



The research landscape on the artificial intelligence: a bibliometric analysis of recent 20 years

Hui Gao¹ · Xiuhao Ding²

Received: 30 April 2021 / Revised: 22 August 2021 / Accepted: 10 January 2022 /
Published online: 23 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Artificial intelligence (AI), a general term that implies the imitation of information process of intelligent behavior and sense with minimal intervention, is one of the most promising research areas and has received a considerable attention with coexisting pros and cons. In order to understand the research status quo and future trends on AI technology, this work uses bibliometric analysis method to obtain this objective. By analyzing the datasets including journal article data collected from Web of Science (WOS), conference paper data retrieved from Scopus and the patent data extracted from Derwent Innovations Index (DII) in the period of 2000–2019, we primarily provide a comprehensive overview to better understand the research status of AI. Bibliometric analysis results can also shed light on the evolution and trends in AI.

Keywords Artificial intelligence (AI) · Bibliometric methods · Patent analysis · Visualization

1 Introduction

Since the term “artificial intelligence” (AI) was introduced at the Dartmouth in 1956, it has attracted considerable attention and has made significant inroads in both theory and practice. The researchers have developed series of principles and methodologies to expand the concept of AI, especially Turing, who proposed the famous “Turing Test” to define “Machine Intelligence” and laid the theoretical foundation of computer science and AI [3, 34]. Although it develops slowly in its emerging stage, AI has a significant growth prospect with the aid of computers. Moreover, the field of AI has shown a dramatically upward trend of growth in the 21st century [18]. Nowadays, the growing ubiquity of AI is changing our daily life

✉ Hui Gao
gaohui-hust@qq.com

¹ School of Business, Hubei University, Youyi Avenue 368, Wuhan, China

² School of Management, Huazhong University of Science and Technology, Luoyu Road 1037, Wuhan, China

imperceptibly and is on its way to reshape the world [38]. For example, some intelligent devices such as Siri or Alexa are well known by ordinary beings with voice- and thought recognition [31]. However, until the remarkable victory of AlphaGo, a Google DeepMind's self-learning algorithm, beat the champion of the board game in the living world, people have to re-consider the future of AI, including its advantages and threats whose consequences are unforeseeable.

Although there have been sounding alarm that AI may pose grave danger to humans, it plays a crucial role in improving human welfare in diverse sectors [24]. Due to the convenience and the social well-being that AI brings, the growing importance of AI has been emphasized by many countries around the world. For example, the 2015 Strategy for American Innovation established high-priority research areas related to AI, meanwhile the 5th Science and Technology Basic Plan of Japan listed AI as a vital technology for smart society [13]. Moreover, the Chinese government issued the New-Generation Artificial Intelligence Development Plan in July 2017 as a guideline to develop the breakthroughs of AI. Additionally, AI is also seen as a key role by the UK government in its 2017 Industrial Strategy, with an announcement of achieving the ambition of being a world leader in AI and data technologies [20], evidenced by £300 million investment in AI research.

The backdrop of the acceleration of AI development across the globe and the importance of AI in innovation and knowledge economics promote the boom of academic research. A Google search on the term “artificial intelligence” returns hundreds of thousands of responses, which demonstrates the prosperity of AI. However, AI has been widely utilized into several independently subjects that each branch area is highly technological and professional and extremely incompatible. This leads to a need of holistic picture of AI domain. Although there exist some review works, most of them focus on a certain field of AI application, such as the applications in Photovoltaic systems [37], the usages in respiratory sound analysis [23], the implementation in precision agriculture [25], the combination with block chain [30], little attention has been paid to a systematic review of the AI literature. With the exponentially increasing number of literature, bibliometric analysis has become a valuable approach to appraising current status and predicting development trends of AI domain. Hence, there is a need to study the evolution of AI research and find the trends in the future. At the same time, as a domain with high-tech, the development of AI technology application is accompanied with the increasing number of granted patents. Therefore, not only the scholarly big data but patent data related AI selected as the base to analyze the overview in this field by using the scientometric methods and tools, which can provide the intellectual turning points and research direction for subsequent study [4].

In order to get a thorough understanding of the state of art in the field of AI, we conduct a bibliometric analysis to map the underlying intellectual base over time in a visualized way and portray its possible trends through co-citation methods. Our study performed on a large-scale dataset which consists of two subsets, a scholarly dataset and a granted patent dataset. Figure 1 shows the distribution of publications (including journal articles and conference papers) and patents in recent two decades and reveals that the enthusiasm for AI runs high both in academia and practice, especially in recent five years. There is a consistent trend form the different datasets. We can see a stable trend from 2000 to 2012 and a rapid growth stage in the following years in the figure. The slight fluctuation of conference data is mainly triggered by some biennial conferences. In this paper, we analyze the journal data in the period of 2000–2019 to explore the inner mechanism of AI domain and meanwhile make a comparison between two periods to identify the evolution of AI. Furthermore, a comparative analysis is conducted by using conference data. And the patent data is used for the analysis of AI

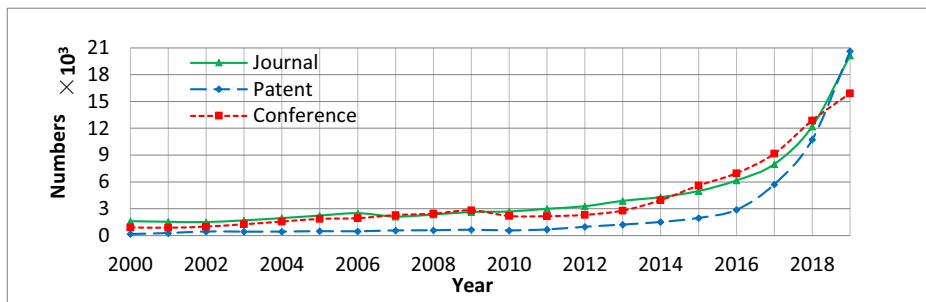


Fig. 1 The distribution of publications and patents in recent two decades

application in practice. Although the extracted articles are not all publications in AI area, the further information revealed by references can make the outcomes more reliable. Results of statistical analyses of bibliometric data make it possible to overview the current status, to outline the future research directions, and to acquire the managerial implications.

2 Methodology

As mentioned above, the aim of this work is to provide a systematic review of AI research, investigate the evolution of this field, detect the main subfields, and predict emerging trends and future directions. The increasing number of academic publications has evoked interest in bibliometrics. And the features of bibliometric data make it possible to generate a holistic view for further analysis. Specifically speaking, the analyses such as citation networks, cooccurrence networks and coupling networks can show the evolution of hot topics, identify the impact of milestone studies, and analyze the relationship between articles and references. Hence, a systematic review with bibliometric analysis was adopted in this study. Moreover, referring to the medical science review process, this research follows the steps of literature retrieval, data cleaning, descriptive analysis, bibliometric analysis, and comparative analysis in a systematic, transparent and reproducible manner.

2.1 Data collection

The development of AI is ultimately about severing people. Hence, to make the results more complete, the data mainly include two types, publications and patents. Accordingly, there are two datasets used in our study, named PUB and PAT respectively. Moreover, we divide PUB into two subsets JOU (journal articles) and CON (conference papers) with the intension of showing the difference of influences between journal and conference outputs.

To ensure the transparency and repeatability of the review process, we should develop a high quality data collection procedure that can minimize the biases and limitations. Data sources, search queries, and coverage are three basic criteria adopted to collect the initial sample. In the first place, many studies have compared mainstream databases, such as Web of Science (WOS), Scopus, Google Scholar, and PubMed, and found that WOS and Scopus are interchangeable and produce similar results for bibliometric analyses. Meanwhile, the WOS Core Collection can provide data with high quality standards, including full records and cited references for each article, which makes data processing more efficient. So the WOS Core

Collection is the data source of publications. And the conference papers are retracted from Scopus. Besides, for the granted patents metadata, the Derwent Innovation Index (DII) patent database, which has a high authority by integrating Derwent World Patent Index and Patent Citation Index, is selected as the data source. Each patent is full recorded with the information of patent number, title, assignees, inventors, abstract, and patent code.

In the second place, we should confirm an extensive search queries to identify related articles and patents. First of all, we must figure out the key issue: what is AI? Although the term has no uniformed definition, one prevailing definition of AI is making the machinery behavior look like the intelligent behavior and it can be used to develop systems that think like humans, act like humans, think rationally, and act rationally [33]. AI research is the study of intelligent agents to perceive the environment, identify complex information processing problems, and make decisions by computational technologies supported with human intelligence [19]. And the near term goal of AI is to create intelligent machines. Furthermore, the prevailing research topics related AI mainly in the aspects of learning, inference, and cognitive. In terms of the definitions and implicit important concepts, we confirmed the terms to select the journal articles. Moreover, Venture Scanner, a research firm of emerging technology industry, classifies the AI sector into 13 categories, which were the selection basis of search terms of AI patents.

In line with other bibliometric studies [21], we use the following criterion to collect records: a record was included when it had one of the search terms in its title, keywords or abstract. Although this may exclude some publications, the co-citation analysis of cited references can offset this shortcoming. Journal articles are considered as more certified since they had been reviewed and approved by peer researchers [27], hence, “article” was the only document type. We set the time span as “2000-2019” and concentrate on the identification of the keywords according to the definition of the term “artificial intelligence” and important AI concepts to search the publications. Furthermore, to reduce the data noise and ensure the representativeness, we clean the data manually by integrating different types of author name and institution, combining diverse forms of keywords, and deleting the records with missing information. The search queries and returned records from searching in September 2020 are shown in Table 1.

Moreover, given the enormous influence of top-tier conferences on AI field, we take the papers published in AI conferences in the same period into consideration to get a thorough understanding of AI domain. We select 15 top-tier conferences on AI (list in Table 2) recommended by China Computer Federation (CCF) and meanwhile at the top of the conference ranking created by Australian deans and the Australian Computing Research and Education Association of Australasia (CORE) [18]. The publication metadata of conference papers is primarily retrieved from Scopus, the largest abstract and citation database, and also

Table 1 Topic search queries used for data collection

Set	Records	Search queries
JOU	88,771	TS=(“artificial intelligence” OR “AI” OR “machine intelligence” OR “machine learning”)
CON	81,136	TS=(“artificial intelligence” OR “AI” OR “machine intelligence” OR “machine learning”)
PAT	51,400	TS=(“artificial intelligence” OR “computer vision” OR “deep learning” OR “machine learning” OR “natural language processing” OR “gesture control” OR “virtual assistant*” OR “smart robot*” OR “voice translation” OR “content recognition” OR “speech translation” OR “video recognition” OR “text embed*” OR “data mining” OR “text mining”)

Table 2 Statistics of top-tier conferences

Conference name	Abbr	Papers	Citations	C/ P
AAAI Conference on Artificial Intelligence	AAAI	6,225	83,541	13.42
Neural Information Processing Systems	NIPS	5,096	245,730	48.22
Association of Computational Linguistics	ACL	4,134	92,667	22.42
IEEE Conference on Computer Vision and Pattern Recognition	CVPR	11,255	631,272	56.09
IEEE International Conference on Computer Vision	ICCV	3,863	241,064	62.40
International Conference on Machine Learning	ICML	3,838	152,201	39.66
International Joint Conference on Artificial Intelligence	IJCAI	5,973	93,650	15.68
International Conference on Automated Planning and Scheduling	ICAPS	825	11,597	14.06
Conference on Empirical Methods in Natural Language Processing	EMNLP	2,947	66,735	22.65
International Conference on Autonomous Agents and Multiagent Systems	AAMAS	4,999	56,999	11.40
Conference in Uncertainty in Artificial Intelligence	UAI	1,219	23,876	19.59
European Conference on Computer Vision	ECCV	8,580	144,602	16.85
European Conference on Machine Learning	ECCML	2,138	23,920	11.19
International Conference on Computational Linguistics	COLING	2,152	27,813	12.92
IEEE International Conference on Robotics and Automation	ICRA	14,849	287,246	19.34

supplemented by the Database Systems and Logic Programming (DBLP) computer science bibliography in case of missing records.

However, we find that the retrieved data are too broad-brushed to ensure the representativeness, especially for journal articles. Hence, we set some exclusion criteria to reduce bias. We firstly remove the duplicates in the datasets assisted by software and clean the data manually by integrating different types of author names and institutions. Then, we omit the items with missing information. Last but not least, we browse the contents of each record, including title, abstract, author keywords, and keywords plus, and delete the records totally unrelated to AI.

Finally, we created a JOU dataset containing 71,407 journal articles, a CON dataset including 67,131 conference papers and a PAT dataset involving 24,822 patent records. Although the extracted data may not include all publications in AI field, we deem that above datasets are massive enough to represent the academic and practical aspects of AI field, especially when the analysis object not only includes these records but also references the publications cited. We primarily focus on the JOU data to outline the status quo and reveal the inner mechanisms of AI academic research through bibliometric analysis. Additionally, we conduct a comparative analysis between JOU and CON to show the different influences of journal and conference outputs on the AI field. As a complement, the PAT data is mainly used to describe the current state of AI applications.

2.2 Methods

The increasing growth of literature has motivated the interest in bibliometrics, an important role in analyzing and assessing academic research outputs [10] with such items as titles, authors, affiliations, journals, keywords, and references. In this paper, bibliometric methods and visualization tools have been used to investigate an overview of AI domain. These have been used to explore current status of scientific fields, evaluate the impact of scholars, describe the collaboration of institutions [12], and identify the evolution trajectory of publications over time [26]. The large amounts of scholarly data make it possible to generate bibliometric

networks such as co-citation networks, co-occurrence networks, coupling networks, and direct citation networks for further analyses.

Traditionally, the granted patent can be used as a proxy of invention. In the age of turbo-charged technology economics, patent data has an increasing trend of growth and considered to have potential value for competitive advantage [32]. Currently, patent analysis method plays a vital role in patent management and firm strategy and has been used for various objectives [11]. For example, patent citation network analysis can provide a novel perspective for understanding the innovation landscape and discovering the technology opportunity [28]. In our study, the patent data is used to conduct a quantitative analysis for overviewing the current status and predicting the possible emerging technologies in AI field. Combining big data of both scholarly publications and granted patents with multi-methods such as bibliometrics and meta-analysis, a systematic review can be developed [2].

Furthermore, we need to make a briefly introduction of CiteSpace, which can support several types of bibliometric analyses that meet the needs of visual analytic tasks. This software models the intellectual structure of the underlying domain through synthesized networks derived from the time series of publications based on the bibliographic records [5]. A time slicing technique is used to build an overview network of relevant literature that is synthesized with a time series of individual network models. In the network generated by this software, the nodes (or vertices) represent each of the items in the dataset, and the links between pairs of nodes denote the co-occurrence relationships [9]. The importance of concepts or cited references is shown by the size of the circle and shown as a citation tree-ring. The different colors indicate the time spot when links occurred for the first time. The time period in the network changes from cool (purple) to warm (yellow) when moving from an early time period to the most recent one. The rule of colors is applied to all networks throughout this paper.

Additionally, two basic structural metric terminologies, occurring in subsequent sections frequently, are betweenness centrality (BC) and citation burst. BC is an indicator to show the importance of the node, marked with purple circles in the network. Nodes that are highly connected to one other and their position between different groups may have high BC values, which tend to identify boundary-spanning potentials that may result in transformative discoveries [6]. Moreover, the citation burst is another indicator calculated by the burst detection technique that can identify abrupt changes of events. The node with high citation burst could be a turning point or landmark in the development of the field.

In the following sections, we mainly present a holistic view of AI domain through CiteSpace by using the time-slicing technique. The visualization results are divided into two parts based on the input data, named publication analysis results and patent analysis results, respectively. The journal articles and conference papers can represent the efforts in academic and the granted patents mainly on behalf of the intellectual property in practice. For JOU, we first conduct a descriptive analysis to portray the general situations of academic efforts, including the macroscopic overview generated by co-occurrence analysis of country and journal and the intuitive presentation of subject matter information showed by dual-map overlays method, which simultaneously showing citing and cited base map. And then we present the co-citation analysis of references and keywords to reflect the trend of research hotspots and find the milestones in the development process. Furthermore, a comparative analysis of CON is made for purpose of showing different influences between journal and conference outputs on AI field. As a complement and comparison, we also proceed a quantitative analysis by using the patent data to provide a more comprehensive and persuasive overview.

3 Results of publication analysis

3.1 Descriptive analysis

Firstly, we conduct country-wise analysis to depict the impact of different countries on AI research. In the perspective of publication counts, USA plays a dominant role with 22,702 counts that almost twice as far ahead of China, the second ranked country with 10,264 counts. The following ranked productive countries with no less than 3,500 occurrence counts are England (5,279), Germany (4,379), Spain (3,597), and Canada (3,570). This phenomenon shows the strong research passion of these countries for AI. In the light of betweenness centrality (BC) value, the top five ranked countries are England (0.34), Australia (0.21), France (0.20), Saudi Arabia (0.18), and Austria (0.13). The node with a high BC value plays a gatekeeper role in the network, hence the countries with both relatively high indicators like England (5,279/0.34), Australia (2,749/0.21), France (3,023/0.20), USA (22,702/0.09), and Spain (3,597/0.09) have significant positions in AI domain. Moreover, countries with high BC values and relatively low publication counts, such as Saudi Arabia (506/0.18), Austria (657/0.13) and Egypt (303/0.09), also play important roles in the development of AI field. On the contrary, the country with high publication counts but low BC values like China (10,264/0.02), Canada (3,570/0.01), India (3,015/0.01), and Italy (2,976/0.02) should pay more attention on the quality rather than quantity.

Moreover, the institution analysis can be seen as an alternative evidence of geographic distribution. Top ten productive institutions are Chinese Acad Sci (1,169; 1.59%), Stanford Univ (597; 0.81%), MIT (579; 0.79%), Univ Michigan (529; 0.72%), Univ Illinois (516; 0.70%), Harvard Univ (509; 0.69%), Univ Florida (508; 0.69%), Univ Wisconsin (498; 0.68%), Univ Washington (482; 0.66%), and Univ Sao Paulo (452; 0.62%). This phenomenon also demonstrates the high production and predominant roles of USA and China in AI research, and gets almost the same conclusion with country analysis.

Secondly, the journal analysis is conducted. The top highly cited journals with their H-index and 5 year impact factor (IF) are shown in Table 3, ranked by citation frequency. The number of focal articles published in these journals also listed and only journals with both high number of citations and publications are shown in the table. Although the sources of literature are distributed almost 9,800 publications, the articles published in listed journals (0.3% of total journals) accounted for approximately 12% of total publications. These journals, hence, can be regarded as mainstream journals and make great contributions to AI research field. However, the journals with high citations (over 4,000) but low publications (less than 70), such as *Nature* (11,091/35), *Science* (10,965/24), *Neural Computation* (4,624/66), and *New England Journal of Medicine* (4,377/2) that may laid the theoretical foundation are excluded. Moreover, some productive journals like *IEEE Access* (476/557), *Sensors* (368/1,440), *Journal of Dairy Science* (328/1,886), *AI Magazine* (318/1,378), *Remote Sensing* (215/724), *Applied Intelligence* (183/810), *Soft Computing* (171/827), and *Journal of Intelligent and Fuzzy Systems* (150/325) are also excluded from the list. These productive journals are emerging journals in AI theory and application that should be paid more attention and may make more contribution in the future. In general, the listed journals are indexed with high H-indexes and impact factors and widely recognized by academia and managers.

Lastly, a publication portfolio analysis with dual-map overlay method was conducted to overview major disciplines in AI domain. The publications constructed by topic search in our dataset can be used for generating a subject matter overlay. A publication represents the output

Table 3 Top 20 highly-cited journals with the number of publications in datasets

Rank	Journal	Citations	Publications	H-index	IF (5 year)
1	Mach Learn	14,691	201	144	3.157
2	P Natl Acad Sci USA	10,225	184	737	10.62
3	J Mach Learn Res	8,950	289	188	8.759
4	PLoS One	7,774	1,073	300	3.227
5	Pattern Recogn	7,038	240	195	6.28
6	IEEE T Pattern Anal	7,003	182	344	15.395
7	Bioinformatics	6,290	432	363	9.853
8	Expert Syst Appl	5,986	988	184	5.448
9	IEEE T Neur Net Lear	5,901	157	196	8.823
10	Artif Intell	4,751	245	142	5.944
11	Neurocomputing	4,577	669	123	4.01
12	BMC Bioinformatics	4,164	555	196	3.213
13	Pattern Recogn Lett	3,432	192	149	3.077
14	IEEE T Knowl Data En	3,404	197	162	5.201
15	INFORM SCIENCES	2,973	310	169	5.563
16	Neuroimage	2,460	182	344	6.682
17	Theriogenology	2,411	590	127	2.288
18	Sci Rep-UK	2,409	466	179	4.576
19	Appl Soft Comput	2,313	338	124	5.39
20	Knowl-Based Syst	2,071	306	107	6.075

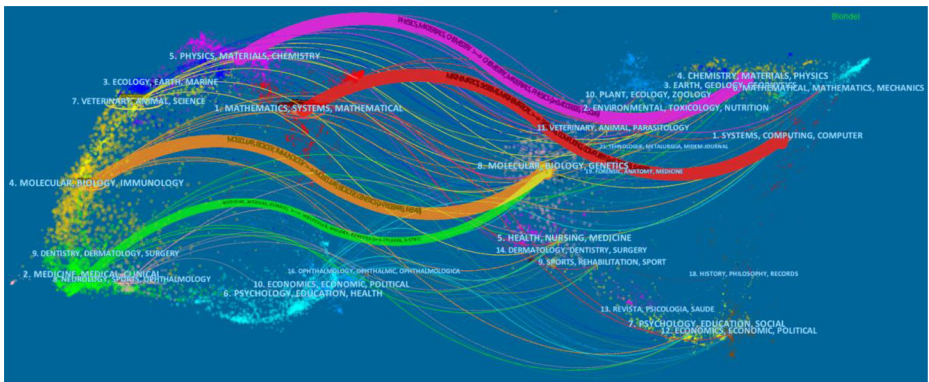
Note: Minimum of 150 publications and 2,000 citations. Rank by the number of citations. Indicators in 2019

of research whereas its references as a whole represent the knowledge base on which the research is built [8]. Figure 2 shows the dual-map overlay result with the citing map in the left and the cited map in the right. The curves between them show the citation context. Each journal is portrayed by one ellipse, of which the longitudinal axis represents the number of articles and the horizontal axis represents the number of authors. To find the trends vividly, the data is divided into two periods. We can see from the Fig. 2(a) that AI related articles published in first decade are mainly involved in limited disciplines. Meanwhile, the citation trajectories remain in the same discipline except for the disciplinary region (shown in the map as curves in green) labeled by the terms such as MEDICINE, MEDICAL, CLINICAL is built on another discipline on the right hand of the map. In the decade to 2019, Fig. 2(b) shows five new trajectories that indicate the different knowledge bases and interdisciplinary trends. This reveals that AI studies have become to an open ecosystem, in which different knowledge bases can collide new sparks for further research. Particularly, there emerges a fresh trajectory named ECOLOGY, EARTH, MARINE in recent decade. It means that AI is beginning to be used in new areas. As a complement, we also conduct a category analysis and find that AI research totally covered 216 subdivision research areas. Although the category of computer science has a dominant role, this area is absolutely interdisciplinary.

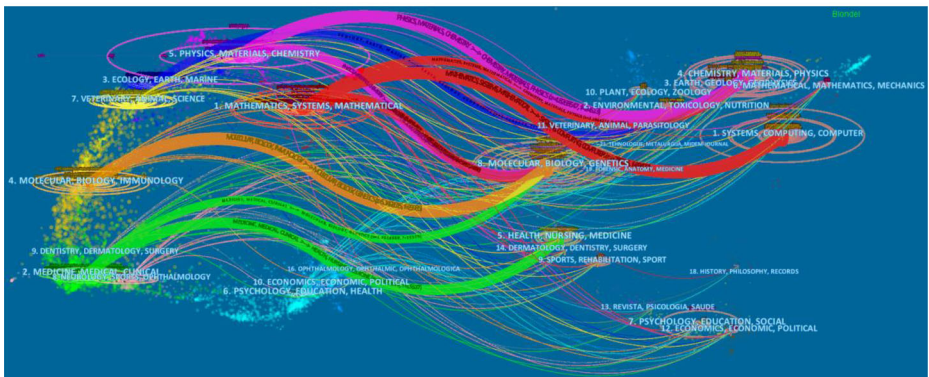
3.2 Co-citation analysis

3.2.1 Visualization results

Reference analysis is the most common type of analysis in the field of bibliometrics. The number of AI publications has a boom increasing in recent decade so that we generate co-citation networks of journal articles in two decades separately. In case of missing important information, we select top 200 levels of most cited publications in every two years with the



(a) The dual-map overlay of 2000-2009

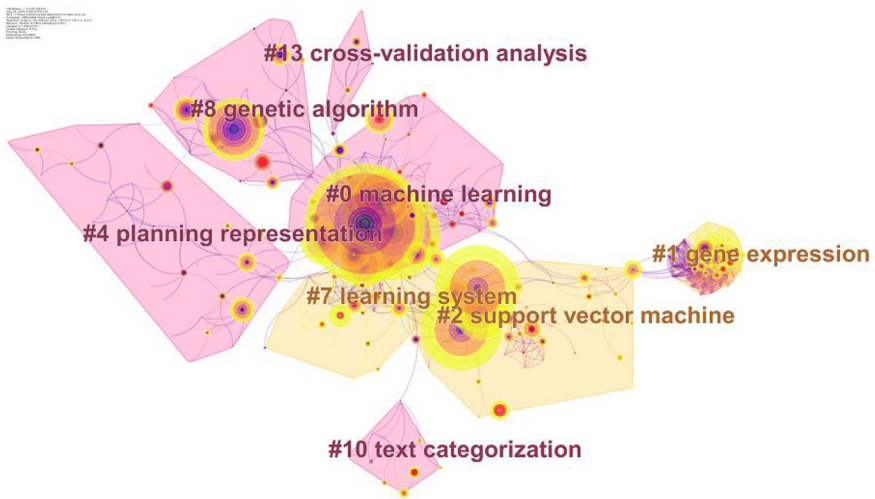


(b) The dual-map overlay of 2010-2019

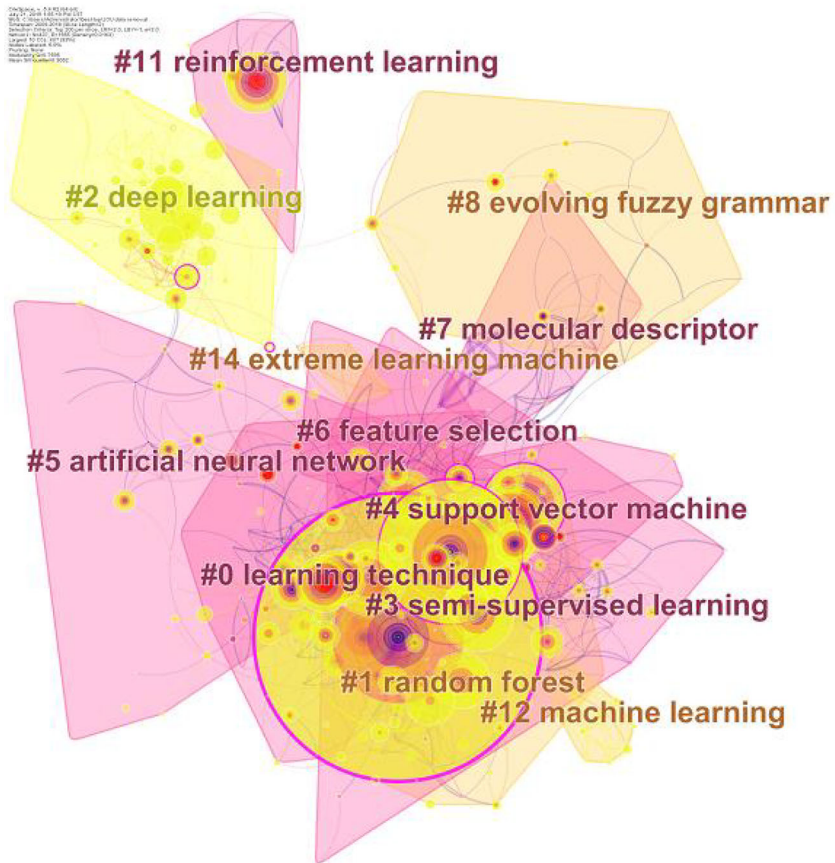
Fig. 2 The dual-map overlays of the AI literature

pruning method of pathfinder to construct the network of references. And then individual networks are synthesized to shape the merged network. The clustered networks of co-cited references are shown in Fig. 3 in a landscape view. The nodes with red tree-rings or large sizes are references required more attention because of their high citation, citation bursts, or both [7]. The color of cluster indicates the time of occurrence and the color rule remains the same as mentioned above. The brighter the color, the closer the time is. In addition, the cluster labels are keeping the same color with their clusters.

Specifically speaking, Fig. 3(a) displays the network in the period of 2000-2009, with 616 nodes and 1907 links. The network has a modularity of 0.8, which is considered relatively high, suggesting that the specialties in AI are clearly defined in terms of co-citation clusters. The 2010-2019 co-citation network under the same selection criteria is shown in Fig. 3(b), including 437 nodes and 1,555 links. It also has a high modularity of 0.76. These two figures visually display the research preference of different stages and also vividly depict the transformation of subdomains. For example, the active emerging cluster (#2 support vector machine) in Fig. 3(a) is becoming a mature field (#4 support vector machine) without activeness in Fig. 3(b). The main cluster (#0 machine learning) in first decade becomes more active in following period because the long term basic research of machine learning starts to affect other areas through application.



(a) The co-citation network of 2000-2009



(b) The co-citation network of 2010-2019

Fig. 3 The landscape view of co-citation network

We can see from the network that the emerging clusters with activeness in recent decade mainly include the themes of learning technique, deep learning, nature language processing, reinforcement learning, and machine learning. The temporal properties of the largest five clusters of two networks are listed in Table 4. The first two columns show the ID of clusters and their sizes. Notably, the value of the third column, “Silhouette”, which is based on the comparison of tightness and separation of cluster, indicates the matching degree between members and clusters. The closer the value is to 1, the higher the homogeneity of the network [29]. In general, the major clusters with high homogeneities are extracted from the co-citation network in terms of cluster size. These clusters present the main sub-domains in AI study and are closely related to each other. Among them, the active ones that portrayed in vivid colors (Fig. 3) are the popular specialties in AI research area. Half of them are active, meaning its attractiveness to scholars.

3.2.2 Major active specialties

In this section, we will specifically focus on the four active clusters in recent decade (see the brighter ones in Fig. 3-b). These specialties, including random forest, deep learning, evolving fuzzy grammar, and machine learning, may not be the largest but the hottest clusters that indicating the trends and hot topics of AI research. To better narrate the main content of each cluster, we respectively process these subsets and find the highly cited articles of the most active clusters in recent decade, listed in Table 5. In a nutshell, these references are the most influential publications that laid foundation and form the basis of emerging trends in AI area.

Specifically speaking, cluster #1 is labeled *random forest* and contains 732 publications in the dataset during last decade. In this cluster, the article with commanding lead is *Random forests* written by Breiman and published in the journal of *Machine Learning*, in which explicitly explore the mechanism and principle of random forest. As one of the ensemble learning algorithm, this learning method has a wide range of application in the fields such as medicine, biological information, management, etc. because of its flexibility and accuracy.

The second largest active cluster (#2) is labeled *deep learning* and contains 599 articles. This specialty is a subset of machine learning and mainly focuses on the theme of neural network. The article with highest citations is *Deep learning*, a more recent publication composed by LeCun et al. in *Nature*. That introduces the development of deep learning and its future. Generally speaking, deep learning is a technology that builds deep neural networks to carry out various analytical activities. The articles in this cluster are mostly about the four key elements of deep learning practice, namely computing capabilities, algorithms, data, and application scenarios. Deep neural network needs a lot of training data and estimated

Table 4 Temporal properties of the largest five clusters

2000-2009				2010-2019			
#	Size	Silhouette	Theme	#	Size	Silhouette	Theme
0	466	0.827	Machine learning	0	739	0.839	Learning technique
1	253	0.897	Gene expression	1	732	0.887	Random forest
2	234	0.963	Support vector machine	2	599	0.91	Deep learning
4	147	0.91	Planning representation	3	422	0.84	Supervised learning
7	129	0.896	Learning system	4	403	0.865	Support vector machine

Table 5 Top highly cited articles in the most active clusters (ranked by citation)

Cluster	Topics	Citations	Title	Year	C/Y	Source
#1 Random forest	Numerical Computation Distribution Prediction Regularization Regression tree	25,550	Random forests	2001	1419.44	Mach Learn
		4,043	Novel methods improve prediction of species' distributions from occurrence data	2006	311	Ecography
		3,163	Greedy function approximation: A gradient boosting machine	2001	175.72	Ann Stat
		2,047	The NumPy array: A structure for efficient numerical computation	2011	255.88	Comput Sci Eng
		1,836	A working guide to boosted regression trees	2008	166.91	J Anim Ecol
		1,467	Random forests for classification in ecology	2007	122.25	Ecology
		1,437	ChEMBL: A large-scale bioactivity database for drug discovery	2012	205.29	Nucleic Acids Res
		1,095	Gene selection and classification of microarray data using random forest	2006	84.23	BMC Bioinformatics
		1,028	Building predictive models in R using the caret package	2008	93.45	J Stat Softw
		913	Extremely randomized trees	2006	70.23	Mach Learn
#2 Deep learning	Neural network Dimension reduction Tree search Visual/speech recognition	8,058	Deep learning	2015	2014.5	Nature
		4,834	A fast learning algorithm for deep belief nets	2006	371.85	Neural Comput
		4,810	Reducing the dimensionality of data with neural networks	2006	370	Science
		4,534	Dropout: A simple way to prevent neural networks from overfitting	2014	906.8	J Mach Learn Res
		3,872	ImageNet large scale visual recognition challenge	2015	968	Int J Comput Vision
		3,109	Faster R-CNN: Towards real-time object detection with region proposal networks	2017	1554.5	IEEE T Pattern Anal
		2,765	Deep neural networks for acoustic modeling in speech recognition	2012	395	IEEE Signal Proc Mag
		2,587	Deep learning in neural networks: An overview	2015	646.75	Neural Networks
		1,763	Mastering the game of Go with deep neural networks and tree search	2016	587.67	Nature
		1,677	Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion	2010	186.33	J Mach Learn Res
#8 Evolving fuzzy grammar	Image features Face detection Pattern recognition	23,526	Distinctive image features from scale-invariant keypoints	2004	1568.4	Int J Comput Vision
		12,464	Image quality assessment: From error visibility to structural similarity	2004	830.93	IEEE T Image Process

Table 5 (continued)

Cluster	Topics	Citations	Title	Year	C/Y	Source
#12 Machine learning	Natural language processing Transfer learning Text expression	9,595	Latent Dirichlet allocation	2003	599,69	J Mach Learn Res
		7,628	Speeded-Up robust features (SURF)	2008	693,45	Comput Vis Image Und
		6,765	Multiresolution gray-scale and rotation invariant texture classification with local binary patterns	2002	397,94	IEEE T Pattern Anal
	5,924	Robust real-time face detection	2004	394,93	Int J Comput Vision	
	5,837	Robust face recognition via sparse representation	2009	583,7	IEEE T Pattern Anal	
	3,715	Fast approximate energy minimization via graph cuts	2001	206,39	Int J Comput Vision	
	2,980	Content-based image retrieval at the end of the early years	2000	156,84	IEEE T Pattern Anal	
	1,139	Enhanced local texture feature sets for face recognition under difficult lighting conditions	2010	126,56	IEEE T Image Process	
	81,601	Generalized gradient approximation made simple	1996	3547,87	Phys Rev Lett	
	7,316	Scikit-learn: Machine learning in python	2011	914,5	J Mach Learn Res	
	6,402	An introduction to ROC analysis	2006	492,46	Pattern Recogn Lett	
	4,163	Statistical comparisons of classifiers over multiple data sets	2006	320,23	J Mach Learn Res	
	2,931	Machine learning in automated text categorization	2002	172,41	ACM Comput Surv	
	2,657	A survey on transfer learning	2010	295,22	IEEE T Knowl Data En	
	2,605	Data clustering: 50 years beyond K-means	2010	289,44	Pattern Recogn Lett	
2,081	Representation learning: A review and new perspectives	2013	346,83	IEEE T Pattern Anal		
2,000	Learning from imbalanced data	2009	200	IEEE T Knowl Data En		
1,660	Top 10 algorithms in data mining	2008	150,91	Knowl Inf Syst		

parameters. Hence, the numeracy, which supports and promotes deep learning, is irreplaceable. Moreover, the algorithm and data are key factors in the development of deep learning. Nowadays, the applications of deep learning are still limited to the most successful fields, speech recognition and image processing. But it has more possibilities in the future.

Cluster #8, labeled *evolving fuzzy grammar*, is the third active specialty. The essence of EFG is that fuzzy grammar evolves incrementally when a change in the learnt model is detected. By using the underlying structure and pattern that represented as fuzzy grammar rules in each evolving cycle, the model becomes more generalized after each change. In this cluster, top cited paper is *Distinctive image features from scale-invariant keypoints* in *International Journal of Computer Vision*, in which presents a method for extracting distinctive invariant features from images that widely used for object recognition. Articles in this subfield mainly discuss the application of this method and compare it with other conventional machine learning techniques such as support vector machine, k-nearest neighbor, and decision tree. Besides, this method is often used in the area of natural language processing to identify, categorize, and express text or image features.

The last active cluster is labeled *machine learning*, which is an ongoing specialty at different stages. Machine learning is a generic term involving a class of algorithms, tasks, and learning theories. Hence, the topics in this sub-domain are pervasive and chiefly focus on various methods and scenarios. In cluster #12, the most popular publication is *Generalized gradient approximation made simple* published in *Physical Review Letters* that shows a simple derivation of generalized gradient approximation. Moreover, we compare this cluster with cluster #2 through the content analysis and find that machine learning is a way to implement AI by using algorithms to parse data, learn from it, and then make decisions and predictions of events in the real world. Deep learning is a technique to realize the application of machine learning and extend the field of AI, through which makes all machine assistance functions possible.

In a nutshell, the top cited articles in above clusters are knowledge source and theoretical foundation of future research. Meanwhile, as complements of above journal analysis, the sources of these articles are also the mainstream journals and provide the intellectual base for AI. Additionally, the main topics confirmed by content analysis in most active clusters are also listed in Table 5. Other topics such as supervised learning, unsupervised learning, semi-supervised learning, artificial neural networks, support vector machine (SVM), feature selection, extreme learning machine, molecular descriptor, gene expression, reinforcement learning, computer vision, intelligent robotics and multi-agent learning are not in the list but still attract a lot of attention. These topics are mainly focused on the theories and applications of algorithms and models that can be classified into two groups, “theory and method” and “application”. There is no doubt that AI as a thriving research field relies on numerous methods [22]. This can be demonstrated by the condition that most topics are algorithms, such as “genetic algorithm”, “support vector machine”, “artificial neural network”, “random forest”, etc.

Additionally, considering the partition of multidisciplinary journals and specialized journals, we divide the JOU data into two parts to show the precise influences. We carry out the co-citation and co-occurrence analyses respectively to discover the differences between general and specialized journals. On the one hand, the results of multidisciplinary data show their influence on machine learning and learning algorithm. When it comes to application, the outcomes discover the prosperity of medicine, biology, and energy. On the other hand, the networks of specialized data obtain similar results with previous JOU data, which indicates the dominant role of specialized journals in AI research. But they pay more attention on feature

selection, deep learning, supervised learning, and reinforcement learning. To sum up, the specialized journals are vital in driving AI research whereas the general journals put more efforts on promotion and application of the frontier research.

Besides, we detect the citation burst of articles in AI research and find that some books with both high citations and burst strength play brilliant roles in this research field. The first one, *Data Mining: Practical Machine Learning Tools and Techniques*, has 1,345 citations and 112.16 burst strength with the burst duration between 2009 and 2013. The second one is named *Pattern Classification*, with citations of 745 and burst strength of 94.3 during 2009–2014. The very title of the third book is *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, a classic of machine learning with 1,306 citations and 68.01 burst strength in the period of 2014–2019. The fourth and fifth ranked books are *Pattern Recognition and Machine Learning* and *Reinforcement Learning: An Introduction*. Undoubtedly, these outstanding books also had giant influence on AI field and laid a firm foundation for the future research.

3.3 Keyword analysis

Keyword analysis, considered as an important tool for analyzing hot topics, can be used to identify the research frontiers of the knowledge domain because author keywords and keywords plus contain core information of articles. The keyword co-occurrence analysis can identify highly valuable keywords, which can indicate past and current research hotspots of a particular domain. Figure 4 shows the keyword network of recent decade in which the links represent the co-occurrence of keywords in one sample, and the thickness corresponds to the relationship intensity. We select top 50 levels of most occurred items from each year to construct the co-occurrence network, finally including 88 nodes and 339 links. Only the nodes with high frequency (over 500) are appeared in the figure. These keywords play a vital role in the network and are proportional to the size of labels. There is no doubt that these hot topics attract the most attention of scholars in AI research.

In line with the outcome of co-citation analysis, the keywords except the search queries “artificial intelligence” and “machine learning” can also be categorized into two groups mentioned above, “theory and method” and “application”. Among the keywords, *neural network*, *learning algorithm*, *support vector machine*, *artificial neural network*, *logistic regression*, *random forest*, *Bayesian analysis*, *genetic algorithm*, *fuzzy logic*, and *particle swarm optimization* belong to the class of “theory and method”. The methods and models have been widely applied in many fields because that AI is an application-driven discipline [1]. As shown in Fig. 4, *classification*, *prediction*, *feature selection*, *identification*, *optimization*, *pattern recognition*, *expert system*, *knowledge-based system*, *design*, *diagnosis*, *performance*, *knowledge management*, *data mining*, and *case-based reasoning* are a part of application fields of AI. Besides, some keywords with strong bursts but not reach the threshold of co-occurrence frequency, such as *multi-agent system*, *natural language processing*, *apoptosis*, *computer vision*, *face recognition*, *image analysis*, *image classification*, *sentiment analysis*, etc. are also main AI concepts.

In the perspective of centrality, the keywords of *classification* (0.33), *diagnosis* (0.22), *performance* (0.18), *machine learning* (0.16), *prediction* (0.15), *risk factor* (0.13), *management* (0.11), *neural network* (0.10), and *support vector machine* (0.09) should be paid more attention in the future. These topics can play gateway roles between sub-fields. Besides, in

view of the citation bursts, the keywords with longer duration such as *gene expression*, *pattern recognition*, *support vector machine*, and *prediction accuracy* are valuable for the trends.

Moreover, to identify the thematic evolution of AI, we carry out a co-word analysis of noun phrases in four stages and list temporal pattern of top 20 frequently used terms with their occurrence frequency in Table 6. This also can be seen as an alternative evidence of keyword analysis to reveal the hot issues and research evolution trends. The result shows the predominance of *artificial intelligence*, *machine learning*, *learning algorithm*, *neural network*, and *artificial neural network* in research domain. The terms such as *support vector machine*, *classification*, and *prediction* have a dramatically increasing trend through the whole period while other terms like *genetic algorithm*, *expert system*, *decision tree*, and *fuzzy logic* have an opposite trend. In addition, the terms with steady ranking (e.g., *data mining*, *pattern recognition*, *feature selection*, etc.) are main application of AI technology.

What we need to focus on is the terms occurring on the list in recent years. For example, *random forest*, *deep learning*, *deep neural network*, *reinforcement learning*, *supervised learning*, *computer vision*, *big data*, and *multi-agent system* are prevailing themes in recent decade and will still deserve extensive attention in the future. The long-term basic research of theory provides numerous methods that will have increasing applications in more areas.

3.4 Comparative analysis

In this subsection, we conduct a comparative analysis between two types of literature, journal article and conference paper, to get a thorough understand of AI field. Firstly, we focus on the different evolutions of publications in the period of 2000-2019. By means of descriptive statistical analysis, Fig. 5 shows the evolution of AI literature. References lay the theoretical and knowledge foundation of the publication. Figure 5(a) depicts the changes of average number of references during 2000-2019, in which shows a similarly and stably upward trend between journal and conference. Generally, journal articles have more references (almost twice) than conference papers. This can be explained by the different acceptance criteria of two types of literature. Conference papers focus more on novelty while journal articles concentrate more on contribution and logicity.

The number of authors can also shed some light on the development of AI field. From Fig. 5(b), we can see that more and more scholars are putting their energy into AI research, which stimulates the rise of publications. The increasing trend of average numbers of authors per paper yields insight into the pervasive research cooperation to a certain degree. Journal articles have more authors than conference papers which demonstrate the collaboration is more needed by journal articles.

Citations can partly reflect the value of literature. We find that the average number of citations per paper, shown in Fig. 5(c), presents a fluctuation trend before 2012 and then appeared a descending trend in recent years. This downtrend is mainly because the latest papers are too new to obtain significant citation rates. And it will change in the future when the articles reach maturity. Moreover, the turning points of journal articles have one year lag than that of conference papers. And these points may be affected by the seminal articles making milestones. Take journal data for example, the first turning point is triggered by *Random forest* published in 2001. Additionally, Fig. 5(d) plots the average numbers of citations per year, in which the growth of citations of journal articles donates the increasing number of references and publications worldwide from another side. And the decline of the average number of citations of conference papers after 2016 can also explain by the immaturity of papers.



Fig. 4 Co-occurrence network of high-frequency keywords

Secondly, we carry out co-citation analysis of conference data to find diverse sub-areas and influential papers in AI domain. We elect top 100 levels of most cited items from each year to shape the co-citation network with a modularity of 0.7959, including 1,499 nodes and 4,624 links, shown in Fig. 6. We can see from the network that the clusters have more papers with citation bursts than that of journal articles (Fig. 3). This partially indicates that conference papers play vital roles in leading the way of research. The largest five clusters contain *face recognition*, *human detection*, *map building*, *semantic segmentation*, and *word sense disambiguation*. While the most active clusters include *graph cut*, *visual tracking*, *deep reinforcement learning*, *entity recognition*, *generative adversarial network*, and *action recognition*. These topics play predominant roles in AI study. In a nutshell, In general, although the subdomains of AI are finely divided, they are primarily belongs to several areas, covering *computer vision*, *natural language processing*, *reinforcement learning*, *unsupervised learning*, and *humanoid robot*.

When a conference paper in our datasets has received the most citations in the period of 2000–2019, we consider it influential. Top 30 cited papers in CON dataset are listed in Table 7, including their publish year, average citations per year, and source. Apparently, CVPR and NIPS play significant roles and lead research trends to some extent. We find that about half of the papers published in 2014 and after, which reveals the powerful influence of top-tier conferences. Compared to journal articles, conference papers especially emerging series of highly cited paper in recent five years can better reveal research hotspots. This can be attributed to the purpose of conferences that communicate fresh ideas and new findings. In addition, journal articles usually have longer peer review process that may lead to a lag of research focus. When we pay attention to the contents of high cited publications in the same period, we find that journal articles lay the knowledge base while conference papers indicate the frontier. For example, the listed papers published in latest five years detect the research

Table 6 Top 20 frequently used terms

R	2000-2004			2005-2009			2010-2014			2015-2019		
	Term	Freq.	Term	Term	Freq.	Term	Term	Freq.	Term	Term	Freq.	
1	Artificial intelligence	1,261	Machine learning	Machine learning	1,864	Machine learning	Machine learning	3,725	Machine learning	Machine learning	15,832	
2	Machine learning	792	Artificial intelligence	Artificial intelligence	1,522	Learning algorithm	Learning algorithm	2,922	Learning algorithm	Learning algorithm	9,481	
3	Learning algorithm	456	Learning algorithm	Learning algorithm	1,405	Artificial intelligence	Artificial intelligence	1,828	Artificial intelligence	Artificial intelligence	4,577	
4	Neural network	445	Neural network	Neural network	702	Support vector machine	Support vector machine	1,369	Support vector machine	Support vector machine	3,668	
5	Genetic algorithm	344	Support vector machine	Support vector machine	666	Classification	Classification	908	Classification	Classification	2,784	
6	Artificial neural network	208	Classification	Classification	473	Neural network	Neural network	823	Neural network	Neural network	2,428	
7	Expert system	193	Genetic algorithm	Genetic algorithm	388	Artificial neural network	Artificial neural network	761	Artificial neural network	Artificial neural network	2,155	
8	Apoptotic index	182	Artificial neural network	Artificial neural network	377	Feature selection	Feature selection	494	Prediction	Prediction	2,122	
9	Decision tree	157	Data mining	Data mining	305	Prediction	Prediction	489	Random forest	Random forest	2,075	
10	Data mining	147	Training data	Training data	241	Reinforcement learning	Reinforcement learning	476	Feature selection	Feature selection	1,649	
11	Fuzzy logic	133	Decision tree	Decision tree	233	Genetic algorithm	Genetic algorithm	436	Deep learning	Deep learning	1,156	
12	Training data	104	Apoptotic index	Apoptotic index	219	Data mining	Data mining	434	Data mining	Data mining	1,047	
13	Classification	102	Feature selection	Feature selection	184	Training data	Training data	400	Deep neural network	Deep neural network	1,038	
14	Support vector machine	91	Pattern recognition	Pattern recognition	150	Random forest	Random forest	316	Supervised learning	Supervised learning	996	
15	Case-based reasoning	79	Fuzzy logic	Fuzzy logic	122	Decision tree	Decision tree	279	Computer vision	Computer vision	923	
16	Inductive logic programming	79	Gene expression	Gene expression	111	Pattern recognition	Pattern recognition	240	Reinforcement learning	Reinforcement learning	879	
17	Pattern recognition	71	Expert system	Expert system	96	Supervised learning	Supervised learning	192	Big data	Big data	776	
18	Knowledge representation	69	Prediction	Prediction	93	Multi-agent system	Multi-agent system	178	Genetic algorithm	Genetic algorithm	768	
19	Decision support system	63	Natural language processing	Natural language processing	81	Fuzzy logic	Fuzzy logic	162	Decision tree	Decision tree	714	
20	Intelligent system	50	Case-based reasoning	Case-based reasoning	81	Particle swarm optimization	Particle swarm optimization	158	Natural language processing	Natural language processing	617	

focus in AI domain. As we can see, *deep neural network*, *convolutional neural network*, *deep residual learning*, and *generative adversarial network* are prevalent models in AI, which were widely applied in real world such as image classification, face recognition, object detection, semantic segmentation, word representation, machine translation, etc.

Thirdly, we examine the evolutions of topics over time for the further study. The co-occurrence network of keywords and terms can be used to identify the popular themes of conferences. Consistent with the keyword analysis of journal articles, we set the same threshold to select the most prevailing keywords/terms to identify the thematic evolution of AI. As comparison, we plot a stacked figure of 20 hot topics over four stages that can clearly reflect the evolution, shown in Fig. 7, in which the vertical axis says the occurrence frequency of topics. The frequency difference between them is mainly due to the length of the publication.

Apparently, there are identical and different topics in both of them. They both concentrate on methods and applications, but with different focuses. The same topics *machine learning*, *learning algorithm*, *neural network*, *data mining*, *pattern recognition*, *training data*, *multi-agent system*, and *reinforcement learning* are prevailing in AI during last two decades. Journals pay more attention to the methods such as *genetic algorithm*, *artificial neural network*, and *support vector machine*, whereas conferences focus far more on applications like *robotics*, *image processing*, *object detection*, and *3d reconstruction*. Furthermore, we can also see a lag effect of journal articles. For instance, *multi-agent system*, *pattern recognition*, and *reinforcement learning* persist in conferences over four stages, but are appeared in journals recently. Hence, what we need to care more about is the topics emerging recently, especially in conferences. We predict that the burgeoning topics like *convolutional neural network*, *deep learning*, *deep neural network*, *generative adversarial network*, and *recurrent neural network* will be more mature in journals and then facilitate the applications of AI technologies in various fields in the future.

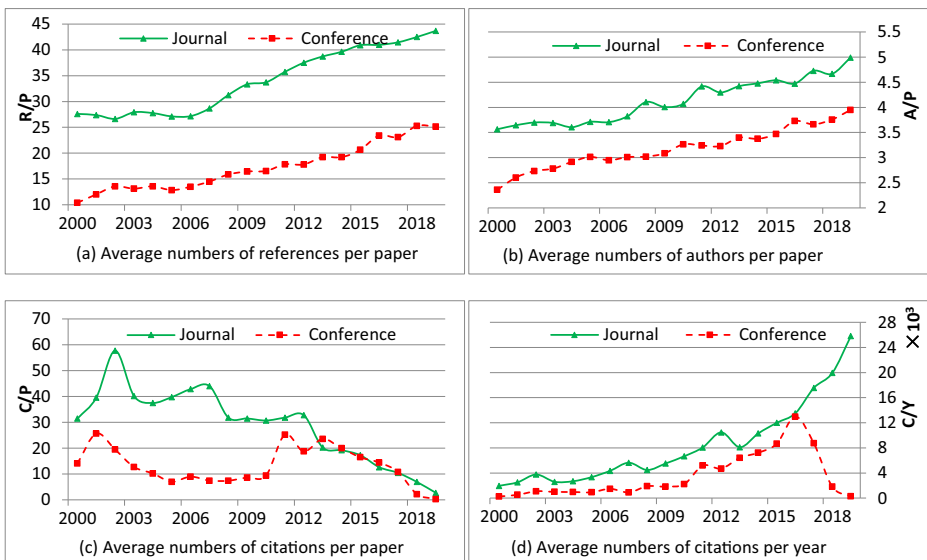


Fig. 5 The evolution of the AI publications

In addition, we still detect the burst terms to recognize the possible blossoming topics in future research. The terms with strong strengths mainly include *supervised learning*, *transfer learning*, *mobile robot*, *human-agent interaction*, *humanoid robot*, *3d architecture*, *image segmentation*, *action recognition*, *machine translation*, *smart grid*, etc. These topics may be attracting more efforts of future researchers.

4 Results of patent analysis

4.1 Descriptive analysis

If the publications primarily represent the state of the art in academia of AI, the patent data can reflect the applications in practice. As one of the most important sources of science and technology, patent information is undoubtedly a key indicator reflecting the level of technological innovation and competitiveness of enterprises and countries. More importantly, patent



Fig. 6 The co-citation network of conference papers in AI

Table 7 Top 30 cited conference papers (ranked by citations)

R	Citations	Title	Year	C/Y	Source
1	24,471	ImageNet classification with deep convolutional neural networks	2012	3495.9	NIPS
2	11,492	Deep residual learning for image recognition	2016	3830.7	CVPR
3	7,230	Going Deeper with Convolutions	2015	1807.5	CVPR
4	6,550	Distributed representations of words and phrases and their compositionality	2013	1091.7	NIPS
5	5,458	Rich feature hierarchies for accurate object detection and semantic segmentation	2014	1091.6	CVPR
6	5,171	Fully convolutional networks for semantic segmentation	2015	1292.8	CVPR
7	5,015	SURF: speeded up robust features	2006	385.8	ECCV
8	4,339	Video Google: atext retrieval approach to object matching in videos	2003	271.2	ICCV
9	4,223	Generative adversarial nets	2014	844.6	NIPS
10	3,599	GloVe: global vectors for word representation	2014	719.8	EMNLP
11	3,523	Algorithms for non-negative matrix factorization	2001	195.7	NIPS
12	3,506	Rectified linear units improve restricted boltzmann machines	2010	389.6	ICML
13	3,498	Batch normalization: accelerating deep network training by reducing internal covariate shift	2015	874.5	ICML
14	3,273	ORB: an efficient alternative to SIFT or SURF	2011	409.1	ICCV
15	3,000	Sequence to sequence learning with neural networks	2014	600	NIPS
16	2,793	Visualizing and understanding convolutional networks	2014	558.6	ECCV
17	2,579	Delving deep into rectifiers: surpassing human-level performance on ImageNet classification	2015	644.8	ICCV
18	2,518	On spectral clustering: analysis and an algorithm	2002	148.1	NIPS
19	2,478	Fast R-CNN	2015	619.5	ICCV
20	2,455	Locality-constrained linear coding for image Classification	2010	272.8	CVPR
21	2,416	You only look once: unified, real-time object detection	2016	805.3	CVPR
22	2,294	Learning realistic human actions from movies	2008	208.5	CVPR
23	2,153	Semi-supervised learning using gaussian fields and harmonic functions	2003	134.6	ICML
24	2,147	Greedy layer-wise training of deep networks	2007	178.9	NIPS
25	2,113	Graph-based visual saliency	2007	176.1	NIPS
26	2,079	Are we ready for autonomous driving? the KITTI vision benchmark suite	2012	297	CVPR
27	2,039	Online object tracking: a benchmark	2013	339.8	CVPR
28	2,014	DeepFace: closing the gap to human-level performance in face verification	2014	402.8	CVPR
29	1,987	Large-scale video classification with convolutional neural networks	2014	397.4	CVPR
30	1,912	3D is here: Point Cloud Library (PCL)	2011	239	ICRA

literature can effectively reveal the progress and trend of technology research and development. Hence, in this section, we conduct a patent analysis in terms of the similar steps of publication analysis.

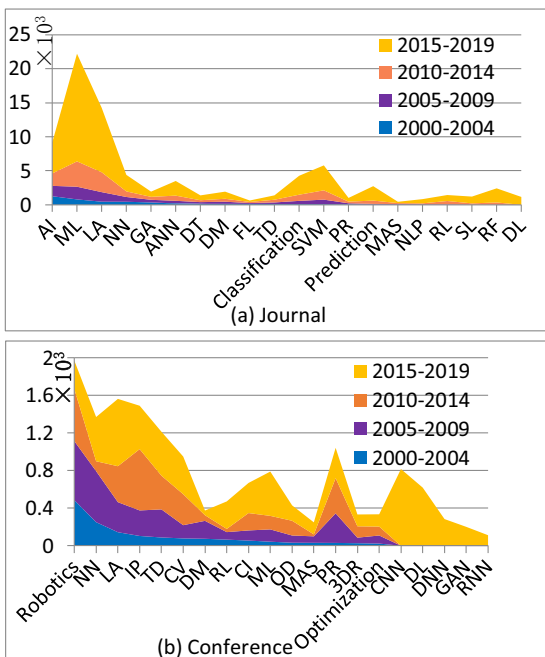
First of all, the regional distribution and the number of patents can reveal the R&D strength of a country and patentee to a certain extent. The country analysis shows that patents of AI technology mainly center on USA, Japan, China, Korea, and Germany. Table 8 displays the top ranked companies possessing AI patents in the period of 2000-2019, with their patents number, burst values, and countries. To be specific, USA plays a leading role in AI domain, followed by Japan. Although the related researches in China started late and most patents pertained to the universities, the development trends of Chinese company are still anticipated with the industry-university cooperation strategy. This conclusion can also be certified by the measure of burst value.

Apparently, the companies listed in Table 8 have predominant positions on the practice of AI technology and also make brilliant contributions for the competitive advantages of their

country. USA shows its dominant position in AI along with the highest number of patents. IBM has been successfully commercializing the AI technologies. Take Watson, a platform of AI, for example, it can provide solutions for various institutions using AI technologies like language process, machine learning, data mining, and information analysis. MICROSOFT put its effort on speech recognition, image identification, and computer vision. The product named Microsoft Xiaoice has realized the natural interaction through understanding the context and semantics of the conversation. Both of the first two companies have high burst values that also demonstrate their position. Although the burst value is relatively low, GOOGLE also has become one of the most powerful companies in AI technology through merger and acquisition, investing and launching AI-related companies and projects. Japan ranked second in this field, especially in the sub specialty of image and voice recognition. FANUC plays a vital role in AI robots and FUJITSU makes a breakthrough in data analysis.

Significantly, the patent applications of China have shown an explosive trend in the field of AI since 2010. BAIDU is at the forefront of Chinese companies with a high burst value that reveals its importance in this field. It is also the first one that set up AI research institute and AI platform in China. Relying on the advantage of Chinese big data, its R&D strengths of Chinese speech recognition, automatic drive, and computer vision are in a leading status. Other Chinese companies not listed in the table, for example, ALIBABA, TENCENT, HUAWEI are also important practitioners in AI technique.

Moreover, in terms of the subject area to which the patent belongs, the top ranked disciplines are engineering, computer science, neurosciences, instruments instrumentation, general internal medicine, telecommunications, biomedical, imaging science, and automation control systems. This result is consistent with the above analysis of publication data and indicates the mutually reinforce between theoretical research and practice.



- AI-Artificial intelligence,
- ML-Machine learning,
- LA-Learning algorithm,
- NN-Neural network,
- GA-Genetic algorithm,
- ANN-Artificial neural network,
- DT-Decision tree, DM-Data mining,
- FL-Fuzzy logic, TD-Training data,
- SVM-Support vector machine,
- PR-Pattern recognition,
- MAS-Multi-agent system,
- NLP-Natural language processing,
- RL-Reinforcement learning,
- SL-Supervised learning,
- RF-Random forest, DL-Deep learning,
- IP-Image processing,
- CV-Computer vision,
- CI-Computational intelligence,
- OD-Object detection,
- 3DR-3d reconstruction,
- CNN-Convolutional neural network,
- DNN-Deep neural network,
- GAN-Generative adversarial network,
- RNN-Recurrent neural network

Fig. 7 The temporal trend of topics in AI

Table 8 Top 15 ranked assignees

Company	Number	Burst	Country	Company	Number	Burst	Country
IBM	1,265	87.97	USA	INTEL	243	14.17	USA
MICROSOFT	607	83.6	USA	AMAZON	162	28.04	USA
GOOGLE	434	36.22	USA	NEC	154	7.65	Japan
BAIDU	397	35.3	China	QUALCOMM	136	8.91	USA
FANUC	292	7.59	Japan	APPLE	129	12.46	USA
SAMSUNG	288	3.91	Korea	FUJITSU	124	3.77	Japan
FACEBOOK	263	28.35	USA	SIEMENS	114	5.47	Germany
				CISCO	111	15.17	USA

4.2 Co-occurrence analysis

Based on the patent data, we conduct keyword and category co-occurrence network to analyze the hotspots and trends in AI technology. Firstly, we compute the BC values of patents with high frequency sorted by the Derwent manual code. The high BC value donates the vital position in the network and the frequency is proportional to the breath of research. The key technologies with high BC values are shown in Table 9.

We find that the most popular technologies belong to T01 (digital computers), especially T01-J (data processing systems), T01-N (information transfer), T01-E (data processing), and T01-S (software content). This demonstrates that most of the key technologies in AI are involved in data processing that based on the breakthrough of algorithm. Moreover, the patents related to educational equipment and multimedia systems are valuable technologies in AI. The pervasive application of AI in education can promote the further improvement of intelligence level. Although less frequency, the techniques applied in chemical and medicine also have high BC values that donate the impact of AI on these industries. In the perspective of frequency, especially that with low centrality but higher than 1,500 counts, such as W04 (audio/visual recording and systems), W01 (telephone and data transmission systems), T04 (computer peripheral equipment), S05 (electrical medical equipment), T06 (process and machine control), P86 (musical instruments acoustics), are also remarkable technologies in practice.

Secondly, we select top 100 levels of occurred keywords from each year to visualize the keyword co-occurrence network of 2000–2019, with 168 nodes and 323 links. The modularity of 0.7587 and mean silhouette of 0.9484 depict the high homogeneity of the network. The largest five clusters are construction site, voice recognition engine, face detection, digital signal processor, and driving arrangement structure, indicating the hot application field in recent years. Specifically, data mining, machine control, gesture control, and voice control are hot technologies and will deserve further research. Production systems, traffic systems, electrical instruments, semiconductor materials, surgery diagnosis, broadcasting, marine and radar systems are major areas of application of AI technique.

The burst value can measure the evolution of turning point in AI and the keyword with high strength of burst donates its hot position. Hence, thirdly, we visualize the occurrence burst history to find the change trends of technology fronts, sorted by the strength value of burst. Considering the lag between theory and practice, we choose a lag phase for burst detection. We can see from Fig. 8 that the dramatic topics are mainly involved in digital computers, data transmission, and audio/visual recording. Specifically, the technology with highest burst strength is the logic operation (T01-E01) that is the foundation of other technology innovation

Table 9 Key technologies in AI domain (ranked by BC)

Technology	Freq.	BC	Category	Technology	Freq.	BC	Category
T01-N01B3	3,848	0.41	Online education	T01-J16C	4,214	0.16	Knowledge processing
T01-S03	5,897	0.29	Software products	T01-J04	2,656	0.16	Mathematical model
T01-J05B	11,715	0.28	Information retrieval	T01-J10B2	4,333	0.15	Image analysis
T01-E01	2,950	0.20	Basic data analysis	T01-N01D	4,436	0.14	Data transfer
T01-J30A	4,743	0.18	Educational aids	T04-D04	1,262	0.12	Recognition & identification
W01-C01D3C	1,165	0.17	Portable terminal	T01-J06A	1,144	0.11	Medical equipment

activities. Other technologies can be roughly divided into three categories, including portable terminals and smartphone, language/speech process, program management and inner mechanism. Besides, the keywords with long impact duration should give more attention, especially those with impact duration lasting to recent are the focus and fronts of ongoing research. For example, social media/virtual communities (T01-N01A2D), neural networks (T01-J16C1),

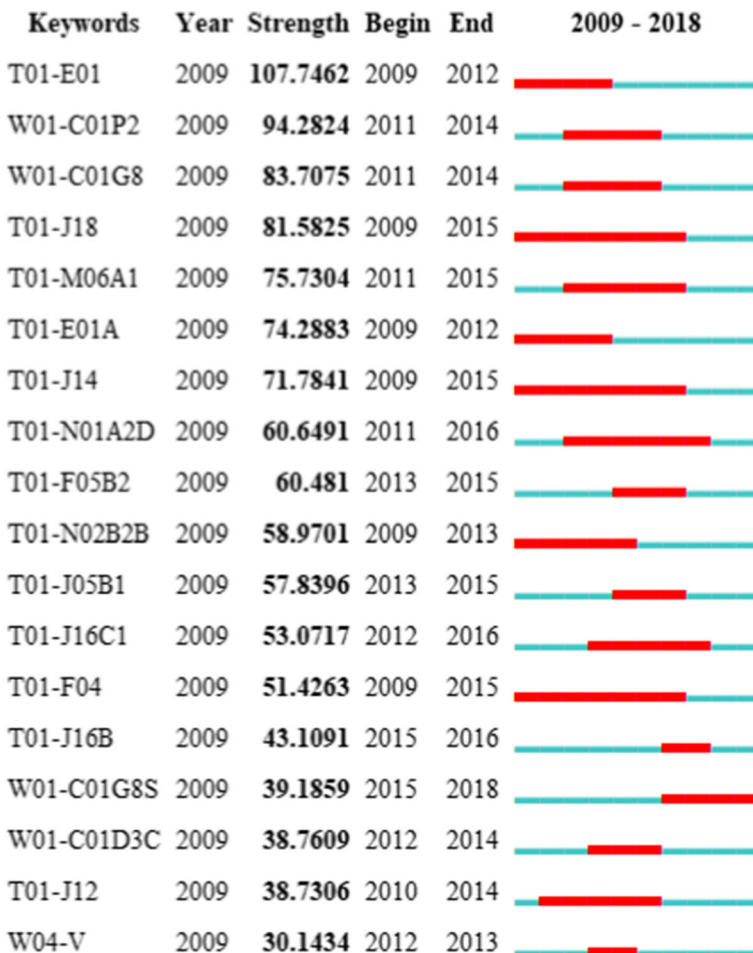


Fig. 8 Patent keywords with the burst values over 30

fuzzy logic systems (T01-J16B), and smartphone (W01-C01G8S) maybe play important roles in the future.

In short, AI technology is becoming more mature and its applications are expanding. From the perspective of patent analysis, the results show that the patents basically cover all directions of AI research. In particular, face recognition, social media, autonomous driving, humanoid robot are hot directions in the practical application. Meanwhile, the patents in the fundamental area, such as database, machine learning, human-computer interaction, internet of things, big data, voice technology, virtual reality, deep learning, natural language processing are particularly active.

5 Conclusion and discussion

5.1 Research findings

AI, as a technology that will lead the future, is profoundly changing the world. This term was originally coined in the 1950s and develops slowly for a long time. Until the beginning of 20th century, machine learning and, later, deep learning, as tools for solving complex problems in the world of data, has sparked interest in AI domain. In the past six decades, AI has entered in a real explosion after two ups and downs. Its rapidly development brings both opportunities and challenges to the industry and social change. By using the bibliometric analyses, this study portrays a holistic overview of AI domain in academic and practice with the data retrieved from WoS, Scopus and DII spanning over the period of 2000-2019. To better quantify the development, we have used journal articles, conference papers, and granted patents metadata. Furthermore, in order to remove the subjectivity error and to ensure measurement accuracy, we conduct a data cleaning and adopt Citespace to analyze and visualize these datasets. Combining the descriptive analysis, comparative analysis, and the results of co-citation and co-occurrence networks, the fundamental (intellectual base) and evolution (research fronts) of AI research has been present and the conclusions can be drawn as follows.

In a nutshell, analyses based on these datasets show the complementary results. This reveals the mutually reinforcement between journal and conference, academic and industry. The descriptive analyses show the dominance and significant impact of USA both in theory and practice. Moreover, there still exists a big step between developed and developing economics. China, though, has already at the top with increasing number of publications and patents in recent years, the researchers and companies should pay more attention to quality than quantity.

On the one hand, from the perspective of academic research, we find that the application of AI is becoming more open-minded and cross-disciplinary. The dual-map overlay reveals that AI has mainly applied in the disciplines of physics, materials, chemistry, ecology, systems, mathematics, computer, molecular, biology, genetics, animal, medicine, psychology, education, etc. The co-citation analysis explores the inner structure of AI domain, including various sub-fields and massive topics. The changing of the most popular keywords and terms over time reflects the trajectories of research hotspots. Moreover, the comparative analysis between journal data and conference data shows distinctions and connections among them. Conference papers have bigger impact on research focus, while journal articles lay more foundation for knowledge base. Both of them work together to advance the development of AI.

Furthermore, the findings based on PUB data also reveal some blossoming sub-fields and topics where researchers could apply their efforts. For example, the area of deep learning,

natural language processing, and reinforcement learning continue their heat in research, especially in applied research. The emerging algorithms such as *convolutional neural network*, *deep neural network*, *generative adversarial network*, and *recurrent neural network* will deserve more attention in next. Moreover, AI technologies now widely applied in the industries like robot, medicine, healthcare, energy, and retail. Future applications will be more widespread and deeper. In addition, AI technology is in a stage of rapid development with the increasing trend of commercialization in practice. The pervasive application to various industries also bring the challenges so the next-generation AI would be make breakthroughs both in basic theories and overcoming existing limitations.

On the other hand, in terms of granted patents, we find that currently advantaged technologies and hot areas in AI include intelligent education, portable smart terminal, biological recognition and identification, image analysis, voice/gesture control, automatic drive, and application of cloud service. AI technology is in its stage of rapid development and has commercialized and penetrated to various industries such as education [16], finance [17], engineering [36], medicine [14, 15], transportation [35], agriculture, et al.

5.2 Future research trends and limitations

This paper contributes to AI by clarifying the existing literature and proposing some research frontiers and directions. At present, AI is still in the stage of perceptive intelligence with the core technologies of speech and visual recognition and is achieving huge breakthroughs with deep learning algorithms. In the future, AI will continue to undergo rapid development inspired by more advanced methods. Currently, AI is mainly based on the technological breakthrough in deep learning, which heavily relies on the data mining technology to acquire a large amount of effective training data. That is to say, computing power and data are the foundations of the development of AI industry. Hence, in the coming years it will be possible to observe the emergence of AI chip for implementation of AI solutions. The specialized calculating terminal is needed to candle various intelligent tasks.

From the technology point of view, it is expected that breakthroughs at algorithms will further promote the development of AI. In accordance with the results of the research, it is reasonable to think of a future that intelligent algorithms will remain the core position. Algorithms illustrate the behavioral pattern of AI. And the algorithms such as deep learning, semi-supervised learning, deep reinforcement learning, are the focus of future research. Take deep learning for example, it is an important driving force for the development of AI on the supportive premise of big data in current stage. Under this condition, scientific works have achieved some progress in the technological areas of computer vision, speech interaction and natural language processing. Nonetheless, there is still no generally suitable deep learning framework with common information processing structures that can deal with various application tasks. Undoubtedly, these fields still need more efforts for further research, especially in the aspect of improving the generalization of deep learning network frameworks and applications.

Generally, AI yields deeper insights by three basic steps of modeling and pattern detection, prediction and optimization. Specifically speaking, modeling is the key prerequisite for reasoning to resolve complex problems and pattern detection is a direct way of extracting hidden information from large scale of data. Prediction task can be considered as a process that learns from historical data to make precise predictions for new conditions. And optimization is a decision-making process for seeking optimal solution. At present, deep learning has

encountered huge challenges in all these three steps due to the complexity of big data and the lack of mature model dealing with non-convex optimization either in theory or practice. In other words, AI technology is still driven by processing data rather than knowledge-based methods such as logical reasoning and predicting. Weak AI has made breakthroughs while strong AI is still to be further explored.

AI is making our lives more convenient through intelligent interaction. With arrival of the third wave of AI, its application in various industries is becoming more and more widespread. In the perspective of technology applications, the most extensive and leading industries are high-tech fields like automotive, humanoid robot, financial, healthcare, telecommunications, followed by logistics, transportation, retail and media industries. AI technology has entered the large-scale commercial stage, and AI products will enter the consumer market in the future. Take humanoid robot for example, it has widely used in customer service, supply chain and smart homes. The comprehensive application in the field of commercial services is opening a new way for the large-scale commercial use and intensifying the pragmatic tendency of AI.

However, there is limited scholarly research investigating the risks of wider AI utilization. Although it is too early to worry about the threats in the development of AI, there is no doubt that it is taking jobs away from workers in every industry. The mass unemployment will be the most pressing problem, especially in labor-intensive industries. Moreover, upcoming disruptions of AI in cities and societies have not been adequately examined.

Besides, we also find some implications for technology management. First of all, the majority of countries promote the development of AI to the national strategy level. Hence, it is imperative for policy makers to introduce more practical promotion policies. The deep combination of AI technology and other industries should be taken into consideration when implementing policies. In the next place, AI-related companies should keep pace with the development trends. AI is a rapidly developing area that needs more resources and abilities. Therefore, the companies should pay more attention to the input of human and financial resources. And the management of intellectual property is a key to success.

Inevitably, this study, like many others, has its own limitations. Firstly, while we construct search queries to retrieve the datasets, the data source may be simple. We cannot guarantee that all related records are in our database, especially conference data. Secondly, the bibliometric methods have an intrinsic weakness that assuming a document is cited when thought important without considering other motivations that may drive authors in citing prior publications. Hence, the results generated by CiteSpace software may be affected by algorithm noise and sampling bias. More precise data and efficient algorithms should be constructed to avoid this. Last but not least, the analyses maybe too descriptive to uncover the inner structure of the networks and more insightful perspective should be needed in further research.

References

1. Browne M et al (2007) Near-shore swell estimation from a global wind-wave model: spectral process, linear, and artificial neural network models. *Coast Eng* 54(5):445–460
2. Carvalho MM, Fleury A, Lopes AP (2013) An overview of the literature on technology roadmapping (TRM): contributions and trends. *Technol Forecast Soc Chang* 80(7):1418–1437
3. Castelfranchi C (2013) Alan Turing's "computing machinery and intelligence" *Topoi* 32(2):293–299
4. Chen C (2004) Searching for intellectual turning points: progressive knowledge domain visualization. *Proc Natl Acad Sci USA* 101(Suppl 1):5303–5310

5. Chen C (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inform Sci Technol* 57(3):359–377
6. Chen C et al (2009) Towards an explanatory and computational theory of scientific discovery. *J Informet* 3(3):191–209
7. Chen C (2017) Science mapping: a systematic review of the literature. *J Data Inf Sci* 2(2):1–40
8. Chen C, Leydesdorff L (2014) Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. *J Am Soc Inf Sci* 65(2):334–351
9. Chen C, Ibekwe-SanJuan F, Hou J (2010) The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. *J Am Soc Inform Sci Technol* 61(7):1386–1409
10. Cobo MJ et al (2015) 25years at Knowledge-Based Systems: A bibliometric analysis. *Knowl Based Syst* 80: 3–13
11. Érdi P et al (2013) Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics* 95(1):225–242
12. Fernandes C et al (2017) The dynamic capabilities perspective of strategic management: a co-citation analysis. *Scientometrics* 112(1):529–555
13. Fujii H, Managi S (2018) Trends and priority shifts in artificial intelligence technology invention: A global patent analysis. *Econ Anal Policy* 58:60–69
14. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. *Metabolism* 69:S36–S40
15. He J et al (2019) The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 25(1):30–36
16. Hinojo-Lucena F et al (2019) Artificial intelligence in higher education: a bibliometric study on its impact in the scientific literature. *Educ Sci* 9(1):51
17. Koyuncugil AS, Ozgulbas N (2012) Financial early warning system model and data mining application for risk detection. *Expert Syst Appl* 39(6):6238–6253
18. Liu J et al (2018) Artificial Intelligence in the 21st Century. *IEEE Access* 6:34403–34421
19. Marr D (1977) Artificial intelligence—a personal view. *Artif Intell* 9(1):37–48
20. Mikhaylov SJ, Esteve M, Campion A (2018) Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. *Philos Trans Math Phys Eng Sci* 376(2128):1–21
21. Najmi A et al (2016) Reviewing the transport domain: an evolutionary bibliometrics and network analysis. *Scientometrics* 110(2):843–865
22. Niu J et al (2016) Global research on artificial intelligence from 1990–2014: spatially-explicit bibliometric analysis. *ISPRS Int J Geo Inf* 5(5):66
23. Palaniappan R, Sundaraj K, Sundaraj S (2014) Artificial intelligence techniques used in respiratory sound analysis—a systematic review. *Biomed Eng-Biomed Tech* 59(1):7–18
24. Parkes DC (2015) Wellman, Economic reasoning and artificial intelligence. *Science* 349(6245):267–272
25. Patricio DI, Rieder R (2018) Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput Electron Agric* 153:69–81
26. Prasad S, Tata J (2005) Publication patterns concerning the role of teams/groups in the information systems literature from 1990 to 1999. *Inf Manag* 42(8):1137–1148
27. Ramos-Rodríguez A, Ruiz-Navarro J (2004) Changes in the intellectual structure of strategic management research: a bibliometric study of the *Strategic Management Journal*, 1980–2000. *Strateg Manag J* 25(10): 981–1004
28. Rodríguez A et al (2016) Patent clustering and outlier ranking methodologies for attributed patent citation networks for technology opportunity discovery. *IEEE Trans Eng Manage* 63(4):426–437
29. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
30. Salah K et al (2019) Blockchain for AI: review and open research challenges. *IEEE Access* 7:10127–10149
31. Sikdar S (2018) Artificial intelligence, its impact on innovation, and the Google effect. *Clean Technol Environ Policy* 20:1–2
32. Soranzo B, Nosella A, Filippini R (2016) Managing firm patents: a bibliometric investigation into the state of the art. *J Eng Tech Manag* 42:15–30
33. Tseng C, Ting P (2013) Patent analysis for technology development of artificial intelligence: a country-level comparative study. *Innovation* 15(4):463–475
34. Turing AM (1950) Computing machinery and intelligence. *Mind* 59(236):433–460
35. Wang F (2017) Artificial intelligence and intelligent transportation: driving into the 3rd axial age with ITS. *IEEE Intell Transp Syst Mag* 9(4):6–9
36. Wang M et al (2018) A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Appl Energy* 220:480–495
37. Youssef A, El-Telbany M, Zekry A (2017) The role of artificial intelligence in photo-voltaic systems design and control: a review. *Renew Sustain Energy Rev* 78:72–79

38. Zeng Y, Wang L (2017) Fei-Fei Li: Artificial intelligence is on its way to reshape the world. *Natl Sci Rev* 4(3):490–492

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.