Research on the Construction of a Precision Guidance Model for Collaborative Planning of College Students' Academic and Career Driven by Big Data

Lijuan Xi, Feifei Xue
School of Economics, Shenzhen Polytechnic University, Shenzhen 518055, Guangdong, China
Corresponding author: Lijuan Xi (e-mail: 13728707550@163.com)

ABSTRACT Addressing the separation between academic management and career planning in higher education, which causes data disconnection and delayed intervention, this paper develops a big data-driven precision guidance model for collaborative academic-career planning. First, multisource data from academic affairs systems, campus behavior records, student affairs systems, and job-market platforms are integrated to construct comprehensive student profiles. Second, deep neural networks are used to mine implicit associations between academic trajectories and career competency requirements. Third, a hybrid recommendation engine combining collaborative filtering and knowledge graphs is developed to dynamically match academic warnings, course-selection suggestions, skill-development resources, internships, and career guidance. Finally, a visualization platform is designed to support tiered intervention by counselors and students. Experimental results show that the model reduces the response time of high-risk academic warnings from 12.5 days to 4.8 days, achieves a Top-5 recommendation accuracy of 81%, increases resource adoption by 72.6%, and improves student GPA by 0.31. The findings verify the model's effectiveness in proactive intervention, academic-career alignment, and precision education support.

INDEX TERMS big data, academic-career planning, precision guidance, knowledge graph, recommendation system

I. INTRODUCTION

CURRENTLY, the employment situation for college graduates is becoming increasingly complex, as employers in the advanced materials sectors are constantly raising their requirements for the composite abilities of talents. The connection between academic performance and professional competence is becoming increasingly close [1-2]. However, in the traditional student guidance work of universities, academic management and career planning are usually handled separately by the academic affairs department and the employment guidance center. The two have long been separated in terms of data systems, workflows, and intervention timing. Counselors find it difficult to keep abreast of students' academic fluctuations and career preparation dynamics, and often can only conduct "remedial" conversations based on experience during the end-of-term warning or graduation season. Students also lack a clear understanding of the connection between their academic trajectory and career goals, resulting in insufficient motivation for learning or confusion in career choices [3]. This reality of "data disconnection, incomplete information, and delayed intervention" urgently requires the introduction of new technologies to break down the data barriers between academics and careers, and to realize the transformation from "experience-driven, passive response" to "data-driven, proactive intervention".

In recent years, scholars have conducted a large number of studies on the application of big data technology in the education of college students. Shofiyyah et al. used qualitative literature review to sort out the innovation of Islamic education management in higher education institutions, analyze its implementation challenges and evaluate its development prospects [4]. Zhou et al. examined the use and cognition of artificial intelligence tools by higher education students, revealing its diverse application scenarios and highlighting its advantages in improving efficiency, supporting personalized learning and enhancing language skills [5]. Suherlan et al. sorted out the efforts of students in higher education to improve their academic performance with the help of digital technology through literature review, pointing out that digital technology can innovate and change the way of learning, making it more in line with national goals and ideals [6]. Vindigni pointed out that traditional lecture-based teaching may be difficult to meet the needs of modern professional practice for cultivating comprehensive abilities, and focused on majors such as design, media and communication to provide reform guidance for optimizing the professional ability development of different disciplines [7]. However, most existing studies are limited to single-dimensional analysis: academic early warning models can identify risks, but it is difficult to link risks with career development needs; career recommendation systems can provide job suggestions, but fail to fully consider students' academic foundation and ability growth trajectory.

In response to the above problems, the rapid development of multi-source data fusion and deep learning technology has provided new ideas for solving the problem of academic-career collaborative guidance. Li Yi took the exhibition major as an example to promote the digital reform of talent training and curriculum system under the background of artificial intelligence, and built an "intelligent plus" training model to cope with the "dehumanization" challenge [8]. Liu relied on the internship scene of the School of Information Engineering of Guangdong Asia Television Performing Arts College to study and build a higher vocational internship quality intelligent evaluation system based on multi-source data fusion to address the problems of strong evaluation subjectivity, data fragmentation and feedback lag [9]. Shang believes that artificial intelligence is reshaping the governance of postgraduate education, but it needs to address the challenges of data governance and technology ethics, and promote digital transformation with technology integration, collaborative governance and data closure [10]. However, the existing methods still need to be improved in terms of deep coupling modeling of academic characteristics and career characteristics, generation mechanism of dynamic intervention strategies and visualization of collaborative development. Based on this, this paper proposes a precise guidance model for academic-career collaborative planning based on multi-source data fusion and deep learning, which aims to solve the problems of single data dimension, insufficient correlation mining and lag in intervention strategies in the existing research.

The research purpose of this paper is to establish a precise guidance model for academic-career collaborative planning based on big data, specifically tailored to the evolving technical demands of science. By integrating academic performance with the specialized competencies required in advanced materials, the model achieves precise, whole-process support for students entering the high-performance fiber industry. First, we integrate the data from multiple sources, such as academic affairs system, campus behavior perception and online recruitment platform, to establish a data basis of whole life cycle of students, including students' academic information, career information and behavior information on campus. Then, we model the academic route and career competency graph based on deep neural network, and find out the implicit coupling between the two. On this basis, we propose a collaborative filtering-based and knowledge graph-based hybrid recommendation engine to realize the dynamic matching of academic warning, suggestions of course selection and career development resources such as internship and skills training. Finally, we construct a visualized "academic-career collaborative planning" decision support platform, and provide precise intervention tool for counselors and students at different levels.

II. MULTI-SOURCE DATA FUSION AND CONSTRUCTION OF A COMPREHENSIVE STUDENT PROFILE

In order to obtain the full picture of the academic performance and potential of career development in students, the given study combined heterogeneous data obtained through different internal business systems and external recruitment platforms in the first place. Four main sources of data were used in the study: first is the academic affairs management system which included basic student information, course grades, credit completion progress, course selection history, and status of curriculum implementation; second, the campus card system through which the behavioral logs like canteen consumption, supermarket shopping, library access control and dormitory entry/exit were recorded; third, the student affairs management system which included awards and punishments, participation in club activities, student cadre positions and records of conversation with their counselors; and fourth, the campus employment site and partner company recruitment sites where the data about student resumes Once the data were collected, preprocessing was done in three phases. To begin with, entity alignment technology was applied on the basis of student ID numbers to match data between various systems, thus removing the duplication of records and creating cross system mapping relationships. Cleaning of the raw data to remove clear outliers, including data that have a daily spending greater than three standard deviation of the mean, incorrectly entered records, or missing timestamps, were then removed. The few missing values were filled based on data sparsity: mean imputation was applied to low-variance demographic fields where missingness was less than 1%, while the K-nearest neighbor (KNN, k=5) algorithm was utilized for time-series behavioral logs to maintain the local continuity of the student profiles. Lastly, the time-series data were synchronized and aligned. While complex feature vectors for long-term trend analysis were aligned to a weekly granularity, the system maintains a parallel high-frequency processing stream for critical daily behavior logs (e.g., attendance and access control) to ensure that specific high-risk triggers can be identified and acted upon within a 24-hour window.

III. ACADEMIC-CAREER FEATURE EXTRACTION AND DEEP ASSOCIATION MINING

On the academic dimension, GPA sequence, professional ranking sequence, course passing rate, distribution of courses by failing, and sequence of re-taking or re-taking courses were extracted from students' course transcripts. On the other hand, the ratio of advanced courses, tendency of taking courses across disciplines, and average difficulty coefficient of elective courses were extracted from students' course selection behaviors to reflect students' academic expansion intention and learning performance. Behavioral dimension is an important complement to learning performance. Based on student ID card and library data, the learning regularity indicators were extracted from student behaviors, including average library time per week, the number of times of studying at night, the holiday study activity, and the regularity of daily life before exam weeks. The correlation between these indicators and academic performance was validated using Pearson correlation analysis, yielding coefficients (r) of 0.68 for library regularity and 0.74 for morning class attendance ($p < 0.01$). These coefficients provide the empirical basis for using behavioral fluctuations as leading indicators—rather than just lagging grade-based data—to trigger early-stage interventions.

In terms of career dimension, structured tags are extracted from student resumes and recruitment platform logs. First, professional skills keywords, language proficiency, internship experience, project experience, and awards are analyzed from the resume text to construct a student skills list. Second, intended industries, intended job levels, expected salary range, and application activity are extracted from job browsing and application logs. Finally, the students' job competency matching degree is assessed in conjunction with their internship experience to form a career orientation feature vector.

After feature extraction, following deep neural network is applied to mine the academic-career correlation. Due to the nonlinear regression capability of deep neural network, compared with traditional linear regression method or simple association rules method, it is hard to learn the high-dimensional nonlinear implicit relationship between academic performance and career development. Therefore, in this paper, we design the joint learning architecture to input the academic feature vector and career feature vector into the deep neural network, and then learn the interaction pattern between two feature vectors through multiple hidden layers nonlinear transformation. The deep neural network contains 3 hidden layers with 256, 128, and 64 neurons respectively. This pyramid-shaped architecture was selected following a grid search optimization to balance the high-dimensional complexity of multi-source input features with the need to prevent overfitting. Preliminary sensitivity analysis indicated that this specific distribution yielded the lowest validation loss compared to flatter or deeper configurations. To avoid the over-fitting situation, the Dropout strategy is introduced after each layer, and the dropout rate is 0.3.

IV. PERSONALIZED RESOURCE RECOMMENDATION AND DYNAMIC INTERVENTION STRATEGY GENERATION

The recommendation engine employs a dual-layered hybrid architecture. The Collaborative Filtering (CF) layer utilizes latent factor modeling to identify 'person-job' similarities based on historical cohorts, while the Knowledge Graph (KG) layer utilizes TransE embedding and semantic path-tracing to map explicit 'course-skill-job' dependencies. This integration ensures the engine provides both data-driven efficiency and pedagogically sound interpretability. The collaborative filtering part mainly focuses on "person-job" and "person-course" information matching. Through the historical behavior data of student groups, calculate the similarity between the target students and similar student groups in terms of course selection, skill mastery and job application, and then recommend advanced courses taken by similar student groups or internship positions applied for similar jobs. Take student A as an example, find out the top 20 students with the most similar academic history to student A and skill list, and then find out the most popular courses taken by these 20 students during their junior year and the jobs they applied for after graduation, and add the courses with a high degree of matching in student A's job intention into the recommendation list.

The knowledge graph component addresses the interpretability and path planning issues of the recommendation results. This study constructs a knowledge graph with courses, skills, and jobs as core entities, including triple relationships such as "course-prerequisite courses," "course-corresponding skills," and "skill-job requirements." When the system identifies a student who wants to apply for a data analysis position but whose Python programming skills are marked as lacking, the knowledge graph module automatically infers that a "Python Programming" course is required beforehand, and then recommends that course to the desired position. This knowledge graph-based path planning helps students clearly see the gap between their "current status" and their "target position" and how to bridge it.

The generation of the dynamic intervention strategy is anchored on the differentiated design depending on the real-time status and risk level of the student. The system places students into three categories:

1) Academic Risk Type, which is marked by a great variation in grades and warnings on more than one course;
2) Career Confusion Type, where there is a lack of resume information, accomplishing scattered courses jobs, and few applications;
3) Collaborative Development Type, where academic and career development is moving positive.

The system produces differentiation intervention strategies in every type. In case of academic risk among students, it is prioritized to offer academic support tools, such as information about retake classes, peer tutoring appointments, and learning methods courses. At the same time, the counselors are warned, and it is recommended to discuss this issue in person. In the case of student with career uncertainty, they have given priority to career assessment product, industry awareness lectures, and low threshold internship opportunities, to help them define their career path. In the case of students who are concerned with collaborative growth, demanding advanced courses and company referrals that are of high quality are offered to enable students to develop better.

The delivery time of the intervention strategies is also determined dynamically. According to real-time tracking of student behaviors data, the system determines the important intervention points. E.g. when the system detects a reduction in library time of a student over a consecutive week by more than half, cross-referenced with a simultaneous decline in course-platform login activity or a pattern of missed assignments, it is categorized as a high-risk intervention point. This multi-factor verification minimizes false positives caused by students choosing to study in alternative locations. The system will notified the student of a point where a student has shown a strong interest in a type of job but is not confident enough to apply and so job application tips and resume optimization tips will be automatically sent when the system notices that this student has been browsing this type of job more than five times without applying.

V. EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

A. OVERALL EXPERIMENTAL SETUP

1) Data Sources and Preprocessing

We take the undergraduate students from 2018–2022 in a key provincial university of science and technology as the subjects of this experiment. The time span of data collection is four academic years in total. The data sources include the academic affairs management system (about 3.2 million academic records in total), the campus behavior perception system (about 18 million campus card and access control logs in total), and the employment platform (about 500,000 career behavior records in total). In the preprocessing stage, the entity alignment across different systems were implemented based on student ID number as the unique identifier firstly. Then, the outlier were deleted directly, and the missing values were processed by mean imputation or KNN nearest neighbor method. Finally, the multi-source data were transformed into weekly time series. To account for the external anomalies in campus behavior during the 2018–2022 period, we applied a z-score normalization to each academic semester independently. This seasonal adjustment ensures that 'learning regularity indicators' are measured relative to the cohort's average behavior within a specific timeframe, mitigating the skew caused by periods of remote or hybrid learning. The total number of experimental samples were 12,568 students in total and split into training set (8,798 students) and test set (3,770 students) randomly, with the ratio of 7:3. This study was approved by the Ethics Committee of Shenzhen Polytechnic University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

2) Experimental Environment and Parameter Settings

The experimental environment and model parameter configuration of this study are shown in Table 1. On the hardware side, high-performance servers are used to support deep learning model training, and the software environment is built on an open-source ecosystem, with model parameters determined through multiple rounds of optimization.

TABLE 1. Experimental Environment and Parameter Settings

Category	Configuration Item	Parameter
Hardware	CPU	Intel Xeon Gold 6248R (48 cores, 3.0GHz)
	Memory	256GB
	GPU	NVIDIA Tesla V100 (32GB)×4
Software	OS	Ubuntu 20.04
	Framework	TensorFlow 2.6, PyTorch 1.9
	Computing Engine	Spark 3.1
	Database	MySQL 8.0, MongoDB 4.4
Model	DNN Structure	256-128-64, ReLU+Softmax
	Collaborative Filtering	Latent factor 50, reg 0.01, lr=0.001
	Knowledge Graph	TransE, embedding 100-dim, γ=1.0
	Training Parameters	epoch=200, batch=256, Adam optimizer, early stopping patience=10

B. COMPARATIVE EXPERIMENT ON THE TIMELINESS OF ACADEMIC EARLY WARNING

To verify the timeliness advantage of this model in academic early warning, students from the second to fourth years of a university were selected and randomly divided into an experimental group and a control group. The experimental group used the early warning module of this study to monitor multi-source behavioral data in real time, while the control group used the traditional final exam grade early warning method. The average response time from the occurrence of the risk to the first intervention by the counselor was statistically analyzed for low, medium, and high-risk events to compare the early warning efficiency of the two groups.

As shown in Figure 1, the academic early warning model constructed in this paper is significantly better than the traditional method in terms of response timeliness. In low-risk scenarios, the average response time of the experimental group was 2.1 days, which was 59.6% shorter than the 5.2 days of the control group; in medium-risk scenarios, the experimental group's response time was 3.5 days, while the control group's was 8.1 days, an improvement of 56.8%; in high-risk scenarios, the experimental group could complete the intervention in an average of 4.8 days, while the control group took as long as 12.5 days, an improvement of 61.6%. Data shows that this model, through real-time monitoring of multi-source behavioral data, has shifted risk detection from "end-of-term discovery" to "daily capture," significantly reducing the intervention window, especially demonstrating a superior response to sudden high-risk events.

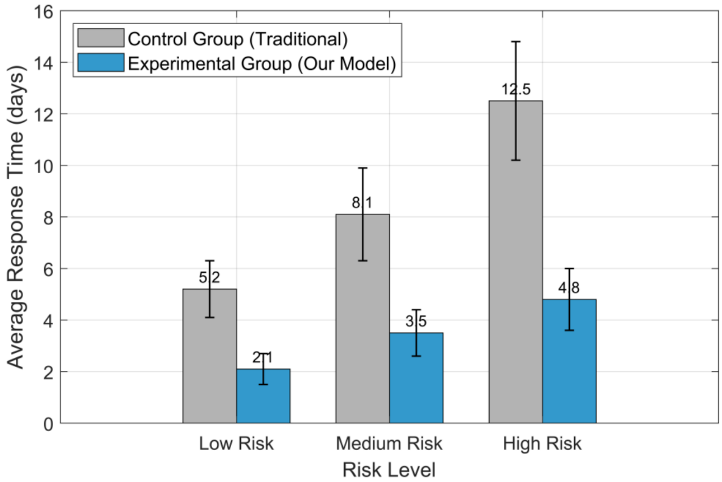

FIGURE 1. Evaluation of the Timeliness of Academic Early Warning

C. CAREER PLANNING RECOMMENDATION ACCURACY COMPARISON EXPERIMENT

This paper randomly choose 5000 recent graduates as test sample. We use following 4 models to give job recommendation.

1) Baseline 1: Rule-based model (Base line 1 just matches major).
2) Baseline 2: One Dimensional Model (Base line 2 is based on resume only).
3) Baseline 3: XGBoost Machine Learning Model.
4) Multi-Source Data Fusion Model Proposed in this paper.

Take the students' real contract position as the benchmark, we calculate the recommendation accuracy of above 4 models at Top-1, Top-3, Top-5 and Top-10. And then we use Base line 1, Base line 2 and Base line 3 to prove the effectiveness of this model.

As shown in Figure 2, the traditional rule-based model had an accuracy of only 0.31 at Top-1, rising to 0.58 at Top-10 as the recommendation list expanded, showing the weakest overall performance. The single-dimensional model uses resume information for matching, achieving a Top-1 accuracy of 0.39 and a Top-5 accuracy of 0.63, but subsequent growth slows. The machine learning model (XGBoost) shows significant improvement, achieving a Top-3 accuracy of 0.61 and a Top-10 accuracy of 0.77. Our model achieves a Top-1 accuracy of 0.62 and a Top-5 accuracy of 0.81. While some latent career competencies remain difficult to quantify directly, the high accuracy is primarily attributed to the model's ability to use multi-source behavioral proxies and deep association mining to infer these hidden variables. This allows the deep neural network to outperform Baseline 3 (XGBoost) by capturing non-linear relationships that single-dimensional models overlook.

D. EVALUATION OF PERSONALIZED RESOURCE RECOMMENDATION AND DYNAMIC INTERVENTION EFFECTS

To validate the efficacy of personalized resource recommendation and dynamic intervention strategy based on multi-source data fusion model, we randomly selected 800 students who had academic warning or career planning requirements as experimental group and control group. The experimental group got personalized recommendation and intervention plan based on multi-source data fusion model, while control group only got uniform push information. At last, we analyzed the intervention efficacy of our model by resource recommendation adoption rate, improvement of academic performance and career planning clarity score between experimental group and control group.

TABLE 2. Evaluation of Personalized Resource Recommendation and Dynamic Intervention Strategies

Indicator	Control Group (n=400)	Experimental Group (n=400)
Resource Recommendation Adoption Rate (%)	35.2 ± 8.1	72.6 ± 6.3
Academic Performance Improvement (GPA)	0.12 ± 0.05	0.31 ± 0.07
Career Planning Clarity Score (1-5 scale)	3.1 ± 0.6	4.2 ± 0.5
Intervention Response Satisfaction Rate (%)	41.5 ± 10.2	83.7 ± 7.8

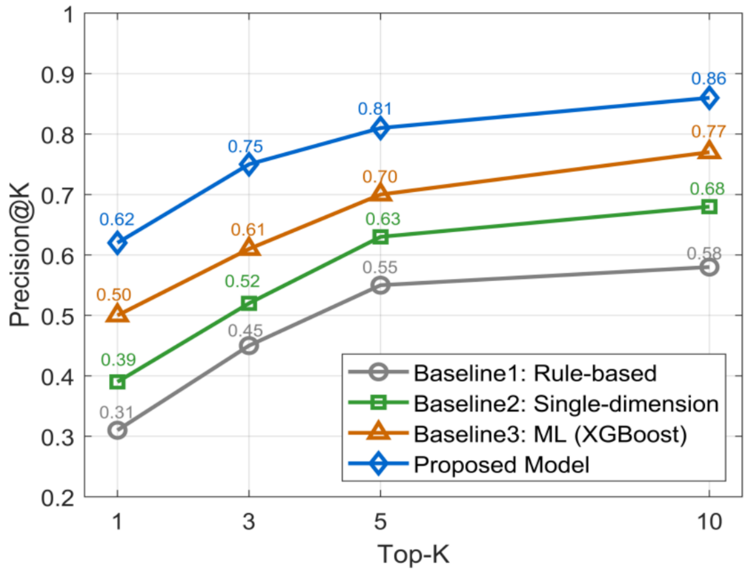

FIGURE 2. Career Planning Recommendation Accuracy Evaluation

In Table 2, the resource recommendation adoption rate of the experimental group reached 72.6%, significantly higher than the 35.2% of the control group; the average improvement in academic performance was 0.31, better than the 0.12 of the control group; the career planning clarity score was 4.2 points, compared to 3.1 points in the control group; and the intervention response satisfaction rate reached 83.7%, compared to only 41.5% in the control group. All indicators demonstrate that the personalized recommendation and dynamic intervention strategy constructed in this study can accurately match students' needs, effectively stimulate students' willingness to act, and significantly improve their academic performance and career planning abilities, thus verifying the practical value of data-driven precision guidance.

VI. CONCLUSION

The academic-career collaborative planning precision guidance model constructed in this study dismantles the separation barriers between multi-source data—including academic affairs, behavior, and employment—to better align student competencies with the technical demands of fiber science and advanced materials. By integrating these datasets, the model provides a streamlined pathway for students to transition from theoretical study into specialized roles within the functional and high-performance fiber industries. It uses deep learning to jointly model the deep relationship between academic achievement and career choice and applies the collective filtering and knowledge graphs to pursue dynamic resource matching, which eventually constitutes a visualized collaborative planning platform. It has been proven through experiment that this model has got considerable benefits in enhancing the efficiency of early warning response, effectiveness of the recommendations and the effectiveness in student development. Nevertheless, the research is also burdened with some limitations: the data is gathered at only one university, the model requires cross-university testing; some of the indicators of career abilities cannot be entirely measured, which can negatively impact the completeness of recommendations; and the dynamic adaptive ability of the intervention strategy can also be improved. Future studies will introduce cross-institutional data sharing and multimodal large models to evaluate student competencies in fiber science and advanced materials, enhancing the adaptive human-machine collaborative mechanism of intervention strategies. This integration will further elevate the intelligence level of precision guidance, ensuring that academic development aligns with the sophisticated technical requirements of the functional and high-performance fiber industries.

AUTHOR CONTRIBUTIONS

Conceptualization – Lijuan Xi and Feifei Xue; methodology – Lijuan Xi and Feifei Xue; investigation – Lijuan Xi and Feifei Xue; writing-original draft preparation – Lijuan Xi and Feifei Xue. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Shenzhen Polytechnic University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

REFERENCES

[1] A. Detgen, F. Fernandez, A. McMahon, et al., “Efficacy of a college and career readiness program: Bridge to employment,” The Career development quarterly, vol. 69, no. 3, pp. 231-247, 2021, doi: 10.1002/cdq.12270.

[2] A. Belle M, O. Antwi C, Y. Ntim S, et al., “Am I gonna get a job? Graduating students psychological capital, coping styles, and employment anxiety,” Journal of Career Development, vol. 49, no. 5, pp. 1122-1136, 2022, doi: 10.1177/08948453211020124.

[3] T. Ibeneme O and M. Ebubechi U, “Practical skill needs of technical college graduates of MVMW for employment in Anambra and Enugu States,” African Journal of Educational Management, Teaching and Entrepreneurship Studies, vol. 3, no. 1, pp. 150-157, 2021, doi: 10.1177/089484532110201.

[4] A. Shofiyyah N, S. Komarudin T, and S. Hasan M, “Innovations in Islamic Education Management within the University Context: addressing challenges and exploring future prospects,” Nidhomul Haq: Jurnal Manajemen Pendidikan Islam, vol. 8, no. 2, pp. 193-209, 2023, doi: 10.31538/ndh.v8i2.3625.

[5] X. Zhou, J. Zhang, and C. Chan, “Unveiling students experiences and perceptions of artificial intelligence usage in higher education,” Journal of University Teaching and Learning Practice, vol. 21, no. 6, pp. 126-145, 2024, doi: 10.53761/xzjprb23.

[6] H. Suherlan, A. Basir, W. Syakhrani A, et al., “The Roles of Digital Application Innovates Student Academic in Higher Education,” Nazhruna: Jurnal Pendidikan Islam, vol. 5, no. 2, pp. 672-689, 2022, doi: 10.31538/nzh.v5i2.2246.

[7] G. Vindigni, “Adaptive and re-adaptive pedagogies in Higher Education: A comparative, longitudinal study of their impact on professional competence development across diverse curricula,” European Journal of Theoretical and Applied Sciences, vol. 1, no. 4, pp. 718-743, 2023, doi: 10.59324/ejtas.2023.1(4).66.

[8] W. Li, “Research on the Digital Transformation Path of Exhibition Major in Higher Vocational Education under the Background of Artificial Intelligence: Taking Guangdong Light Industry Vocational and Technical College as an Example,” Journal of Guangdong Open University, vol. 31, no. 1, pp. 93-97, 2022, doi: 10.3969/j.issn.1008-9764.2022.01.18.

[9] Y. Liu, “Research on Intelligent Evaluation System for Internship Quality in Higher Vocational Colleges Based on Multi-Source Data Fusion,” International Journal of Emerging Technologies and Advanced Applications, vol. 2, no. 9, pp. 1-8, 2025, doi: 10.62677/IJETAA.2509139.

[10] B. Shang, “AI Empowerment: Digital Transformation of Graduate Education Governance,” Journal of Computing and Electronic Information Management, vol. 18, no. 3, pp. 5-10, 2025, doi: 10.54097/yq1swa37.