A Review of Artificial Intelligence Risks in Social Science Research

Yan Du[1](✉) and Chihping Yuan[2]

[1] Institute of Guangdong, Hong Kong and Macao Development Studies, Sun Yat-Sen University, 135 Xingang West Road, Guangzhou, China
duyan5@mail2.sysu.edu.cn

[2] School of Economics and Trade, Guangzhou Xinhua University, 7 Yanjiangxiyi 1st Road, Dongguan, China
yuanchip@mail.sysu.edu.cn

Abstract. This paper uses CiteSpace to analyze the knowledge map of Artificial Intelligence (AI) risks in social science research. Scholars mainly analyze the technical, ethical, social and existential risks brought about by AI technology from the perspective of computer science, economics, medicine and philosophy. It is foreseeable that with the advancement of AI technology, AI risks will become more complex. As the real world changes, regulations and safeguards for AI need to be further explored and improved in the future.

Keywords: artificial intelligence · risks · CiteSpace · knowledge map

1 Introduction

With the progress of AI technology and the expansion of applications, the risk of AI has gradually become a hot spot in academia, resulting in a lot of related literature in Computer Science, Philosophy, Economics and other disciplines. According to the results of the Web of Science Core Collection, more than 80% of the literature involving "artificial intelligence" and "risk(s)" has been in Computer Science since the establishment of the database, while the highest proportion of social science subjects is Business & Economics, accounting for about 26%.

The risks presented by AI are understood to be social problems in need of social as much as technical solutions. Social scientists have done much to raise awareness of the complex problems encountered and produced by AI outside the laboratory. This has involved a material and discursive push into the organisations, disciplines and debates through which these technologies are being developed. Substantial contributions have been made to characterise AI and propose measures for regulating the risks they generate [1].

2 Subject Categories Analysis

While the research on AI risks in natural sciences mainly focuses on the theoretical and technical problem, this paper highlights the important role of social science research.

A. Luqman et al. (Eds.): PMIS 2022, AHCS 6, pp. 273–293, 2023.
https://doi.org/10.2991/978-94-6463-016-9_30

Fig. 1. A Pathfinder network of subject categories.

A topic search for "artificial intelligence AND risk" results in 809 English articles published before November 20, 2021. CiteSpace is used to generate and analyze the knowledge map of AI risks based on bibliographic records retrieved from SSCI database. The networks clearly show the intellectual structure of AI risks, including five aspects such as disciplines, keywords, co-cited references, significant clusters, and citing papers analysis, which has great enlightenment significance for us to conduct AI risks research.

Which disciplines are involved in the social science research of AI risks? Figure 1 shows the distribution of disciplines in the retrieved papers. In each year from 2002 to 2021, we select the 20 most frequently cited papers and use Subject Category Analysis in Citespace to extract their subject information and combine them into the chart above. The size of the circle represents the number of the papers in this category, and the color represents the time of publication of the paper. In Fig. 1, Business & Economics, with the largest circle, is the most common category. Disciplines that appear with at least 50 frequencies are listed in Table 1. We can intuitively see that Business & Economics, Computer Science and Engineering are the most popular categories, which is roughly the same as the dual-map above.

CiteSpace characterizes emerging trends and patterns of change in such networks in terms of a variety of visual attributes. The citation rings in purple shows the structural properties of the node, and its thickness indicates the degree of its betweenness centrality [2]. Although some rings are very small, they are also marked purple because they are more centrally mediated. Table 2 lists 8 categories with a centrality of more than 0.1. As can be seen, Engineering, Business & economics are not only research hotspots, but also play a unique role in connecting different disciplines.

Red circles indicate articles with citation bursts, that is, rapid increases of citation counts. 12 categories with strong citation bursts are listed in Fig. 2. The early outbreak research related to AI risks mainly focus on Operations Research & Management Science, Economics, etc. However, what the dual-map did not indicate was that in recent years, more papers have been on Law, Medical Informatics and Health Care. It may be because they are emerging fields and have not yet been able to accumulate rich research to become the focus.

Table 1. Major Subject Categories.

Rank	Freq	Category
1	150	BUSINESS & ECONOMICS
2	133	COMPUTER SCIENCE
3	99	ENGINEERING
4	84	ENVIRONMENTAL SCIENCES & ECOLOGY
5	73	ENVIRONMENTAL SCIENCES
6	66	SCIENCE & TECHNOLOGY - OTHER TOPICS
7	59	PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
8	55	MANAGEMENT
9	54	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
10	53	SOCIAL SCIENCES - OTHER TOPICS
11	50	BUSINESS
12	50	ECONOMICS

Table 2. Subject Categories with the highest betweenness centrality (>= 0.1).

Rank	Centrality	Freq	Category
1	0.36	99	ENGINEERING
2	0.28	150	BUSINESS & ECONOMICS
3	0.2	2	WATER RESOURCES
4	0.19	133	COMPUTER SCIENCE
5	0.15	25	PSYCHOLOGY
6	0.11	32	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS
7	0.1	53	SOCIAL SCIENCES - OTHER TOPICS
8	0.1	2	METEOROLOGY & ATMOSPHERIC SCIENCES

Subject Categories	Year	Strength	Begin	End
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	2002	11.5	**2002**	2018
ECONOMICS	2002	4.05	**2006**	2011
MATHEMATICS	2002	3.92	**2006**	2017
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	2002	9.83	**2009**	2018
ENGINEERING, ELECTRICAL & ELECTRONIC	2002	7.77	**2009**	2018
COMPUTER SCIENCE	2002	3.59	**2011**	2012
REGIONAL & URBAN PLANNING	2002	3.39	**2013**	2018
NEUROSCIENCES & NEUROLOGY	2002	3.54	**2016**	2018
PSYCHIATRY	2002	3.93	**2017**	2019
LAW	2002	5.44	**2018**	2021
HEALTH CARE SCIENCES & SERVICES	2002	9.22	**2019**	2021
MEDICAL INFORMATICS	2002	6.69	**2019**	2021

Fig. 2. Top 12 Subject Categories with the strongest citation bursts. Photo credit: Original

3 Keywords Analysis

In addition to focusing on Subject Categories, this paper also show a more intuitive scope through Keyword analysis. CiteSpace divides the co-citation network into a number of clusters of co-cited references, where references are tightly connected within the same clusters. Figure 3 shows the 12 most significant clusters labeled by keywords from citing articles of this cluster.

In Fig. 3, we label clusters with keywords by the LLR (log likelihood ratio) algorithm to emphasize the unique of clusters [3]. Table 3 lists the 12 major clusters by their size. Cluster #1 and #10 mainly focus on the acceptance and application of AI, while Cluster #11 concentrates on bias and discrimination caused by the AI algorithms.

We can roughly divide these 12 clusters into four fields (Table 4). The first mainly focuses on the risks existing in AI theoretical algorithms, including predictive models, machine learning, iterative and neural networks, etc. (#0, #3, #, #4, #12); the second concentrates on AI application in industries, including medical industry, automobile industry and industry planning, etc. (#8, #9, #10); the third field includes AI application scenarios in common company management, including supply chain management, financial health and project management etc. (#2, #5, #6); the fourth is mainly questions related to AI social recognition and acceptance (#1, #7).

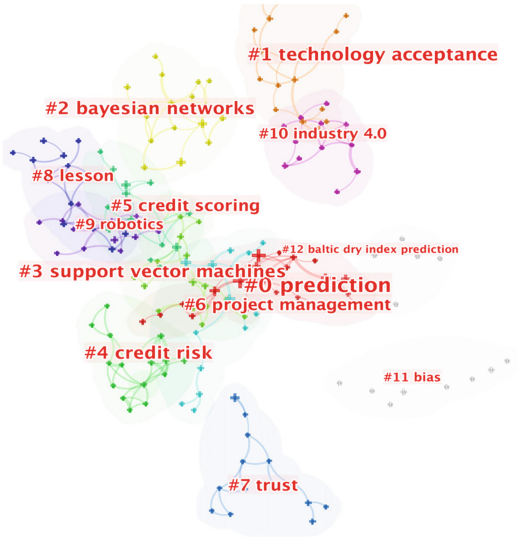

Fig. 3. A landscape view of the keyword network.

Table 3. Major clusters of co-cited references.

Cluster ID	Size	Silhouette	Label (LLR)	Mean (Year)
0	22	0.972	Prediction	2016
1	18	0.971	technology acceptance	2020
2	17	0.931	Bayesian networks	2019
3	17	0.93	support vector machines	2018
4	16	0.914	credit risk	2017
5	14	0.944	credit scoring	2017
6	14	0.929	project management	2017
7	14	0.979	Trust	2019
8	13	0.778	Lesson	2018
9	12	0.928	Robotics	2019
10	12	0.892	industry 4.0	2020
11	8	0.912	bias	2020
12	6	0.986	Baltic dry index prediction	2018

Table 4. Classification of clusters by research fields.

ID	Fields	Clusters
1	algorithm	prediction (#0) machine learning (#3) iteration (#4) neural network (#12)
2	industry applications	lesson (#8) Robotics (#9) industry 4.0 (#10)
3	company management	customer trust (#2) financial health (#5) project management (#6)
4	attitude	acceptance (#1) Trust (#7)

4 Co-citation Analysis

Due to the ambiguity of the interdisciplinary terms, there are apparently unrelated topics in a cluster in Keyword analysis. In addition to the analysis of Subject Category and Keyword, this study also conduct a more targeted co-citation analysis in this section. Co-Citation analysis means that two papers appear together in the bibliography of the

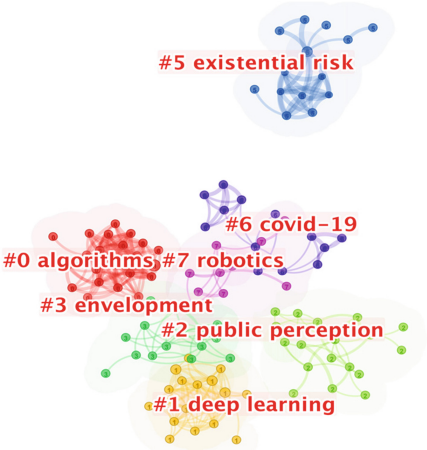

Fig. 4. A landscape view of the co-citation network.

third citing paper, and the two documents form a co-citation relationship [3]. Through co-citation analysis, it is easier for us to grasp the main research basis and research hotspots in this field. The references cited in one paper are represented as nodes in a co-citation network. The lines between the nodes of these references represent how often they were cited in the same paper. It is a hypothesis that if two references are frequently cited together, they are related in some way [2].

The following landscape view is generated based on co-cited references included in the retrieved bibliographic records (Fig. 4). The seven largest connected components include 111 nodes, which account for 76% of the entire network. Clusters in different colors indicate the time when co-citation links appeared for the first time. Clusters in red were generated earlier than that in purple as shown in the legend at the bottom left. Each cluster is labeled by keywords of citing papers. Table 5 lists the 7 main clusters by size with labels extracted by three different algorithms. As in the previous section, we use the keywords extracted in the LLR algorithm by default.

The development of research can also be identified from the list of papers with strong citation bursts in major clusters (Fig. 5). The larger the "strength", the more the paper is cited in a short period of time, which is likely to be an important milestone in its field. It can be seen that the time sequence of citation bursts is roughly consistent with the timeline view of the co-citation network, and there is a lag between publication time and citation bursts. In 2015, a paper that emphasized the threat of advanced artificial agents briefly appeared [4]. Then in 2018, the impact of deep learning on the medical field began to attract attention [5], and the attention continues until now. In the same year, the legal correction of data discrimination became the second hot spot of the year [6]. In the past two years, AI has significantly facilitated applied research in the Medicine [7], and even clinical predictive analytics through electronic health records [8]. At the same time, scholars began to pay attention to the moral hazard of AI [9], and began to emphasize the importance of government regulation [10].

Table 5. The five largest clusters of co-cited references.

Cluster ID	Size	Silhouette	Label (LLR)	Mean (Year)
0	22	0.964	Algorithms	2016
1	19	0.977	deep learning	2017
2	18	0.923	Public perception	2018
3	15	0.645	Envelopment	2018
5	14	1	existential risk	2013
6	13	0.913	covid-19	2018
7	10	0.743	robotics	2017

Cluster ID	References	Year	Strength	Begin	End
5	Bostrom N, MIND MACH	2012	1.98	**2015**	2015
1	LeCun Y, 2015, NATURE	2015	3.56	**2018**	2021
0	Barocas S, 2016, CALIF LAW REV	2016	3.40	**2018**	2018
1	Topol EJ, 2019, NAT MED	2019	4.15	**2020**	2021
7	Taddeo M, 2018, SCIENCE	2018	2.68	**2020**	2021
1	Rajkomar A, 2018, NPJ DIGIT MED	2018	2.14	**2020**	2021
7	Scherer, 2016, HARV JL TECH	2016	2.04	**2020**	2021

Fig. 5. Top 7 References with the Strongest Citation Bursts. Photo credit: Original

But these papers with strong citation bursts are not representative of the cluster they belong to. The following section will particularly focus on the typical largest clusters, where key literatures in each cluster will be analyzed to summarize its intellectual base.

5 Major Clusters Analysis

Cluster analysis helps us to understand the major specialties associated with the co-cited network. In this section, we will primarily focus on the largest seven clusters to summarize their intellectual base, which is a collection of scholarly achievements that have been collectively cited.

5.1 Cluster #0 Technical Risk

Through screening and co-citation cluster analysis, there are 22 representative references in Cluster #0. 5 high-impact references of them are listed in Table 6, across a 5-year period from 2014 to 2018. The cluster has a silhouette value of 0.962, which is a very high level of homogeneity, indicating that the highly cited articles within the cluster reveal a fairly high degree of topic consistency. Labeled as “algorithms”, these papers

Table 6. High-impact members of Cluster #0.

Rank	Centra.	Author	Year	Source
1	0.11	Barocas S	2016	CALIF LAW REV
2	0.05	Burrell J	2016	BIG DATA SOC
3	0.18	Caliskan A	2017	SCIENCE
4	0	Tutt A	2017	ADMIN LAW REV
5	0	Diakopoulos N	2016	COMMUN ACM

focus on the technical risks behind AI, which can be divided into two perspectives: algorithm discrimination (1, 3, 4) and algorithm opacity (2, 5).

From the algorithmic discrimination perspective, the core of AI is algorithm, and algorithm discrimination is the primary risk in the application of AI technology. The most influential literature in this group divides discrimination into intentional discrimination and unintentional discrimination. Although algorithms avoid intentional discrimination in decision-making, they may produce unconscious discrimination that depends on data. The latter is difficult to define and regulate legally. Some scholars have verified the semantic biases formed by machine learning through specific text training models, and proposed the possibility of technically eliminating such cultural biases. As AI becomes increasingly integrated into our daily lives, with it comes the potential risks of increasing complexity. In this regard, some scholars have established and created a new professional consumer protection agency to supervise algorithms.

From another point of view, the opacity of the algorithm is a serious problem in Law and Sociology. It is too tough to improve laws and social supervision on the "black box". Some scholars have divided opacity into three categories: confidentiality, technicality, and effectiveness. Machine learning algorithms are the third black box based on efficiency. The decisions generated by algorithms through black boxes appear in various scenarios of our lives, such as detection systems, news editing, dynamic pricing, etc., so we must be aware of the necessity of algorithm regulation. A paper in 2016 emphasized the establishment of a classification management system for algorithms based on transparency to clarify the attribution of responsibilities in algorithmic decision-making.

Degree centrality (Degree) and betweenness centrality (Centrality) are also listed in Table 6. Degree centrality measures the degree of connection between nodes and other nodes while betweenness centrality indicates the importance of a node's location in the network, that is, the importance of nodes connecting different groups of nodes. We focus more on papers with high betweenness centrality, which are often important hubs connecting different fields. It is generally believed that the literature whose betweenness centrality is not less than 0.1 has a good bridge role4.

As we see, there are two papers with high betweenness centrality, which play an important role in connecting different co-citation clusters. The paper in 2015 pointing out the existence and serious consequences of unintentional discrimination by algorithms, which also had a citation burst in 2018. Another is a study that simulates the process of semantic bias formation in machine learning.

5.2 Cluster #1 Breakthrough in Medicine

As the second largest cluster, Cluster #1 was active from 2015 to 2019. Table 7 lists the 8 most influential papers in this cluster. The cluster is labeled as "deep learning" with a very high level silhouette value (0.977). The highly cited references are all medical applications of deep learning.

On the one hand, deep learning has greatly promoted medical research such as clinical diagnosis and drug discovery, realizing targets such as detection of skin lesions through deep convolutional neural network algorithms and identification of diabetes through retinal film. The clinical prediction results of the deep learning model achieved high accuracy in in-hospital mortality, 30-day unplanned readmission, prolonged hospital stay, and discharge diagnosis, and the results even outperformed traditional, clinically used prediction models.

On the other hand, although deep learning has significantly improved the diagnostic efficiency of clinicians, health systems, and patients, there are still many limitations in practical applications, including issues of privacy, bias, security, and lack of transparency. AI algorithms are likely to cause major harm to patients and cause major medical malpractice in clinical practice, and their opacity and interpretability have always been controversial. Because the data lacks information about minority groups, optimized algorithms may exacerbate existing inequalities, such as the U.S. health care system paying more for whites. In addition, AI in medical applications also has the privacy problem of personal data leakage and the security problem of hackers harming patients.

There are also three references (1, 2, 8) with citation bursts from 2018 to 2020 in this cluster. In 2018, scholars began to pay attention to the promotion of deep learning in the medical field 6. Clinicians, health systems, and patients all benefit significantly from the application of AI, especially deep learning. The prediction of deep learning models based on electronic health records is even more accurate than traditional models. These studies have all received attention in the past two years.

Table 7. High-impact members of Cluster #1.

Rank	Centra.	Author	Year	Source
1	0.03	Topol EJ	2019	NAT MED
2	0.03	LeCun Y	2015	NATURE
3	0.07	Esteva A	2017	NATURE
4	0.07	Gulshan V	2016	JAMA-J AM MED ASSOC
5	0.01	Obermeyer Z	2019	SCIENCE
6	0.03	Chen TQ	2016	PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE
7	0	Schmidhuber J	2015	NEURAL NETWORKS
8	0	Rajkomar A	2018	NPJ DIGIT MED

5.3 Cluster #2 Social Risk

Cluster #2, labeled as "public perception", is the third largest cluster with the active period from 2017 to 2021. Table 8 lists the 7 most influential references. The silhouette value of this cluster is 0.923, which is a relatively high level of homogeneity. Based on the perspective of public perception, these references propose ethical guidelines for AI applications (2, 5, 6), and emphasize the social risks in AI applications (1, 3, 4, 7).

AI ethics guidelines is the bottom line for AI technology applications. AI brings new opportunities along with new challenges. Examples include the risk of misuse or abuse, degeneration of human skills, of being out human control, and the risk of affecting human autonomy. Since the establishment of the AI4People Scientific Committee in 2018, it has been aimed to promote the establishment of a friendly AI society, emphasizing how to use ethical constraints to reduce the potential risks of AI. How to explain ethical artificial intelligence? Analyzing and comparing 84 studies on the ethics of AI around the world, some scholars found that the ethical AI standards have converged in five basic issues: transparency, fairness and justice, non-maliciousness, responsibility, and privacy. Advocating ethical AI means we need to explore explainable AI techniques (XAI). There is still a long way to achieve this.

Even if AI technology meets ethical standards, we still need to pay attention to its potential societal risks. AI technology has penetrated into all aspects of life, and people are increasingly worried that humans will be replaced by intelligent machines. We must pay attention to the division of labor between humans and machines. AI decision-making systems should be designed to augment human capabilities, not replace them. At this stage, AI can indeed have a significant impact on the labor market. One literature has evaluated the probability of 702 occupations in the United States being replaced by computers in detail, and theoretically confirmed the concern of "technology-induced unemployment". In specific practical cases, AI has also proved its flexibility and high efficiency, such as taking over many human jobs in universities, governments, enterprises and other scenarios. In addition to the impact on individuals in the labor market, the digital revolution brought about by AI will intensify competition among companies, and our society is at risk of widespread unemployment and a widening wealth gap.

Table 8. High-impact members of Cluster #2.

Rank	Centra.	Author	Year	Source
1	0.05	Frey CB	2017	TECHNOL FORECAST SOC
2	0.04	Floridi L	2018	MIND MACH
3	0.04	Makridakis S	2017	FUTURES
4	0.02	Jarrahi MH	2018	BUS HORIZONS
5	0.05	Jobin A	2019	NAT MACH INTELL
6	0	Arrieta AB	2020	INFORM FUSION
7	0	Kaplan A	2019	BUS HORIZONS

5.4 Cluster #3 Ethical Risk

Cluster #3 is the fourth notable cluster, being active from 2016 to 2021, the seven most influential papers are listed in Table 9. The silhouette value of this cluster, 0.645, is the lowest among the main clusters, most likely because of too many research subjects in this cluster, but it still reaches a reasonable level (0.5). The cluster label for this group is "envelopment", which is an encapsulation concept in machine learning. This cluster also includes medical "black box" issues. Table 10 lists the 7 high-impact papers, including the frontiers of medical applications of AI (2, 3), the "black box" problem of algorithms (1, 7), and the ethical risks of clinical applications (4, 5, 6).

The application of AI in Medicine is mainly divided into two directions: virtual and physical. The former is a deep learning algorithm that is continuously optimized based on experience, and is mainly used in electronic medical records. The latter mainly includes medical equipment, nursing robots and other physical entities directly involved in medical work. For example, medical image diagnostic systems have been able to take on tasks that were previously only done by human experts. This trend will also continue to expand to other areas of medicine, such as clinical practice, translational research, and basic biomedical research.

With the continuous breakthrough of AI technology, the range of traditional medical problems that can be solved is also larger and larger, but we sacrifice the interpretability of the process for accurate diagnosis results. Predictive systems based on deep learning cannot tell us about the complex associations and operations inside. Given the technical difficulty of explaining "black-box" algorithms, some scholars have emphasized that we should adopt interpretable models from the outset, especially in high-risk fields such as healthcare and justice.

In addition to the problem of algorithm "black box", there are also many risks and hidden dangers in clinical application, including data management, data standardization, patient privacy and security, patient acceptance, etc. If AI technology is to be deeply applied in medicine, the privacy and security of patients is a great potential problem, which may lead to legal risks and ethical charges. In addition, the patient's own attitude is also an important factor to consider. In the case of consistent medical cost and accuracy, patients prefer to choose a doctor over an AI machine, because the machine cannot take

Table 9. High-impact members of Cluster #3.

Rank	Centra.	Author	Year	Source
1	0.01	London AJ	2019	HASTINGS CENT REP
2	0.08	Yu KH	2018	NAT BIOMED ENG
3	0.07	Hamet P	2017	METABOLISM
4	0.07	He JX	2019	NAT MED
5	0.02	Price WN	2019	NAT MED
6	0.02	Longoni C	2019	J CONSUM RES
7	0	Rudin C	2019	NAT MACH INTELL

Table 10. High-impact members of Cluster #5.

Rank	Centra.	Author	Year	Source
1	0	Bostrom N	2012	MIND MACH
2	0.02	Muller VC	2016	SYNTH LIBR
3	0	Russell S	2015	AI MAG
4	0	Zlotowski J	2017	INT J HUM-COMPUT ST

into account the uniqueness of the case, which arouses the anxiety and resistance of patients. When the machine just assists the doctor to complete the diagnosis, the patient does not reject the participation of the machine.

5.5 Cluster #5 Existential Risk

Cluster #5 is the 5th-ranked cluster, but it is the earliest cluster with a period of high activity from 2010 to 2017. Table 10 lists the four most influential references. The silhouette value of this cluster is 1, indicating that the topics of highly cited articles are very consistent. The common keyword of these references is "existential risk", that is, the risk that AI may cause the end of humanity. These references discuss the threat that super AI may pose to humans from four perspectives: long-term planning (3), philosophical theory (1), expert opinion (2), and intuitive perception (4).

The conference summary report of "The Future of Artificial Intelligence: Opportunities and Challenges" (2015) pointed out that the successful practice of AI technology has brought tangible benefits to human beings, but at the same time, there are many risks and pitfalls that need to be avoided. To ensure the robust research and valuable practice, the effectiveness, legitimacy, safety, and controllability of AI must be emphasized in long-term research. We cannot take it for granted that super AI will definitely share the wisdom and values of human beings, nor can it be assumed that when it achieves its goals, it will definitely take not infringing on the interests of human beings as the bottom line. Relying on instrumental rationality to ensure human safety is unreliable 5.

A survey of expert group on AI research pointed out that super AI may appear in a few decades, threatening the existence of human beings, and it is important to establish a regulatory system from now on. A social experiment on robot attitudes also showed that autonomous robots can pose real-world threats to humans (including job, resource, and security threats) and identity threats (threat to human uniqueness), which reinforces negative attitudes toward robots.

5.6 Cluster #6 Motivation for Adoption

Cluster #6 is the sixth-ranked cluster with the active period from 2016 to 2021. There are no particularly prominent high-cited or emerging papers. Table 11 lists 10 of these papers with high impact. The silhouette value of this cluster is 0.913, which is a relatively high level of homogeneity. Labeled as "Covid-19 (new coronavirus)", most of the papers

Table 11. High-impact members of Cluster #6.

Rank	Centra.	Author	Year	Source
1	0	Qiu HL	2020	J HOSP MARKET MANAG
2	0.05	Hengstler M	2016	TECHNOL FORECAST SOC
3	0	Baryannis G	2019	INT J PROD RES
4	0.08	Sun TQ	2019	GOV INFORM Q
5	0.04	Wirtz J	2018	J SERV MANAGE
6	0.02	Cubric M	2020	TECHNOL SOC
7	0.01	Lin HX	2020	J HOSP MARKET MANAG
8	0.01	Lu L	2019	INT J HOSP MANAG
9	0	Tung VWS	2018	INT J CONTEMP HOSP M
10	0	Ivanov D	2020	TRANSPORT RES E-LOG

citing references in this cluster are related to the epidemic. But in fact, as the research basis of the citing papers, most of these literatures focus on the pre-pandemic, and mainly study three aspects related to the drivers of AI adoption.

The first is considering the pros and cons of adopting AI technologies, including the motivations and barriers to AI adoption in business management (6), the pros and cons of public sector stakeholders (4), and how to foster trust in AI within the enterprise (2).

In the past ten years, the application of AI in business and management has increased significantly, and the motivation for adopting AI technology is mainly at the economic level, such as reducing costs and improving efficiency. But there are also technical and social barriers such as lack of data, high cost, and lack of trust. Within the same sector, different stakeholders may have opposing attitudes on whether to adopt AI. Based on a survey of the adoption of AI systems (IBM Watson) in China's public healthcare system, three types of stakeholders, including government policymakers, hospital directors, and IT company managers, have different perceptions in several dimensions of AI adoption, and most of the differences they recognize focus on non-technical issues such as policy and data. For enterprises, commercialization projects that adopt AI may be selected after comprehensive consideration, but the content of consideration should not only include economic, technical and social factors, but also the trust and acceptance of AI applications within the enterprise.

The second is to explore AI application models in supply chain risk management (SCRM) (3, 10). SCRM is designed to identify, assess, mitigate and monitor unexpected events that could disrupt any part of the supply chain, and its data multidimensionality and decision-making adaptability are a good fit for AI. However, researchers in the field of SCRM currently seldom apply technologies such as machine learning, and there are still many gaps in the fusion research of SCRM and AI, including decision automation, dynamic supply chain forecasting, etc. Epidemic outbreaks are special cases of supply chain risk (SC). Combined with the current severe epidemic situation, if this coronavirus

outbreak is defined as a unique SC disruption risk, the impact of the outbreak on SC performance can be predicted through simulation.

Finally, there is the consideration of whether to adopt service robots. Not only the potential of service robots (5), but also the survey on the acceptance of hotel service robots (1, 9), and the willingness of users to use robots (7, 8).

The combination of robotics and technologies such as big data and biometrics will have the potential to drastically change the operational structure of the service industry, and the popularization of service robots may achieve huge economies of scale. Many scholars have carried out experiments and researches on service robots in hotel scenarios. Studies have shown that humanized robots can promote the friendly relationship between customers and robots, and have a positive impact on customers' living experience. Among them, social influence, hedonic motivation, anthropomorphism, performance expectations, and subjective emotions will all affect users' willingness to use robots.

5.7 Cluster #7 AI Challenges in Governance

Cluster #7 is the smallest of the outstanding clusters, and the cluster was active from 2016 to 2019. Table 12 lists the 8 most influential papers. The silhouette value of this cluster is 0.743, a relatively low level among all clusters, but higher than the criterion of significance (0.7). The clustering keyword is "robotics", which is an important carrier of AI applications. These references focus on the governance of AI applications, and can be divided into two perspectives: to explain AI (1, 6, 7, 8) or to govern its application (2, 3, 4, 5).

The first is to explain AI. We also highlighted in Cluster #1 that algorithmic risk can have serious consequences for individuals, groups, and society. Often we expect better and better algorithms that can rule out obvious cognitive deficits and even help us discover and explain moral hazard (e.g. discrimination). But many high-accuracy algorithms are black boxes, and explaining these algorithms often requires sacrificing accuracy. But without trying to explain, it can also create problems of discrimination and trust. Explainable Artificial Intelligence (XAI) proposes a move to more transparent AI,

Table 12. High-impact members of Cluster #7.

Rank	Centra.	Author	Year	Source
1	0.33	Mittelstadt BD	2016	BIG DATA SOC
2	0.02	Scherer	2016	HARV JL TECH
3	0.02	Wirtz BW	2019	INT J PUBLIC ADMIN
4	0	Taddeo M	2018	SCIENCE
5	0.01	Cath C	2018	PHILOS T R SOC A
6	0.01	Adadi A	2018	IEEE ACCESS
7	0	Miller T	2019	ARTIF INTELL
8	0.02	Guidotti R	2019	ACM COMPUT SURV

and considerable effort will be required in the future to address XAI challenges and outstanding issues. From the perspective of social science, most XAI researchers at present simply use the subjective judgment of researchers on "good" explanations, and there may be some cognitive biases and social expectations in the explanation process, which requires the introduction of relevant knowledge in the fields of philosophy, psychology, cognitive science and human-computer interaction.

The second is the governance of AI, which mainly includes two aspects: ethical standards and government governance. AI technology is reshaping our daily life, but there are still practical challenges in delegation and responsibility, invisibility and influence, translational ethics and so on. By regulating the design, regulation, and use of AI, ethics can reduce risks while we realize the full potential of AI. For government departments, the advancement of AI technology has opened up ten usage scenarios such as knowledge management, office automation, and predictive analysis, but at the same time, we must be alert to possible public risks. For example, most AI technologies are emerging in a regulatory vacuum, it is difficult to identify legal liability for accidents 11. AI is increasingly permeating every aspect of our society, and its proliferation in high-risk areas makes governance a priority. How to manage AI chaos and produce responsible, fair and transparent AI still has many ethical, legal and technical challenges.

In this cluster, A literature (1) that elaborates on the ethical issues arised from algorithmic decision-making has high betweenness centrality. In addition, two papers (2, 4) in this cluster have citation bursts. The burst timeline shows that since 2018, scholars have begun to pay attention to the ethical risks of AI, and have begun to emphasize the importance and urgency of government governance of AI.

6 Citing Papers

The intellectual base is a collection of references in the same research field, while the research front is the latest research inspired by the intellectual base. Given the lag of co-citations, we need to focus on the latest citing papers in these fields in addition to studying co-citations as a knowledge base. Citing literature is further developed on the basis of cited literature, and can be regarded as the frontier of research.

We manually screened citing papers according to three criteria: publication time, citations, and literature source, and obtained 26 papers published in 2021 with at least three citations and indexed by SSCI, which can be roughly divided into five research directions: richer application scenarios, more comprehensive interactive experience, broader social impact, more adequate regulatory experience, and new trends in the post-epidemic era.

Richer application scenarios. We divide the main application scenarios into medical applications and other applications. In the exploration of medical applications, AI technology, especially deep learning, has not only expanded from traditional applications to other scenarios such as drug discovery, personalized medicine, cancer grading, and neurological disease prediction, but also promoted scholars' exploration in pathology and clinical medical practice (Table 13). In clinical diagnosis, human experts can resolve uncertainties through rich practical experience, but AI tools cannot, which may lead to errors in diagnosis results. To this end, some scholars have proposed a method (FairLens)

Table 13. Richer application scenarios (medicine).

	Author	Tittle
19	Panigutti, Cecilia	FairLens: Auditing black-box clinical decision support systems
20	Harrison, James H. Jr Jr	Introduction to Artificial Intelligence and Machine Learning for Pathology
21	Prakash, Ashish Viswanath	Medical practitioner's adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study
22	Davahli, Mohammad Reza	Controlling Safety of Artificial Intelligence-Based Systems in Healthcare
23	Piccialli, Francesco	A survey on deep learning in medicine: Why, how and when?
24	Lebovitz, Sarah	Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what

Table 14. Richer application scenarios (other).

	Author	Tittle
10	Baneres, David	A predictive analytics infrastructure to support a trustworthy early warning system
13	Fridgeirsson, Thordur Vikingur	An authoritative study on the near future effect of artificial intelligence on project management knowledge areas
17	Eling, Martin	The impact of artificial intelligence along the insurance value chain and on the insurability of risks
18	Buckmann, Marcus	Comparing minds and machines: implications for financial stability
26	Kazim, Emre	Systematizing audit in algorithmic recruitment

to explain this difference to revise the clinical decision-making system. The study found that AI's high-precision can only be realized when the adoption rate of AI in clinical diagnosis reaches a certain level. How to improve the adoption rate and give full play to the advantages of AI clinical application is also an important topic. In addition, in order to address safety challenges in healthcare services and reduce accident risks, it is also necessary to establish a safety control system (SCS) that includes application guidelines for policy, training, incentives, etc.

Scholars also focused on the overall impact of AI on other industries and on traditional scenarios (Table 14). For example, AI may affect the business model and scope of insurable risk in the insurance industry and reduce systemic risk in the financial industry. In more specific application scenarios, AI can not only be applied within companies to improve the accuracy of project management (PMI), and effectively manage project

Table 15. More comprehensive interactive experience.

	Author	Tittle
1	Chi, Oscar Hengxuan	Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery
4	Larkin, Connor	Paging Dr. JARVIS! Will people accept advice from artificial intelligence for consequential risk management decisions?
6	Chuah, Stephanie Hui-Wen	Unveiling the complexity of consumers' intention to use service robots: An fsQCA approach
8	Ribeiro, Manuel Alector	Customer acceptance of autonomous vehicles in travel and tourism
15	Yeh, Shin-Cheng	Public perception of artificial intelligence and its connections to the sustainable development goals

Table 16. Broader social impact.

	Author	Tittle
9	Nazareno, Luisa	The impact of automation and artificial intelligence on worker well-being
16	Schiff, Daniel S.	Assessing public value failure in government adoption of artificial intelligence
25	Foster-McGregor, Neil	Job automation risk, economic structure and trade: a European perspective

costs, schedules and risks, but also in schools to improve teachers' teaching efficiency through mastering and alerting students' learning progress (David et al., 2021). But the impact of AI may not all be positive. For example, the application of AI in recruitment may present the risk of discrimination, and fairness and impartiality must be ensured through technical review.

A more comprehensive interactive experience. Academics are increasingly focusing on public attitudes and willingness towards AI (Table 15). A survey in Taiwan shows that the general public's attitude towards AI is rational and optimistic. Some scholars have found that trust tendency, functional design and the background of the task will affect consumers' trust in social robots, so they established the Social Service Robot Interaction Trust Scale (SSRIT) in turn. We also focus on what factors govern consumers' willingness to use AI devices such as service robots and autonomous vehicles. At this stage, the public has mostly taken a wait-and-see attitude towards AI-assessed medical or financial issues.

Broader social impact. For countries, automation risk varies widely across industries in the same country, but very little across industries, and the impact of trade is very limited

Table 17. More adequate regulatory experience.

	Author	Tittle
2	White, James M.	Ignorance and the regulation of artificial intelligence
11	Mantelero, Alessandro	An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems
12	Ulnicane, Inga	Good governance as a response to discontents? Deja vu, or lessons for AI from other emerging technologies
14	Jia, Kai	Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines

(Table 16). For governments, we find that the negative impact of government automated decision-making systems (ADS) on public values such as fairness, transparency, and responsiveness may trigger citizen dissatisfaction. But unexpectedly, for individuals, unlike the risks of "machine substitution" that we emphasize, scholars have found that employees in industries with high automation risk seem to have less work pressure, suggesting that complementary technologies may have positive effects on employees.

More adequate regulatory experience. In recent years, AI has developed rapidly in various fields, and hundreds of norms and ethical guidelines have emerged (Table 17). Through the classic dichotomy of "probability and severity" in the field of risk management, some scholars have established a four-dimensional theoretical framework to encode and compare existing AI guidelines, and found that the existing guidelines focus more on individual risk and cross-generational risk. The experience of other emerging technologies shows that for a new technology to develop in a beneficial way, it is inseparable from inclusive governance, active government participation, common social interest, international scientific and technological cooperation, and responsible innovation. In the specific content supervision, some scholars have established an assessment model of the impact of AI on human rights based on international laws, emphasizing that under different cultural, ethical and legal backgrounds, only human rights can provide a standard paradigm for the supervision of AI. Some scholars have also emphasized the importance of social science. The social nature of AI risks can help to promote the diversification and rationalization of supervision and greatly alleviate the public resistance. Implementing and strengthening the regulation of AI applications cannot rely solely on technical knowledge, but must also combine social knowledge and social non-knowledge (i.e., ignorant knowledge) in the field of social sciences. Knowledge about ignorance helps explain unobservable barriers to AI applications such as complexity, inhumanity, and intractability 1.

In the post-epidemic era, AI has also had a certain impact on human life, including supply chain management in industries and the acceptance of AI to consumers (Table 18). In India, the application of AI in supply chain management (SCM) is still in its infancy, and process factors, information sharing, and supply chain integration (SCI) all influence the willingness of SMEs to adopt AI. Supply chain disruptions and uncertainty caused by the spread of coronavirus (COVID-19) have severely impacted the stability

Table 18. New trends in the post-epidemic era.

	Author	Tittle
3	Nayal, Kirti	Are artificial intelligence and machine learning suitable to tackle the COVID-19 impacts? An agriculture supply chain perspective
5	Kim, Seongseop (Sam)	Preference for robot service or human service in hotels? Impacts of the COVID-19 pandemic
7	Nayal, Kirti	Exploring the role of artificial intelligence in managing agricultural supply chain risk to counter the impacts of the COVID-19 pandemic

of agricultural supply chains (ASCs), and artificial intelligence and machine learning (AI-ML) has had a significant impact on supply chain risk mitigation (SCRM) positive effects.

In addition to shining in production, AI technology is also increasingly accepted in the tourism industry. Unlike most studies conducted prior to the COVID-19 pandemic, a survey revealed that consumers now prefer hotels with service robots, possibly due to concerns about safety and social distancing. This case highlights the importance of perceived threat in risk assessment, and perceived threat to the new coronavirus has greatly influenced customer preferences for hotels with robotic staff.

7 Conclusions

In recent years, people have paid more and more attention to what risks the widespread application of artificial intelligence technology will bring, and how we should face and manage evolving AI technologies. This paper uses the CiteSapce to analyze international research on AI risks, including their subject co-occurrence, keyword co-occurrence, co-citation references and citing papers. We reveal the intellectual base of these research, and analyze the future trends based on the citing papers (Fig. 6).

From the international research hotspots of AI risk research, scholars mainly analyze the technical risks, ethical risks, social risks and existential risks brought by AI technology from the fields of Computer Science, Economics, Medicine and Philosophy. First of all, we not only pay attention to the risks of discrimination and opacity brought about by the algorithm technology itself, and the "black box" problems in the industry represented by medicine, but also pay attention to the impact that AI may have on the industry, on micro-employment, and on human beings. Secondly, in addition to the direct risks brought by AI, scholars are also concerned about the social acceptance of AI and major breakthroughs in AI applications (represented by medicine). Finally, existing research emphasizes that a comprehensive understanding of the potential risks of artificial intelligence, and the regulation and implementation of important measures to manage artificial intelligence technology are of great significance to promoting friendly technological progress, maintaining social stability, and safeguarding human development.

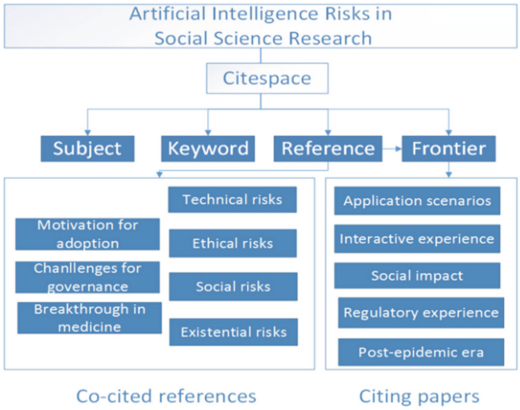

Fig. 6. The framework of the review. Photo credit: Original

From the latest research on artificial intelligence application risk, the literature mainly focuses on five research directions: "richer application scenarios", "more comprehensive interactive experience", "broader social impact", "more adequate regulatory experience", and "new trends in the post-epidemic era". It is foreseeable that, with the continuous progress of AI technology, the application of AI in life will become more extensive and in-depth, and its risk issues will become more and more complicated. The changing environment requires more adequate regulatory experience and more comprehensive safeguards.

Acknowledgements. This work was funded by the "Public Management" Construction Project of Characteristic Key Discipline from Guangdong Province, China in 2016 (F2017STSZD01).

References

1. White JM, Lidskog R (2022) Ignorance and the regulation of artificial intelligence, J. Risk Research, 25(4):488–500. https://doi.org/10.1080/13669877.2021.1957985
2. Chaomei C, Zhigang H et al (2012) Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Expert Opin Biol Ther 12(5):593–608
3. Jie L, Chaomei C (2016) CiteSpace: text mining and visualization in scientific literature. Capital University of Economics & Business Publishing House, Beijing
4. Nick B (2012) The superintelligent will: motivation and instrumental rationality in advanced artificial agents. Mind Mach 22(2):71–85
5. LeCun Y, Bengio Y et al (2015) Deep learning. Nature 521(7553):436–444
6. Solon B, Selbst AD (2016) Big Data's disparate impact. California Law Rev 104(3):671–732
7. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine 25(1):44–56

8. Alvin R, Eyal O et al (2018) Scalable and accurate deep learning for electronic health records. NPJ Digit Med 1(1):18
9. Mariarosaria T, Luciano F (2018) How AI can be a force for good. Science 361(6404):751–752
10. Scherer MU (2016) Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harv J Law Technol 29