2024 5th International Conference on Electronic Communication and Artificial Intelligence (ICECAI) | 979-8-3503-8694-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICECAI62591.2024.10674896

Application of Data Mining and Visualization Techniques in Analyzing the Demand for Artificial Intelligence Jobs

Renzhi Luo[1,a*]
[1]School of Mathematics and Computer Science, Chuxiong Normal University, Chuxiong, China
[a] *Corresponding author: 20070402210@cxtc.edu.cn

Zhiming Qian[2,b]
[2]School of Mathematics and Computer Science, Chuxiong Normal University, Chuxiong, China
[b] qzhiming@cxtc.edu.cn

Ziying Li[3,c]
[3]School of Mathematics and Computer Science, Chuxiong Normal University, Chuxiong, China
[c] lzy1102667631@gmail.com

***Abstract*-In the context of labor market transformation led by technological advancement, this study uses data mining and visualization techniques to analyze the information of AI-related jobs crawled on 51JOB website and explore the key features of job demand. Through the techniques of decision tree, support vector machine, K-mean algorithm and TF-IDF algorithm, combined with visualization libraries such as Python's Matplotlib, Seaborn, and Plotly, and visualization tools such as bar charts and box-and-line diagrams, this study demonstrates the key information on job distributions, salary levels, and skill requirements. The results of the study show that intermediate talents with 3-4 years of work experience and those with bachelor's degree or above are more popular in the AI industry, and there is a significant geographic clustering of job demand, especially in cities such as Wuhan, Chengdu and Hangzhou. In addition, the study finds that education and work experience are important influences on salary levels, and that private companies dominate recruitment activities in the AI field. Educational institutions should offer market-oriented AI courses, such as focusing on the combination of theory and practice, key skills in algorithm development, data mining, and natural language processing, and the development of key skills through real-world projects. The findings of this study provide a scientific basis for policy makers, enterprises and educational institutions in optimizing recruitment strategies and developing education and training programs.**

Keywords-Artificial intelligence jobs; Data mining; Visualization techniques; Salary analysis; Job demand.

I. INTRODUCTION

In the technological revolution of the 21st century, Artificial Intelligence (AI) has become a central driver of technological progress and shaping the future society. In particular, by 2023, the labor market is undergoing an unprecedented transformation as a result of the continuous innovation and expansion of the application scope of AI technology. According to the World Economic Forum's Future of Work report and Employment Outlook 2023 report, approximately 23% of global jobs are expected to be significantly impacted by the transformation of the AI industry over the next five years. This transformation not only signals a high demand for AI professionals, [1] but also exposes a number of challenges such as an imbalance between talent supply and demand, unclear job requirements, and opaque industry information.

Data mining refers to the process of extracting or mining knowledge from a large amount of data through algorithms and statistical techniques. [2] In AI job demand analysis, data mining can help identify trends and patterns in the job market. Data visualization involves converting data into graphical or visual formats to make the information easier to understand and analyze. [3] While the existing literature has explored the macroeconomic impact of AI technologies on the labor market, [4] there is a relative lack of specific analysis on the characteristics of AI job demand, particularly in terms of pay levels, the relationship between work experience and educational requirements, and the characteristics of the geographic distribution of jobs. This lack of research highlights the urgency of an in-depth study of the demand for AI jobs in order to provide a solid scientific basis for relevant policy making and education and training.

Therefore, this study uses data mining and visualization techniques to conduct a comprehensive analysis of AI job recruitment information, aiming to reveal the association between salary and work experience and education; explore the characteristics of the distribution of AI jobs in different regions, especially the major talent clusters; compare the specific requirements of different types of enterprises for work experience and education, and analyze the differences of these requirements among various types of enterprises; and finally, based on the job market data, it proposes how to adjust the talent training and education strategies of the AI industry to accurately meet the market demand.

The innovation of this study lies in the systematic application of comprehensive data analysis methods to deeply explore the multidimensional demand characteristics of AI jobs, and to propose targeted education and training and talent development accordingly. Through these findings, this study not only provides decision-making support for enterprises and job seekers, but also provides an empirical basis for higher education institutions to adjust their curricula and training programs, and contributes new perspectives and strategies for the sustainable and healthy development of the AI field.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on September 26,2024 at 18:49:29 UTC from IEEE Xplore. Restrictions apply.

II. METHODS

A. Data collection and pre-processing

51JOB, a leading recruitment platform in China, was chosen as the data source for this study, considering its wide coverage and representativeness in the recruitment market, which can effectively reflect the current real-time demand trend of AI jobs. Using a self-developed Python web crawler program, combined with Mitmproxy and Selenium technology, we successfully crawled 10,347 job postings related to "artificial intelligence" on October 17, 2023, which included job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, job titles, and job titles. The data captured includes job title, company name, work experience, region, education requirements, posting date, company industry, employee skills and benefits, nature of the organization, salary range, and salary type.

Before data mining and visualization and analysis, this study conducted thorough data preprocessing. First, in the data cleaning stage, various techniques such as regular expression, string matching, strip method, duplicated function and find_outliers_iqr function were used to deal with errors, missing, duplicated or outliers. Specific operations include deleting 2,293 duplicates, removing 291 records with empty salary ranges, and excluding 664 anomalous salary data through the interquartile range (IQR) method. In the data type conversion session, fields such as education and work area are converted to classification type, job posting date is converted to date-time type, and salary range data is converted to numeric type to prepare for subsequent analysis. In the data integration stage, the salary range was processed with unit unification and the work area was split so as to analyze the salary level and geographic distribution in a more detailed way. After these steps, the total amount of processed data was 7,199.

B. Data Analysis Methods

In terms of data mining, various techniques such as decision tree, support vector machine (SVM), and K-mean algorithm were used to effectively classify jobs and analyze geographic location clustering to identify major job categories and demand hotspot regions within the AI field. In addition, the text in the job descriptions is deeply mined using the Word Frequency-Inverse Document Frequency (TF-IDF) algorithm to extract key information and analyze skill demand trends.

In terms of data visualization, this study utilizes a variety of chart forms such as bar charts, pie charts, box-and-line charts, radar charts, and word cloud charts, and displays key information about AI jobs, including the distribution of the number of jobs, the salary level, and the demand for skills through visualization libraries such as Python's Matplotlib, Seaborn, and Plotly. The application of these tools not only improves the clarity and accuracy of the analysis results, but also enhances the visual effect and strengthens the interpretation and expression of the data, which facilitates better understanding and utilization of the research results.

Through these methods, this study aims to deeply analyze the talent and skill demand in the field of artificial intelligence, and provide scientific basis and practical suggestions for relevant decision-making.

III. RESULTS

A. Salary level in the field of artificial intelligence

In the field of artificial intelligence, the salary level is closely related to work experience and education. (Fig. 1 to Fig. 5) The analysis results show that the number of positions with 3-4 years of work experience is the largest, accounting for 26.62% of the total positions, highlighting the market's high demand for intermediate-experienced talents. Entry-level positions (1-2 years of experience) also provided a considerable number of job opportunities, while the demand for senior positions (5 years of experience and above) was relatively low. In terms of educational qualifications, bachelor's degree jobs have the highest demand at 51.54%, followed by college and master's degree, with less demand for doctoral and lower high school and other qualifications. As far as salary is concerned, the more work experience one has, the higher the maximum salary level usually is. In addition, there is a significant difference in the highest salaries under different educational requirements, with jobs with doctoral and master's degrees offering much higher highest salaries than those with bachelor's and college degrees.

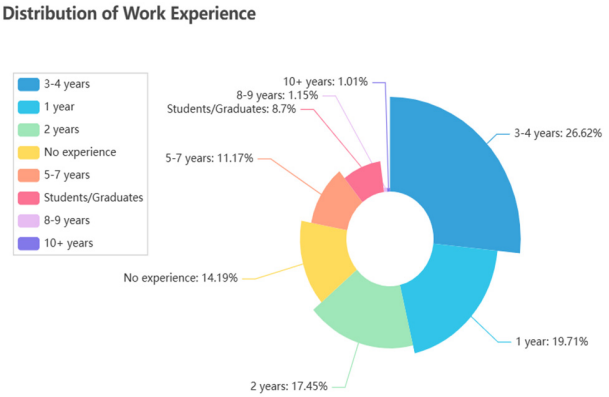

Fig. 1. Distribution of Work Experience

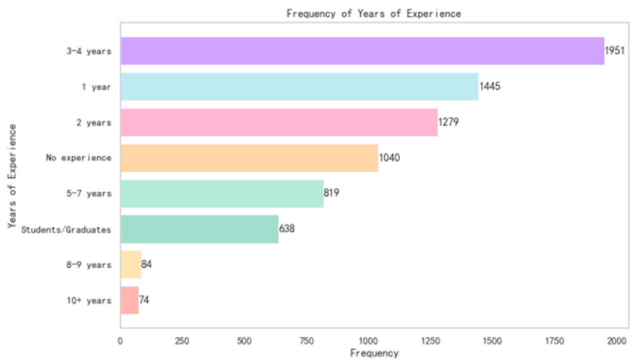

Fig. 2. Frequency of Years of Experience

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on September 26,2024 at 18:49:29 UTC from IEEE Xplore. Restrictions apply.

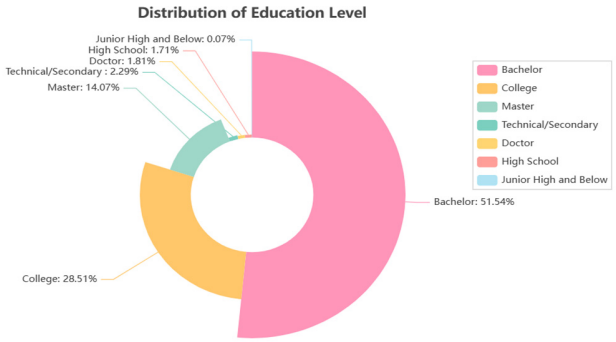

Fig. 3. Distribution of Education Level

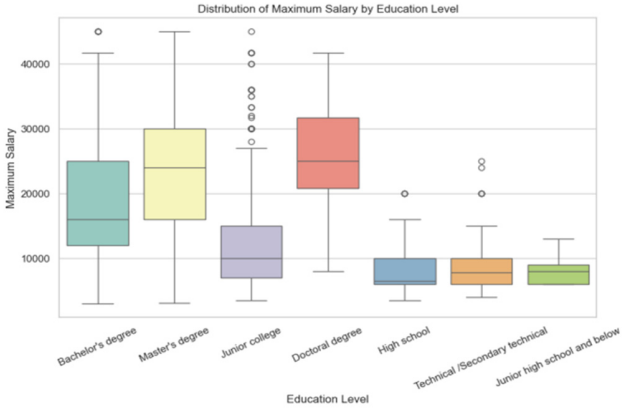

Fig. 4. Distribution of Maximum Salary by Education Level

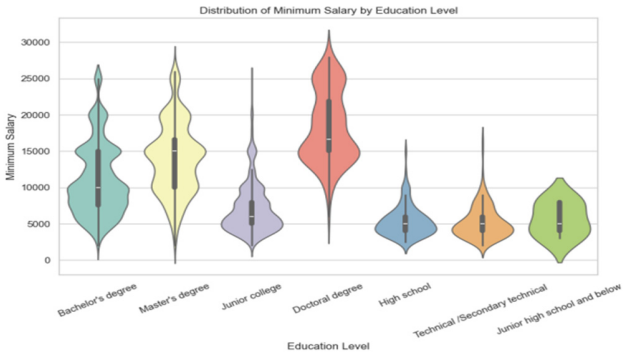

Fig. 5. Distribution of Minimum Salary by Education Level

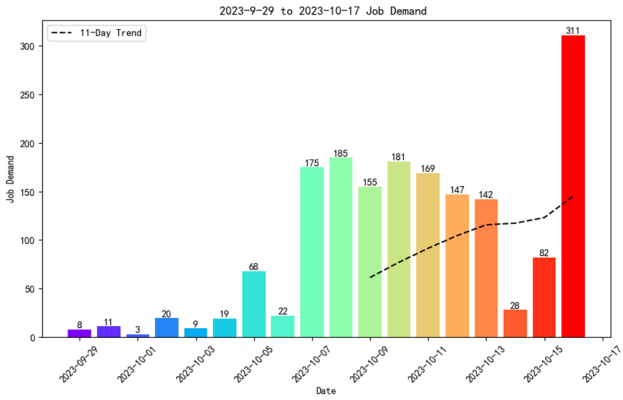

Fig. 6. 2023-9-29 to 2023-10-17 Job Demand

In addition, the data shows a significant fluctuation in demand for AI jobs from September 29 to October 17, 2023. The National Day and Mid-Autumn Festival were the main influences on demand during this period. Demand was lowest on October 1 for the National Day (3), highest on October 16 for the Mid-Autumn Festival (311), and higher on October 7 and 8 for the last two days of the National Day holiday (175 and 185, respectively). Demand fluctuated between 10 and 200 on the other dates. Holiday vacations led to a drop in demand and a relative increase in demand on weekdays.(Fig. 6)

B. Geographical Distribution Characteristics

In the geographical distribution of the demand for AI jobs in China, Wuhan, Chengdu and Hangzhou highlight their leadership with their high demand in the AI field, accounting for 20.3%, 18.5% and 10.7% of the total demand respectively. (Fig. 7) Demand in these cities is concentrated in specific areas corresponding to high-tech industrial parks and economic development zones, such as Hongshan District in Wuhan, High-Tech District in Chengdu, and Binjiang District in Hangzhou. In addition, Guangzhou, Xi'an, Suzhou, Changsha, Shanghai, Shenzhen and Nanjing also show high demand for AI jobs, each exhibiting unique geographical advantages.

Artificial Intelligence job postings reflect differences in the application and development of AI technology between regions and industries. Economically developed and technologically advanced regions and industries have strong demand for AI technology, intense competition, and abundant opportunities. On the contrary, economically backward and technologically lagging regions and industries have less demand, relatively mild competition, and limited opportunities. This variation also reveals the impact of the level of technological development and market maturity on AI jobs.

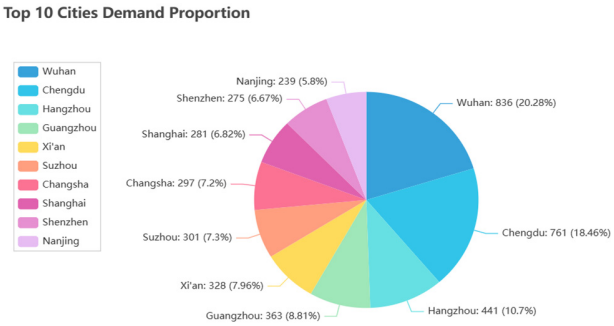

Fig. 7. Top 10 Cities Demand Proportion

C. Enterprise Recruitment Demand

Distribution of enterprise nature: funnel chart analysis shows (Fig. 8) that among enterprises recruiting for AI-related positions, private enterprises occupy an absolutely dominant position with a demand of 4,748 positions, far exceeding other types of enterprises. State-owned enterprises (SOEs) and listed companies have a demand of 833 and 543 respectively, showing that these enterprise types also have a significant demand for professionals in the field of AI. In contrast, government agencies and representative offices of foreign companies have relatively low demand for hiring in the field of AI, with only 3 and 4 positions, respectively.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on September 26,2024 at 18:49:29 UTC from IEEE Xplore. Restrictions apply.

The distribution of companies in the AI sector reflects the trend towards commercialization and privatization, with private companies demonstrating innovation and dynamism. State-owned enterprises (SOEs) and listed companies play an important role in large-scale investments and long-term R&D projects, providing a stable employment environment. Professionals can choose to diversify their employment paths, such as startups, SOEs or global opportunities., or look globally for opportunities in listed and foreign-owned companies.

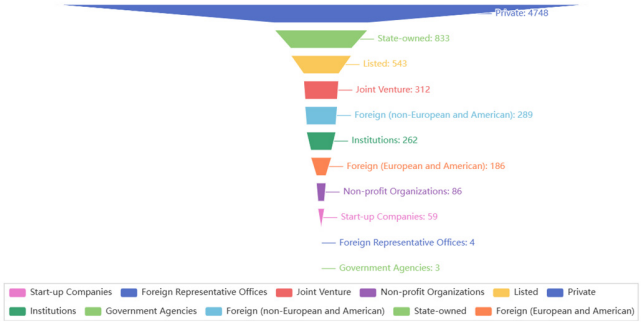

Fig. 8. Distribution of enterprise nature

D. Demand for human resources and skills

Among the talent cultivation and education and training programs in the field of artificial intelligence, the market shows a clear preference for the demand for specific skill positions. The analysis results show that Algorithm Engineer, AI Artificial Intelligence Development Engineer and Intelligent System Architect are the three most in-demand positions, with 1,050, 628 and 451 hiring demands, respectively, highlighting the importance of these technical positions and the high demand for these skills in the market. Meanwhile, Data Mining Engineer and Natural Language Processing Engineer positions also demonstrated considerable demand, confirming the breadth and diversity of employment opportunities in AI. (Fig. 9) In addition, word cloud graph analysis reveals the hotness of skills and concepts such as graphic designers and algorithm engineers, although the market demand for these skills may not be as high as for positions directly related to AI technology.(Fig. 10)

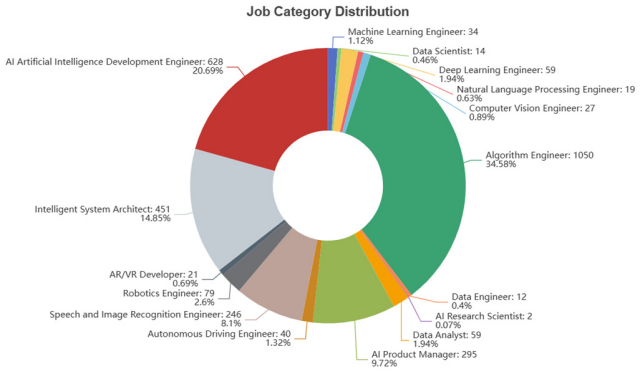

Fig. 9. Job Category Distribution

Fig. 10 WorldCloud of Job Categories

IV. DISCUSSION

A. Salary Level Status and Trends in the Artificial Intelligence Field

In the analysis of salaries and career development in the field of artificial intelligence, professionals with mid-level experience (3-4 years) were found to be the most sought-after, [5] emphasizing the importance of real-world project experience in boosting employment competitiveness and salaries. At the same time, educational qualifications also have a significant impact on salary; [6] while undergraduate positions are the most numerous, master's and doctoral degree holders typically enjoy a higher salary advantage. For example, if the graph shows that the average top salary for doctoral degrees is significantly higher than other degrees, this could indicate that advanced research and specialized skills are valued in the AI industry. It is recommended that first-time entrants actively build up their professional skills and project experience, and consider upgrading their education to a master's degree or above. Educational institutions should provide practical courses and encourage internship experience in response to market demand, so as to help students and job seekers make informed career planning and education investment decisions.

B. Geographical Distribution Characteristics

The demand for jobs in China's AI industry is geographically concentrated in cities leading in technology development and application, such as Wuhan, Chengdu and Hangzhou, creating a clear phenomenon of clustering of talent and companies. [7] This clustering trend has led to a favorable industrial ecology, concentrating capital, policy and technology resources and attracting more talents and enterprises. For job seekers, understanding these clustering cities can help them choose the right place to develop their careers, and especially for those who are new to the industry, may significantly improve their chances of finding a satisfying job. Meanwhile, policymakers can adjust industrial policies based on these data, such as strengthening support for high-tech zones to promote the further development of AI technology and industrial upgrading, creating more opportunities for companies and job seekers. These insights into industry development from geographical characteristics provide an important reference for enterprises and policy makers.

C. Enterprise Recruitment Demand

Although state-owned enterprises and listed companies are fewer in number, they play an important role in key technology research and development by virtue of their strong capital and

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on September 26,2024 at 18:49:29 UTC from IEEE Xplore. Restrictions apply.

brand advantages. Government agencies and foreign companies, though small in recruitment, offer special job opportunities, such as public service applications and internationalized work environments. This diversity of companies provides job seekers with a wealth of employment options and helps individuals choose the right work environment and development path according to their career plans. [8]

D. Suggestions for Talent Cultivation and Education and Training Programs

Educational institutions and training centers should design market-oriented curricula that focus on cultivating key skills in AI fields such as algorithm development, data mining, and natural language processing, and emphasize the combination of theory and practice so that students can accumulate experience through actual projects. [9] The curriculum should adopt an interdisciplinary education model that integrates knowledge of computer science, mathematics, statistics and industry applications to cultivate composite talents who can adapt to the application of technologies in multiple fields. At the same time, the curriculum needs to support lifelong learning, [10] be continuously updated to keep up with technological developments, and increase the weight of practical applications, such as providing real-world learning experiences through internships and project collaborations, to ensure that learners are able to transform their knowledge into practical working capabilities. Knowing which skills and jobs are most in demand can help them make more targeted decisions about their education and career development. For example, if a person is interested in "deep learning" skills and there is a high demand for them, they may choose to pursue further specialized training or a related degree program.

V. CONCLUSION

This study uses data mining and visualization techniques to deeply analyze the demand characteristics of AI jobs, including multiple dimensions such as salary level, work experience, educational requirements and geographical distribution. Key findings include: the market demand for mid-level talents with 3-4 years of work experience is high, reflecting that professionals with this level of experience are particularly important in the field of AI; despite the fact that there are more jobs for job seekers with bachelor's degree, those with higher education enjoy better salaries, showing the significant role of higher education in obtaining quality jobs. In addition, the demand for jobs shows a trend of geographic clustering, especially in cities such as Wuhan, Chengdu and Hangzhou, where high-tech industrial parks play a key role in promoting industrial clustering and technological innovation, providing job seekers with an important guideline for geographic positioning. An analysis of the nature of enterprises shows that private companies are hiring the most in the AI field, while state-owned and listed companies also show significant demand for AI talent. Overall, these findings provide practical advice for job seekers, educational institutions, and policy makers in the AI field, and demonstrate the potential for the application of data mining and visualization techniques in job market analysis. With the demand for specialized talent expected to continue to grow as AI technology further develops and its applications expand, it is extremely important for all industry participants to stay on top of industry developments and the latest trends in the market.

VI. LIMITATIONS

Despite the progress made in this study in analyzing the demand for AI jobs, there are some limitations. First, the data were mainly sourced from the 51JOB website and did not cover other recruitment platforms, which may have affected the comprehensiveness and accuracy of the analysis. Second, the data is based on snapshots at specific points in time, which may not accurately reflect long-term trends and cyclical changes. In addition, despite the application of data mining and visualization techniques, the in-depth analysis of data, especially the processing and parsing of text data, is still inadequate. Future research can improve the representativeness, reliability and depth of the study by integrating multiple data sources, conducting long-term analysis and adopting advanced techniques such as machine learning, so as to better serve policy formulation and talent cultivation in the field of artificial intelligence.

REFERENCES

[1] Di Battista, A., Grayling, S., Hasselaar, E., Leopold, T., Li, R., Rayner, M., & Zahidi, S. (2023, May). Future of jobs report 2023. In World Economic Forum, Geneva, Switzerland. https://www. weforum. org/reports/the-future-of-jobs-report-2023.

[2] Jackson, J. (2002). Data mining; a conceptual overview. Communications of the Association for Information Systems, 8(1), 19.

[3] Khan, M., & Khan, S. S. (2011). Data and information visualization methods, and interactive mechanisms: A survey. International Journal of Computer Applications, 34(1), 1-14.

[4] Bonsay, J. O., Cruz, A. P., Firozi, H. C., & Camaro, P. J. C. (2021). Artificial intelligence and labor productivity paradox: The economic impact of AI in China, India, Japan, and Singapore. Journal of Economics, Finance and Accounting Studies, 3(2), 120-139.

[5] Liu, J., & Guan, X. (2020). Job information data analysis and visualization system implementation based on Python crawler. Softw Eng Appl, 9(4), 317-325.

[6] Ashenfelter, O., & Mooney, J. D. (1968). Graduate education, ability, and earnings. The Review of Economics and Statistics, 78-86.

[7] Xiangfeng, L. (2007). SME development in China: A policy perspective on SME industrial clustering. Asian SMEs and Globalization, ERIA Research Project Report, 5, 37-68.

[8] Garai-Fodor, M., Vasa, L., & Jäckel, K. (2023). Characteristics of segments according to the preference system for job selection, opportunities for effective incentives in each employee group. Decision Making: Applications in Management and Engineering, 6(2), 557-580.

[9] Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., ... & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy.

[10] Shiohira, K. (2021). Understanding the Impact of Artificial Intelligence on Skills Development. Education 2030. UNESCO-UNEVOC International Centre for Technical and Vocational Education and Training.

Authorized licensed use limited to: Hellenic Mediterranean University. Downloaded on September 26,2024 at 18:49:29 UTC from IEEE Xplore. Restrictions apply.