Tracing the evolution of AI: conceptualization of artificial intelligence in mass media discourse

Yujia Zhai
Department of Information Resource Management, Tianjin Normal University, Tianjin, China and School of Information Management, Wuhan University, Wuhan, China
Jiaqi Yan and Hezhao Zhang
Department of Information Resource Management, Tianjin Normal University, Tianjin, China, and
Wei Lu
School of Information Management, Wuhan University, Wuhan, China

Abstract

Purpose – This study/paper aims to understand the public perceptions of AI through mass media discourse. In the past few years, significant progress has been made in the field of artificial intelligence (AI). The benefits of AI are obvious, but there is still huge uncertainty and controversy over the public perception of AI. How does the mass media conceptualize AI?

Design/methodology/approach – In this paper, the authors analyze the evolution of AI covered by five major news media outlets in the past 30 years from 7 dimensions: scientific subject, keyword, country, institution, people, topic and opinion polarity.

Findings – First of all, different subjects are competing for and dividing up the right to speak of AI, leading to the gradual fragmentation of the concept of AI. Second, reporting on AI often includes reference to commercial institutions and scientists, showing a successful integration of science and business. Moreover, the result of topic modeling shows that news media mainly defines AI from three perspectives: an imagination, a commercial product and a field of scientific research. Finally, negative reports have focused on various issues relating to AI ethics.

Originality/value – The results can help bridge various conversations surrounding AI and promote richer discussions, increase the participation of scientists, businesses, governments and the public and provide more perspectives on the functions, prospects and pitfalls of AI.

Keywords New technology, Artificial intelligence, Mass media, News coverage, Public perception, Conceptualization

Paper type Research paper

Introduction

Artificial intelligence (AI) is a term that is both widely used and loosely defined (Kanal and Lemmer, 2014). Even in academia, experts in different fields have different opinions on what AI is, what AI can do and how to design, standardize and integrate it into social implications (Müller and Bostrom, 2016). Over the past 30 years, significant progress has been made in this area because of the increase in computing processing power, the development of algorithms and, perhaps most importantly, the availability of large data sets that can help train AI systems. Today, most AI systems involve machine learning or deep learning. These algorithms can both recognize patterns in the data set without human guidance and evolve over time through self-learning.

Having attracted a lot of investment from governments and enterprises, technology companies integrate AI into a wide range of products, from driverless cars and weapons to TVs and provide daily living services, such as health care, news production and social media management. In the past few years, governments have also adopted AI as an important policy initiative. Obviously, these fast-evolving technologies will completely change the life of the entire world. But how AI does this and what role the news media may play in shaping this transformation, all remain open questions.

Public perceptions of AI come mainly from mass media. News reports can explain to the public what AI is, what AI can do and what AI means to them. However, research has largely ignored the role of media discourse in the construction of AI concepts. At the same time, media reports often oscillate between two sensational poles: a utopian dream of a useless future and eternal life and a nightmare against the uprising of robots and the end of the

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/2398-6247.htm**

Information Discovery and Delivery
48/3 (2020) 137–149
© Emerald Publishing Limited [ISSN 2398-6247]
[DOI 10.1108/IDD-01-2020-0007]

This work was supported by National Social Science Foundation of China (18CTQ027).

Received 12 January 2020
Revised 3 March 2020
8 March 2020
Accepted 9 March 2020

world (Cave and ÓhÉigeartaigh, 2019). Although the advantages of AI are obvious, there is still huge uncertainty and controversy regarding the positioning of AI, not only related to the possible impact of AI in the entire society but also the way to regulate and develop AI systems. News coverage raises an essential, dynamic and vital public discussion for addressing this emerging public issue (Jensen *et al.*, 2017). Therefore, we need to better understand the news discourse of AI: public narratives, expectations, hopes and fears.

In this paper, we analyze the evolution of AI covered by five major news media outlets in the past 30 years from seven dimensions: scientific subject, keyword, country, institution, people, topic and opinion polarity. Our results can help bridge the various conversations surrounding AI and promote richer discussions, increase the participation of scientists, businesses, governments and the public and provide more perspectives on the functions, prospects and pitfalls of AI.

Related works

History of artificial intelligence

Human beings have never stopped imagining and exploring the realization of AI. However, the development of AI has not been as smooth as people hoped; on the contrary, it has been advancing in twists and turns, and once faced collapse. The history of AI can be divided into five stages from the timeline of scientific research.

From the 1950s to the early 1960s, it was the embryonic stage of AI. In 1950, Alan Turing first proposed a benchmark question (later called the Turing test) for judging machine consciousness – if a machine can imitate human conscious behavior, would it not be conscious? Turing's question shaped the philosophy of AI (French, 2000). In the summer of 1956, McCarthy, Minsky and other scientists held a meeting at Dartmouth College in the USA to discuss "how to use machines to simulate human intelligence". They first proposed the concept of "artificial intelligence (AI)", marking the birth of the subject of AI. This meeting regulated and promoted the development of AI as a research subject for many years. Turing's work and the Dartmouth Symposium created a necessary framework for the field of AI research. In this stage, AI made some initial progress: the appearance of the perceptron and the birth of the first patent for the industrial robot "Unimate", all mark the first spring of AI (Rosenblatt, 1958).

During the second stage, from the 1960s to the early 1970s, the development of AI encountered some obstacles. The first "winter" of AI came (Buchanan, 2005). Due to some breakthroughs made in the field of AI in the early days, people began to hope that AI could achieve some higher-level tasks. Researchers also began to exaggerate to ensure that their asset chain would not break; however, the successive failures and the failure of the expected goal (for example, it is impossible to prove that the sum of the two continuous functions is still a continuous function by machine), which led to a continuous decrease in funding for AI, and a subsequent lack of development.

The first "winter" of AI lasted until the 1970s and continued to spread into the 1980s. However, although the development of AI was not as smooth as expected in the early 1970s to the mid-1980s, there were still some gratifying achievements. At this stage, the emergence of expert systems enabled AI to move to the practical application field, which automatically helped people to solve specific problems. During this period, expert systems were widely used in biology, medicine, ecological environment and other fields and achieved success (Larkin *et al.*, 1980; McDermott, 1980; Turban and Watkins, 1986).

From the mid-1980s to the mid-1990s, some defects of the expert system appeared with the expansion of its application scope. At the same time, as funds were concentrated in the hands of only a few researchers, AI stalled. It was not possible to fund other AI projects, and some scientists even wanted to give up their research work (Searls, 2007).

In the fifth stage, after the mid-1990s, with the arrival of the big data era, the emergence of various technologies and computer algorithms promoted the fast growth of AI, where it became more and more "humanized". The number of news reports about AI also ushered in the peak of its history. The entire industry began investing heavily in AI, with derivatives and start-ups emerging at an unprecedented rate.

Public perceptions and concerns about artificial intelligence

Public perceptions and concerns about AI are important because the success of any emergent technology depends in large part on public acceptance. It is necessary for us to understand the historical context and future development trend of a new technology which is growing and infiltrating into all aspects of human life.

People's understanding of AI may be vague and one-sided. Because AI contains complex mathematical algorithms and theories and involves all aspects of society, it is difficult for us to have a clear and comprehensive understanding of this concept that originated in the scientific field. But we are no strangers to the word AI because the media's news reports are full of recent developments in and applications of AI. The prototype of people's understanding of emerging technologies is usually shaped by the mass media, and differences in media reports will also affect the public's understanding (Cacciatore *et al.*, 2012; Shih *et al.*, 2008).

Fast and Horvitz (2017) used the reports of AI in *The New York Times* over a 30-year period to explore the public understanding and discussion of AI. They analyzed the levels of engagement, pessimism and optimism toward AI, the prevalence of specific hopes and concerns and AI-related topics. The study found that news discussions on AI increased dramatically after 2009, and that these discussions were more optimistic than pessimistic. However, in recent years there have been growing concerns about the loss of control over AI, ethical issues and the negative impact on work.

Optimism about AI expresses people's hope for the development of AI technology. "I trust the machine more

Information Discovery and Delivery
Volume 48 · Number 3 · 2020 · 137–149

than I trust the doctor," said Dominguez, a computer scientist at the University of Washington in Seattle (Ananthaswamy, 2011). However, when considering the utility value of news reports, negative news has a greater impact on public perception than positive news, so both the public and the news media pay more attention to negative news (Lee *et al.*, 2005). Negative news spreads faster among the crowd, which has a negative effect on the understanding of AI (Naveed *et al.*, 2011). Owing to the increasing frequency of news media reports on security issues caused by AI, people's worries about the existence of AI have also increased (Shakirov, 2016). According to Stewart Russell, an authoritative expert in robotics and other fields, AI is as dangerous as nuclear weapons (Bohannon, 2015). This kind of worry may come from many aspects, not only the security risks brought by AI, but also the related problems caused by the security aspects (privacy issues, e.g. sensitive data sources are leaked; safety issues, e.g. the harm humans can cause to machines and vice versa), which are also the focus of public concern.

The public speculates that in the near future, AI systems that are smarter than humans will be invented. For human well-being, it is essential that these powerful systems will be used for good. The role of AI ethics is to ensure that relevant laws and regulations of AI reduce their risks while realizing their potential. With the increasing application of AI in various industries, from logistics to medical diagnosis, understanding which tasks and decisions can be handed to AI, and how to attribute when problems arise, are important research questions for AI ethics. In a content analysis of 84 AI ethical documents, Jobin *et al.* (2019) summarized eleven overarching ethical values and principles: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability and solidarity. Most of these are related to our daily lives. At the same time, new ethical challenges will continue to emerge. Therefore, media reports on related issues will inevitably affect the public's perception of AI.

The media agenda setting has a strong impact on public understanding of new technology (Russell Neuman *et al.*, 2014). Media reports and public attention are an interactive process, where a topic with more related news often attracts more public attention. In the process of news reporting, the language structure used in a news report to highlight the description of AI can more widely clarify the concept of AI and its relevance to the field of communication (Lee *et al.*, 2005). Media outlets can make use of their own characteristics of communication, such as language combination, topic setting, emotional guidance, post tracking and other ways to report the content and news in the field of AI.

From this point of view, this article uses a variety of natural language-processing technologies to process AI news data and analyze the concept of AI in media reports. The results of our research will reveal what exactly is meant by AI in media discourse, as well as concerns and expectations about AI. This will guide academics and experts engaged in AI research to better transform AI technology services into social applications.

Methodology

Data set

We used the LexisNexis database to search for AI-related articles, mainly from *USA Today*, CNN, *The Washington Post*, *The New York Times* and *The Guardian*. The choice of media is mainly based on three factors. First, the number of readers: five media outlets are recognized as the mainstream news media in the world, with a wide audience and high attention. Second, these media outlets have a long history that can trace the long-term conceptual construction process of AI. Third, past studies often chose these media outlets to study the coverage of new technologies, thus facilitating our comparison of research results.

A total of 11,956 news articles were obtained through the keyword search using "AI" or "artificial intelligence" or "A.I.". We removed those with any tag in the following list: Web Log; Text; Summary; Special Report; Series; Schedule; Book Review; Quote; Question; Paid Death Notice; Series; Op-Ed; Obituary (Obits); List; Letter; Interview; Editorial; Correction; and Biography. In addition, we removed news where the title was empty or the content length was less than three sentences. After screening, we obtained 9,914 news items in total, from June 16, 1977 to December 31, 2018. After that, we extracted news sentences containing "AI" or "artificial intelligence" or "A.I." from the title and content of the news articles. We also kept the two sentences before and after. Finally, we had a total of 22,094 sentences. Then, we carefully read 5,578 sentences containing "AI" to ensure that the word "AI" in them does not mean anything other than artificial intelligence.

Therefore, we established two corpora. Corpus #1 consists of all the AI-related sentences and Corpus #2 includes the title and full text of the news Table I, Figure 1.

Method

As shown in Figure 2, we first used regular expressions to obtain the annual frequency of different AI subjects and keywords. Then, we used the Stanford Named Entity Recognizer (NER) to extract the name of the institution, person and country from Corpus #1. Next, we used latent Dirichlet allocation (LDA) to extract the topics and analyzed news sentiment by the Stanford CoreNLP toolkit (Manning *et al.*, 2014).

Scientific subjects

To understand which academic subjects are related to AI in media reports, we need to extract the frequency of the subject names from all the news reports and calculate their annual

Table I Number and time range of AI news articles for each medium

News media	No. of newspapers	Timeline
CNN	503	1992/5/5 to 2018/12/31
USA Today	655	1989/7/10 to 2018/12/11
The Guardian	2,637	1993/5/6 to 2018/12/31
The Washington Post	2,372	1977/6/16 to 2018/12/31
The New York Times	3,747	1980/9/5 to 2018/12/31

Figure 1 Distribution of the number of newspapers for different media

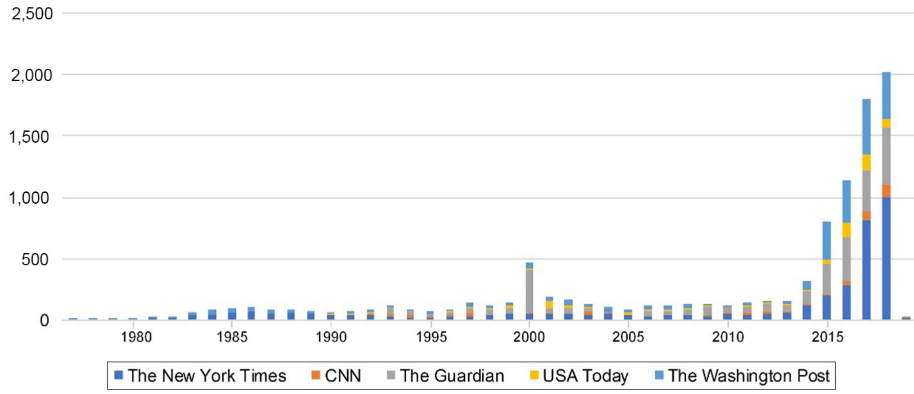
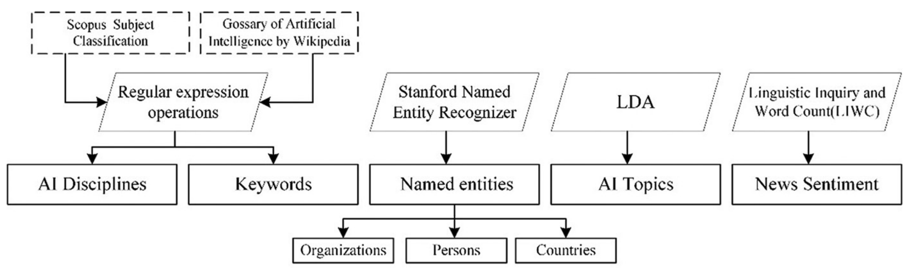


Figure 2 Methodological framework for analyzing the concept of AI in the news



Notes: The dotted box is the supplementary material we used, the diamond box is the method we used, and the box is the unit of conceptual analysis

distribution. Many studies choose to adopt subject categories to represent scientific subjects (Zhu and Yan, 2015; Zhai *et al.*, 2018). Scopus uses All Science Journals Classification for Science classification, which defines 27 subjects and 4 big supergroups: life sciences, social sciences, physical sciences and health sciences.

Here, we need to matching subject names with the sentences from the news reports and count their frequency. However, names such as "art" and "energy" are not necessarily expressed as subjects. Therefore, we manually re-coded the names of 27 subjects to reduce the ambiguity of their representation. The subject names are as shown in Table II. Then we converted all sentences and names to lowercase, removed punctuation and used regular expressions for text matching. It should be noted that the representation of a subject is diverse in news reports. Our list cannot contain all the names that appear.

Keywords and named entities co-occurring with artificial intelligence

Our next task is to discover the composition of AI in different periods of news media; that is to say, what specific concepts are often associated with AI and how they shift through time. We used the glossary of AI provided by Wikipedia[1] to extract the top 10 phrases for each 5-year period from the news. This term list is about AI, its sub-subjects and related fields. To match the keywords more accurately, we removed all punctuation and segmentation from the text and convert all words to lowercase.

We adopted the Stanford NER to extract AI-related entities. Stanford NER labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names (Finkel *et al.*, 2005). Here we focused on the distribution of three entities: institution, person and country. For the first two, the results of the automatic extraction are not accurate. As newspapers generally use last name or first name to indicate the corresponding person, we carefully checked and determined who the person was based on the original news and changed it to his/her full name. Dealing with the institution's name also followed this process and principles. Also, we classified the types of people and institutions manually.

Topic modeling with latent Dirichlet allocation

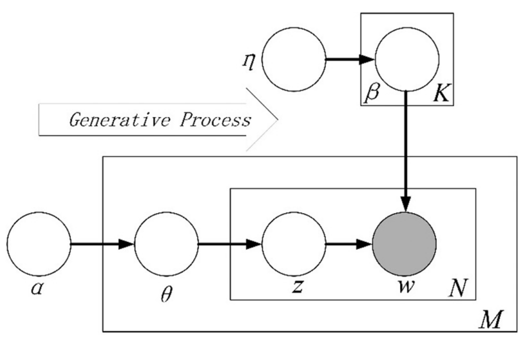
For understanding the content of news reporting AI, we used LDA for our topic modeling task. Proposed by Blei *et al.* in 2003, LDA is an unsupervised algorithm to extract meaningful topics from a large size of unlabeled documents. It treats each document as a vector of words and each of them can be represented as a probability distribution over some topics. As output, LDA produces a list of word-sets that represent different topics and topic distribution of each document in the corpus.

Information Discovery and Delivery
Volume 48 · Number 3 · 2020 · 137–149

Table II Re-coded scientific subject names based on ASJC

ID	Description
1	Agricultural science
2	Art science
3	Astronomy
4	Biochemistry
5	Biological science
6	Biology
7	Chemical engineering
8	Chemistry
9	Computer science
10	Decision science
11	Earth science
12	Engineering
13	Environmental science
14	Genetics
15	Health science
16	Humanities science
17	Immunology
18	Life science
19	Material science
20	Mathematics
21	Medical science
22	Microbiology
23	Multidisciplinary
24	Neuroscience
25	Physical science
26	Planetary science
27	Psychology

Figure 3 Graphic model presentation of LDA



With a corpus of M documents $\{w_1, w_2, \ldots w_m\}$ containing words from a list of N terms, LDA assumes that documents are produced from a set of K topics. In a document, each word w_i is associated with a hidden variable $zi \rightarrow \{1, \ldots, K\}$ indicating the topic from which w_i was generated. The probability of word w_i can be expressed as:

$$P(w_i) = \sum_{j=1}^{K} P(w_i|z_i = j)P(z_i = j)$$

where $P(w_i \mid z_i = j) = \beta_{ij}$ is a probability of word w_i in topic j and $P(z_i = j) = \theta_i =$ is a document-specific mixing weight indicating the proportion of topic j in the document.

The multinomial parameters β and θ are sampled, respectively, as latent random variables from a Dirichlet prior with parameters α and η. Each document is obtained using the following generative process (as shown in Figure 3):

- sample a K-vector θ of document specific mixing weights from the Dirichlet distribution $P(\theta \mid \alpha)$; and
- for each word, sample topic assignment j according to mixing weights $P(z) = \theta$ and draw a word according to $P(w \mid z) = j$.

Here, Gensim, a Python library, is used for the LDA topic modeling (Rehuřek and Sojka, 2010). We used the standard parameters provided by Gensim (alpha =“symmetric”, eta = None, decay = 0.5, eval_every = 10, iterations = 50, gamma_threshold = 0.001, update_every = 1).

Table III An example of topic popularity for four topics

Topic ID	Doc 1	Doc 2	Doc 3	Doc 4	*Popularity*
Topic 1	0.31	0.02	0.09	0.11	0.55
Topic 2	0.22	0.12	0.39	0.04	0.85
Topic 3	0.01	0.80	0.22	0.03	1.49
Topic 4	0.12	0.61	0.43	0.51	1.67

Table IV LDA topics with top keywords in AI news As the keywords of topics 9 and 21 are meaningless, we did not take them into further discussion

Topic ID	Top 5 keywords
Topic 0	Job-work-market-business-industry
Topic 1	Movie-story-film-character-novel
Topic 2	Robot-human-machine-robotic-researcher
Topic 3	Weapon-military-official-war-security
Topic 4	Student-course-online-university-mooc
Topic 5	Mail-system-program-software-network
Topic 6	Human-game-chess-program-player
Topic 7	News-social-content-site-government
Topic 8	Plane-flight-airline-pilot-passenger
~~Topic 9~~	Se-chuxing-uni-didi-game
Topic 10	Drone-aircraft-hawk-system-unmanned
Topic 11	App-device-voice-home-video
Topic 12	Game-player-video-team-play
Topic 13	Game-gpu-virus-hacker-system
Topic 14	Software-system-business-industry-chip
Topic 15	Datum-job-government-machine-system
Topic 16	Space-mars-human-mission-planet
Topic 17	Car-vehicle-driver-self-autonomous
Topic 18	Film-show-minute-director-child
Topic 19	Music-song-musical-sound-composer
Topic 20	Student-school-program-university-high
Topic 21	Uni-ne-se-building-sw
Topic 22	Market-percent-business-fund-stock
Topic 23	User-bot-app-service-tech
Topic 24	Growth-economist-economy-worker-job
Topic 25	Art-museum-free-artist-gallery
Topic 26	School-student-education-high-woman
Topic 27	Human-science-brain-book-life

The data set we used here is the combination of titles and full texts (Corpus #2). To obtain meaningful words for topic inference, we adopted several language-preprocessing techniques to prune the corpus:

- converted all words to lowercase and the plural to its singular form;
- used the Stanford CoreNLP Toolkit for part-of-speech analysis and named entity recognition, retaining only nouns, verbs and adjectives; and
- removed words with less than three letters and words from a stop word list.

Topic popularity

A popular topic is one that appears more often in the articles than others (Griffiths and Steyvers, 2004) (Table III). Here, we calculated topic popularity by summing up the topic proportion in all news. For example, as illustrated in Table IV, four articles are assigned to four topics. For each topic i, the popularity of topic $Top\_Pop(i)$ can be calculated through aggregating $\theta_{d,i}$.

The topic popularity for topic i in year t can be expressed as:

$$Top\_Pop(i\,|t) = \sum_{d|year(d)=t} \theta_{d,i}$$

where $year(d)$ denotes the year of article d.

Sentiment analysis

News media coverage has an important impact on the diffusion of technology. For example, actively promoting the superiority of technology in the news can better help technology spread, while reporting the negative impact of technology on daily life will make people refuse to adopt it Shariff *et al.* (2017). In this research, we use three categories, "sad," "anger," and "anxiety" from the Linguistic Inquiry and Word Count (LIWC) to understand the relationship between negative reports and various topics. LIWC is a text analysis program and is often used in news coverage and sentiment studies. It calculates the degree to which various categories of words are used in a text. By finding the distribution of three categories and topics in the news, we calculate Pearson's correlation coefficient between the degree of each negative category and topic proportion in every news report.

Results

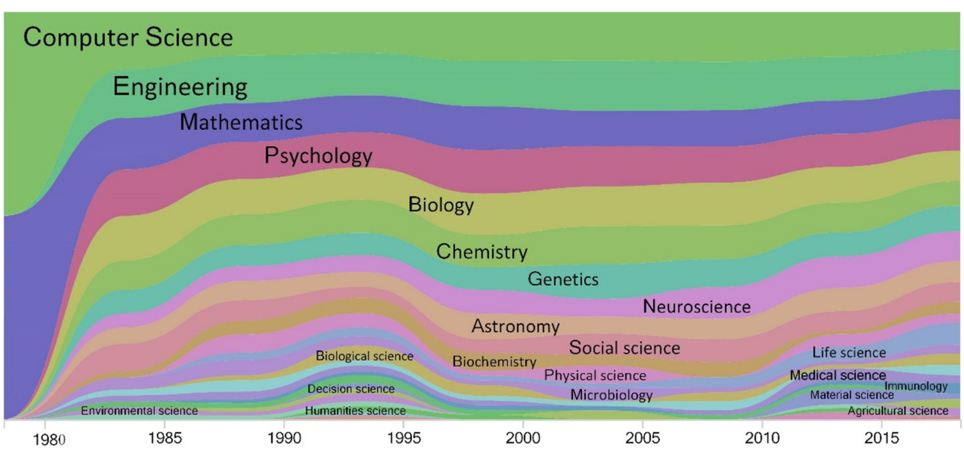
Subjects and keywords in artificial intelligence news

A large number of new AI technologies have been adopted in different areas of modern society, and these applications have been widely reported. Therefore, by analyzing the frequency of different scientific subjects and AI-related keywords, we can show the evolving impact of AI on scientific discovery. The first question that needs to be answered is which subject and keywords are more relevant to AI in the news Figure 4.

AI is not driven by a single subject. We found that computer science, engineering and mathematics are the main subjects of AI with the top three highest frequencies. So far, the number of subjects involving AI has reached 22. Nowadays, different subjects are competing for and dividing up the right to speak of AI, leading to the gradual fragmentation of the concept of AI.

The development of AI has gone through a tortuous process, which is not only manifested in related subjects but also reflected in the evolution of neighboring concepts. Figure 5 shows the top 10 keywords of AI that occupy the hot spots in each time period in the news media. First of all, robots, as a concrete entity from imagination to realization, have been a hotspot in news since 1980s. Speech recognition and autonomy, as the keywords related to robotics, have also attracted the attention of the mass

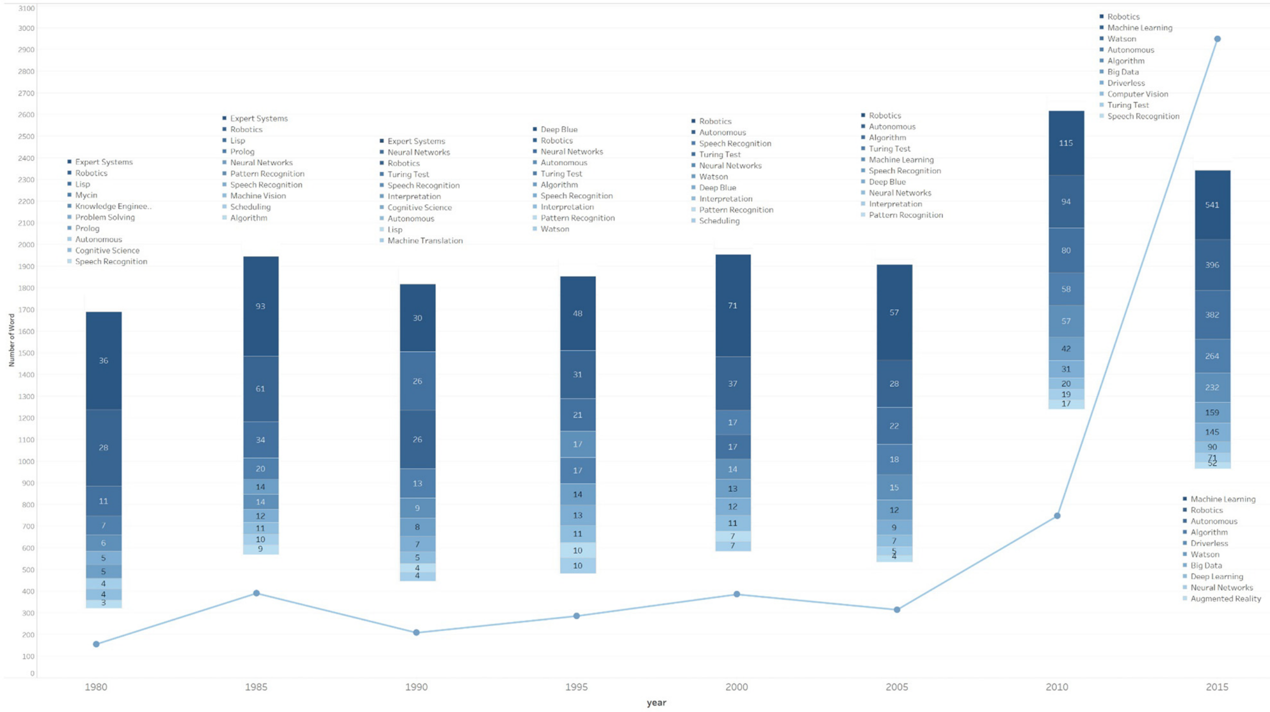
Figure 4 Time distribution of the frequency of AI-related subjects in the news



Notes: Each stream in the river map represents a subject, and the width of the stream represents the frequency w of the subject in the corresponding year. Considering the weight of computer science and engineering, which is much higher than the others, we calculate the logarithm of w (log(w) + 1) to reduce the difference to show the evolution and emergence of subjectsIpsum

Information Discovery and Delivery
Volume 48 · Number 3 · 2020 · 137–149

Figure 5 Top 10 keywords co-occurring with AI in news in a 5-year window period



Notes: The line represents the total frequency of the top 10 keywords

media in the past 40 years. Second, LISP and Prolog, as two major computer languages, attracted much attention in the early stage. In the follow-up process, new languages have been constantly produced, making these two languages gradually fade from sight in the media. Finally, driverless cars, big data and machine learning have gradually become hot news in recent years.

Starting from the late 1970s, AI news corresponds in time to the last three development stages of our review of AI history. In addition, according to the subjects appearing in the news over time, computer science was the leading one, and then subjects closely related to engineering such as biology and chemistry appeared, and finally life science and agricultural science appeared. Therefore, we divided the development of AI news into three stages.

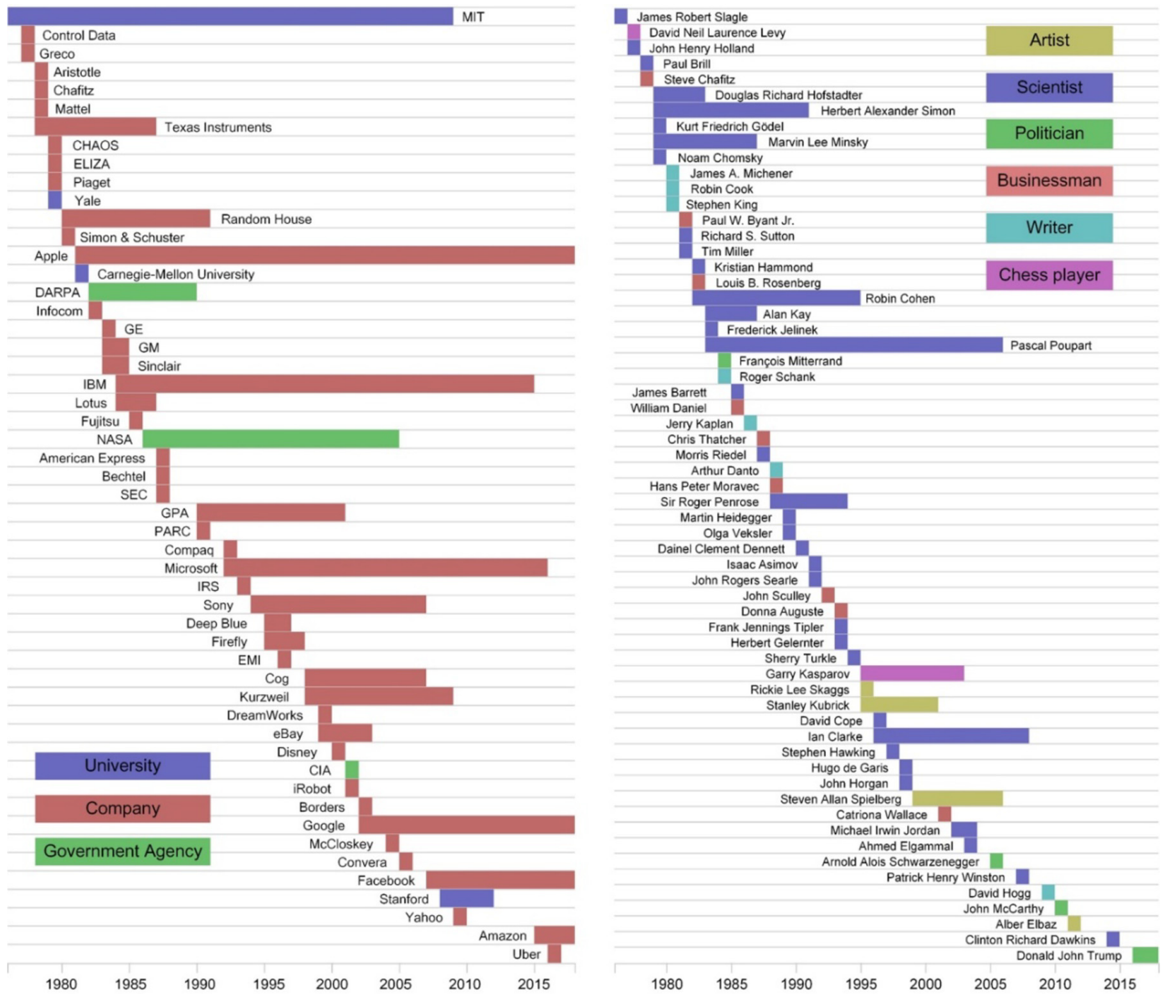
1970s was the first stage of the development of AI. Expert systems were the most important AI applications at this stage. AI standardize specific knowledge and integrates it into system, so it is possible to reach the professional level of human experts in certain specific fields (Duda and Shortliffe, 1983). MYCIN was one of the initial expert systems to perform with the level of expertise of a human expert and to provide users with a complete explanation of its logical reasoning. Designed to aid doctors in proper diagnosis and treatment of bacteremia infections, MYCIN was the most successful application example of an early expert system. It achieved a major breakthrough in AI from theoretical research to practical application, from a general reasoning strategy to the use of specialized knowledge (Shortliffe, 2012). Because of its wide application fields and its success in medical, chemical, geological and other fields, AI has been continuously promoted into application development.

Because the development of an expert system needs the support of a large number of digital operations and logical reasoning, a number of computer languages caught the attention of the news media, especially Prolog and LISP (Colmerauer and Roussel, 1996). The former is the abbreviation of programming in logic, which is a logic programming language. The latter is the abbreviation of list processing, which is an early developed and significant table processing language. However, over time, a series of problems with the expert system began to appear, such as narrow field, lack of common-sense knowledge, difficulty in knowledge acquisition, single reasoning method, lack of distributed function and incompatibility with existing databases (Liao, 2005).

In the second stage, the expert system gradually faded from the public's attention and research into AI fell into a low ebb. At that point, cognitive science, psychology, chemistry and genetics began to arise. Experts were constantly exploring AI's application scenarios, leading to the transformation from Narrow AI (also called applied AI) to Strong AI (also called artificial general intelligence) (Searle, 1980). The proportion of the first two basic subjects (computer science and mathematics) in AI news decreased significantly.

Thanks to the continuous breakthrough of computing power under Moore's law, AI's application scope became more extensive, and the first wave peak appeared in this period.

Figure 6 Institutions and people co-occurring with AI in the news



Cognitive science, astronomy, biological science, decision science, social science and other fields developed rapidly and began to occupy certain positions in AI reports. Some milestones also emerged at this stage. The Turing test is a famous test suggested by Turing in 1950 to judge whether a machine can think or not. It tests whether a machine can show intelligence equivalent to or indistinguishable from human beings. Watson and Deep Blue, as computers developed by IBM, are typical examples of robots beating human beings. The former became the world's first machine to beat Garry Kasparov, the world chess champion at that time, on May 11, 1997 (Campbell *et al.*, 2002). People even began to wonder whether AI had the unique thinking ability of human beings (Parnas, 2017).

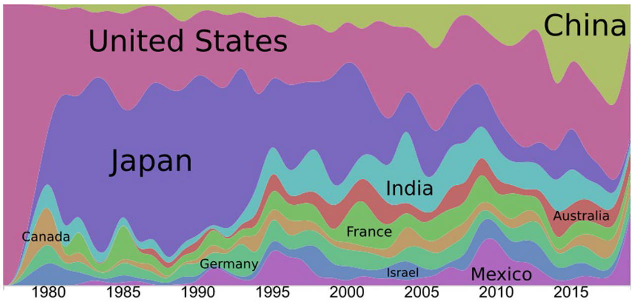
In the current, third stage, agricultural science, materials science and immunology have gradually appeared in AI-related reports. AI has ushered in its own spring. As a rising star, machine learning has become the headlines, increasing from 94 in 2010 to 541 at the end of 2018. Parallel computing technology, image classification, speech recognition, knowledge Q and A, human-computer games, driverless vehicles and other technologies have also achieved a breakthrough from immature to mature (Jordan and Mitchell, 2015). It is worth noting that, in the 20 years from 1985 to 2005, the ranking of neural network increased first and then decreased. Research funding and the decline in public interest made it impossible to carry out further research on neural networks. Around 2010, the topic was no longer hot news, but in recent years it has appeared again.

Institutions, people and countries in artificial intelligence news

First, we extracted the institutions and people from the news, then selected the top 10 names respectively that appeared frequently each year as candidates, and finally retained only the institutions and people that appeared in the top 10 for two consecutive years. The results are shown in Figure 6.

There are three main types of institutions: university, company and government agency. AI-related institutions reported in the news are mainly focused on companies. Of all the universities involved, the Massachusetts Institute of Technology is not only the earliest to emerge but also has more than a decade of history, followed by Stanford University, Yale University and Carnegie Mellon University.

Figure 7 Top 10 countries co-occurring with AI in the news



Notes: Each stream in the river map represents a country, and the width of the stream represents the frequency w of the country in the corresponding year

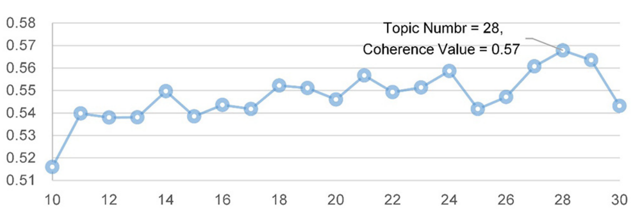
For AI news, company stories have always been the focus of coverage. Apple appeared in 1980 and continues to be newsworthy to this day, followed by IBM, Microsoft, Google and Facebook. However, most companies in AI news are ephemeral and fleeting. This is mainly because of fierce market competition and technological upgrading. And the ones that have always occupied the hot spots are closely related to everyone's daily life.

For government agencies, AI has always been related to the functions of military, aerospace and intelligence agencies, such as Defense Advanced Research Projects Agency, National Aeronautics and Space Administration and Central Intelligence Agency. This shows that the government and related institutions attach great importance to the application and development of AI in a military confrontation, high-tech field and national security.

In AI news, the relevant people are mainly scientists, followed by artists, politicians, businesspeople, writers and chess players. In scientific communication, although AI is covered by commercial values, the foundation of the concept of AI needs support from science. Therefore, in the mass media, the portrait of AI mapping on characters is mainly concentrated on scientists. At the same time, in the field of art, celebrities in the movie and television industry have also received a lot of attention, such as Stephen Allen Spielberg and Stanley Kubrick, who are both famous directors Figure 7.

In general, the USA and China are the hottest countries in the media coverage. Countries that are also highly mentioned in the field of AI are Japan, Australia, India, Canada, France, Germany, Israel and Mexico.

Figure 8 Coherence value for different topic numbers



USA was the first country to develop the AI industry and is also the leader in this field. Japan, known as the "kingdom of robots," was at the forefront of the world in terms of the production, export and use of robots in the 1980s and for the following 30 years, when its robot industry developed rapidly. In recent years, China and other countries, including India and Australia, have also begun to occupy the AI market.

Japan has experienced a very pronounced decline since 2010 and is gradually being replaced by China. So far, many Japanese companies have been relatively slow at taking advantage of the opportunities presented by AI. Owing to China's huge market size and vibrant online commerce and social networks, the country is awash in data, providing the optimal foundation for AI's deep learning systems. The State Council of China issued an ambitious policy blueprint calling for the nation to become "the world's primary AI innovation center" by 2030, and predicted that, by then, China's AI industry could be worth $150bn (Larson, 2018).

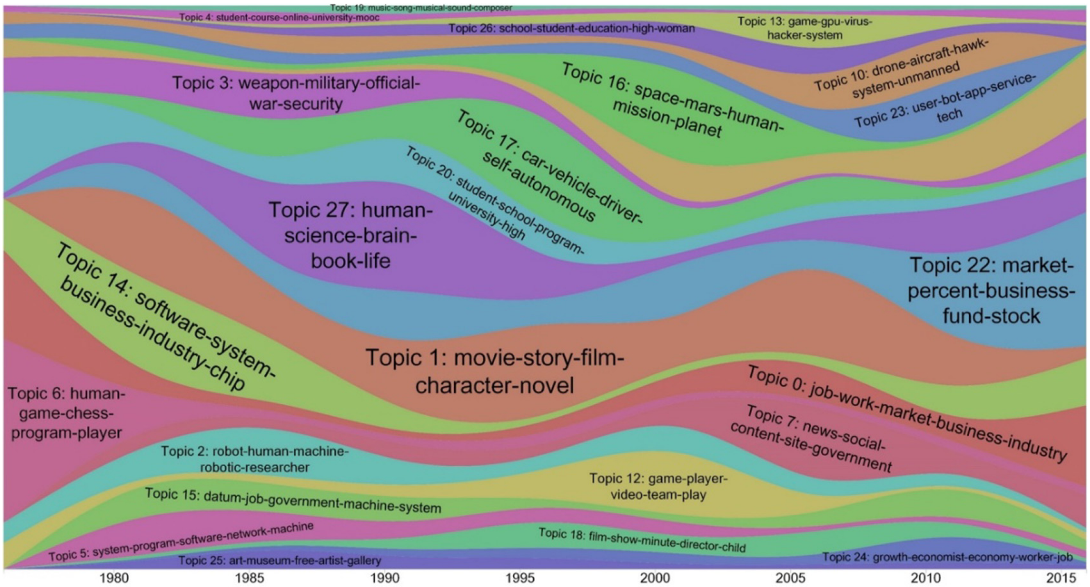
Topics in artificial intelligence news

To find the optimal number (K) of topics, we built LDA models with different numbers of K and picked the one with the highest coherence value. The coherence score is for assessing the quality of the learned topics. It measures the degree of semantic similarity between words in the topic. A topic model with a higher coherence value is more semantically interpretable and in line with statistical inference (Stevens *et al.*, 2012). We valued K from 10 to 30. The result in Figure 8 shows that, when K is 28, the coherence value is the largest. Therefore, 28 topics are extracted from the news, and each topic is represented by five main keywords, as shown in Table IV Figure 9.

The news media defines AI from three perspectives: an imagination, a commercial product and scientific research. During the lifetime of AI evolution, three topics are the most important and account for the largest proportion. Topic 1 is about movies and novels on AI, Topic 22 concerns the commercial market and Topic 27 focuses on AI research. These topics set different tones for the public to understand the concept of AI.

Although AI has shown promise in many fields, many scientists believe that computers will not surpass human capabilities in our lifetimes. Well-known AI expert Wolfgang Wahlster said, "Artificial intelligence is nowhere near the real thing." On the one hand, scientists believe that AI should be conscious and it is unknown whether conscious machines can actually be realized. Although scientists have been studying detailed biophysical computer simulations of the cerebral cortex, they are not optimistic that modeling the brain will provide the insights needed to build conscious machines (Koch and Tononi, 2008). Even the scientific community is still discussing what a conscious machine is Carter *et al.* (2018). On the other hand, movies have imagined a bright future for the development of AI and growing news coverage has biased people's understanding. It is very easy for people to think they know far more about AI than they actually do. The term "Artificial Intelligence" is often misguidedly used by marketing and product promotion specialists in relation to

Figure 9 Evolution of topic popularity



Notes: Each stream in the river map represents a topic, and the width of the stream represents the popularity the topic in the corresponding year. Only 24 topics are marked; the other 4 topics are ignored because their popularities are too small

selling more products. The tech-hype machine splutter has built a reputation for AI of making huge promises and then failing to deliver on them (Brynjolfsson and Mcafee, 2017).

With the development of AI, topics are always changing. In the first stage, as a topic closely related to people, games, topic 6 occupies the main content of news reports. From 1970 to 1980, reports of man-machine battles during this time greatly stimulated public interest in AI. In 1968, after two AI researchers, John McCarthy and Donald Michie, predicted that the computer would beat the world chess champion in 10 years (Berliner, 1978). International Master David Levy made a famous bet that no chess computer would be able to beat him within 10 years. For a long time after that, the topics related to AI and chess have been of interest to the media and the public. In addition to Topic 6, Topic 20, which is related to education, and Topic 14 and topic 0, which are that related to the business and industrial fields, are followed.

When the second stage comes, the media begins to change the focus to topics such as topic 1, Topic 17, Topic 16 and Topic 27. Topic 1 is closely related to the field of entertainment. AI in this field is mostly reflected in movies and novels, which also reflects people's imagination and expectation of AI. Topic 17 is related to transportation, Topic 16 is closely related to aviation, and Topic 27 is related to brain science.

In the third stage, AI has entered reality, and the popularity of topic matter related to movies and novels has declined. Commercial applications and the promotion of technology have become mainstream. At the same time, issues related to industry jobs and social media have become hot topics in recent years.

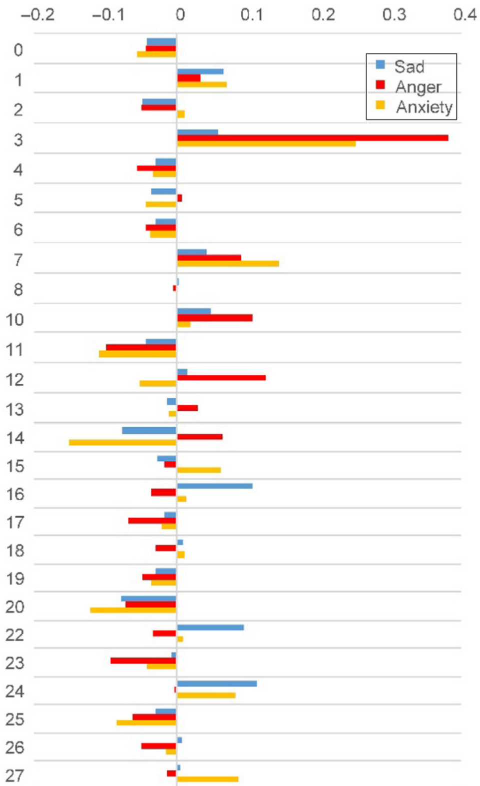
Sentiment analysis

The development of cutting-edge technology is always full of risk-taking elements, zigzagging forward in disputes and contradictions. The social, economic and scientific fields that easily trigger negative emotions are closely related to the abuse and uncertainty of AI. Next, we want to understand which aspects of AI have triggered negative emotions. By calculating the correlation between the proportion of different topics in each article and the frequency of different types of sentiment words contained in each article, we show the relationship between three main types of negative emotions (sad, anger and anxiety) from LIWC and different topics, and the results are shown in Figure 10.

Nowadays, the forefront of scientific literature and media reports pay close attention to the rising discussion of AI ethics (Jobin *et al.*, 2019). AI might jeopardize jobs for human workers. From our results, sad news stories cover Topics 16, 22 and 24, which focus on space, business, and economic and employment issues. We are actively trying to eliminate repeatable physical work from society without considering the employment of the displaced. Despite the ongoing "unemployment recovery" in developed countries, the AI community is developing intelligent mechanization to reduce labor consumption. From algorithms for writing reviews and sports stories, to smart machines that can do everything from spraying crops to acting as taxis, to transporting materials around mines (Makridakis, 2017), AI is eliminating jobs on a previously unseen scale.

The abuse of AI by malicious actors may trigger societal rage and anxiety (Russell *et al.*, 2015). Topics 3, 10 and 12 are most likely to be associated with an angry expression. These three

Figure 10 Correlation between different topics and the different types of negative emotions



topics focus on military, weapons, security and gaming. In addition, Topics 7 and 27 are more prominent and are closely related to anxiety, focusing on social news and human rights issues. The development of AI has reached a point at which the deployment of lethal autonomous weapons systems is feasible within years. These systems have been described as the third revolution in warfare, after gunpowder and nuclear arms (Altmann and Sauer, 2017). Beyond that, AI may challenge different aspects of human rights, such as freedom, equality and justice (van Est *et al.*, 2017).

Conclusion

It can be seen from the media's reports on AI that the impact of AI on human beings is more profound than ever before. Its application has involved all aspects of people's lives and has gradually begun to overcome past defects. With the attention of all sectors of society, the support of government departments and the promotion of the global industry, AI has become an important force in the new era.

There are great similarities between the application of early AI technology and many contemporary applications. People's conception of AI took shape 40 years ago. This conclusion is also a kind of inspiration for researchers in other fields. When developing new scientific research, it has a certain reference function for innovation success to trace back and explore its original ideas.

The public's understanding of AI has been controversial. On the one hand, as the core force of the new round of scientific and technological revolution and industrial transformation, AI is promoting the upgrading of traditional industries, driving the rapid development of the "unmanned economy," and has a positive impact on the fields of people's livelihood such as intelligent transportation, intelligent home, intelligent medical care, etc. On the other hand, AI is not a completely "humanized" technology. When we transfer the right of judgment to computer systems, there will be many moral and ethical dilemmas involved, personal privacy issues caused by big data information disclosure, traffic accidents caused by driverless cars on the road, racial discrimination and gender discrimination in AI recognition, etc. These issues will not only lead to the introduction of more relevant laws, but also spark a discussion on the responsibility for AI, network security and privacy protection. It can be expected that, with the increasing number of negative reports on AI in the media, researchers will have more and more discussions on AI ethics, and more people will join the discussion on AI design.

Our research contributes to a clear understanding of the evolution of the concept of AI in the news media coverage and provides instructive guidance for researchers, business investment and policymakers in related fields. In the future, we will discuss in depth the development and evolution of AI in specific fields and the impact of different media perspectives on the public's perception of AI. This study has some limitations. On the one hand, news data mainly comes from several important international media, not all media, especially non-Western media. On the other hand, using natural language processing has biased the results compared to using manual annotation.

Note

1. https://en.wikipedia.org/wiki/Glossary_of_artificial_intelligence

References

Altmann, J. and Sauer, F. (2017), "Autonomous weapon systems and strategic stability", *Survival*, Vol. 59 No. 5, pp. 117-142.

Ananthaswamy, A. (2011), "I, algorithm: a new dawn for AI", *New Scientist*, Vol. 209 No. 2797, pp. 28-31.

Berliner, H.J. (1978), "A chronology of computer chess and its literature", *Artificial Intelligence*, Vol. 10 No. 2, pp. 201-214.

Bohannon, J. (2015), "Fears of an AI pioneer", *Science*, Vol. 349 No. 6245, pp. 252-252.

Brynjolfsson, E. and McAfee, A. (2017), "The business of artificial intelligence", *Harvard Business Review*, pp. 1-20.

Buchanan, B.G. (2005), "A (very) brief history of artificial intelligence", *AI Magazine*, Vol. 26 No. 4, pp. 53-53.

Cacciatore, M.A., Anderson, A.A., Choi, D.-H., Brossard, D., Scheufele, D.A., Liang, X., Ladwig, P.J., Xenos, M. and Dudo, A. (2012), "Coverage of emerging technologies: a comparison between print and online media", *New Media & Society*, Vol. 14 No. 6, pp. 1039-1059.

Campbell, M., Hoane, A.J., Jr. and Hsu, F. (2002), "Deep blue", *Artificial Intelligence*, Vol. 134 No. 1-2, pp. 57-83.

Carter, O., Hohwy, J., Van Boxtel, J., Lamme, V., Block, N., Koch, C. and Tsuchiya, N. (2018), "Conscious machines: defining questions", *Science*, Vol. 359 No. 6374, pp. 400-400.

Cave, S. and ÓhÉigeartaigh, S.S. (2019), "Bridging near- and long-term concerns about AI", *Nature Machine Intelligence*, Vol. 1 No. 1, pp. 5-6.

Colmerauer, A. and Roussel, P. (1996), "The birth of prolog", *History of Programming Languages–II*, ACM Digital Library, pp. 331-367.

Duda, R.O. and Shortliffe, E.H. (1983), "Expert systems research", *Science*, Vol. 220 No. 4594, pp. 261-268.

Fast, E. and Horvitz, E. (2017), "Long-term trends in the public perception of artificial intelligence", *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, *International Joint Conferences on Artificial Intelligence*, *Menlo Park, CA*, pp. 963-969.

Finkel, J.R., Grenager, T. and Manning, C. (2005), "Incorporating non-local information into information extraction systems by Gibbs sampling", *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, *Association for Computational Linguistics*, pp. 363-370.

French, R.M. (2000), "The turing test: the first 50 years", *Trends in Cognitive Sciences*, Vol. 4 No. 3, pp. 115-122.

Griffiths, T.L. and Steyvers, M. (2004), "Finding scientific topics", *Proceedings of the National Academy of Sciences*, Vol. 101 (suppl 1), pp. 5228-5235.

Jensen, J.D., Pokharel, M., Scherr, C.L., King, A.J., Brown, N. and Jones, C. (2017), "Communicating uncertain science to the public: how amount and source of uncertainty impact fatalism, backlash, and overload", *Risk Analysis*, Vol. 37 No. 1, pp. 40-51.

Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Vol. 1 No. 9, pp. 389-399.

Jordan, M.I. and Mitchell, T.M. (2015), "Machine learning: trends, perspectives, and prospects", *Science*, Vol. 349 No. 6245, pp. 255-260.

Kanal, L.N., and Lemmer, J.F. (Eds)., (2014), *Uncertainty in Artificial Intelligence*, Vol. 4, Elsevier.

Koch, C. and Tononi, G. (2008), "Can machines be conscious? ", *IEEE Spectrum*, Vol. 45 No. 6, pp. 55-59.

Larkin, J., McDermott, J., Simon, D.P. and Simon, H.A. (1980), "Expert and novice performance in solving physics problems", *Science*, Vol. 208 No. 4450, pp. 1335-1342.

Larson, C. (2018), "China's AI imperative", *Science*, Vol. 359 No. 6376, pp. 628-630.

Lee, C.J., Scheufele, D.A. and Lewenstein, B.V. (2005), "Public attitudes toward emerging technologies: examining the interactive effects of cognitions and affect on public attitudes toward nanotechnology", *Science Communication*, Vol. 27 No. 2, pp. 240-267.

Liao, S.H. (2005), "Expert system methodologies and applications–a decade review from 1995 to 2004", *Expert Systems with Applications*, Vol. 28 No. 1, pp. 93-103.

McDermott, J.P. (1980), "RI: an expert in the computer systems domain", *AAAI*, Vol. 1, pp. 269-271.

Makridakis, S. (2017), "The forthcoming artificial intelligence (AI) revolution: its impact on society and firms", *Futures*, Vol. 90, pp. 46-60.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. and McClosky, D. (2014), June). "The stanford CoreNLP natural language processing toolkit", *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.

Müller, V.C. and Bostrom, N. (2016), "Future progress in artificial intelligence: a survey of expert opinion", in Muller, V.C. (Ed.), *Fundamental Issues of Artificial Intelligence*, Springer, Berlin, pp. 555-572.

Naveed, N., Gottron, T., Kunegis, J. and Alhadi, A.C. (2011), "Bad news travel fast: a content-based analysis of interestingness on Twitter", *Proceedings of the 3rd International Web Science Conference*, *ACM*, p. 8.

Parnas, D.L. (2017), "The real risks of artificial intelligence", *Communications of the ACM*, Vol. 60 No. 10, pp. 27-31.

Rehuřek, R. and Sojka, P. (2010), "Software framework for topic modelling with large corpora", in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, *ELRA*, *Valletta*, *Malta*, pp. 45-50.

Rosenblatt, F. (1958), "The perceptron: a probabilistic model for information storage and organization in the brain", *Psychological Review*, Vol. 65 No. 6, p. 386.

Russell, S., Hauert, S., Altman, R. and Veloso, M. (2015), "Ethics of artificial intelligence", *Nature*, Vol. 521 No. 7553, pp. 415-416.

Russell Neuman, W., Guggenheim, L., Mo Jang, S. and Bae, S. Y. (2014), "The dynamics of public attention: agenda-setting theory meets big data: dynamics of public attention", *Journal of Communication*, Vol. 64 No. 2, pp. 193-214.

Searle, J.R. (1980), "Minds, brains, and programs", *Behavioral and Brain Sciences*, Vol. 3 No. 3, pp. 417-424.

Searls, D.B. (2007), "A view from the dark side", *PLoS Computational Biology*, Vol. 3 No. 6, p. e105.

Shakirov, V. (2016), "Review of state-of-the-arts in artificial intelligence with application to AI safety problem", ArXiv:1605.04232 [Cs], available at: http://arxiv.org/abs/1605.04232

Shariff, A., Bonnefon, J.-F. and Rahwan, I. (2017), "Psychological roadblocks to the adoption of self-driving vehicles", *Nature Human Behaviour*, Vol. 1 No. 10, p. 694.

Shih, T.J., Wijaya, R. and Brossard, D. (2008), "Media coverage of public health epidemics: linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics", *Mass Communication and Society*, Vol. 11 No. 2, pp. 141-160.

Shortliffe, E. (Ed.) (2012), *Computer-Based Medical Consultations: MYCIN*, Vol. 2, Elsevier, New York, NY.

Stevens, K., Kegelmeyer, P., Andrzejewski, D. and Buttler, D. (2012), "Exploring topic coherence over many models and many topics", *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, *Association for Computational Linguistics*, pp. 952-961.

Turban, E. and Watkins, P.R. (1986), "Integrating expert systems and decision support systems", *MIS Quarterly*, Vol. 10 No. 2, pp. 121-136.

Van Est, R., Gerritsen, J. and Kool, L. (2017), "Human rights in the robot age: challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality", *Expert Report Written for the Committee on Culture,* *Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)*, *Rathenau Instituut*, The Hague.

Zhai, Y., Ding, Y. and Wang, F. (2018), "Measuring the diffusion of an innovation: a citation analysis", *Journal of the Association for Information Science and Technology*, Vol. 69 No. 3, pp. 368-379.

Zhu, Y. and Yan, E. (2015), "Dynamic subfield analysis of disciplines: an examination of the trading impact and knowledge diffusion patterns of computer science", *Scientometrics*, Vol. 104 No. 1, pp. 335-359.

Corresponding author

Yujia Zhai can be contacted at: zhaiyujiachn@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**