[1*]Baohua Zhang

A Study of Structural Change Trends in the Online Labor Market: A Quantitative Analysis Based on Big Data Techniques

Journal of Electrical Systems

***Abstract:* -** The global labor market is experiencing unprecedented structural changes as the field of artificial intelligence continues to advance under the rapid development of information technology. This study aims to reveal the dynamic changes in market demand and their impact on the education and training system by quantitatively analyzing the online labor market, especially the recruitment information of AI-related jobs. The study employs text mining techniques to analyze large amounts of online recruitment data in order to build a skills lexicon and identify job skill needs. By analyzing job postings in the field of artificial intelligence, the study reveals important changes in skill demand and salary structure. The results of the study show that the skill demand for machine learning and data analytics is as high as 229.05 and 130.36 frequencies, indicating an increased demand for highly skilled labor. In addition, the salary analysis showed that nearly half of the positions in the image processing field paid between $100,000 and $200,000 per year. The results of the study provide an important reference for higher education institutions and policy makers in adapting to technological change, emphasizing the need to update educational curricula and enhance skills training.

Keywords: artificial intelligence, big data, online labor market, skill demand, text mining

1. Introductory

In the past decade, the new generation of information technology, centered on artificial intelligence, has led the

[1] *Corresponding author: Baohua Zhang, laughtogether@163.com

Author 1 Affiliation: Talent Service Section, Jiaxiang County Public Employment and Talent Service Center, Jining City 370829, China

opening of the digital era [1]. The global economy has ushered in a new technological revolution driven by industrial intelligence and digitalization [2-3]. The United States, Germany and other Western countries have implemented strategies such as "Industrial Internet" and "Industry 4.0", aiming to promote industrial re-innovation through intelligent technology, and trying to lead the global industrial division of labor again [4]. The advancement of intelligence has not only become a new driving force for the economic growth of various countries, but the global artificial intelligence market has also shown a trend of rapid development [5]. At present, it is becoming more and more common for enterprises worldwide to apply intelligent technologies, and the integration of production processes with intelligent technologies is deepening [6]. Global labor productivity continues to increase due to the application of smart technologies, the development of smart technologies will significantly promote the formation of a data-driven smart economy, and the structure of the labor market in the era of the smart economy will experience structural changes [7]. With the increase in the number of enterprises, the benefits of technological progress may not be fully realized in this structural change, and may even face underemployment and deterioration of the distribution structure if the knowledge and skills of job seekers cannot be converted and upgraded in a timely manner to adapt to the new job requirements [8].

In the current stage of research on online job postings, scholars are mainly divided into two schools. The first type of research is devoted to extracting subject terms from online recruitment data in order to clarify the details of the requirements of different positions [9-11]. While the second type uses text mining techniques to analyze job descriptions and then build job skill dictionaries, which helps to reveal the core skills required in various industries [12-13]. With the rapid development of the field of artificial intelligence, the number of related positions has gradually increased, attracting the attention of many job seekers. Research for the field of artificial intelligence is particularly important, not only to clarify the specific skill requirements of the industry, but also to help job seekers understand the knowledge they must acquire and the employment prospects of the industry in different regions [14]. Although research on online job postings has been carried out extensively, relatively few studies have been conducted on the AI industry. This study chooses to focus on job postings for AI positions, aiming to gain a deeper understanding of the industry's needs and provide prospective job seekers with an analysis of job requirements, as well as hopefully providing a reference for institutions of higher education in the construction of an AI education system. Through comprehensive analysis, this paper can enable potential job seekers to have a more comprehensive grasp of the career needs in the field of artificial intelligence and be fully prepared for finding the right position.

In this study, a skill dictionary for the field of artificial intelligence is constructed and relevant job postings are analyzed in-depth using text mining techniques to reveal the job needs and trends in the field. A wider range of data resources were accessed through Python. In the information processing session, automation techniques were used to significantly optimize the data processing process and improve efficiency. In addition, methods such as correlation analysis were used to explore the interactions of different dimensions in the job information, providing job seekers with a deeper understanding of the position. By converting unstructured data into a structured format and analyzing it in detail using tools such as skill dictionaries and word cloud maps, this study helps job seekers accurately understand and adapt to job requirements. The application of text mining techniques not only highlights key employability skills, but also supports students and new entrants to the industry to adjust their learning and skill development strategies according to market needs and to fully prepare for their future careers.

2. Literature Review and Theoretical Foundations

2.1 Online Labor Market

With the rapid advancement of Internet technology, corporate human resource planning has experienced significant changes, and the traditional offline labor market has been impacted by the Internet. The gradual rise of online recruitment platforms, such as job boards and apps, has spawned a transition from traditional offline to online labor markets [15].

The online labor market eliminates geographical and time constraints and improves the matching mechanism between recruiters and candidates, making it more flexible and precise. Advances in information technology, especially the popularization of the Internet, have changed the way recruiters and candidates search and match with each other. This web-based recruitment method allows both parties to quickly browse information at any time and any place, mediated through an online platform, to realize efficient job search and recruitment activities [16].

The study also shows that there are significant differences between traditional labor market and online labor market in terms of virtualization of participating subjects and flexibility of employment relationship. Internet recruitment websites optimize the talent matching process, reduce intermediate links and improve recruitment efficiency through recommendation, search and pre-screening functions [17]. Internet recruitment also reduces information asymmetry, enhances the intelligence and personalization of recruitment services, and makes the matching between employers and employees more qualitative and efficient [18].

Typical features of online labor markets include timeliness, dynamism, and lack of spatial limitations, making them a gathering place for emerging technologies to post recruitment information. In particular, recruitment activities in high-tech industries such as computers, network communications, and e-commerce are concentrated here, and the recruitment targets usually have a high level of literacy [19].

Based on the above analysis, this study adopts a narrow perspective to define "online labor market", focusing on job postings on the Internet. The target of the study is emerging technology jobs such as big data and artificial intelligence, and recruiting companies tend to release information through online platforms. Considering the needs of the study, the online labor market is selected as the data collection site, which is convenient for the rapid and extensive collection of AI and big data related job information nationwide, while the traditional offline market is not applicable to the data collection needs of this study.

2.2 Impact of the Application of Smart Technologies on the Labor Market

Since the Industrial Revolution, technological progress has been the core driver of economic growth, which has led to a continuous change in the skill demand of global talent. Many scholars have studied the relationship between technological progress and workers' skill demand, among which Acemoglu proposed the Skill-Biased Technical Change (SBTC) theory, which emphasizes that the adoption of new technologies raises the demand for high-skilled and highly educated personnel, which further promotes the popularization and application of new technologies and forms a This further promotes the popularization and application of new technologies, forming a complementary relationship between technology and skills [20].

By analyzing data on skill changes in the United States from 1980 to 1990, Acemoglu reveals that skill-preferred technological advances not only promote the complementarity of technology and skills, but also cause changes in the structure of employment [21]. In recent years, with the rise of new-generation technologies such as big data and artificial intelligence, scholars have also begun to pay attention to the impact of these technologies on the labor market. The changes in the labor market are reflected in three aspects: first, the rapid development of industries based on scientific and technological knowledge, leading to changes in the industrial structure; second, the increased demand for talents with high skills and high education systems; and third, the changes in the employment model, with a more flexible market, richer benefits, and explicit remuneration for knowledge and skills [22].

Rickardo & Meiriele's (2023) study shows that big data and artificial intelligence, as representatives of the sixth technological revolution, have changed traditional labor patterns and promoted innovation, informatization, digitization, and intelligence in the production process, and the demand for skills in the labor market tends to be diversified, and the knowledge system of laborers tends to be more complex [23].Xie et al. .(2021) argued that such technological advances will continue to drive the labor market to shift in a technology-biased direction, with traditional jobs being replaced by automation, thus creating new positions and a significant increase in the demand for high-skilled personnel [24].

The theory of technology-preferred technological progress suggests that the use of new technologies increases the demand for highly skilled and educated personnel and promotes the development and popularization of new technologies, which, in turn, create a high degree of dependence on skilled personnel. The impact of technological progress on the quality of the labor force is manifested in the increased demand for highly skilled and highly educated labor. The tendency of technological progress to shift the labor market demand towards higher skills has also been confirmed through studies in different countries [25].

Although current research has explored the impact of the widespread use of information technology on the labor market based on Acemoglu's theory of skill-preferred technological advancement, there is a relative lack of quantitative research addressing the impact of big data and artificial intelligence technologies on the online labor market. This study aims to quantitatively analyze intelligent technology job information through the topic modeling approach in text mining in order to reveal the structural impact of information technology on the talent recruitment market when it is widely used in companies.

2.3 Application of Text Mining Technology in the Study of Talent Market Demand

Text Mining (TM) is the process of analyzing unstructured text data for key information and patterns. The definition of this field usually includes extracting unknown, understandable and useful knowledge from large amounts of text and using this knowledge to optimize the organization of information for future referencing and use [26]. Unlike traditional analyses for structured data, the free-flowing and highly unstructured nature of textual data makes it difficult to apply standard analytical methods directly, so it is often necessary to transform textual data into some form of structured data [27].

The popularization of the Internet has made online recruitment platforms an important channel for posting job information. These platforms not only facilitate the exchange of information between enterprises and job seekers, but also provide scholars with opportunities to study market demand and improve the problem of information asymmetry in the job market. By analyzing online job postings, researchers can reveal market demand for specific skills, trends in job requirements, and other important career characteristics, thus providing graduates with better career guidance and helping them cope with the competitive job market more effectively [28]. The application of text mining techniques in the recruitment market is reflected in frequency analysis, keyword extraction, and pattern recognition. These techniques can help extract useful data from a wide range of job postings, such as basic requirements and specialized skill requirements in job descriptions. In addition, more advanced analysis methods, such as sentiment analysis and topic modeling, can provide further insights into deeper market trends and candidate needs behind job information [29].

Currently, many foreign researchers use text mining techniques to analyze the needs of the talent market. Earlier, Todd et al. (1995) analyzed information systems-related job postings in U.S. and Canadian newspapers by extracting keywords and frequencies [30].Lee (2006) studied information technology management positions and built a preliminary skills categorization directory and dictionary [31].Sodhi & Son (2009), on the other hand, collected from online recruitment platforms Operations Research related job postings and built a professional skills dictionary, which not only performed frequency analysis, but also extended to cross-tabulation and correlation analysis, making the study more in-depth [32].Smith & Ali (2014) collected information on software development positions and analyzed the demand trend of programming languages using keyword indexing techniques, based on which they made recommendations for related professional skills courses [33]. recommendations [33]. These studies show that an in-depth study of recruitment data is feasible in foreign countries. However, due to the complexity of the Chinese language, these English analysis methods are not fully applicable to Chinese data processing.

3. Research methodology

3.1 Information Collection and Selection

At present, network recruitment is mainly carried out in two ways: first, through the company's official website to obtain recruitment information, although this way of information is more reliable, but in view of the numerous and chaotic job advertisements released by many enterprises, which is more inconvenient for job seekers; second, with the help of 51job, BOSS direct employment, Lahou and other online recruitment platforms. These platforms provide not only a large number of recruitment information, but also updated rapidly, which can reflect the current needs and trends of the talent market. In view of this, this study chooses to collect recruitment information data on 51job website. On the website of MileagePlus, the posted job information is mainly divided into three main content blocks: firstly, enterprise-related information, covering the name of the enterprise, the nature of the enterprise, the size of the company and the industry in which it is located, etc.; secondly, job-related information such as geographic location and salary level; and lastly, the details of the position, including the job duties and qualifications. After determining the objectives of data collection, the next step is the actual information collection process. The method used in this study is the web crawler technique, specifically using the Python programming language to write scripts to automate the process of crawling the relevant data on the MileagePlus website. This method is not only efficient, but also capable of processing and analyzing recruitment data on a large scale to support the research. The specific steps are as follows.

In certain cases, we can manually obtain all the URLs. using Selenium automation tools can handle complex operations such as automatic page flipping that require logging in, so we only need to obtain the URL of the first page. considering that the research objective is AI-related positions, to simplify the crawler code, we can filter the job information directly through the website, selecting nationwide as the workplace and full-time as the nature of the job, and obtain the page URLs after these conditions. We use different parsing techniques for the HTML structure of the page: first, we review the layout structure of the page, and then we use tools such as regular expressions or XPath to parse it. For example, open the job page of the MileagePlus website and view its source code.

In Spyder software, the following Python code is executed to crawl the AI-related job postings published on the MileagePlus website between March 1 and March 18, 2023, and finally 8100 data are crawled from the website and stored in an Excel table. The data we obtained includes both structured data, such as job title, company name, work location, salary, posting time, salary type, company size, work area, experience, education, and unstructured job information. The specific algorithm code is as follows:

```
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
# Configure the Selenium WebDriver
driver = webdriver.Chrome()
# Define the URL for the initial page to start scraping
start_url = "http://www.51job.com"
driver.get(start_url)
# Here, input the relevant job position and set other filters like location to 'national' and job type to 'full-time'
# After setting the filters, get the page URL where the data resides
# For demonstration, we're assuming the filtered URL is directly accessed below
filtered_url = "http://www.51job.com/ai-jobs/"
driver.get(filtered_url)
# Use BeautifulSoup to parse the page content
soup = BeautifulSoup(driver.page_source, 'html.parser')
# Create a list to hold all the job data
```

```
job_list = []
# Find all job postings based on the div class where they are located
jobs = soup.find_all('div', class_='tHeader tHjob')
for job in jobs.
    job_info = job.find('div', class_='company main')
    company_info = job.find('div', class_='tHeader sidebar')
    # Extract structured data
    job_title = job_info.find('h1').text.strip()
    company_name = company_info.find('h2').text.strip()
    location = job_info.find('span', class_='lname').text.strip()
    salary = job_info.find('strong').text.strip()
    # Extract unstructured data
    job_description = job_info.find('div', class_='bmsg job_msg inbox').text.strip()
    # Store the job data in a dictionary and append to the list
    job_data = {
        'Job Title': job_title,
        'Company Name': company_name,
        'Location': location,
        'Salary': salary,
        'Job Description': job_description
    }
    job_list.append(job_data)
# Convert the list of dictionaries to a DataFrame
df = pd.DataFrame(job_list)
# Save the DataFrame to an Excel file
df.to_excel('AI_Jobs_51job.xlsx', index=False)
# Close the browser
driver.close()
print("Data scraping completed and saved to Excel.")
```

3.2 Information Data Pre-processing

Since companies may repeatedly post the same job information at different points in time, the dataset may therefore contain a large number of duplicate entries. The existence of these duplicate data poses an adverse effect on the accuracy and efficiency of the study. Therefore, cleansing of job posting data is crucial, including the culling of duplicates as well as the removal of those records with incomplete information. Specifically, the de-duplication process refers to the removal of duplicate job postings by the same company, while the culling of invalid information involves the removal of those records with incomplete job descriptions. Through this series of data cleansing steps, a total of 5989 valid data were finally compiled, which will provide more accurate and reliable data support for subsequent studies.

3.2.1 Preprocessing of Structured Data

Important preprocessing steps in job posting data analysis include ensuring data consistency and comparability. First, the name of the city to which the job belongs needs to be standardized, retaining only the name of the city to facilitate geographic location analysis. Second, company types need to be standardized as private companies, startups, state-owned enterprises (including institutions), foreign companies (subdivided into foreign (European and American), foreign (non-European and American), and joint ventures), and listed companies, in order to analyze the hiring conditions and salary levels of different types of companies. Company size should be divided into six ranges: less than 50, 50-149, 150-499, 500-999, 1,000-4,999, and more than 5,000, in order to conduct correlation analysis between company size and factors such as recruitment demand and salary. Work experience needs to be unified into five levels of no experience required, 1-3 years, 3-5 years, 5-10 years, and more than 10 years, in order to facilitate the analysis of the correlation between experience and salary package and job type. Salary levels should be standardized into 10,000/year units, and divided into six ranges of less than $100,000, $100,000-200,000, $200,000-300,000, $300,000-400,000, $400,000-500,000, and more than $500,000 per year based on the average of the upper and lower limits of the salary to facilitate the comparison of the salary levels and talent demand for different positions. These steps ensure the accuracy and validity of recruitment information data when performing descriptive and correlation analysis.

3.2.2 Preprocessing of Unstructured Data

In online recruitment, unstructured data mainly refers to the job description text, which details the job duties and requirements. Effective mining of these textual data requires firstly a lexical processing, i.e. breaking continuous text into keywords or phrases. The job requirements of recruitment information can be subdivided into basic conditions (e.g., academic qualifications, majors, etc.), professional skills (e.g., specific algorithmic frameworks, mastery of programming languages), and comprehensive personal qualities (e.g., motivation, communication skills, ability to analyze problems, etc.). By analyzing these data, the key skills and quality requirements of a specific industry or position can be revealed, providing direction to job seekers and helping them target to improve relevant skills and comprehensive quality to increase the chances of getting the desired position. This kind of data analysis not only helps individual career planning, but also provides a scientific basis for decision-making in human resource management.

In this paper, we chose to use the jiebaR package of R language for text segmentation, and the segmentation engine adopted is MixSegment, which effectively combines the dictionary-based maximum probability method and the statistics-based Hidden Markov Model, thus improving the accuracy and efficiency of segmentation. Given the specificity of the research data, which includes a variety of statistical software titles as well as algorithms related to artificial intelligence, the article further customizes the segmentation lexicon according to the research needs. For this purpose, AI-related lexicons were downloaded from Sogou in order to more accurately identify and process these specialized terms. This approach not only optimized the precision of the participle dictionary, but also ensured that the results were more closely aligned with the domain characteristics of this study.

The presence of elements such as numbers, the auxiliary "of" and conjunctions can be observed in the results of the above word division, as well as the words "skillful" and "mastery", The words "skilled", "master", "relevant", "priority" and "function" appear frequently in the text. These high-frequency words do not have substantial value when analyzing job information, but may interfere with the accurate identification and interpretation of key information. For this reason, it is important to perform lexical filtering to exclude these non-critical words. This process helps to remove the noise and extract the information directly related to the job more accurately, thus improving the efficiency and accuracy of information mining.

4. Analysis of results

4.1 Word Frequency Analysis

The Chinese word segmentation tool jiebaR in R language is used to segment the "job description text data" in the field of big data and artificial intelligence. On this basis, the existing Sogou thesaurus is used as a specialized thesaurus, and the deactivated dummy word database comes with the R language. In addition, according to the research needs of this paper, the corresponding specialized thesaurus and deactivated thesaurus are constructed. As shown in Table 1 below.

Table 1 Self-checking thesaurus

form	specialized thesaurus
self-constructed word	Data Analytics, Big Data, Artificial Intelligence, Business Applications, Internet of Things, Mobile Internet, VR, Language Processing, Machine Learning, Data Mining

In the word frequency statistics of the text, the TF-IDF algorithm is used to convert the frequency values to freg values, and the advantage of the conversion is that the common meaningless words can be eliminated.

In the AI job postings shown in Table 2, the subwording and sorting of the job descriptions show that the keywords that ranked high in keyword frequency include machine learning, engineer, internet, robotics, Python, data analysis, image processing, and natural language processing. The high frequency of these terms indicates that the current AI industry has very clear requirements for candidates' expertise, with particular emphasis on machine learning and deep learning skills. In addition, the job posting highlights essential knowledge of programming languages such as Python, as well as expertise in areas such as data analysis, image processing and natural language processing.

Table 2 Artificial Intelligence Job Description Word Frequency

serial number	Word	freq
1	machine learning	229.0497
2	hiring out	173.1942
3	the Internet	158.1382
4	calculators	157.1572
5	mechanical person	156.8368
6	python	143.1780
7	data analysis	130.3645
8	image processing	127.5362
9	java	120.8441
10	Full Content	120.5337
11	sense of responsibility	119.3766
12	data mining	117.9341
13	master	109.2067

14	prescription	103.2663
15	linux (computer)	98.5794
16	automatic	96.4398
17	natural language processing (NLP)	94.9684
18	comprehensive database	93.1765
19	logical thinking	91.5307
20	tensorflow	91.4740

Figure 1 below shows a more intuitive and in-depth word cloud display of the AI word frequency ranking.

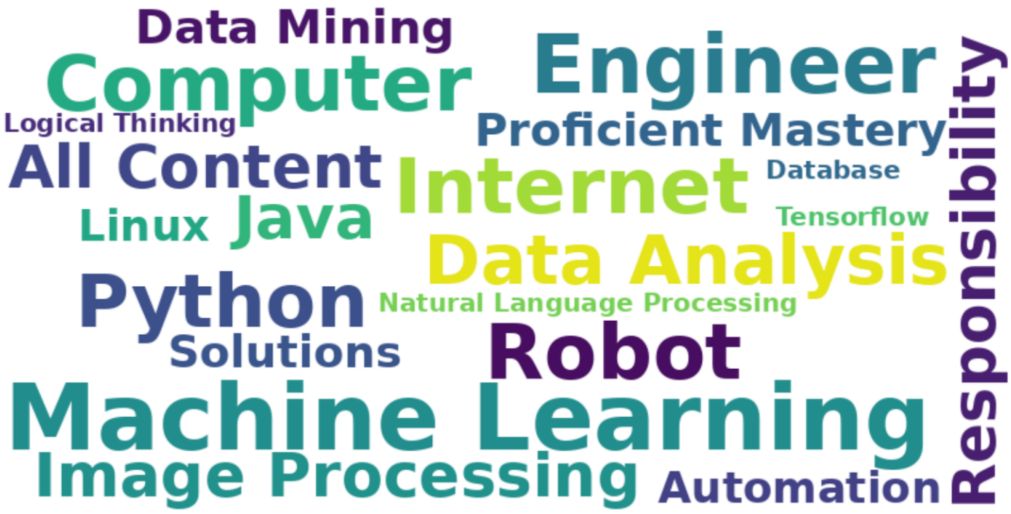

Fig. 1 Word frequency cloud of artificial intelligence job descriptions

From the word frequency table and word cloud diagram of "job description" of AI job information, it can be seen that the demand for AI positions not only needs to be proficient in machine algorithms, but also needs to be equipped with big data-related tools and analytical capabilities of multidisciplinary interactive talents.

4.2 Analysis of Artificial Intelligence Job Requirements

This article provides a detailed analysis of job requirements in three technical areas of artificial intelligence jobs: image processing, speech recognition and drones. Of the data collected, the image processing field had the most job postings, totaling 1,209. The number of job postings in the area of speech recognition technology was 654, while there were 632 job postings for positions related to drone technology, bringing the overall job posting data to a cumulative total of 2,495.

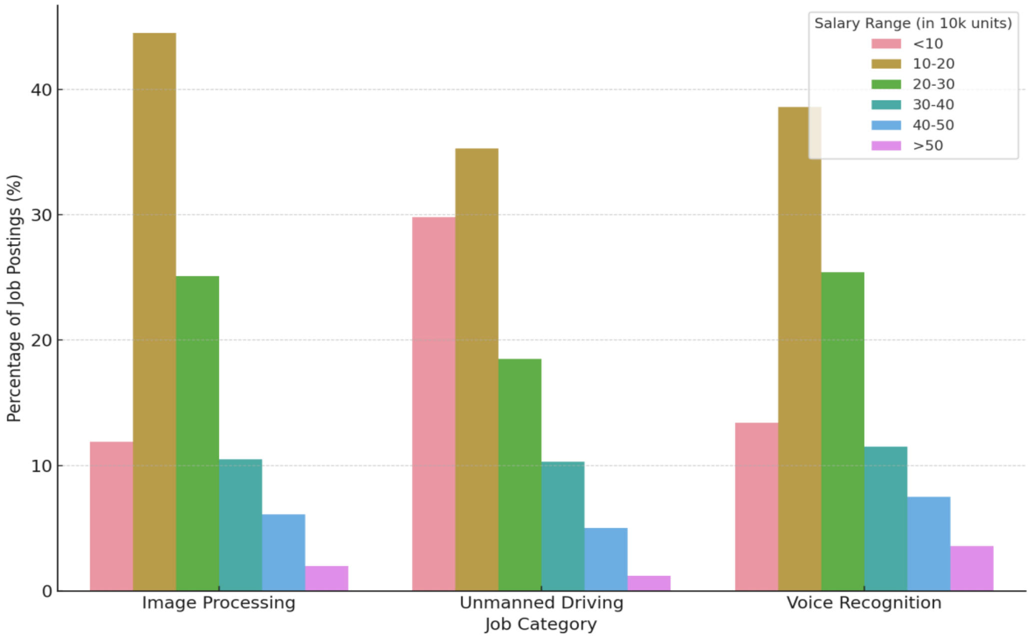

Figure 2 Salary levels between specific job categories in Artificial Intelligence

Figure 2 reveals the relationship between different positions and their corresponding salary levels. In the image processing field, nearly half of the positions have an annual salary between $100,000 and $200,000; 25.1% of the positions lie in the range of $200,000 to $300,000, 11.9% of the positions have an annual salary of less than $100,000, and about 20% of the positions have an annual salary of more than $300,000. In the field of drones, 35.3% of the positions have an annual salary between $100,000 and $200,000, 29.8% have an annual salary of less than $100,000, 18.5% are in the range of $200,000 to $300,000, and more than $300,000 account for 17% of the positions. As for the speech recognition field, 38.6% of the positions have an annual salary between $100,000 and $200,000, 25.4% are in the range of $200,000 to $300,000, 13.4% have an annual salary of less than $100,000, and more than $300,000 account for about 23% of the positions. This indicates a significantly uneven distribution of salaries even within the same category, with the largest number of jobs in the $100,000 to $200,000 salary bracket, followed by those in the $200,000 to $300,000 salary bracket. Especially in the unmanned field, the proportion of salaries below 100,000 is high. SPSS was utilized to conduct the test, and the test results are as follows:

Table 3 Cardinality test

	numerical value	df	Bilateral significance
chi-square (math.)	105.841	10	0.000
similarity ratio	95.236	10	0.000
linear correlation	15.259	1	0.000
Number of observations	2495		

The results of the test in Table 3 show that the p-value is less than 0.05, so we reject the original hypothesis, which indicates that there are significant differences in the minimum educational requirements for different job categories. In general, among the three job categories, image processing-related jobs have the most stringent educational requirements, followed by speech recognition-related jobs, while driverless-related jobs have relatively low educational requirements. This finding may be related to the technical complexity and professional knowledge requirements of each field, reflecting the different degrees of reliance on professional ability and theoretical knowledge in different technical fields. This difference in educational requirements provides job seekers with important information about skills and educational background preparation.

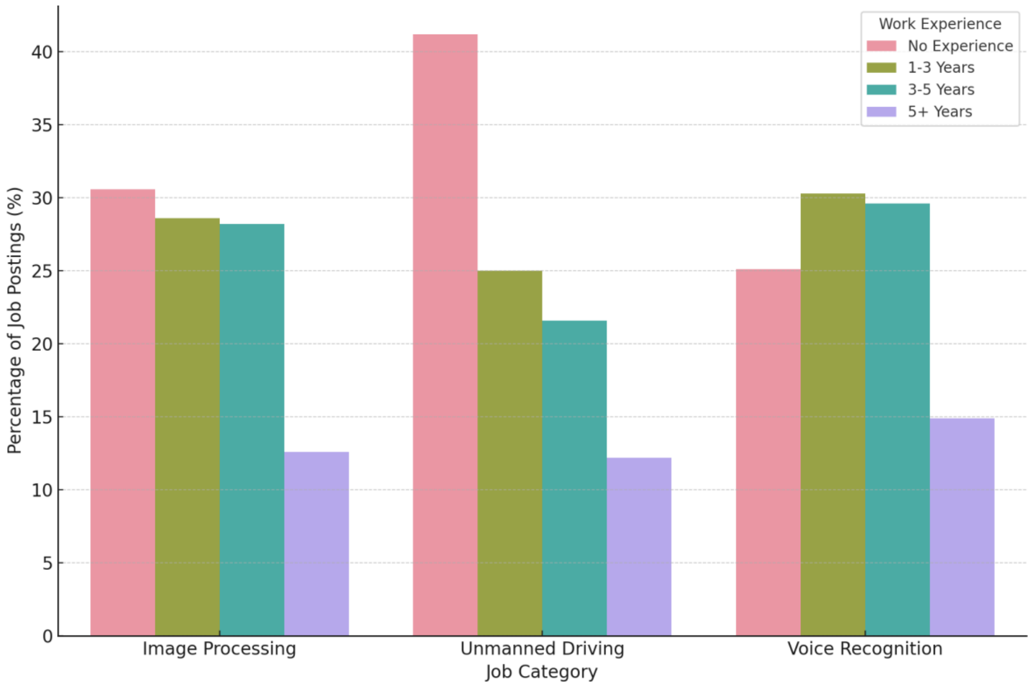

Figure 3 Work experience between specific job categories in Artificial Intelligence

Figure 3 illustrates the relationship between different job categories and work experience requirements. In the image processing domain, positions are close to the demand for no work experience, 1-3 years of experience, and 3-5 years of experience, with 30.6%, 27.6%, and 29.3%, respectively, while for more than 5 years of experience, the percentage of positions is 12.5%. For the driverless field, 41.2% of the positions have no specific requirements for work experience, 24% of the positions require 1-3 years of experience, 22.6% of the positions require 3-5 years of experience, and the percentage of positions with more than 5 years of experience is 13.3%, showing that the overall requirements for experience in this field are low. In the field of speech recognition, most positions require 1-3 years or 3-5 years of experience, and the percentage of positions in these two experience ranges is 30.3% and 285%, respectively. The percentage of positions requiring no experience is 24.1%, while the percentage of positions with more than 5 years of experience is 15.9%, and the data reveals the variability of work experience requirements in different technical fields. The results of the chi-square test are as follows:

Table 4 Chi-square test

	numerical value	df	Bilateral significance
chi-square (math.)	43.554	6	0.000
similarity ratio	41.685	6	0.000
linear correlation	8.274	1	0.000
Number of observations	2495		

The results of the test in Table 4 indicate that the p-value is less than 0.05 and hence the original hypothesis is rejected. This indicates that there is a significant difference in the work experience requirement between the different job categories. Specifically, the driverless related jobs have the lowest requirements for work experience, followed by the image processing field. Comparatively, the speech recognition field has the most stringent work experience requirements.

4.3 Job Relevance Analysis

4.3.1 Data processing

This study focuses on analyzing the interrelationships of multidimensional indicators in job postings, and adopting association rules to reveal the intrinsic connections of factors affecting high salaries. The study covers five dimensions: experience requirement, education requirement, salary range, job type and work location. In order to effectively carry out the association analysis, this paper uses the Apriori algorithm in R language to reprocess and code the data. Specific methods include: artificial intelligence job types are divided into image processing, speech recognition, and unmanned driving; work experience is categorized as no experience required, 1-3 years, 3-5 years, and more than 5 years; salary levels are adjusted to less than \$100,000, \$100,000-\$200,000, \$200,000-\$300,000, and more than \$300,000 per year; and work locations are categorized into north, south, east, and north and south of the country, as well as other cities. Through these measures, this paper aims to dig deeper into the relationship in the recruitment information and provide scientific basis for the recruitment strategy and salary structure.

4.3.2 Correlation analysis

After processing the recruitment data, the Apriori algorithm in R language is used to analyze the association. The initial threshold value of support is 0.03, the initial value of confidence is 0.5, and the minimum number of associations is 2. 265 association rules are obtained through the analysis. In order to further explore the key factors affecting the salary level, the following association rules are filtered with salary as the key right association.

This analysis explores the salary levels of different positions in the AI industry in the context of a variety of factors through association rules. Key findings include: for salaries below \$100,000 per year, this is usually associated with people who have no work experience, are located in an average Tier 1 city, and are working in a drone-related position. For positions between \$100,000 and \$200,000 per year, this salary range was found to be mostly associated with professionals working in average Tier 1 cities, with 1-3 years of experience, and working in image processing. In contrast, annual salaries of more than 300,000 RMB are usually seen among those working in developed cities such as North, Shanghai, Guangzhou and Shenzhen with more than 5 years of work experience.

It can be seen that salary levels are affected by a number of factors, with work experience and direction of employment being the main influences. The richer the work experience, the higher the salary usually is. At the same time, the chosen job category also has a significant impact on salary, such as the salary level of image processing positions mentioned in the analysis is generally higher than that of drone positions. These findings provide an important basis for salary strategies and career planning in the AI industry.

4.5 Analysis of Structural Requirements for Artificial Intelligence Jobs

By analyzing the information of different categories of AI jobs, we found that the job requirements have significant differences. In order to help candidates better understand the skill requirements of various job categories, this study will construct a skill requirement framework for the AI industry. First, textual disambiguation is required to process job information. Given that each industry has its own specific terminology, existing standard word-splitting tools such as the dictionary that comes with iiebaR often fail to meet domain-specific needs. Therefore, this study constructs a specialized skills dictionary on its own based on the AI and computer-related vocabulary in the Sogou cellular library, combined with specific job postings. The construction of this dictionary is crucial for subsequent text data mining.

After completing the construction of the professional dictionary and text partitioning, the next step is the in-depth analysis and mining of job information. As can be seen from existing studies, for example, Wang Ping has constructed a skill indicator system for job information in the e-commerce industry, and Tang Huadong has established a skill indicator system for education technology-related fields. However, the corresponding skill indicator system has not yet been established for the job information of AI industry. This study is based on this practical need and constructs a skill indicator system applicable to the AI industry. The system is mainly divided into two skill dimensions, covering five primary skill indicators and 17 secondary skill indicators, as shown in Table 5 below. The construction of this table not only helps to understand the needs of the industry, but also provides candidates with a clear career development direction and detailed information about the required skills.

Table 5 Specific indicators

Level 1 indicators	Secondary indicators
A Computer skills	A1 programming language
	A2 database system
	A3 Operating systems
	A4 Data processing
B Office automation	B1 Office software
	B2 Document reading
C Interpersonal communication	C1 Communication skills
	C2 Teamwork
	C3 Management coordination capacity
D Professional attitude	D1 Work ethic
	D2 Psychological quality
	D3 Thinking skills
	D4 Capacity for implementation
	D5 Learning capacity
E Professional qualifications	E1 Education
	E2 Work experience
	E3 English proficiency

When analyzing the basic skill requirements of related positions in the image processing field, the data showed that the five most frequent items in the first-level skill indicators were, in order, professional qualifications, computer skills, professional attitudes, interpersonal skills, and office automation skills. When further examining the secondary skill indicators, it is found that in the category of professional qualifications, the need for work experience dominates; in the category of computer skills, the importance of mastering various programming languages is prominent; in the category of professional attitudes, the ability to learn is particularly critical; in the category of interpersonal skills, the ability to work effectively in a team is important; and in the category of office automation skills, the ability to handle documents is the Core. Accordingly, the top five key skills for image processing related positions include: work experience, knowledge of programming languages, relevant education, ability to learn, and ability to work in a team (see Table 6).

Table 6 Relationship between AI job categories and Tier 1 indicators

Job Category	computer skill	office automation	interpersonal communication	professional attitude	professional qualification

image processing	6.94%	1.73%	5.00%	6.02%	11.73%
speech recognition	6.32%	2%	8.47%	7.34%	1.33%
unpiloted	3.77%	0.92%	7.65%	6%	7.85%

In analyzing the skill requirements for the three types of jobs, namely image processing, speech recognition and unmanned driving, we can see that different occupational categories place varying degrees of emphasis on the first-level skill indicators. Image processing jobs focus on professional qualifications, computer skills and professional attitudes; speech recognition jobs emphasize professional qualifications, interpersonal interactions and professional attitudes; and unmanned driving jobs also emphasize professional qualifications, interpersonal interactions and professional attitudes. Although speech recognition and unmanned driving have similar emphasis on the Tier 1 skill indicators, they differ in the specific requirements of the Tier 2 skill indicators. Speech recognition positions primarily require work experience, programming language mastery, communication skills, teamwork, and management coordination; drones highlight work experience, communication skills, teamwork, programming language mastery, and education. The top five skill requirements for image processing positions include work experience, programming language mastery, education, learning ability and teamwork. While the three positions overlap on the Tier 1 skill indicators, they show significant differences in the more specific competency requirements. The image processing position focuses more on experience and technical skills; the speech recognition position requires good communication and coordination skills in addition to technical skills; and the driverless position emphasizes work experience and teamwork skills. This reflects the fact that technical positions in different fields have a clear focus on the specific needs of skills.

5. Discussion

This paper provides an in-depth analysis of AI-related jobs in the online labor market by using modern data technology, especially text mining technology. It is found that with the development of technology, the impact of smart technology on the labor market has gradually become apparent, especially in the field of artificial intelligence, where the demand for highly skilled personnel has increased significantly. This finding is in line with the Skill-Biased Technological Progress Theory (SBTC) proposed by Acemoglu & Restrepo (2018), which emphasizes that technological development promotes the demand for high-skilled personnel. Through extensive data collection and analysis of online recruitment platforms, this study reveals specific needs and market trends for AI jobs. For example, key skills that frequently appear in job descriptions include machine learning, deep learning, and the programming language Python. These results are consistent with research by Xie et al. (2021), which suggests that the growth of big data and AI is driving increased demand for specific skills.

In addition, the findings of this paper demonstrate salary differences between different positions, which are particularly prominent in the field of artificial intelligence. This may reflect the market's emphasis and urgency on different skills, in which the demand for jobs in the fields of image processing and natural language processing is more specialized and technical, and the corresponding salaries are higher. In terms of theoretical contributions, this study not only validates existing theories of skill-preferred technological progress, but also reveals the specific impact of technological progress on labor market structure through empirical data. The methodology of this paper also provides new perspectives and tools for future research, especially in analyzing large-scale online data using text mining techniques. In a practical sense, the results of this study can help companies and educational institutions to better understand the current market demand for skills and thus optimize talent training and recruitment strategies. In addition, for policymakers, understanding how technological advances affect the labor market can help develop more effective employment and education policies to meet the challenges posed by technological change.

6. Conclusion

This study provides an in-depth analysis of structural changes in the online labor market, in the quantitative processing of job postings through big data technology in the field of artificial intelligence, the study finds:

(1) The labor market demand has undergone a dramatic shift . By quantitatively analyzing the recruitment information of AI-related positions, it is found that machine learning and data analysis skills account for the highest proportion of all skill demands, reaching a frequency of 229.05 and 130.36 respectively, which indicates that the demand for high-skilled labor has increased significantly with technological progress. Especially in the fields of image processing and speech recognition, the market's reliance on talents with specialized technical skills is gradually increasing.

(2) Significant differences in salary structure. Salary analysis shows that in the image processing field, nearly half of the positions have an annual salary between \$100,000 and \$200,000, while 35.3% of the positions in the unmanned field have an annual salary in this range. In contrast, about 23% of the positions in the natural language processing field exceeded \$300,000, showing the market's salary differences and emphasis on different technical positions.

(3) Education and training needs are prominent. Changes in the technological needs of the market point to the need for timely adjustments in the education system. For example, the higher demand for work experience and knowledge of programming languages in the field of image processing suggests that relevant education and training should strengthen the teaching of these skills.

While this study reveals important trends in the online labor market, it may not fully cover all the dynamics of the labor market as the data collection is primarily based on online platforms. Future research could capture market changes more comprehensively by expanding data sources and analytical methods. In addition, new job categories and skill needs may continue to emerge as technology advances. Future research should continue to monitor these trends to assess the long-term impact of technological change on the structure of the labor market.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Grygiel-Tomaszewska, A., & Kurkliński, L. (2021). The impact of artificial intelligence on allocative efficiency: insights and gaps in the case of high-frequency trading. in The Economics of Sustainable Transformation (pp. 318-333). Routledge. https://doi.org/10.4324/9781003219958-17.

[2] Wu, L. (2023). The Application and Impact of Artificial Intelligence Technology in International Trade. 2023 6th International Conference on Computer Network, Electronic and Automation (ICCNEA), 228-232. Electronic and Automation (ICCNEA), 228-232. https://doi.org/10.1109/ICCNEA60107.2023.00056.

[3] Xian, F. (2023). Quantifying the Economic Benefits of Artificial Intelligence Technology. Proceedings of the 4th Management Science Informatization and Economic Proceedings of the 4th Management Science Informatization and Economic Innovation Development Conference, MSIEID 2022, December 9-11, 2022, Chongqing, China. https://doi.org/10.4108/eai.9-12-2022.2327579.

[4] Dyba, W., Maria, E., & Chiarvesio, M. (2022). Actions fostering adoption of Industry 4.0 technologies in manufacturing companies in European regions. Investigaciones Regionales - Journal of Regional Research, 2022(53), 27-46. https://doi.org/10.38191/iirr-jorr.22.009.

[5] Wang, L., Sarker, P., Alam, K., & Sumon, S. (2021). Artificial intelligence and economic growth: a theoretical

framework. Scientific Annals of Economics and Business, 68(4), 421-443. https://doi.org/10.47743/saeb-2021-0027.

[6] Yang, T., Yi, X., Lu, S., Johansson, K. H., & Chai, T. (2021). Intelligent manufacturing for the process industry driven by industrial artificial intelligence. engineering, 7(9), 1224-1230. https://doi.org/10.1016/j.eng.2021.04.023.

[7] Ma, Z., & Wang, L. (2021). Identifying the Impacts of Digital Technologies on Labor Market: A Case Study in the Food Service Industry. 2021 IEEE Integrated STEM Education Conference (ISEC), 214-214. https://doi.org/10.1109/ISEC52395.2021.9764118.

[8] Santhosh, A., Unnikrishnan, r., Shibu, S., Meenakshi, K., & Joseph, G. (2023). AI IMPACT ON JOB AUTOMATION. international journal of engineering technology and management sciences, 7 (4), 410-425. https://doi.org/10.46647/ijetms.2023.v07i04.055.

[9] Loth, R., Battistelli, D., Chaumartin, F., Mazancourt, H., Minel, J., & Vinckx, A. (2010). Linguistic information extraction for job ads (SIRE project). , 222-224. https://doi.org/10.5555/1937055.1937114.

[10] Hong, X., Zhang, Y., & Jiong, Z. (2012). Feature Terms Analyzing Strategy for Recruiting Websites. 2012 Second International Conference on Business Computing and Global Informatization,. 451-453. https://doi.org/10.1109/BCGIN.2012.123.

[11] Mahdi, H., Dagli, R., Mustufa, A., & Nanivadekar, S. (2021). Job Descriptions Keyword Extraction using Attention based Deep Learning Models with BERT. 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 1-6. https://doi.org/10.1109/HORA52670.2021.9461296.

[12] Godbole, S., Bhattacharya, I., Gupta, A., & Verma, A. (2010, October). Building re-usable dictionary repositories for real-world text mining. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 1189-1198). https://doi.org/10.1145/1871437.1871588.

[13] Li, N., Kang, B., & Bie, T. (2023). SkillGPT: a RESTful API service for skill extraction and standardization using a Large Language Model. arXiv, abs/2304.11060. https://doi.org/10.48550/arXiv.2304.11060.

[14] Meesters, M., Heck, P., & Serebrenik, A. (2022). What Is an AI Engineer? An Empirical Analysis of Job Ads in The Netherlands. 2022 IEEE/ACM 1st International Conference on AI Engineering - Software Engineering for AI (CAIN), 136-144. https://doi.org/10.1145/3522664.3528594.

[15] Yang, Y., & Wang, Y. (2022). Enterprise Human Resources Recruitment Management Model in the Era of Mobile Internet. mob. inf. syst., 2022, 7607864:1-7607864:6. https://doi.org/10.1155/2022/7607864.

[16] Kokkodis, M., & Ipeirotis, P. G. (2023). The good, the bad, and the unhirable: recommending job applicants in online labor markets. management science, 69(11), 6969-6987. https://doi.org/10.1287/mnsc.2023.4690.

[17] Aleisa, M., Beloff, N., & White, M. (2021). Implementing AIRM: a new AI recruiting model for the Saudi Arabia labor market. Journal of Innovation and Entrepreneurship, 12, 1-41. https://doi.org/10.1186/s13731-023-00324-w.

[18] Biemann, T., Mayrhofer, W., & Koch-Bayram, I. (2023). Embedded in context: How time and distance affect the convergence of personnel selection practices. Human Resource Management Journal, 33(1), 47-68. https://doi.org/10.1111/1748-8583.12339.

[19] Aleisa, M., Beloff, N., & White, M. (2021). Implementing AIRM: a new AI recruiting model for the Saudi Arabia labor market. Journal of Innovation and Entrepreneurship, 12, 1-41. https://doi.org/10.1186/s13731-023-00324-w.

[20] Card, D., & Dinardo, J. (2002). Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. Journal of Labor Economics, 20, 733 - 783. https://doi.org/10.1086/342055.

[21] Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. american economic review, 108(6), 1488-1542. https://doi.org/10.1257/AER.20160696.

[22] Felani, H. O., & Ratnawati, N. (2022). Unveiling The Impact of Technological Advancements on Macroeconomic Stability and Labor Market Dynamics. Syntax Literate; Jurnal Ilmiah Indonesia, 7 (9), 14999-15008. https://doi.org/10.36418/syntax-literate.v7i9.14114.

[23] Rickardo, G., & Meiriele, S. (2023). Artificial Intelligence: Its Impact on Employability. World Journal of Advanced Research and Reviews, 18(3), 198-203. https://doi.org/10.30574/wjarr.2023.18.3.1056.

[24] Xie, M., Ding, L., Xia, Y., Guo, J., Pan, J., & Wang, H. (2021). Does artificial intelligence affect the pattern of skill demand? Evidence from Chinese manufacturing firms. Economic Modelling, 96, 295-309. https://doi.org/10.1016/J.ECONMOD.2021.01.009.

[25] Nogueira, M. C., & Madaleno, M. (2023). Skill-Biased Technological Change and Gender Inequality across OECD Countries-A Simultaneous Approach. Economies, 11(4), 115. https://doi.org/10.3390/economies11040115.

[26] Jadhav, A., Jagtap, P., Gurav, S., Jadhav, S., Jadhav, N., & Akkalkot, A. (2023). A Survey on Text Mining - Techniques, Application. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 338-343. https://doi.org/10.32628/cseit2390391.

[27] Sukumar, S., Lung, C., & Zaman, M. (2023). Knowledge Graph Generation for Unstructured Data Using Data Processing Pipeline. 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), 466-471. https://doi.org/10.1109/COMPSAC57700.2023.00068.

[28] Khaouja, I., Kassou, I., & Ghogho, M. (2021). A Survey on Skill Identification From Online Job Ads. iEEE Access, 9, 118134-118153. https://doi.org/10.1109/ACCESS.2021.3106120.

[29] Singh, S., Krishnan, D., Sehgal, P., Sharma, H., Surani, T., & Singh, J. (2022). Gradient Boosting Approach for Sentiment Analysis for Job Recommendation and Candidate Profiling. 2022 IEEE Bombay Section Signature Conference (IBSSC), 1-6. https://doi.org/10.1109/IBSSC56953.2022.10037443.

[30] Todd, P., McKeen, J., & Gallupe, R. (1995). The Evolution of IS Job Skills: a Content Analysis of IS Job Advertisements from 1970 to 1990. MIS Q., 19, 1-27. https://doi.org/10.2307/249709.

[31] Lee, S. M., & Lee, C. K. (2006). IT managers' requisite skills. Communications of the ACM, 49(4), 111-114. https://doi.org/10.1145/1121949.1121974

[32] Sodhi, M., & Son, B. (2009). Content analysis of OR job advertisements to infer required skills. Journal of the Operational Research Society, 61, 1315-1327. https://doi.org/10.1057/jors.2009.80.

[33] Smith, D., & Ali, A. (2014). Analyzing Computer Programming Job Trend Using Web Data Mining. issues in Informing Science and Information Technology, 11, 203-214. https://doi.org/10.28945/1989.

[34] Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. American Economic Review, 108(6), 1488-1542. https://doi.org/10.1257/aer.20160696