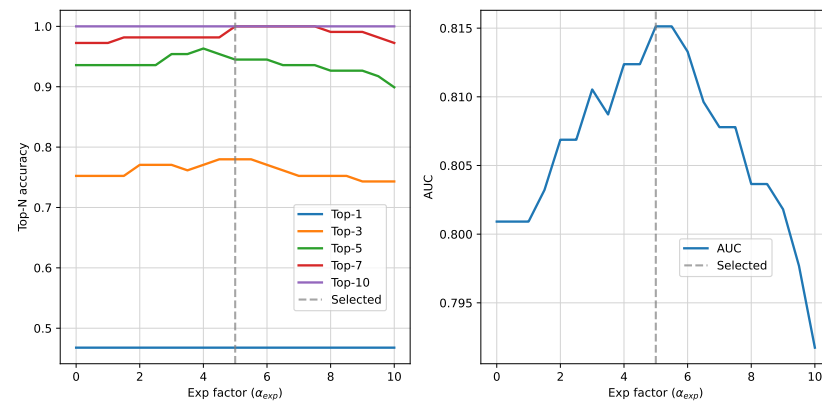


Figure 4. Effect of exponential scaling parameter α_{exp} on model performance.

While the fXGB module establishes an independent baseline, α_{exp} acts as a fine-tuning parameter by determining the sharpness of the CF module's score distribution. When α_{exp} approaches 0, the CF scores flatten toward a uniform distribution, neutralizing the collaborative filtering contribution and leaving the final ranking primarily determined by the fXGB module. Conversely, as α_{exp} increases, the distribution of CF scores becomes increasingly steep. As shown in Figure 4, as α_{exp} increases from 0 to 10, the Top-3, Top-5, and Top-7 accuracies, along with the AUC, initially rise and then decline, with the highest observed AUC occurring around $\alpha_{\text{exp}} = 5$.

Figure 5 further illustrates this behavior through the CF output score sequences for two example users under different α_{exp} conditions. A higher α_{exp} induces a “winner-takes-all” effect, which can suppress potentially valid secondary signals. In contrast, $\alpha_{\text{exp}} = 5$ sharpens the distribution moderately, widening the margin between top candidates and irrelevant options without completely excluding secondary choices. Consequently, this balanced distribution appears to complement the fXGB predictions well. Based on these observations, subsequent experiments uniformly adopt $\alpha_{\text{exp}} = 5$.

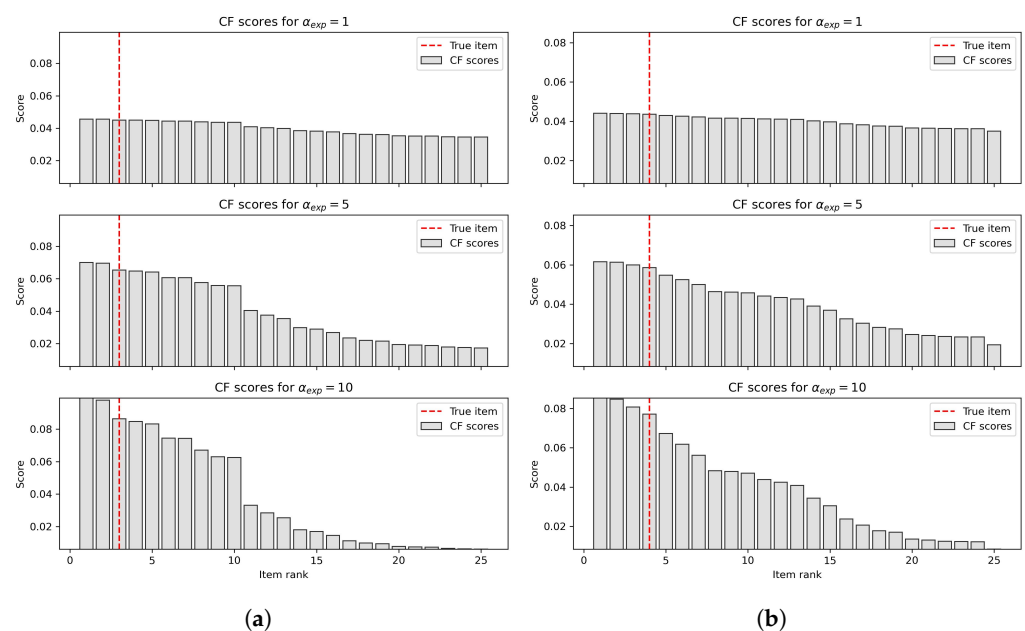


Figure 5. Example of collaborative filtering module output score sequences under different α_{exp} values: (a) output for user A; (b) output for user B.

5.3. Weight Optimization

To scientifically determine the weights of various dimensional components in the recommendation framework, this paper adopted a research strategy combining theory-guided and data-driven approaches. A decision framework with “preference”, “capability”, and “experience” as core criteria was established in Section 4. The AHP pairwise comparison matrix was set based on the following rationale: preference is weighed higher than capability ($a_{12} = 1.5$), aligning with the educational psychology consensus that interest is a key predictor of long-term academic engagement [40]. From a low-risk expectation perspective, the experience dimension was assigned a slightly lower weight than preference ($a_{13} = 1.35$), while capability and experience were considered equally important ($a_{23} = 1$). The calculated weight vector was $\omega = [0.42, 0.29, 0.30]^T$, with a consistency ratio (CR) of 0.001, demonstrating high logical consistency well below the 0.10 threshold.

To systematically investigate the impact of different dimensional component weights on recommendation performance and verify the effectiveness of the AHP adjustment mechanism, a grid search experiment was conducted. Figure 6 shows a heatmap of the recommendation model’s AUC performance under different weight proportion combinations.

The experiment focused on two sets of weight parameters: w_{pref} and w_{cap} , which jointly govern the relative significance of preference and capability in calculating user similarity (see Equation (12)), while w_{fXGB} and w_{cf} collectively determine the relative contribution of the fXGB module and the CF module to the final recommendation score (see Equation (13)). As illustrated in Figure 6, the horizontal axis represents the preference weight proportion, i.e., $w_{\text{pref}} / (w_{\text{pref}} + w_{\text{cap}})$, and the vertical axis represents the fXGB weight proportion, i.e., $w_{\text{fXGB}} / (w_{\text{fXGB}} + w_{\text{cf}})$. Fifteen uniformly distributed points were selected for each weight proportion. Performance across all weight combinations was evaluated using the AUC of Top- N accuracy, providing an integrated assessment across multiple list lengths.

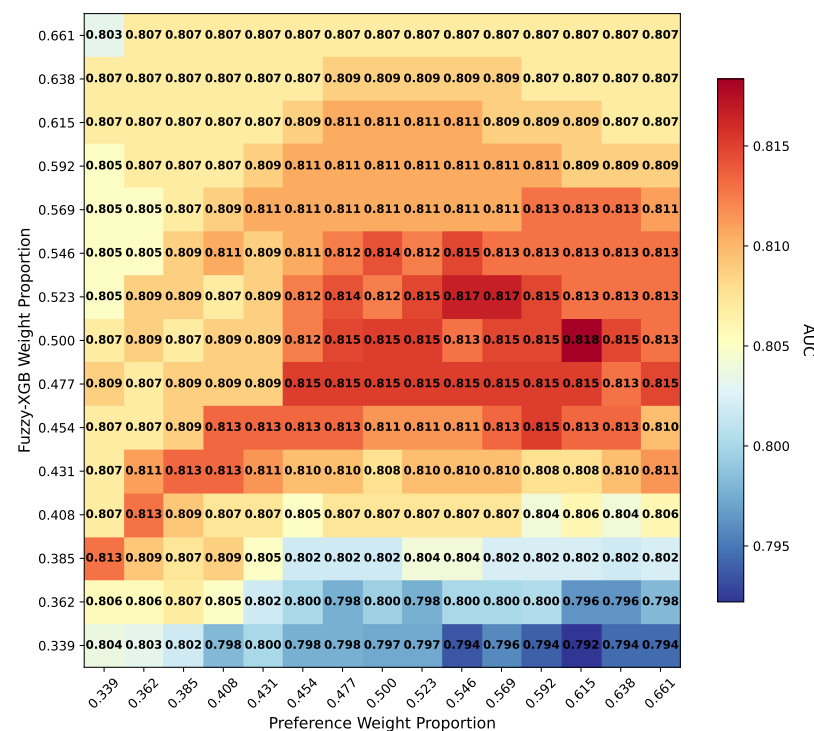


Figure 6. Heatmap of model AUC under different weight proportions.

The heatmap clearly reveals the sensitivity of model performance to weight configuration. Regions with higher AUC (red areas in Figure 6) form a distinct “plateau” primarily

located towards the right-center. This indicates that extreme weight configurations and over-reliance on a single component are detrimental to performance. The optimal performance appears near a preference weight proportion of 0.615 and an fXGB weight proportion of 0.500, i.e., $w_{\text{pref}} = 1.6$, $w_{\text{cap}} = 1$, $w_{\text{fXGB}} = 1$, $w_{\text{cf}} = 1$. This configuration suggests two key insights into the recommendation mechanism: first, within the CF module, preferences appear to be slightly stronger determinants of user similarity than capability scores; second, individual-level feature inference and group-level behavioral consensus provide complementary signals to some extent, which should be balanced comparably. The weights determined via the AHP method (preference proportion 0.60, fXGB proportion 0.54) are reasonably close to this empirical optimal region, supporting the practical effectiveness of the theory-guided AHP adjustment.

5.4. Evaluation on Real-World Data

To further evaluate the real-world applicability of the system trained on LLM-simulated data, this paper recruited a group of current university students from a comprehensive university in China to participate in an online questionnaire-based validation study. All participants were in their third year or above (including both undergraduate and graduate students) and thus had already declared and settled into their current major. They were fully informed of the research purpose and procedure prior to participation and provided their consent. The recruited students represent a diverse range of majors spanning engineering, science, and humanities disciplines, corresponding to the major labels in SimuCareer-25. A total of 24 valid samples were obtained, with 18 being positive instances based on self-rated major suitability (“somewhat suitable” or “very suitable”). The proposed system and control model groups were tested using the positive samples. The aggregation weights for the proposed system were set to the parameters verified above.

According to Figure 7 and Table 6, discrepancies between the simulated training data and real-world responses introduce several minor performance variations. Specifically, the CF model’s drop below Bayes at Top-7 and Top-10 is more pronounced here, reflecting its high sensitivity to such distributional shifts. Furthermore, a comparison between Fuzzy-XGB and standard XGB reveals a performance crossover on the real-world data. Fuzzy-XGB yields a noticeable advantage at Top-1 accuracy, yet it is slightly surpassed by XGB at Top-5 through Top-10, potentially indicating that the benefit of fuzzification lies in smoothing self-assessment noise to protect primary recommendations. Integrating these modules, the proposed framework yields a mean AUC of 0.759 ± 0.057 , exhibiting a marginal quantitative advantage over the baselines in this small-sample setting.

Table 6. AUC comparison on real-world data.

Model	Bayes	CF	XGB	Fuzzy-XGB	Ours
AUC *	0.436 ± 0.042	0.437 ± 0.053	0.730 ± 0.028	0.749 ± 0.062	0.759 ± 0.057

* Mean $\pm$ standard deviation results from models trained on 5 random splits of the simulated data.

To more rigorously evaluate the models’ discrimination ability, samples where users self-rated their current major as “neutral”, “somewhat unsuitable”, or “very unsuitable” were treated as negative candidates with unclear preference or low suitability. A binary classification evaluation was conducted, utilizing a trained and validated instance of the proposed framework and baseline models. Based on the preceding evaluation, $N = 7$ was adopted as the decision threshold to delineate the boundary, meaning each model provides seven items as positive recommendations, with the rest considered “not recommended”.

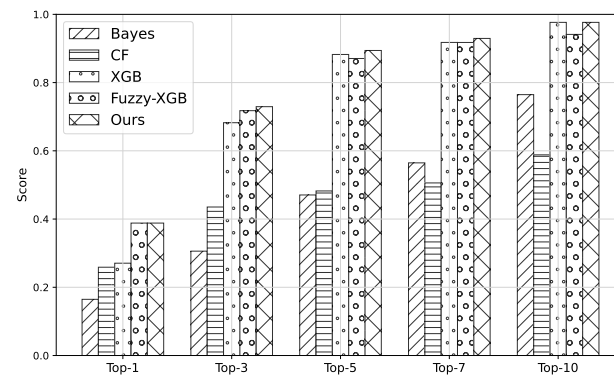


Figure 7. Top-*N* accuracy comparison on real-world data (mean values across 5 random splits).

As shown in Figure 8, evaluating these models using Precision, Recall, and F1-score [38] reveals that the proposed system performs favorably across these metrics. Complementing the data fidelity findings (Section 3.3), these downstream outcomes suggest that while the simulated training data exhibits observable structural deviations from the real-world sample, the systematic features captured by the LLM are transferable for practical utility.

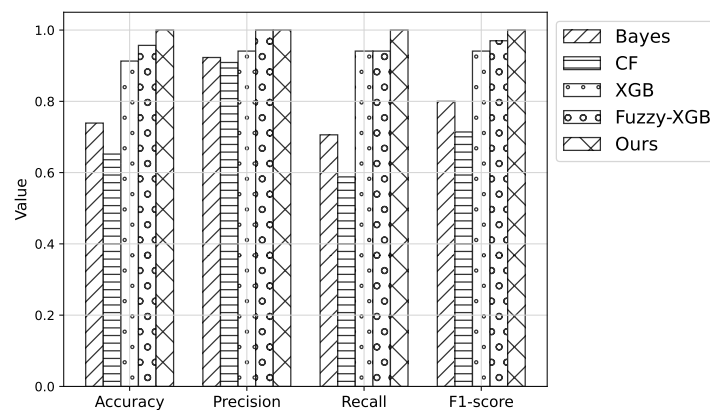


Figure 8. Comprehensive performance metrics on real-world data.

6. Conclusions

This paper proposes a recommendation system framework to assist university applicants in major selection. The framework integrates personal characteristics with group experience, forming a decision triad of “preference, capability, and experience.” This paper constructs a Fuzzy XGBoost module to process high-dimensional user features, applies collaborative filtering to mine group experience, and employs the Analytic Hierarchy Process (AHP) for adjustable weighted fusion of the module outputs. For dataset construction, it exploratorily uses an LLM to simulate human participants, generating training samples. Data fidelity assessments reveal that while the simulated data exhibits structural deviations from the small real-world sample, likely due to strong LLM priors, it successfully maintains systematic and structured features. Downstream real-world evaluations validate both the feature transferability of the simulated dataset and the practical utility of the recommendation framework. Additionally, experiments demonstrate that the framework’s core dimensions, including user preference, capability, and group experience, are all critical components for generating robust recommendations.

Several limitations and directions for future research remain. First, the reliability of the current method is constrained by biases inherent in self-reported data; future enhance-

ments could therefore involve designing psychologically informed questionnaires to better guide user responses. Second, to further complement the individual-level assessment, future work could incorporate broader situational contexts, such as family influence, career prospects, and potential salary. However, such factors should be treated with caution, as their introduction could undermine the core objective of fit-based recommendation. Finally, regarding dataset construction, this paper applied a unified role-prompting scheme on a single LLM to facilitate standardized batch generation. While our data fidelity assessment suggests that the inherent sampling stochasticity of the LLM naturally introduces meaningful inter-participant variance, future research should explore more sophisticated synthesis strategies. Specifically, utilizing multiple distinct LLMs and introducing fine-grained demographic personas or temperature variations could explicitly control and further enrich the diversity of the generated profiles. Investigating how these generation strategies systematically impact downstream recommendation performance remains a promising avenue for subsequent studies.

Author Contributions: Study design, data collection, experimentation, and manuscript writing, S.H.; supervision and manuscript review, Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Our study involved an anonymous questionnaire survey conducted through Wenjuanxing (a widely adopted online survey platform in China). As the survey was purely behavioral, non-interventional, involved no medical or biological samples, and collected no personally identifiable information from participants, formal ethics committee approval was not required under our institutional guidelines. Nevertheless, all participants were fully informed of the survey's purpose prior to participation, and participation was entirely voluntary. Accordingly, we do not have an ethics approval certificate or related documentation to provide.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yao, Z. Analyzing Chinese College Students' Satisfaction with Major Choices. In *Proceedings of the 2022 International Conference on Science Education and Art Appreciation (SEAA 2022)*; Atlantis Press: Dordrecht, The Netherlands, 2022; pp. 1396–1403. [[CrossRef](#)] [[PubMed](#)]
2. Ackerman, P.L.; Heggestad, E.D. Intelligence, Personality, and Interests: Evidence for Overlapping Traits. *Psychol. Bull.* **1997**, *121*, 219–245. [[CrossRef](#)] [[PubMed](#)]
3. Qamhie, M.; Sammaneh, H.; Demaidi, M.N. PCRS: Personalized Career-Path Recommender System for Engineering Students. *IEEE Access* **2020**, *8*, 214039–214049. [[CrossRef](#)]
4. Verma, P.; Sood, S.K.; Kalra, S. Student Career Path Recommendation in Engineering Stream Based on Three-Dimensional Model. *Comput. Appl. Eng. Educ.* **2017**, *25*, 578–593. [[CrossRef](#)]
5. Wang, C.; Wang, K.; Bian, A.Y.; Islam, R.; Keya, K.N.; Foulds, J.; Pan, S. When Biased Humans Meet Debiased AI: A Case Study in College Major Recommendation. *ACM Trans. Interact. Intell. Syst.* **2023**, *13*, 1–28. [[CrossRef](#)]
6. Dang, Y.; Liu, M.; Zhang, B.; Luo, X. Empowering Student Futures: An AI-Driven Approach to Give Personalized Major/Career Recommendations for High School Students. In *Proceedings of the 2024 International Conference on Artificial Intelligence, Digital Media Technology and Interaction Design*; ACM: New York, NY, USA, 2024; pp. 103–113. [[CrossRef](#)]
7. Yadalam, T.V.; Gowda, V.M.; Kumar, V.S.; Girish, D.; Namratha, M. Career Recommendation Systems Using Content Based Filtering. In *Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES)*; IEEE: New York, NY, USA, 2020; pp. 660–665. [[CrossRef](#)]
8. Wang, L.; Fu, Y.; Zhang, Y. A Career Recommendation Method for College Students Based on Occupational Values. *Int. J. Emerg. Technol. Learn.* **2023**, *18*, 201–215. [[CrossRef](#)]

9. Lent, R.W.; Brown, S.D.; Hackett, G. Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *J. Vocat. Behav.* **1994**, *45*, 79–122. [\[CrossRef\]](#)
10. Oh, Y.J.; Jia, Y.; Lorentson, M.; LaBanca, F. Development of the Educational and Career Interest Scale in Science, Technology, and Mathematics for High School Students. *J. Sci. Educ. Technol.* **2012**, *22*, 780–790. [\[CrossRef\]](#)
11. Earl, J.; Iskandar, F.; Elizondo, F. Take a Job, Any Job: Exploring the Importance of Matched Interests to Career Paths and Work Satisfaction. *J. Employ. Couns.* **2019**, *56*, 33–45. [\[CrossRef\]](#)
12. Verma, B.K.; Srivastava, N.; Bharti, A.K. Predictive Models for Career Guidance Based on Academic and Behavioral Data with Data Mining and Machine Learning Technique. In *Proceedings of the 2025 3rd International Conference on Disruptive Technologies (ICDT)*; IEEE: New York, NY, USA, 2025; pp. 419–423. [\[CrossRef\]](#)
13. Roy, M.; Sharma, K. AI-Powered Career Guidance: A Scalable Model for Personalized Recommendations. *IEEE Access* **2026**, *14*, 20927–20941. [\[CrossRef\]](#)
14. Nie, M.; Xiong, Z.; Zhong, R.; Deng, W.; Yang, G. Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. *Appl. Sci.* **2020**, *10*, 2841. [\[CrossRef\]](#)
15. Dhelim, S.; Chen, L.; Aung, N.; Zhang, W.; Ning, H. A Hybrid Personality-Aware Recommendation System Based on Personality Traits and Types Models. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 12775–12788. [\[CrossRef\]](#)
16. Ko, H.; Lee, S.; Park, Y.; Choi, A. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics* **2022**, *11*, 141. [\[CrossRef\]](#)
17. Patel, B.; Kakuste, V.; Eirinaki, M. CaPaR: A Career Path Recommendation Framework. In *Proceedings of the 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*; IEEE: New York, NY, USA, 2017; pp. 23–30. [\[CrossRef\]](#)
18. Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
19. Zadeh, L.A.; Klir, G.J.; Yuan, B. *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*; World Scientific: Singapore, 1996; Volume 6.
20. Saaty, T.L. *Decision Making with Dependence and Feedback: The Analytic Network Process*; RWS Publications: Pittsburgh, PA, USA, 1996.
21. Saaty, T.L. Decision Making—The Analytic Hierarchy and Network Processes (AHP/ANP). *J. Syst. Sci. Syst. Eng.* **2004**, *13*, 1–35. [\[CrossRef\]](#)
22. Wongvilaisakul, W.; Netinant, P.; Rukhiran, M. Dynamic Multi-Criteria Decision Making of Graduate Admission Recommender System: AHP and Fuzzy AHP Approaches. *Sustainability* **2023**, *15*, 9758. [\[CrossRef\]](#)
23. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
24. Lekan, K.; Pardos, Z.A. AI-Augmented Advising: A Comparative Study of GPT-4 and Advisor-Based Major Recommendations. *J. Learn. Anal.* **2025**, *12*, 110–128. [\[CrossRef\]](#)
25. Habbu, K.K.; Pasare, A.A.; Jadhav, P.S.; Khupse, S.; Dedgaonkar, S.; Lal, B. Career Advisory System Using LLM. In *Proceedings of the 2025 Artificial Intelligence and Smart Technologies for Sustainability Conference (AISTS)*, Rajkot, India, 21–23 August 2025; pp. 1–7. [\[CrossRef\]](#)
26. Lyu, Y.; Lu, J.; Guo, J.; Jing, Z.; Qiu, H.; Ouyang, J.; Li, L. GLIDE: Generative LLM-Driven Data Augmentation for Meta-Services With Applications to E-Commerce Scenarios. *IEEE Trans. Comput. Soc. Syst.* **2026**, *13*, 2941–2954. [\[CrossRef\]](#)
27. Peng, Y.; Chang, L.; Tang, Y.; Li, J.; Zhang, T.; Yang, Q. Zero-Shot Recommendation via LLM-Generated Synthetic Interaction Data. In *Proceedings of the 2026 5th International Symposium on Computer Applications and Information Technology (ISCAIT)*, Dongguan, China, 23–25 January 2026; pp. 687–691. [\[CrossRef\]](#)
28. Natividad, M.C.B.; Gerardo, B.D.; Medina, R.P. A Fuzzy-Based Career Recommender System for Senior High School Students in K to 12 Education. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *482*, 012025. [\[CrossRef\]](#)
29. Aher, G.V.; Arriaga, R.I.; Kalai, A.T. Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies. In *Proceedings of the 40th International Conference on Machine Learning*; PMLR: New York, NY, USA, 2023; Volume 202, pp. 337–371.
30. Anthis, J.R.; Liu, R.; Richardson, S.M.; Kozlowski, A.C.; Koch, B.; Brynjolfsson, E.; Evans, J.; Bernstein, M.S. Position: LLM Social Simulations Are a Promising Research Method. In *Proceedings of the 42nd International Conference on Machine Learning*; PMLR: New York, NY, USA, 2025; Volume 267, pp. 81005–81034.
31. Cummins, J. The Threat of Analytic Flexibility in Using Large Language Models to Simulate Human Data: A Call to Attention. *arXiv* **2025**, arXiv:2509.13397.
32. Meade, A.W.; Craig, S.B. Identifying Careless Responses in Survey Data. *Psychol. Methods* **2012**, *17*, 437–455. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Baumgartner, H.; Steenkamp, J.B.E.M. Response Styles in Marketing Research: A Cross-National Investigation. *J. Mark. Res.* **2001**, *38*, 143–156. [\[CrossRef\]](#)
34. Liu, Q.; Khalil, M.; Shakya, R.; Jovanovic, J. Scaling While Privacy Preserving: A Comprehensive Synthetic Tabular Data Generation and Evaluation in Learning Analytics. *arXiv* **2024**, arXiv:2401.06883.

35. Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*; ACM: New York, NY, USA, 1994; pp. 175–186. [[CrossRef](#)]
36. Liu, N.N.; Yang, Q. EigenRank: A Ranking-Oriented Approach to Collaborative Filtering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*; ACM: New York, NY, USA, 2008; pp. 83–90. [[CrossRef](#)]
37. Chen, M.; Beutel, A.; Covington, P.; Jain, S.; Belletti, F.; Chi, E. Top-K Off-Policy Correction for a REINFORCE Recommender System. *arXiv* **2021**, arXiv:1812.02353.
38. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. [[CrossRef](#)]
39. Liu, J.; Dolan, P.; Pedersen, E.R. Personalized News Recommendation Based on Click Behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*; ACM: New York, NY, USA, 2010; pp. 31–40. [[CrossRef](#)]
40. Renninger, K.A.; Hidi, S. *The Power of Interest for Motivation and Engagement*, 1st ed.; Routledge: New York, NY, USA, 2015.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.