International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7820-7830

https://doi.org/10.70517/ijhsa463670

Analyzing the Impact of Artificial Intelligence on China's Economic Development Based on Cite Space Bibliometrics

Jiayong Liu[1,*] and Lan Jiang[2]
[1] School of Economics and Management, Yan'an University, Yan'an, Shaanxi, 716000, China
[2] School of Economics and Management, Xi'an Aeronautical Institute, Xi'an, Shaanxi, 710077, China
Corresponding authors: (e-mail: liujy1688@126.com).

Abstract The rapid development of Artificial Intelligence (AI) technology has aroused widespread concern globally, especially its application in economic development is regarded as an important force to promote economic structural transformation and industrial upgrading. As the second largest economy in the world, China is gradually accelerating the process of deep integration of artificial intelligence and economic development. This paper systematically investigates the impact of AI on China's economic development through the CiteSpace bibliometric analysis tool. The study used 4,562 relevant documents included in the China Knowledge Network Database (CNKI), extracted the topics using the LDA model, and combined with the co-occurrence analysis of the literature to reveal the current status and trends of research in the field of AI and China's economic development. The results show that after 2016, the research on AI and economic development has gradually heated up, and the number of publications reached its peak in 2023, with an annual number of publications close to 931. Furthermore, the keyword analysis of the literature shows that the main research focuses on areas such as "digital economy", "industrial upgrading" and "job market", among which the high-frequency words of "artificial intelligence" and "economic development" are 11,687 and 8,124 times respectively. Artificial intelligence plays an important role in promoting China's economic development, especially in industrial upgrading and technological innovation with significant impact.

Index Terms artificial intelligence, economic development, industrial upgrading, digital economy, job market, technological innovation.

I. Introduction

In recent years, artificial intelligence (AI) technology has received widespread attention and become a hot topic in the field of science and technology [1]. More and more enterprises, organizations, and government departments have started to invest in and adopt AI technologies in an attempt to derive business value and social benefits from them [2], [3]. As a growing economy with a large market, China's application of AI has had a significant impact on China's economic development [4], [5].

The impact of AI on China's economic development is multifaceted, including but not limited to the following aspects: (1) Promoting industrial upgrading and transformation, the application of AI technology can improve production efficiency, reduce costs, optimize resource allocation, and promote the transformation of traditional industries towards intelligence and high-end, which can help to improve China's position in the global value chain, and achieve industrial upgrading and transformation [6]-[9]. (2) Creating new economic growth points, the continuous innovation and application of AI technology can give rise to new industries and business models, such as smart manufacturing, smart home, smart medical care, etc. The development of these emerging industries will bring new growth points for China's economy [10]-[12]. (3) Enhance international competitiveness, the rapid development and application of AI technology has made China more competitive globally, by strengthening the R&D and application in the field of AI, China can enhance its competitiveness in the international market and promote economic development [13]-[16]. (4) Promote the adjustment of employment structure, with the wide application of AI technology, some traditional jobs may be replaced or reduced, but at the same time, new employment opportunities will be created, which requires the labor market to adapt to the new technological changes and improve the skills and quality of workers to adapt to the new employment demand [17]-[20]. (5) Promoting innovation-driven development, artificial intelligence technology is one of the important areas of innovation-driven development, through strengthening the research and development and application in the field of artificial intelligence, it can promote scientific and technological innovation and industrial innovation, and improve the quality and efficiency of economic development [21]-[24].

Artificial Intelligence (AI), as an emerging technological force, is profoundly affecting all levels of the economy globally, especially in a country with huge economic volume and development potential like China. In recent years, with the continuous progress and application of technology, the influence of AI has been expanding, and its role in promoting China's economy has become increasingly obvious. For China's economy, AI is not only a symbol of technological progress, but also a key factor in promoting the optimization of industrial structure and achieving digital transformation. However, although a large number of studies have explored the technological innovation and application of AI, systematic research on how it specifically promotes China's economic development is still insufficient. In this paper, we first selected relevant literature between 2005 and 2024 and conducted a bibliometric analysis using CiteSpace to reveal the impact of AI on China's economic development domain. The main research themes and hot issues of AI and economic development are extracted through LDA theme modeling. Second, the academic focus on AI and economic development and the evolutionary trend of related technologies were analyzed in conjunction with co-occurrence analysis. It is found that the potential of AI in industrial upgrading, job market reconfiguration and digital economy development is becoming increasingly apparent. With the deepening application of AI technology, the digital transformation process of the Chinese economy is accelerating. This paper further explores the future development trend of China's economy under the influence of AI technology and provides theoretical support for policy formulation.

II. Research data and methodology

This study explores the impact of AI on China's economic development by conducting CiteSpace bibliometrics on literature related to AI and economic development.

II. A. Sources of basic data

II. A. 1) Data collection

Literature databases contain a lot of literature information, which is an important factor affecting the results of bibliometric research. Because the research object of this paper is the academic literature on AI and China's economic development, China Knowledge Network Database (CNKI) is chosen as the source of literature data for this paper, which is a resource sharing platform covering many disciplines and is one of the most authoritative literature resource databases in China. Considering that the more important research results on AI and economic development in China are generally published in core journals, this study finally chooses two representative sub-databases on China Knowledge Network (CNKI), i.e., Peking University Core (PUC) and South University Core (CSSCI).

In addition to the database affecting the credibility of the analysis, the search style will also affect the credibility of the analysis results. There are two main forms in the current search style, one is searching by journals and the other is searching based on topics. Search by journal is for research topics that are not highly interdisciplinary and have clear disciplinary boundaries. Searching by subject is to fix the subject terms, mainly searching for keywords, titles and abstracts containing the required search terms. Considering that the research on artificial intelligence and economic development started late and is more interdisciplinary, this paper adopts the search by subject term to obtain research data, and sets the search time range as the period of 2005-2024, and selects the core of Beida and CSSCI for the sub-bank, and finally retrieves 4,562 journal articles.

II. A. 2) Data pre-processing

The raw data of 4562 journals used in this paper were downloaded from CNKI in Refworks format and saved locally as txt files. Due to the limitation of CNKI system, each txt file can only write 500 articles at most. Therefore, 4562 articles were saved as 10 txt files, named as download_01.txt, download_02.txt and so on up to download_10.txt, where each article was stored in the txt file in the form of one field per line, and each article was stored with 17 metadata, including subject, Article Off Summary, Keywords, Author, First Author, Corresponding Author, Fund, Abstract, References, Classification Number, Source, DOI, Institution Publication Time, Citation, Download and other metadata. Because metadata such as corresponding author, classification number, DOI, References, piece off abstracts have less impact on the analysis of the current state of the literature and the study of thematic hotspots, such metadata were excluded. Finally, 11 document metadata were analyzed and studied.

Because all the documents are stored in the form of 17 fields in the txt document, but some of the above fields have less impact on the analysis results. Therefore, the raw data need to be processed: the articles scattered in different txt documents are filtered out and stored in a csv file. Regular expression is a powerful text filtering tool, Python's re library perfectly supports the regular function, using Python's csv library to filter the data in the form of relational data saved to a csv file. The data filtering and writing in this paper are based on re library and csv library.

The data is checked for accuracy, validity and residual values. Firstly, 4562 documents are filtered from 10 txt files, which is consistent with the data downloaded from CNKI database. Secondly, the data preprocessing includes

Chinese word splitting, retaining proper nouns and selecting deactivated words. In this paper, we use data normalization in Python to normalize the data, and the final valid data contains 10348 research authors and 1580 research institutions. The key steps to get the research sample data are as follows:

(1) Analyze the text structure: each article contains “RT.... LA” fragment.

(2) Filter the “RT.... LA” fragments.

(3) Process the filtered data: Split the segments by line breaks (n) to get the list of fields.

(4) Determine whether the list of fields has SN fields, if so, save the list of data in a csv file, if not, then count the statistics.

II. B.Use of research tools

With the help of visual analysis tools, the potential characteristics of the paper's information can be accurately and intuitively reflected. Therefore, in this paper, we use Citespace visual analysis software in the econometric analysis of journal literature, Python language in data processing for disambiguation, and Gensim, a natural language processing toolkit, in topic modeling.

II. B. 1) Citespace, a bibliometric analysis tool

In the bibliometric analysis there are many visualization software can be presented on the results of the analysis, such as VOSviewer, HisCite, BICOMB, etc., but these visualization software have some drawbacks, for the Knowledge Network data analysis is not in-depth and not comprehensive. Therefore, this paper chooses to use Citespace visualization software in the bibliometric analysis, which is able to achieve the required level of this study.

Citespace is a modeling software for bibliometric analysis, through the scientific mapping program can clearly show the intrinsic connection between keywords, authors, institutions, etc. in a certain research field, which is conducive to scientific research, and thus has been widely used.

The actual operation process of Citespace analysis software is shown in Figure 1. First, it is necessary to download the sample data from CNKI, and then analyze the sample data after data conversion. Secondly, the new project is used to store the data, and set the analysis parameters and select the value of the analysis node, such as “Time Slicing”, “Node Types”, “Pruning Pruning” and so on. After setting, click Run to get the visualization graph, and the knowledge graph can be adjusted several times until it is satisfied according to the actual situation. Finally, write the analysis results based on the knowledge graph.

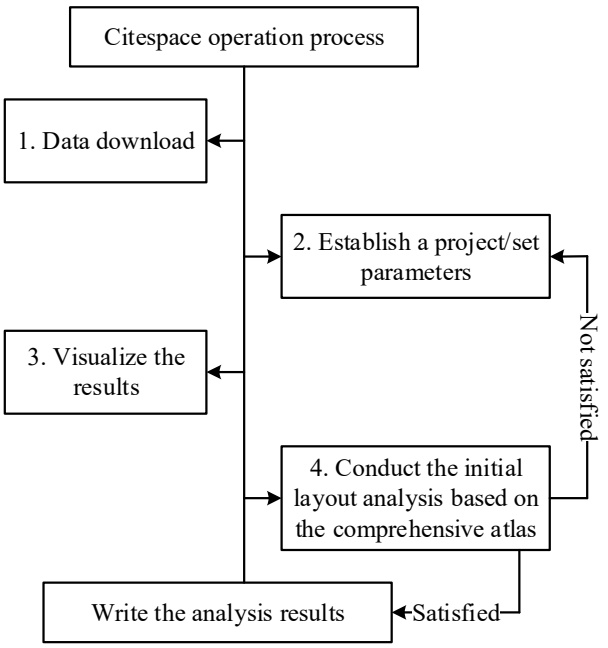

Figure 1: Operation process of Citespace

II. B. 2) LDA theme modeling tool Gensim

Python programming language is popular due to its small size, simple and easy-to-use features, and plays an important role in data mining and analysis. At the same time, the power of Python programming language is the ability to call a variety of toolkits, this study in the natural language processing call Gensim third-party toolkit for

semantic feature recognition.Gensim toolkit does not have a special format requirements for the analysis of the corpus, according to this feature can be created for modeling and analysis of a separate corpus. In the model training, Gensim toolkit and expressed into Model toolkit to convert the document set into a two-dimensional vector set, Gensim conversion logic is shown in Figure 2.

In this paper, we choose to use Python language for topic modeling, and in data preprocessing, the data obtained from CNKI is cleaned and binned to get the corpus analyzed in this paper. Regular expressions are used for judgment in cleaning, and Jieba library is used in word separation to get the final sample data. The toolkit Gensim is used for natural language processing in LDA topic modeling, and the WordCloud toolkit is invoked for visual presentation when analyzing topics. Finally, the word frequency statistics function is used to count the number of documents contained in the themes and the number of themes contained in each year, in order to prepare for revealing the law of theme evolution.

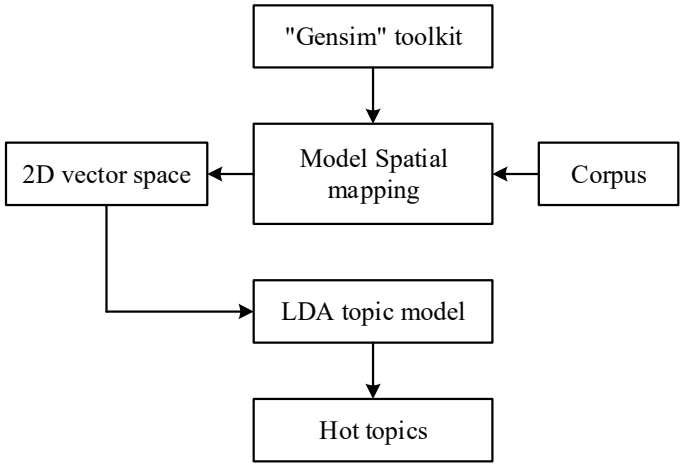

Figure 2: Gensim's natural language processing logic

II. C.Choice of research methodology

II. C. 1) LDA Subject Modeling

LDA (Latent Dirichlet Allocation) [25], a three-layer Bayesian probabilistic model for document topic generation, is an unsupervised machine learning algorithm that can be used to identify latent topic information in a document set or corpus, and in this paper, it is mainly used to infer and output the topic distribution and the corresponding probabilities in the corpus of academic literature. The LDA model is based on the probabilistic latent improvement of the PLAS model for semantic analysis [26]. The number of parameters of the PLAS model grows linearly with the number of documents, which is prone to overfitting phenomenon and makes the trained model lack of generalization ability. Introducing the Delikeray prior parameters, the problem of the PLAS model missing probability model at the document level is solved, and the lexical probability distribution of topics and the document topic probability distribution are realized.The LDA model is a fully probabilistic topic generating model, which is widely used in text processing tasks, and it performs well on texts with a large corpus size.

The basic idea of LDA is to assume that the document is randomly generated by a number of implicit topics, which are composed of corresponding feature words, with a three-layer structure of word words, topic topics and document documents. The document-word is converted from a high-dimensional vector space matrix mapping to two low-dimensional matrices: document-topic matrix, topic-word matrix, in which document-topic and topic-word obey polynomial distribution. The conditional probability formula for the occurrence of each word in a document is:

$$P(\text{Word}\mid\text{Document})=\sum_{\text{Theme}}P(\text{Word}\mid\text{Theme})\times P(\text{Theme}\mid\text{Document}) \tag{1}$$

In Equation (1), $P(\text{Word}\mid\text{Document})$ on the left side of the equal sign is known, indicating the probability value of the word appearing in the document, while $P(\text{Word}\mid\text{Theme})$ on the right side of the equal sign is unknown, and $P(\text{Theme}\mid\text{Document})$ are unknown, $P(\text{Word}\mid\text{Theme})$ is the probability of denoting the word under the corresponding theme, the higher the probability value indicates the greater the relevance of the feature word to the theme. $P(\text{Theme}\mid\text{Document})$ is the probability value that the document corresponds to the theme, the larger the probability value indicates that the theme is more closely associated with the document, and thus the theme structure can be reflected.

Suppose there are M documents and K topics can be extracted. Each document has its own topic distribution, and the topic probability distribution parameter α obeys the Dirichlet distribution. θ is the topic distribution for each article, and for the i th document the topic distribution is θ_i . Each topic has its own lexical item distribution, and the lexical item distribution probability parameter β also obeys the Dirichlet distribution. α, β and the number of topics K are hyper-parameters that need to be determined prior to the modeling. the LDA schematic diagram is shown in Fig. 3.

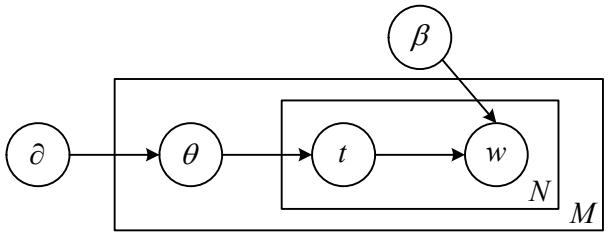

Figure 3: Principle of the LDA model

For the *n*th word in the document, the topic is first extracted in the topic distribution of the document, and then in the distribution of lexical items corresponding to this topic, and the stochastic process is repeated until the full traversal of the document is realized, with a joint probability of:

$$P(\theta, t, w \mid \alpha, \beta) = P(\theta \mid \alpha) \prod_{n=1}^{N} P(t \mid \theta) \times P(w \mid t, \beta) \tag{2}$$

where t denotes the topic TOPIC, w is the word word in the document, α is the topic probability distribution, β is the word-item probability distribution, θ and t are both implicit variables, and N denotes the number of all words in the document.

II. C. 2) Hyperparameter estimation methods

In topic modeling, two hyperparameters, the probability distribution of lexical items under each topic and the topic probability distribution of each document, are crucial. In the determination of the two parameters of a and β , in addition to continuous debugging iterations based on the default parameters, Gibbs sampling as well as VI variational inference can be used to train the model and determine the parameters. Among them, Gibbs sampling [27] is an algorithm used in Markov chain Monte Carlo theory to obtain a series of observation samples approximately equal to a specified multiple probability distribution. It operates on the principle of fixing one dimension of the probability vector randomly each time, sampling the current dimension value based on the variable values of other dimensions, iterating until convergence, and finally outputting the parameters to be estimated.VI is the selection of the variational distribution to approximate the a posteriori probability distribution of the whole model, which can convert Bayesian inference into an optimization problem with a clearer objective, but the selection of the variational distribution has a great impact on the whole model.

Most LDA-related studies are based on MCMC-based Gibbs sampling, which does not need to estimate the actual parameters, and the parameters can be calculated by itself as long as the appropriate hyperparameter K, i.e., the number of topics, is determined.The parameter inference method of Gibbs sampling is simple and easy to understand, and the feature of easy parallelization makes it perform better in training the topic extraction from a large number of documents. In this paper, Gibbs sampling is used as the learning algorithm for model training.

At this stage, LDA topic model is widely used for corpus hidden topic mining, the advantage is that it can effectively extract the topic feature words on the structured text environment, but if the text data is too complex and the topic is not clear, the classification effect may not be good. In this paper, we strictly carry out manual screening in the process of data preprocessing, and sieve out the comments that have nothing to do with artificial intelligence and economic development as well as those that are interfering with text analysis as much as possible on the premise of guaranteeing the number of samples, so that the content features are all similar, and therefore, we choose the LDA model as the research method of topic word extraction in this paper.

III. Visual analysis of the impact of artificial intelligence on China's economic development

In this chapter, the literature data on AI and China's economic development searched from the China Knowledge Network Database (CNKI) are measured and a knowledge map is constructed so as to explore the impact of AI on China's economic development.

III. A. Chronological analysis of the literature

Academic research and development will also go through the process of initial, growth and maturity, etc. In the academic field, the publication of literature is one of the important ways to display the experimental process and research results, so the number of outputs of papers varies in different time periods of academic research. Therefore, it is necessary to explore the relationship between the year and the number of papers to determine the evolutionary process of the field for the development of academics.

The evolution of the number of publications with increasing years is shown in Figure 4. Before 2016, the research field of AI and economic development was in the beginning development stage, and the overall number of papers was in a very low state.Between 2016 and 2023, the number of papers increased year by year, and reached a peak of 931 papers in 2023, which can be regarded as a period of rise in the research of AI and China's economic development in this stage.After 2023, the number of papers showed a decreasing trend However, it can still be seen that although the overall trend slows down, the number of articles is still higher than the average number of articles in these two decades. Therefore, it can be predicted that in the next few years, the annual number of articles will not necessarily rise, but there is a greater likelihood that it will stabilize at more than 700 articles, which means that the research on AI and economic development will be in a stable development stage.

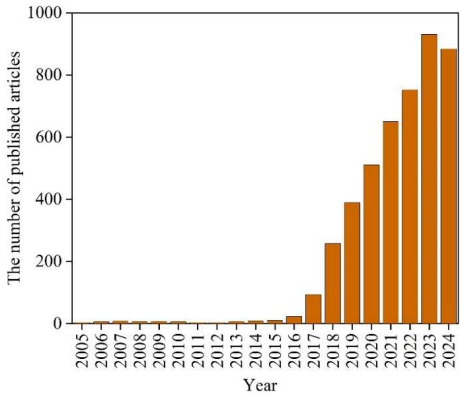

Figure 4: Number of published articles

Table 1: Statistics of High-Yield Institutions on CNKI

Sort	Issuing institution	Frequency	Sort	Issuing institution	Frequency
1	Jilin University	83	11	Central Communist Party School	39
2	Northwestern University	61	12	Peking University	36
3	Southwest University of Finance and Economics	59	13	Sichuan University	35
4	Zhongnan University of Economics and Law	54	14	Zhejiang University	34
5	Jiangxi University of Finance and Economics	47	15	Tsinghua University	34
6	Central University of Finance and Economics	46	16	University of Foreign Economic and Trade	33
7	Renmin University of China	45	17	Nanjing University	32
8	Wuhan University	42	18	East China Nnormal University	31
9	Nankai University	42	19	Shandong University	31
10	Shanghai University of Finance and Economics	40	20	University of Chinese Academy of Social Sciences	31

III. B. Analysis of issuing bodies

In this section, the issuing institutions of the literature are analyzed, and the statistics of the relevant academic literature of Knowledge Network's highly productive institutions are shown in Table 1. Eighteen of the top twenty institutions are Chinese institutions of higher education, with a total of 855 articles respectively, accounting for 18.74%

of the total journal literature, and 107 of the 126 core authors are also from institutions of higher education, accounting for 84.92% of the total number of core authors, which shows that the main areas of research in the field of AI and economic development are concentrated in institutions of higher education.

In addition to applying the metric of volume of issuance to the analysis of highly productive institutions, it is also possible to perform a burst analysis of issuing institutions. Burst analysis, i.e., analyzing the rapid growth in the number of articles issued by an institution in a certain period of time, is also another indicator that deserves attention. On the basis of constructing the cooperative co-occurrence map of institutions, adjusting the threshold value of "burstiness" of Citespace, and sorting according to the time of burstiness, we can get the ranking of burstiness institutions in the field of Artificial Intelligence and Economic Development of Knowledge Network Academic Journals, as shown in Figure 5.

In Figure 5, strength represents the strength of emergence, begin is the beginning year of emergence, end represents the end year of emergence, and the orange line represents the time period from emergence to the end. According to the emergence year of each institution, four institutions with emergence time during 2005-2016 can be classified as pre emergence institutions, and six emergence institutions with emergence time during 2016-2024 can be classified as recent emergence institutions. Institutions U1~U10 denote Tsinghua University, Peking University, Zhejiang University, Fudan University, Shanghai Jiaotong University, University of Science and Technology of China, Chinese Academy of Social Sciences, Xi'an Jiaotong University, Harbin Institute of Technology, and Renmin University of China, respectively.

Through the analysis of the above 10 emergent institutions, it can be found that some institutions in the early emergent institutions, such as Tsinghua University and Peking University, have a more stable volume of publications, indicating that these two institutions were the core of research in the field during 2005-2016. Among the recent emergent institutions, Renmin University of China and Harbin Institute of Technology, although they did not publish much in the early period, have contributed increasing academic results in the field of AI and economic development research in recent years, and are likely to become the main force of research in this field in the future.

Figure 5: Top10 Emerging Institutions on CNKI

III. C. Analysis of Issuing Journals

The data of high-yield journals with publications involving the field of artificial intelligence and economic development were organized, and the statistics of high-yield journals in KnowledgeNet are shown in Table 2.

Brad's law, one of the three laws of bibliometrics, points out that according to the number of specialized papers in a certain field included in a journal, arranged according to the number from the largest to the smallest, the journal can be divided into the core zone, the related zone, and the non-core zone oriented to this field. And at this time, the number of literature in each zone is roughly equal, and the number of journals is close to $1:n:n^2$. According to this law, the journals in which the literature is published are divided, and the results of the distribution of journals in KnowledgeNet are shown in Table 3.

The number of journals in the three regions is 43:225:971, which is approximated to $1:5:5^2$, when n is about 5. It can be seen that there exists a more stable core journal region for journals in the field of artificial intelligence and economic development. Most of the journals that publish the top five numbers of literature are journals related to economy, informatization, and business, and it can be assumed that there exists a wide range of applications of artificial intelligence and economic development in these fields.

Table 2: Statistics of high-yield journals on CNKI

Sort	Journal of Publication	Frequency	Sort	Issuing institution	Frequency
1	Informatization Construction	129	14	Marketing Field	67
2	Internet Economy	84	15	Modern Business	67
3	The Information Industry in China	84	16	People's Forum	65
4	China's Informatization	80	17	Communication World	65
5	China Economic Weekly	80	18	Collective Economy in China	63
6	Mall Modernization	79	19	Peking University Business Review	63
7	Software and Integrated Circuits	78	20	National circulation economy	62
8	Service Outsourcing	75	14	Research on Industrial Innovation	60
9	Shanghai Informatization	71	15	Science and Technology China	59
10	Chinese Business Theory	68	16	New Economy Guide	58

Table 3: Distribution of CNKI journals

Region	Classification criteria	The number of periodicals	The number of documents	Proportion of the total literature
Core area	>13	43	1082	23.72%
Related areas	3-13	225	1776	38.93%
Unrelated area	<3	971	1704	37.35%

III. D. Keyword co-occurrence analysis and cluster analysis

III. D. 1) Keyword co-occurrence analysis

The co-occurrence analysis of keywords in 4,562 valid literatures revealed that the word frequencies of the main search terms "artificial intelligence" and "economic development" were as high as 434 times and 389 times respectively, and the mediating centrality was 0.24 and 0.32 respectively, indicating the credibility of the data collation. Secondly, there were 277 cases of "China", with a mediating centrality of 0.24. This indicates that when Chinese scholars study the impact of artificial intelligence on economic development in the Chinese context, they pay more attention to their own national conditions and conduct more research from the aspects of "artificial intelligence" and "economic development" themselves. Other high-frequency keywords also include "industrial upgrading" 224 times, "innovation-driven" 217 times, and "digital economy" 217 times.

By drawing the co-occurrence graph of key words, it is found that the high-frequency nodes of research related to artificial intelligence and China's economic development are mostly gray, that is, mostly before 2018. A further examination of the red links reveals that the current research keywords mostly revolve around words such as "digital transformation", "sustainable development", "data security", "regional coordination", and "talent cultivation".

III. D. 2) Keyword clustering analysis

In order to more intuitively show the affinity between high-frequency keywords and explore the theme of the study on the impact of artificial intelligence on economic development, CiteSpac was used to further make a keyword clustering map. Selected from February 2005 to October 2024, Year Per Slice is 1, Types=Keyword, and the threshold setting method of g-index, the keywords are clustered by the logarithmic great likelihood algorithm, and the Q-value=0.9124>0.3, and the S-value=0.5182>0.5, which indicates that this clustering view is significant and reasonable. Finally, 23 clustering labels are generated, of which 10 are more obvious, and the results of keyword clustering analysis are shown in Table 4. In order to increase the specific information of clusters, this paper selects the main keywords of each cluster for in-depth analysis of the specific content contained in the name of each cluster, and summarizes them into the following 3 categories.

(1) Cluster #1 Industrial Upgrading, #5 Digital Economy, #6 Technological Innovation, #10 Production Efficiency, centered on technology-driven and industrial change, the main keywords are intelligent manufacturing, industry 4.0, digital transformation, smart factories and so on. Reflecting the direct transformation of industries and production methods by AI technology itself.

(2) Clustering #2 economic growth, #3 job market, #4 policy regulation, focusing on economic effects and macro impacts, the main keywords include GDP growth, structural unemployment, industrial policy and so on. The development and application of artificial intelligence has a systematic impact on macroeconomic indicators, market mechanism and policy environment.

(3) Cluster #7 Data Security, #8 Sustainable Development, #9 Social Equity, the main keywords include privacy protection, technological ethics, green economy, technological universality, etc., which interprets the challenges and adaptive adjustments of AI to human society, ethics and long-term development.

Table 4: Keyword cluster analysis

#	Tightness	Cluster name	Main key words
1	0.97	Industrial upgrading	Intelligent manufacturing, industry 4.0, emerging industries, transformation of traditional industries
2	0.99	Economic growth	GDP growth, total factor productivity, economic growth rate, long-term growth momentum
3	0.94	Employment market	Structural unemployment, skill requirements, labor force transfer, human-machine collaboration
4	0.98	Policy regulation and control	Industrial policy, regulatory framework, government intervention, national strategy
5	0.98	Digital economy	Digital transformation, data element, internet+, new infrastructure
6	0.95	Technological innovation	Technological progress, investment in research and development, technology spillover, patent layout
7	0.94	Data security	Privacy protection, algorithm bias, technical ethics, data governance
8	0.95	Sustainable development	Green economy, carbon neutrality, resource optimization, circular economy
9	0.92	Social equity	Income distribution, digital divide, technology inclusion, protection of vulnerable groups
10	0.95	Production efficiency	Automation, smart factories, cost optimization, productivity improvement

III. E. LDA Analysis

III. E. 1) Word frequency analysis

Through machine learning 4562 research documents, the keywords in the documents are extracted, and the Top20 high-frequency keywords are finally obtained after eliminating invalid keywords as shown in Table 5.

It can be seen that the main hotspots in the field of AI and economic development research are concentrated in the fields of artificial intelligence, economic development, employment, industrial structure, technological progress, new infrastructure, talent cultivation, smart economy, and digital transformation. These keywords reflect the diversity and complexity of the research on AI and economic development, covering a wide range of aspects such as industrial structure upgrading, employment situation, AI technology application, and social construction. The research in these fields is of high value in promoting industrial structure optimization, intelligent talent cultivation, and intelligent and digital construction of engineering construction. The diversity of keywords reflects the development and innovative integration of AI and economic development related research in the field of social disciplines.

Table 5: Keyword cluster analysis

Ranking	Key words	Frequency	Ranking	Key words	Frequency
1	Artificial intelligence	11687	11	New infrastructure	2173
2	Economic development	8124	12	Double circulation	1962
3	Industrial upgrading	5351	13	Sustainable development	1825
4	Innovation-driven	3925	14	Data security	1508
5	Digital economy	3668	15	Regional coordination	1483
6	Production efficiency	3273	16	Talent cultivation	1479
7	Employment	3214	17	Intelligent economy	1412
8	Industrial structure	3081	18	Financial technology	1375
9	Policy	2339	19	Globalized competition	1308
10	Technological progress	2196	20	Digital transformation	1247

III. E. 2) Determination of optimal number of topics

Before performing LDA topic modeling, topic consistency under different topics needs to be calculated to ensure the accuracy of the LDA topic model training results. Too small a number of topics may make the number of keywords after the LDA model's disambiguation insufficient to depict all the current status of research in the research area, and too many topics may make too many keywords, resulting in the inability to condense the topics for analyzing and researching.

The pyLDAvis topic visualization results are shown in Figure 6. Combining Figure 6, the optimal number of topics for this literature set is determined based on the topic confusion model in LDA and Word2Vec models, and the results of topic confusion analysis are shown in Figure 7. It can be concluded that the optimal number of themes is 5.

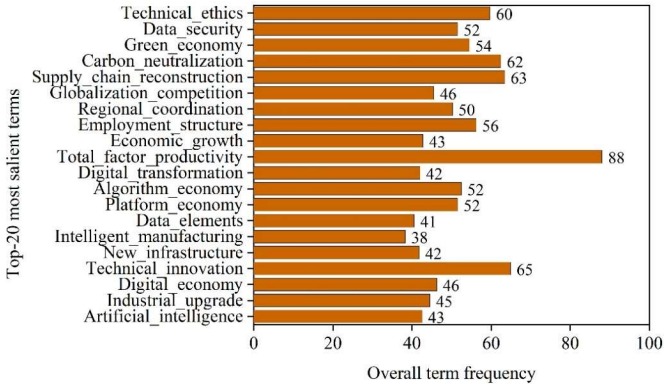

Figure 6: Visualization of the pyLDAvis theme

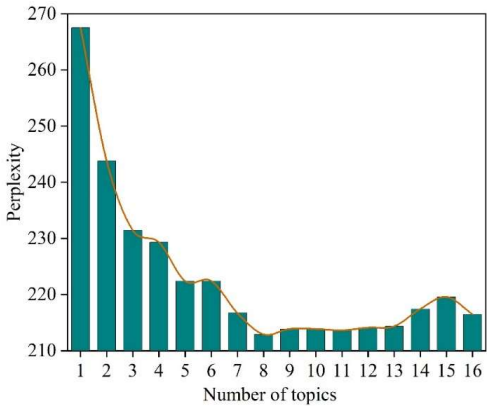

Figure 7: Topic perplexity

The 20 keywords under the same theme are summarized to finally get 5 thematic hotspots: AI-driven industrial upgrading and digital transformation, AI and employment restructuring, AI and economic policy synergistic innovation, AI-enabled green and sustainable economic development, and socio-economic risks and governance challenges of AI.

IV. Conclusion

The application of artificial intelligence in China's economic development has gradually become an important force that cannot be ignored. From the results of the analysis of literature data, the research on AI and economic development has shown rapid growth after 2016, with the number of publications reaching a peak of 931 in 2023. In addition, the literature analysis reveals that AI makes the most significant contribution to industrial upgrading, especially playing a key role in smart manufacturing and digital transformation. According to the keyword co-occurrence analysis, digital economy and technological innovation are the main hotspots of current research, and the frequency of related research literature is high, indicating that the academic attention in this field is increasing. Specifically, the keywords “artificial intelligence” and “economic development” have a frequency of 11687 and 8124 times respectively, demonstrating their central position in the research. With the advancement of AI technology, China's economy will gain greater development potential in industrial optimization, employment structure adjustment and technological innovation in the future. Based on the results of the current study, it can be speculated that in the future, AI will further contribute to the digital transformation of the Chinese economy and promote green and sustainable development. However, the popularization and application of AI technology is also accompanied by issues such as data security and social fairness, which require the government and relevant departments to introduce a more comprehensive policy framework to ensure the benign operation of AI technology in economic development.

Funding

This research was supported by Yulin Municipal Science and Technology Bureau 2020 industry-University-research project “Rural Science Popularization, human capital Improvement and economic growth - taking Yulin as an example”(No. CXY-2020-068);Yulin Municipal Science and Technology Bureau 2019 industry-university-research

project “Baiyu Mountain ecological compensation mechanism and industrial structure optimization strategy research”(No. 2019-171);Project of Education Department of Shaanxi Province, "Research on Brand Construction of Agricultural Products of Farmers' Specialized Cooperatives in Northern Shaanxi" (No.20JK0453).

References

[1] Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224.
[2] Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29.
[3] Raj, R., & Kos, A. (2023). Artificial Intelligence: Evolution, Developments, Applications, and Future Scope. Przeglad Elektrotechniczny, 99(2).
[4] Zhou, G., Chu, G., Li, L., & Meng, L. (2020). The effect of artificial intelligence on China’s labor market. China Economic Journal, 13(1), 24-41.
[5] Wu, F., Lu, C., Zhu, M., Chen, H., Zhu, J., Yu, K., ... & Pan, Y. (2020). Towards a new generation of artificial intelligence in China. Nature Machine Intelligence, 2(6), 312-316.
[6] Zou, W., & Xiong, Y. (2023). Does artificial intelligence promote industrial upgrading? Evidence from China. Economic research-Ekonomska istraživanja, 36(1), 1666-1687
[7] He, Y. (2023). Path and Mechanism of Industrial Internet Industry Promoting the Transformation and Upgrading of Small and Medium-sized Enterprises with Artificial Intelligence. Mobile Information Systems, 2023(1), 3620662.
[8] Deng, Y., & Wang, H. (2019, October). Research on industrial integration and upgrading of artificial intelligence and real economy. In 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA) (pp. 692-695). IEEE.
[9] Cheng, Y. (2024). Research on the Transformation and Upgrading of Manufacturing Industry in the Era of AI Empowerment. Journal of Artificial Intelligence Practice, 7(2), 182-187.
[10] Kraus, N., Kraus, K., Shtepa, O., Hryhorkiv, M., & Kuzmuk, I. (2022). Artificial intelligence in established of industry 4.0. WSEAS transactions on business and economics, (19), 1884-1900.
[11] Mateo, F. W., & Redchuk, A. (2021). The emergence of new business and operating models under the industrial digital paradigm. Industrial Internet of Things, Platforms, and Artificial Intelligence. Machine Learning. Journal of Mechanics Engineering and Automation (JMEA), 11(2), 54-60.
[12] Bharadiya, J. P., Thomas, R. K., & Ahmed, F. (2023). Rise of artificial intelligence in business and industry. Journal of Engineering Research and Reports, 25(3), 85-103.
[13] Adigwe, C. S., Olaniyi, O. O., Olabanji, S. O., Okunleye, O. J., Mayeke, N. R., & Ajayi, S. A. (2024). Forecasting the future: The interplay of artificial intelligence, innovation, and competitiveness and its effect on the global economy. Innovation, and Competitiveness and its Effect on the Global Economy (February 26, 2024).
[14] Medvedeva, A. M. (2019). Artificial intelligence as a new tool for growth of innovation and competitiveness of the digital business. Revista Espacios, 40(35).
[15] Dai, T. (2024, May). Trade International Competitiveness Systems Based on Artificial Intelligence. In 2024 Second International Conference on Data Science and Information System (ICDSIS) (pp. 1-5). IEEE.
[16] Senadjki, A., Ogbeibu, S., Mohd, S., Hui Nee, A. Y., & Awal, I. M. (2023). Harnessing artificial intelligence for business competitiveness in achieving sustainable development goals. Journal of Asia-Pacific Business, 24(3), 149-169.
[17] Wang, X., Chen, M., & Chen, N. (2024). How artificial intelligence affects the labour force employment structure from the perspective of industrial structure optimisation. Heliyon, 10(5).
[18] Shen, Y., & Zhang, X. (2024). The impact of artificial intelligence on employment: the role of virtual agglomeration. Humanities and Social Sciences Communications, 11(1).
[19] Liang, H., Fan, J., & Wang, Y. (2025). Artificial Intelligence, Technological Innovation, and Employment Transformation for Sustainable Development: Evidence from China. Sustainability, 17(9), 3842.
[20] Shen, Y. (2024). Future jobs: analyzing the impact of artificial intelligence on employment and its mechanisms. Economic Change and Restructuring, 57(2), 34.
[21] Ancajima Miñan, V. Á., Suxe Ramírez, M. A., Infante Saavedra, C. L., Rubio Cabrera, W. F., More Reaño, R. E., & García Merino, L. S. (2023). Innovation Driven By Artificial Intelligence: Exploring The Opportunities And Challenges For Technological And Business Development. Journal of Namibian Studies, 33.
[22] ALIYEV, N. (2025). Artificial Intelligence in Digital Silk Road: Driving Innovation and Economic Transformation. EUROASIA JOURNAL OF SOCIAL SCIENCES & HUMANITIES, 12(1), 95-102.
[23] Piao, Y. (2024). The Promotion of Artificial Intelligence and the Formation of New Quality Productivity. Advances in Economics and Management Research, 12(1), 74-74.
[24] Han, F., & Mao, X. (2024). Artificial intelligence empowers enterprise innovation: evidence from China’s industrial enterprises. Applied Economics, 56(57), 7971-7986.
[25] Liqing Jiang, Jie Wang, Yihua Wang, Hang Yang, Lingwang Kong, Zhongjun Wu... & Yingsong Jiang. (2025). Bibliometric and LDA analysis of acute rejection in liver transplantation: Emerging trends, immunotherapy challenges, and the role of artificial intelligence. Cell Transplantation,34,9636897251325628.
[26] Alemayehu Eyor & Fang Yi. (2024). Supervised probabilistic latent semantic analysis with applications to controversy analysis of legislative bills. Intelligent Data Analysis,28(1),161-183.
[27] Yetunde Esther Ogunwale & Micheal Olalekan Ajinaja. (2023). Application Research on Semantic Analysis Using Latent Dirichlet Allocation and Collapsed Gibbs Sampling for Topic Discovery. Asian Journal of Research in Computer Science,16(4),445-452.