Discover

Quo vadis artificial intelligence?

Yuchen Jiang[1] · Xiang Li[1] · Hao Luo[1] · Shen Yin[2] · Okyay Kaynak[3]

© The Author(s) 2022   OPEN

Abstract

The study of artificial intelligence (AI) has been a continuous endeavor of scientists and engineers for over 65 years. The simple contention is that human-created machines can do more than just labor-intensive work; they can develop human-like intelligence. Being aware or not, AI has penetrated into our daily lives, playing novel roles in industry, health-care, transportation, education, and many more areas that are close to the general public. AI is believed to be one of the major drives to change socio-economical lives. In another aspect, AI contributes to the advancement of state-of-the-art technologies in many fields of study, as helpful tools for groundbreaking research. However, the prosperity of AI as we witness today was not established smoothly. During the past decades, AI has struggled through historical stages with several winters. Therefore, at this juncture, to enlighten future development, it is time to discuss the past, present, and have an outlook on AI. In this article, we will discuss from a historical perspective how challenges were faced on the path of revolution of both the AI tools and the AI systems. Especially, in addition to the technical development of AI in the short to mid-term, thoughts and insights are also presented regarding the symbiotic relationship of AI and humans in the long run.

1 Introduction

Human-created machines are already able to do all-encompassing types of labor-intensive work. Yet, on many occasions, driven by demands for higher productivity and perhaps simply curiosity, humans have been trying to infuse human intel-ligence into machines, which constitutes the original motivation of Artificial Intelligence (AI). AI research has been going on for over 65 years and has made impressive achievements in terms of theoretical study and real-world applications [1, 2]. AI is being used almost everywhere and is considered a core skill for the future. The AI market is projected to grow to $190 billion by 2025, at a CAGR (compound annual growth rate) of over 36% between 2018 and 2025 [3].

There are many definitions of artificial intelligence. In the Turing test, AI is defined as the ability of machines to com-municate with humans (using electronic output devices) without revealing the identity that they are not humans, where the essential judgment criterion is binary. Marvin Minky, one of the pioneers of AI, defined AI as enabling machines to do things that require human intelligence. The symbolic school believes that AI is the operation of symbols, and the most primitive symbols correspond to the physical entities. Although the descriptions of AI are various, the core of AI is widely believed to be the research theories, methods, technologies, and applications for simulating, extending, and expanding human intelligence. Nowadays, the concept of AI has an increasingly profound impact on human life. As the

✉ Hao Luo, hao.luo@hit.edu.cn | [1]Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China. [2]Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology, 7033 Trondheim, Norway. [3]Department of Electrical and Electronics Engineering, Bogazici University, 34342 Istanbul, Turkey.

Springer

roles of steam engines in the Age of Steam, generators in the Age of Electricity, and computers in the Age of Information, AI is the pillar of technology in the contemporary era and beyond.

"AI" has become a buzzword in almost every aspect of our lives. The semantic network graph (Fig. 1) is built based on the search results in Web of Science (2021-08-26) and plotted using the VOS Viewer software. It shows the impact degree and the connections of the keywords that are most related to AI. According to the color of the links, it can be ascertained that the "application" of AI has received great attention in the literature. The concept is closely related to "system" sciences while "neural network," "classification," and "prediction" are the main focuses in terms of algorithms. The research fields of AI include systems and engineering, brain science, psychology, cognitive science, mathematics, computer science, and many other fields. The application fields of AI are extensive, covering (but not limited to) speech recognition [4, 5], image processing [6, 7], natural language processing [8, 9], smart robots [10, 11], autonomous vehicles [12, 13], energy systems [14, 15], healthcare [16, 17], Fintech [18], etc. In limited areas, AI has surpassed humans. The phenomenon of ever-increasing AI-beyond-humans has triggered a new wave of discussion on how AI may change human society. Although many AI applications are stunning, they have not frightened us in the way of the sci-fi movies such as The Terminator or The Matrix because the capability is designed to be area-specific and non-comprehensive. It will be discussed in the subsequent sections that the mainstream success is limited to ANI (artificial narrow intelligence) rather than AGI (artificial general intelligence). Nevertheless, it is not always the case from a long-term historical and forward-looking perspective. Therefore, at these critical times of rapid development of AI technologies, it is believed to be the time to discuss the past, the present, and the future of both AI tools and AI systems.
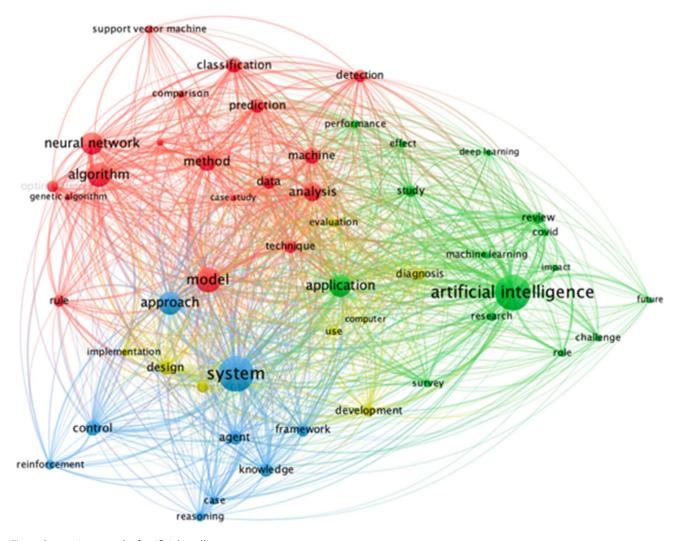

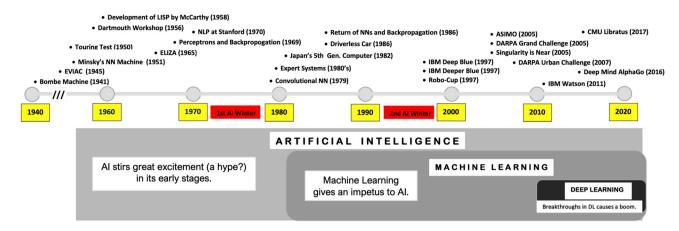Fig. 1 Semantic network of artificial intelligence

Fig. 2 The timeline of AI (copied with permission from [1])

The article is unfolded along the time axis, focusing on the past, the present, and the future. Specifically, the content is organized as follows. The next part introduces the history of AI development, including the Birth of AI, the AI winters, and the resurgence of AI. In Sect. 3, the state-of-the-art AI and the advancements in the discipline are introduced, including Artificial Narrow Intelligence (ANI), Autonomous systems, Symbiotic Systems, and Human-in-the-Loop Systems. Section 4 discusses the future of AI, including Understanding and Cognition, Artificial General Intelligence, Singularity, and Artificial Super Intelligence. The last section concludes the article.

2 Past

Figure 2 shows a timeline that sketches the history of AI by highlighting the important events. In the following part, a discussion will be made aligned with the timeline in general, yet more detailed examples and an inclination to the underlying factors that impact the prosperity of AI research and development will be presented.

2.1 The birth of AI and the golden age

The period before the year 1956 is regarded as the incubation period of AI. Until that time, scientists and engineers tried to replace part of their mental work with machines. In 1936, mathematician Alan Turing proposed a mathematical model of an ideal computer, which laid the theoretical foundation for later to come electronic computers. Neurophysiologists W. McCulloch and W. Pitts built the first neural network model (M-P model) in 1943 [19]. The M-P model is the first mathematical model constructed to mimic the structure and the working principle of biological neurons. It can be regarded as the earliest artificial neural network. Back in 1949, Hebb proposed the learning mechanism based on neuropsychology [20]. "Hebb learning rule" is an unsupervised learning rule, which can extract statistical features of the training sets and classify data according to data similarity. This is the earliest idea of machine learning (ML), and it is very close to the process of human cognition. In 1952, IBM scientist Arthur Samuel developed a checkers program that could learn implicit models from the current position and instruct subsequent moves [21]. In such context, chess programs were among the earliest work of evolutionary computing. The algorithm compares a modified copy to the best version, and the winner becomes the new standard.

It was John McCarthy who originally coined the term AI at the 1956 Dartmouth Summer Research Project on Artificial Intelligence [2]. He is therefore considered to be the father of artificial intelligence. Since this event, the research into AI has rendered many remarkable achievements, including machine learning, theorem proving, pattern recognition, problem-solving, expert systems, and natural language processing.

In 1957, American psychologist Frank Rosenblatt introduced the "Perceptron" model [22]. It can be used to build a system that uses "neurons" for recognition—the capability of learning profoundly impacted the subsequent design of neural networks and the connection mechanisms. The pioneering work on the perceptron is still a trendy topic in the introductory courses for AI today. In 1960, the perceptron algorithm was transmitted to a physical hardware implementation called "Mark 1 Perceptron" as the artificial brain [23], consisting of an array of perceptrons, connected to a camera.

At that time, it was magical and inspiring for the people to see that a machine could be trained to distinguish whether a person in a photograph is a male or a female with reasonable accuracy. In the same year, Newell et al. summarized the thinking rules of humans through psychological experiments. They compiled a general problem-solving program, which could be used to solve 11 different types of problems [24]. A few years later, E. A. Feigenbaum from Stanford University designed an expert system [25]. The system could determine the molecular structure of compounds based on the experimental analysis of mass spectrometers. The study of AI during this period was groundbreaking, with outcomes showing excellent prospects [26]. Meanwhile, the field of computing developed as a field of its own. Transistors and hardware architectures progressed every year, and the essential software and computer programs emerged to implement the AI theories and algorithms.

2.2 AI winters

AI winters refer to the periods when the funding of AI significantly decreased due to disillusionments [1]. There have been two recognized AI winters in history. The first was around the 1970s [27]. Earlier, the researchers had over-confidence in their preliminary results of machine translation. Excessive hype about an optimistic future of machine translation was attracting considerable investments in a context where the US Defense Establishment (the predecessor of DARPA today) expected an automatic Russian-to-English translation during the cold war. It turned out that the challenge of the task was way underestimated. Little progress was made from 1967 to 1976. The failure of machine translation disappointed the investigators and poisoned the development of the whole AI field. In 1966, the Automatic Language Processing Advisory Committee (ALPAC) submitted the report "Language and Machines" to the US government and refuted the feasibility of machine translation [28]. In 1969, Minsky and Papert published the Perceptrons [29], which pointed out the significant flaws of perceptron that perceptron can only solving linear problems. In 1973, the British Scientific Research Council (SRC) presented the Lighthill Report to the British government [30]. These criticisms triggered a significant shrinkage of the funding for fundamental research in AI.

As the effects of the first AI winter slowly faded away, a new wave of interest escalated in AI technologies oriented towards commercial applications. This time, expert systems were extensively implemented based on "if–then" rules set up by domain experts. Once again, its potential was hyped because people had insufficient awareness of its limitations in dealing with highly complex tasks and systems, such as medical diagnosis and vision systems. In 1984, John McCarthy commented that a critical defect of expert systems lies in the lack of knowledge about their limitations [31]. In the late 1980s, another trough in AI development was experienced, accompanied by the withdrawal of funds and the shift of interest to other rising technologies such as general-purpose computers.

Despite the significant differences in the historical backgrounds, several common factors can be identified in the emergence of the AI winters [32]. From a professional viewpoint, researchers had insufficient and biased evaluations about the pros and cons. Consequently, the predictions and the conclusions lacked rational, rigorous, practical, and comprehensive analysis. Furthermore, the news media acted as a catalyst in turning away the public focus from the real challenges that the researchers and the practitioners were facing [33]. Another explicit factor was the underdeveloped tools from the related disciplines, such as the unavailability of algorithmic approaches (mathematics), information resources (electrical engineering), and computing power and information infrastructure (computer science). Furthermore, the investigators were unaware of the risks and underestimated the required time for fundamental research. Large-scale investments and divestments were the triggers of the AI winters.

2.3 The rise of AI

Several vital factors facilitate the resurgence of AI.

First, the success of machine learning is one of the initial and base drives. Since late 1980s, a series of important ML theories and techniques have been proposed, many of which are necessary materials in today's ML textbooks [34, 35]. Typical ones include decision tree, proposed by J. R. Quinlan in 1986; support vector machine, proposed by Vapnik and Cortes in 1995; Adaboost, proposed by Freund and Schapire in 1997; random forests, proposed by Breiman in 2001. As for the neural network-related advancement in this period, it was not until the article promoting back-propagation published in 1986 on Nature [36] that the research community paid good attention. The successes in ML and the other progresses discussed below become complementary factors on the path of AI resurgence.

Second, colossal amounts of data are available as the fuels for model training [37, 38]. The quality and the scale of the datasets are the determinants of the robust performance of prediction/classification [39, 40]. High-quality datasets built

in typical and practical scenarios are encouraged to be shared in the research community. In this context, people devote efforts to designing and carrying out exhaustive experiments to collect real-world data, representing typical working conditions. In this process, the types of variables and how the data are measured constitute a part of the intellectual inputs by humans. As one of the key factors contributing to successful AI training, such experience has an unneglectable influence on the quality of the datasets in terms of comprehensiveness and typicality because it decides how well the data distribution of the constructed dataset can align with the ground truth. Publicly available datasets function as a baseline for defining AI challenges and for fair comparison of novel AI algorithms [41].

Third, the rise of AI is associated with a significant increase in computing power [42, 43]. The expansion in the physical scales of the AI-based machines enables them to deal with highly complicated tasks. In the meantime, it will inevitably slow down information transmission and processing and lead to plenty of energy consumption. A promising breakthrough is expected in terms of photon computing to overcome the bottleneck of today's general practice based on electrons [44]. In addition to dealing with the maximum possible speed constrained by physical laws, numerous implementation challenges may cause significant delays. With advanced techniques of acceleration taking place in both hardware architecture [45] and software algorithms, the training tasks that usually take several months can be shortened to several days or even hours [46–48]. This is possible because different deep neural networks can be decomposed into several blocks where many required operations are the same. For example, a typical CNN (Convolutional Neural Network) block usually involves the processes of convolution, pooling, normalization, activation, and a fully connected layer [49]. In recent years, there have been fruitful outcomes reported about AI chips. Dedicated NPUs (Neural network Processing Units) have been developed and used on Kirin 970 chips and A11-A14 Bionic chips for iPhones. This makes it possible to perform real-time onboard (portable devices) tasks such as video processing for live broadcasts, fast biometric identification, and games with 3D rendering [50–52].

Fourth, several AI systems that outperformed top human players in contests and competitions impressed the public and rebuilt confidence in AI [53, 54]. In 2011, IBM's AI system Watson won the championship at America's popular quiz TV show "Jeopardy!". Although it was not connected to the Internet, it had a four terabytes storage that covered 200 million pages of information, including the full version of Wikipedia. This is a remarkable achievement in implementing a machine with joint abilities ranging from information search to knowledge representation, automatic reasoning, and natural language processing. From 2016 to 2017, AlphaGo challenged professional go players and defeated over 60 human masters of go, including the world champion Lee Sedol and the worlds' top-ranking player Ke Jie. The developer of AlphaGo, i.e., the team with Google's DeepMind, later built a new AI system called AlphaGo Zero, which is entirely self-taught starting from scratch, with zero information from human's game records. The system surpassed thousands of years of human knowledge with reinforcement learning and shockingly discovered novel strategies over a few days [55]. These examples show the dominating role of deep learning in the new AI age and demonstrate the phenomenal speed that AI can grow.

In Table 1, the critical events in AI development during the past 25 years are summarized.

3 Present

3.1 Widespread use of ML and then DL

At the very beginning of Google's Machine learning handbook, it writes, "If you can build a simple rule-based system that does not require machine learning, do that." It can be interpreted in two ways: first, as a type of heuristic algorithm, machine learning still cannot achieve guaranteed performance with full interpretability; and second, ML may have satisfactory performance (but still, not guaranteed performance) in specific scenarios that rule-based methods cannot. There are considerable challenges in figuring out all the rules and steps for highly complex problems. Non-AI schemes designed purely by human intelligence would lead to significant defects in the solutions or decision logic [57, 58]. Likewise, expert systems mimic humans' way of thinking and rely too much on domain knowledge and human experience. Furthermore, the key to designing an expert system lies in utilizing expertise rather than methodology. Due to insufficient generalization ability and low robustness, these methods soon reached a bottleneck in widespread applications.

Rather than specifying the inputs and the human-defined algorithms to obtain the outputs, an alternative means to solve complex problems is to select the inputs and the expected outputs and to use machines to find the patterns between them automatically and thus "learn" the algorithms [59–61]. This is the general idea behind how (supervised) machine learning works. Machine Learning shows superiority in the following aspects. First, ML is good at learning from

Table 1 Important events in AI development: last 25 Years

Year	Event identifier	Description
1995	NavLab-5	During the road test, a long-distance autonomous driving experiment across the United States was conducted for the first time. The autonomous driving distance was 496 km, accounting for 98.1% of the total distance, with speed of about 55 miles/hour
1997	Deep Blue vs Kasperov	Kasparov defeated the AI chess system Deep Blue with a score of 4:2
	Deeper Blue vs Kasperov	Deeper Blue defeated Kasparov with a score of 3.5:2.5, making it the first computer system that successfully defeated the defending world champion in a standard chess game
	Pathfinder	NASA's MARS Pathfinder mission made a successful landing and the first autonomous robotics system
	LSTM	A type of recurrent neural network which shows good potential at processing and predicting time series of long intervals and delays
	First Robocup in Japan	The first official Robot World Cup (RoboCup in short) games and conference was held in Tokyo. Over 40 teams participated (real and simulation combined), and over 5000 spectators attended
2000	ASIMO	Honda's ASIMO robot, is able to walk as fast as a human, delivering trays to customers in a restaurant setting
2004	First DARPA Grand Challenge	A prize competition for autonomous vehicles is held in the Mojave Desert. None of the autonomous vehicles finished the 150-mile route
2005	DARPA Grand Challenge	Five vehicles successfully completed the 212 km (132 mi) course of rough, cross-country Nevada terrain
	Book by Kurzweil	Singularity is Near by Ray Kurzweil is on the shelves. 2045 is the predicted year of reaching singularity
2007	DARPA Urban Challenge	CMU's autonomous vehicle "Boss" won the race, being able to drive safely in traffic at speeds up to 48 km/h (the maximum speed limit according to the rules). Motion planning develops as a relatively independent module in autonomous vehicle
2009	ImageNet	A computer vision project built by Stanford University, based on which many important image recognition algorithms were proposed. Now it has become the largest image recognition database in the world
2011	Apple Siri	Siri was integrated with iPhone 4S, making it the first customer device that can interact with humans through natural voice. Now it can constantly learn new sounds and intonations
	IBM Watson in Jeopardy	The AI system won the championship at America's popular quiz TV show. It was not connected to the Internet but had a four terabytes storage that covered 200 million pages of information, including the full version of Wikipedia
2012	Google brain simulator teaches itself to recognize cats	A neural network of 16,000 computers was let loose on YouTube images for 3 days. By the end, it learned to recognize cats
	Deep learning CNN as used by AlexNet	The introduction of AlexNet convolutional neural network (CNN), designed by Alex Krizhevsky. It won the 2012 ImageNet LSVRC-2012 competition by a large margin (15.3% vs 26.2% (second place) error rates, increasing interest in Deep CNNs
2013	The debut of Atlas Robot	The Atlas disaster-response robot made its public debut on July 11, 2013. In its original form, the 6'2", 330-lb. humanoid robot—developed for DARPA by Boston Dynamics was capable of a range of natural movements
2014	GAN	An unsupervised learning strategy based on the game between a generator and a discriminator. Generator can generate data close to the real data
	TrueNorth	The first digital custom-designed, large-scale neuromorphic processor (Digital 'Big Brain' chip) [56]
2016	DeepMind AlphaGo	AlphaGo 2016 version beats professional Go player Lee Sedol
	Quantum communication	The first quantum communication experiment satellite ("Micius") successfully launched by China. Uses quantum communication encryption technology for secure data transmission
	Google's Firefly	Self-driving car prototype that was designed without a steering wheel, brake or gas pedals, or traditional controls of any kind

Table 1 (continued)

Year	Event identifier	Description
2017	DeepMind AlphaGo Zero	Entirely self-taught using reinforcement learning, with zero information from human's game records. Beats AlphaGo Lee with a score of 100:0 based on 3 days' self-training
	CMU Libratus	CMU built machine played against four top-class human poker players. After 20 days Libratus was ahead by about USD 1.8 million while all human players were negative
	Waymo's driverless car	Fully driverless vehicle operated on a private road in California's Central Valley, which consists entirely of production model cars modified with self-driving technology. Waymo was allowed to deploy its driverless cars on Arizona roads without safety drivers
	AI chips	TrueNorth chip from IBM used for building supercomputers with extremely low power assumption. For mobile terminals, Kirin 970 and A11 Bionic chips realized the pioneering built-in independent NPUs (Neural network Processing Units) for smart phone AI computing
	Transformer	A competitive machine learning architecture for Seq2Seq tasks (natural language processing, computer vision, etc.). Originally put forward in the NPIS 2017 paper Attention is All You Need
2018	BERT	A new masked language model which achieved state-of-the-art results in 11 natural language processing tasks
	Google Duplex	A personal AI assistant that can carry out natural conversations with humans. At Google I/O 2018, it showed impressive functions in understanding the users' demand in booking a restaurant and then making phone calls on behalf of the user
	AI frameworks	The number and the update speed of AI frameworks accelerated significantly in 2018, for example, Pytorch from Facebook, TensorFlow from Google, MXNet from Amazon
2019	Apollo Go Robotaxi	An L4-level driving taxi fleet from Baidu, started trial operation in test sections in Changsha
	Quantum computing	A new computing mode that regulates quantum information units to perform fast calculations. Breakthrough was made regarding the realization of high-performance single-photon sources
	DeepMinds' AlphaStar	A master AI player of "StarCraft 2," trained by reinforcement learning. Defeated many human top-professional players
2020	AI and vaccine	LinearDesign, the first mRNA vaccine gene sequence design algorithm (Help to improve the stability and the protein expression level of vaccine design) developed by Baidu Research Institute
	Applied NLP	Microsoft introduced its Turing Natural Language Generation (T-NLG). OpenAI's GPT-3, a state-of-the-art autoregressive language model that uses deep learning to produce a variety of computer codes, poetry, and other language tasks exceptionally similar, and almost indistinguishable from those written by humans. Its capacity is ten times greater than that of the T-NLG
	Cruise origin	The first production car designed for automated driving and a fully automated vehicle without driver and pedals
	DeepMinds' AlphaFold 2	Successfully predicted the structure of most proteins in the contest of CASP 14, with tiny differences from the real protein structure by one atom width. Such result reached the level that humans use complex instruments such as cryo-electron microscopes to observe and predict
2021	Wu Dao 2.0	The largest neural network ever created and the latest development in NLP, with 1.75 trillion parameters, 10× larger than GPT-3. Wu Dao 2.0 can learn from text and images and deal with tasks that include both types of data (multimodality)

colossal amounts of structured data while humans are not, due to limited memory and brain capacity. Second, having "seen" an adequate number of training samples, the ML solutions obtained show good generalization capability. Third, each newly available data contributes to adjusting the input–output projections [62]. The learning process can be realized incrementally, which enables online and life-long learning [63]. Apart from supervised learning, there are also practical requirements for unsupervised learning. In many cases, the labels for the training data are unavailable due to the high workload and the need for domain experts. In one way, clustering-based approaches can automatically divide data into groups without specifying the exact mapping to the feature space. However, for estimation and prediction tasks, proper encoding/decoding procedures are needed. A similar idea applies to reinforcement learning. As a semi-supervised learning scheme, the subject learns continuously by taking actions and adjusting the acting rules ("policies") based on the evaluation of the feedback ("rewards"). To be implementable for action evaluation in a limited period, the action space should usually be spanned by finite elements. A typical example is the application of collaborative robot control [64].

To deal with highly complex problems and discover insights from extensive data collections, deep structures are necessary [65–67]. In one aspect, by specifying the neural network model structure, a set of functions are parameterized. As a result, the upper limit of learnability has been decided. As an extreme example, any instance from the set of all linear functions can never describe a quadratic relationship regardless of what parameters are chosen. Therefore, it is promising that deep structures can capture more complex nonlinear relations between the inputs and the outputs than shallow and flat structures, given the same number of neutrons. [Some successful examples with deep structures: GoogLeNet (22 layers), AlphaGo (40 layers), etc.]. In another aspect, the learned weights determine which single function from the set can minimize the loss function, whereas how well they are learned determines the upper limit of performance (such as precision, robustness, confusion matrix, etc.). DL has shown strength in dealing with unstructured data, such as a piece of experience, a specific scenario, or skills transferred from other activities. The inputs can be anything that is properly encoded, e.g., an image, a video clip, a sound wave, texts, or readouts from various sensors. At the stage of problem formulation, rules are established to transform unstructured data into structured ones. The outputs can be indicators taken from a finite set of predefined categories (for classification tasks) or instances of variables that have an infinite number of values (for prediction tasks).

In terms of algorithmic development, several researchers (including us) have experimented with techniques from a wide spectrum, ranging from (pure) knowledge-driven solutions to data-driven approaches, more recently expanding the range to include the ML/DL-based techniques with supports of big data and knowledge-based work automation techniques. It can be stated that helpful knowledge is scattered and dispersed in a wide range of raw data or information fragmentation. The challenge is to mine the evidence hidden in hard-to-notice details, so logical reasoning/deduction and subtle attention to the critical aspects based on the subconscious or experience are needed on the ML/DL technical route. Regarding the wide acceptance of such a research route, mainstream research fields include computer vision and pattern recognition, robotics and automation, natural language processing (considered to be the holy grail of AI), knowledge discovery and data mining, strategy- and decision-making, and so forth.

3.2 Artificial narrow intelligence (ANI)

The application of AI techniques has penetrated our daily lives [68]. Most of the existing AI systems are dedicatedly designed for a limited range of predefined tasks. In this context, they are called artificial narrow intelligence (ANI) or weak AI. When we need to go to someplace but do not know the route, all we need is to specify the destination address to the Map apps on the smartphones, and it will provide a list of optimized path planning solutions that minimize the traveling time or cost. When we want to acquire information from the Internet, the search engine does most of the preliminary information collection job, including fine-targeted advertisement. Moreover, reliable real-time bioinformatics analysis has enabled novel means of personal identification, and as a result, novel means of payment based on fingerprints and face recognition become possible. Many ANI applications do not require globally optimal solutions, which are always challenging to resolve with formal approaches. Therefore, heuristic approaches have been widely studied as time-feasible alternatives. For safety–critical/security-critical systems, compromising optimality and achieving a theoretically guaranteed minimum performance by constraining the solution space is favorable. In recent years, quantum mechanics have been suggested to be a promising technology for cyber-security [69, 70]. In August 2016, Chinese scientists launched the first-ever quantum communication satellite. The maturation of the new generation of communication and computing techniques like quantum transmission and quantum computation may establish a new norm in the AI-embedded ecosystems.

In the domain of knowledge work automation, ANI systems have widely been applied to Q&A robots, serving banks, airline hotlines, and online consultation [71]. The key feature of these applications is a precise categorization of the domain knowledge and a minimal range of frequently asked questions. From a domain-specific perspective, ANI tools have been or are being developed for education, medication, transportation, marketing, and game production [72–76]. By contrast, ANI becomes insufficient for highly complex tasks such as the optimization problems for process industry and would lead to unsatisfactory performance in terms of decision-making. The main challenge lies in how to connect the engineering operations limited by the physical laws (objective factors) to human experience and to the socio-economic factors (subjective factors). This requires a set of ANI systems, working jointly for the same target. In this sense, the ability of ANI is constrained to problem solving rather than problem definition and resource exploitation, thus in essence, dependent on how well the problems can be defined by humans.

3.3 Autonomous systems

An autonomous system is a group of components forming a unified whole with the power of acting independently from the external environment [77, 78]. Today, autonomous vehicles and autonomous drones have become typical applications of such. They are targeted to perceive the surroundings, get aware of the real-time status, make plans and decisions, and take actions, all of which are achieved by onboard devices and/or with the aid of resources on the cloud and the edge side, with little human intervention. It is not difficult to find that these behaviors are driven by human-like problem solving and learning abilities, which are the backbones for any autonomous system. In case of emergency, there can be human instruction override for vehicles and drones. However, it is not always feasible in extreme environments such as for deep space and deep-sea exploration. The time delay of information transmission becomes significant, considering the long distances. Therefore, a robust autonomous system needs self-diagnosis and self-restoration capabilities [79].

In another aspect, autonomous learning is necessary for generic AI systems. The learning ability determines how in the long-term the autonomous system self-evolves. On the opposite side of natural beings, most of the engineered autonomous systems have fixed functionality sets that usually cannot be expanded once constructed. It should be noted that not all the learning-based approaches enable autonomous learning abilities. The point lies in whether they are used for offline model training only or for online updates as well. The latter is also referred to as life-long learning. Training with positive examples, commonly understood as teaching and practicing, is a means for learning in the narrow sense. Beyond it, ability migration (transfer learning), critical thinking (adversarial training), imitation (imitation learning), and creativity (heuristic design) have begun to be proven as alternative effective technical routes for AI design [80, 81].

Knowledge, information, and data compose the fuel of the learners. Knowledge describes regular patterns and abstract facts that human understands. Therefore, it is usually semantic and embedded in books and research articles. To be interpretable and useful for machines, it needs to be modelled, transformed, and generated, which corresponds to the tasks of knowledge representation, knowledge reasoning, and knowledge discovery. Information is transmissible content that has timeliness, such as news and contextual data flow. The backbone of the contemporary industrial revolution—the cyber-physical systems (CPSs)—bridges the virtual space and physical space by establishing automatic flow of information and knowledge. In such context, the entities of autonomous systems can maintain digital replicas (digital twins, DT) that evolve along with real-time flow of information/data. The entities and the corresponding digital replicas form a conjugate pair to deal with the real-world complexity and uncertainty [82, 83]. In the future, the digital replica may eventually become a feature subset of the physical entities. Furthermore, subsystems and submodules in the autonomous system can be linked to the knowledgebase and information flow of the entire fleet of assets, combining workflows in a novel closed-loop manner. It is worth to note that the R&D activities on closed-loop empowerment systems based on CPS/DT are still ongoing [84]. A few industrial examples are introduced in [83] and the references therein.

3.4 Symbiotic systems and human-in-the-loop systems

While the AI systems of today have outperformed humans only in a limited range of highly specific tasks, there are still risks and concerns about a future when humans must co-exist with novel forms of AI systems. It may not be too early to start thinking about the role of and the relationship between the human and the machines with intelligence.

Human-AI symbiotes are considered as a possible form of harmonic co-existence of human beings and AI. With the aid of the brain-machine interfaces (BMIs), AI acts as a kind of digital expansion of the central nervous systems. In such context, the boundary of the biological and the artificial minds is blurred. There may also be drastic changes in the way

how we communicate with each other and interact with machines, as information exchange is more efficient via nerve (bio)electric signals alone than using the nerve signals to drive the muscles to generate different facial expressions, voices, postures, or gestures. To this end, there need to be strong connections between the AI systems by human minds and vice versa. Today, fundamental works are being carried out.

According to the forms of information representation, machines can receive human instructions by means of bio-signal recognition and behavior recognition. The former takes EEG (electroencephalography) signals, thermal signals, and even chemical signals and interprets them using signal processing/pattern recognition approaches. The latter takes optical signals and audio signals and interprets them with machine vision techniques and speech recognition techniques. Beyond all these, people today spend much time in living a digital life and are equipped with numerous electronic devices. The information uploaded to various apps is useful for creating digital profiles. For instance, personal social media networks characterize daily behavior and interpersonal relationship while the real-time public medical data can reflect the epidemic situation [85–87]. Smart phones, smart glasses, tablets, and so forth act as auxiliary interfaces to monitor and influence human behavior [88].

Regarding the implementation solutions that bridge human minds and AI, further advancements in BMI (or brain-computer interface, BCI) technologies are necessary [89]. The main tasks include capturing, amplifying, analyzing, and interpreting the brain wave signals, and based on which to control the connected machines/computers. The success necessitates interdisciplinary research, covering neuroscience, data science (signal processing and machine learning), as well as communication and sensing technologies [90]. In the laboratory environment, researchers have managed to achieve non-invasive implementation for writing and spelling [91, 92], prosthetic/exoskeleton control [93], and for general-purpose communication [94]. By contrast, the invasive BMI techniques with electrodes in the brain can achieve better performance in control precision and are more suitable for real-time applications. Additional advancements in the study of brain functions and the development of nanomaterial and implant surgery techniques are necessary for implementations on a broader scale [95, 96]. It is also worth noting that the state-of-the-art practice is still limited to open-loop control and unidirectional information flow, from humans' brains to computers. This is more due to ethical concerns than technical feasibility, as experiments with animals have demonstrated the effectiveness of artificial stimuli to control visual and motion perceptions [97].

With the establishment of effective links, demand-driven application scenarios that benefit from the symbiotic eco-system have started to surface. Over the past decade, human-in-the-loop (HITL) systems have become a mainstream research subject that addresses the issues regarding modeling and predicting human-related factors so that they can be treated as a part of closed-loop systems. On the opposite side of the industrial practice that treats human participants as unfavorable factors [98] and pursues unmanned working conditions with high automation degrees, the main challenge lies in dealing with the uncertainties introduced by the human participants. While it is still impractical to interpret human thinking in real-time, AI may identify the intention by observing our status and behavior, by learning from the changes made to the environment with which we are interacting, or by learning from how we react to previous cases.

HITL technology plays an essential role in training, design, operation, and optimization [99–101]. For instance, pilots are taught on flight training simulators (FTS) where the flight environment is adaptively adjusted and rendered by AI. Similarly, surgeons and dentists can carry out virtual surgeries as a part of training or strategy-making before actual medical activities. As for designation, AI creates groups of (digital) prototypes as candidate product solutions and constantly interacts with the designers or the customers to meet personalized requirements, in a parallel, case-specific, and efficient manner. In industrial assembly processes, smart robots or those known as collaborative robots help human operators (and vice versa) to manipulate tools and give warnings and suggestions [102, 103]. An autonomous driving system and advanced driving assistant system (ADAS) monitor the traffic condition and free the human drivers from intensive attention. Furthermore, with HITL optimization techniques, AI systems can reduce the conservativeness in making recommendations. They focus on the best options and are not concerned about outrageous ones because the human judges will filter them out. Based on these examples, it can be summarized that both humans and AI interact with the environment, perceiving the states and making corresponding decisions and changes. From the perspective of systems and control, there are two controllers in the same closed-loop, driven by the biological and the artificial brains. In this sense, to ensure overall stability and achieve good performance, understanding and predicting human actions is no less important than designing explainable AI. The attractive part is that humans excel at leveraging unstructured data while machines outperform at analyzing structured data, forming a natural complement.

In the long term, there are several advantages if human-AI symbiotic systems become more communal. In one aspect, symbiotic systems may ensure that the intelligence level is improved in a balanced and controllable manner. No individuals or some groups of individuals hold monopolistic information resources in all domains so that AI dictators or human

dictators powered by super-AI might be avoided. In the meantime, the competence in humans and AI is concerted to a collaborative relationship [104]. Additionally, the openness and transparency of AI development can be maintained over the process of establishing a symbiotic ecosystem. It relies on interpretable operation mechanisms and explainable AI (XAI), helping us to understand from the AI's perspective how a problem/task is defined, what information sources are needed, how to exploit knowledge from big data, what factors are taken into consideration, which solution is adopted to solve the problem, and eventually, how a decision or a prediction is made. Furthermore, the in-depth integration with AI systems may unleash the potential of the human brain and elevate overall social efficiency. The limitations in information availability and computing capability might be eliminated. Instead, the bottleneck would lie in creativity, horizons, and non-renewable resources.

3.5 Driving forces of AI development

An increase in the bulk of subjects facilitates the AI R&D ecosystem and the related ecology. In the short term, the development of AI is to a large extent driven by dominant stakeholders, especially with support from major investments and governmental strategic planning [3, 68]. The key lies in whether the novel technologies can be grounded swiftly, contributing to economic harvest and regional/social stability. On the other hand, learning from the historical experience of "AI winters", it is still uncertain and indirect that whether, in the mid to long term, AI development will converge to the basic demands of the general public, for improving the quality of life, the efficiency of working, and the degree of happiness.

The healthcare industry covers all the attributes from patients to doctors, from platform to end users, from groups to individuals [105]. The industrial chain is constituted of medicinal materials, medicine, medical instruments, medical institutions, medical information, and medical examination. Today, healthcare applications are acting as an active driving force of AI development. Colossal amounts of medical data are collected every day. The complexity of medical information lies in the high dimensionality, nonlinearity, and sometimes strong dynamics. In this context, the AI-powered diagnosis systems need to be well-trained to achieve an acceptable performance, which usually consists of a list of indicators such as those characterizing the prediction accuracy and the confusion matrix that characterizes the classification performance.

AI plays a key role in establishing novel norms of modern hospitals towards one-stop medical services and new business models. In the future, the allocation of limited medical resources will still be the major issue. In the cases where the physicians require reliable intelligent preprocessing tools to reduce their workload in dealing with raw data, AI may act as a good assistant. For instance, AI helps to markup suspicious lesion areas on optical images [106], X-ray images [107], or magnetic resonance images [108–110], and can analyze patients' speech to diagnose Alzheimer's disease [111]. Researchers also put hope on shortening the cycle of new drug discovery with the ability of AI in learning from vast volumes of medical literature and datasets [112–115]. Furthermore, embedding AI systems in portable devices will help reduce the shortage of medical resources in remote areas and reduce social inequality. Motivated by these practical demands, there are huge driving forces in the AI-related R&D activities, the production, and the applications of AI-embedded devices, as well as the construction of intelligent platforms and infrastructure. Currently, AI has improved the medical automation, but is mainly used to aid diagnosis. Using AI alone as the backbone of healthcare is very dangerous. In 2015, IBM launched Watson Health, a business unit that uses AI to solve critical problems in healthcare. Their algorithms prescribe drugs with severe side effects and even death to a cancer patient.

Another driving force behind the advancement of AI is related to intelligent transportation applications. Research on autonomous vehicles has been an extremely popular direction. The tasks cover all aspects from perception to decision making, which are all considered to be in need of intelligence [116]. Delightfully, what were considered as challenging problems have now applicable approaches, such as localization, object recognition, and tracking, movement prediction, and path planning [117]. Although these are ANI tasks, the high demands in dealing with the complexity and uncertainty in the real-world environment and the requirements of real-time and onboard implementation are boosting the AI industry. From the viewpoint of commercialization, it is hard to tell whether the traditional automotive companies or the Internet enterprises dedicated to intelligent solutions will be dominating the innovation and the market of smart automobiles. From another perspective, vehicular cyber-physical systems are being studied to connect the information islands of each individual transportation participants [118]. The models of the real-time regional traffic network can be created with the aid of the cloud computing resources and based on the publicly available information collected from distributed agents (such as vehicle sensors, pedestrians' smart phones, and roadside infrastructure). On this basis, intelligent algorithms are required to play the role from a macroscopic perspective, for instance, in the prediction and optimal diversion of the traffic. Today, policies are put forward to provide support for promoting intelligent transportation applications and to provide regulations for testing activities and open test roads.

In addition to the above, wide categories of simulation applications are relying on and in turn driving the development of AI. The typical application scenarios range from manufacturing to education, from smart city to entertainment, and many more beyond [119–121]. The enabling technologies include but are not limited to Internet of things, extended reality (XR), digital twin, robots, smart grids, space technologies, and so forth. In comparison with the AI winters, the historical period of the twenty-first century is full of opportunities, grasping which will lead us to a prosperous future where AI can create values and benefit human lives to the greatest extent.

4 Future

4.1 Understanding and cognition

While AI systems are created by humans, they may evolve in a much faster speed than anticipated and develop high complexity in structure, behavior, and decisions that humans can hardly follow. In the long term, AI systems and human beings need to establish common cognitions in terms of the outlook on the world, life, and values. However, to date, engineered systems cannot be modelled to precisely replicate the real-world entities and the physical environment. That is why a bridge to connect the cyber and the physical worlds is being built in the Industry 4.0 era, hoping to feed the engineered digital replicas with online real-time measurements (for synchronization, correction, and update) and further why Kalman filter is regarded as a cornerstone technique in the domains of sensing/perception/control due to its ability to achieve optimal estimation by compromising measurements (with inevitable error) and model outputs [122, 123].

As discussed in Sect. 3.4, AI will play important roles in symbiotic systems in the future. In the narrow sense, human–machine symbiotes would bring benefits in terms of overcoming disabilities or making humans stronger (also referred to as "human augmentation") and getting more knowledgeable. Especially, in an age of information explosion, it has become increasingly difficult to search through the available resources. Yet, a type of cognitive digital twin (CDT) may be very helpful to collect and pre-screen the related data, and then to perform tasks on behalf of human. An illustrative example of "Google Duplex" was given in Table 1 where AI made phone calls to book a restaurant. With the grounded applications of natural language processing and computer vision, CDT may execute users' much complex demands. Customized optimization is feasible by integrating the knowledge structure of an individual, and it is actually an ongoing research which is usually referred to as personal CDT. In turn, the knowledge structure and the practical skills of humans could be improved in a targeted manner by using courses suggested by AI (just like what was envisioned in the movie The Matrix). Similarly, in the future, scientists could manage the explosion of knowledge better with the aid of machine reading to discover new knowledge that they need most from hundreds of thousands of publications each year.

We see that there are plenty of AI algorithms inspired by the nature, such as the grey wolf algorithm, ant colony optimization, particle swarm, genetic algorithm, and so forth. Neural networks are inspired by biological nerve systems. Nevertheless, the models have been much simplified, and as a result, the mechanisms of learning and acting become different and divergent that the AI systems can hardly understand the nature and the socio-economic worlds like we do. In the future, with super large-scale neural networks such as Wu Dao 2.0 and GPT-3 becoming common, it may be expected that multimodality across the major branches of AI such as computer vision and natural language processing will end up unified [124].

On the other hand, can humans understand the behavior of AI systems and trust their decisions? As aforementioned, XAI (or interpretable AI) is the key towards extensive and generic application of AI systems for reliability-critical tasks. Theoretical guarantee is needed in terms of why it works and what the target is [125–128].

The research on XAI can take two routes. First, AI systems are designed as equivalent or approximate solutions to the conventional non-AI solutions with clear physical meanings. In this way, it is clear what the design target is and what exactly is to be learned, and the heavy dependency on trial-and-error during the design phase can be reduced by learning from data. For instance, the "kernel trick" is usually used to simplify the computation of the inner product of nonlinear functions which map low-dimensional inseparable variables to a high-dimensional separable feature space. The problem is that there has been no consensus reached on how to optimally select the type and how to set the parameters of the kernel function, and the online computational load is quite high. As a result, although the performance is satisfactory and sometimes superior, kernel-based methods are facing great challenges in practical real-time applications. Promisingly, Deep Neural Network (DNN) training enables automating the process of manual tuning of critical parameters in these conventional solutions. The theoretical support is provided by Mercer's Theorem, which proves that any semi-positive function is a valid candidate. Therefore, by using DNN, it can achieve a learnable and faster realization of the

kernel function where the explicit form and the coefficients of the nonlinear mapping function can be obtained [129, 130]. Nevertheless, it is worth noting that the learned functions based on DNN vary greatly during each training epoch.

Second, efforts are dedicated to the study of the characteristics of the AI models to derive human-understandable logic. This can be straightforwardly achieved for knowledge-based AI systems. Otherwise, XAI needs to answer questions including (i) which factors/data contributes most to the results and (ii) how to interpret results using linguistic expressions. As for question (i), just-in-time learning, locally weighted projection regression, and K-nearest neighboring-based methods are typical examples that link decision-making directly to the related data/local models [131–135]. As for question (ii), temporal logic-based methods are promising solutions [136]. Furthermore, as for the "black-box/grey-box" models, the work is to shed light on the characteristics of the sub-blocks of the deep structures [137]. Ablation experiments analyze the necessity of each module and the influence on the overall performance [138]. In addition to the theoretical perspective, explainability also needs to be discussed along with the specific application background. Some tasks require robust performance against data changes, e.g., having good tolerance to disturbance, noise, missing values, and outliers, while the others require high sensitivity to abnormal changes [139–141].

If confusing advice is provided by an AI and unfortunately it cannot be fully interpreted, shall we trust it? For many decision-making processes, there is no right or wrong, but a series of strategic moves are favorable in the long run. However, it is very subtle what factors are taken into consideration and how much weight is assigned to each factor. For instance, for a dying patient, shall we take every chance to extend his/her life regardless of the suffering or turn to terminal care. In addition to facts and targets, the implicit and uncertain factors may play a decisive role. The point is that we may always have a question mark in our hearts when dealing with human-involved issues because machines can hardly cognize and have emotion as we do. Even though the decisions can be interpreted by following the second XAI research route, for such cases, AI systems can be only used as a supervisory tool. "All models are wrong, but some are useful." In this sense, to be compatible with basic laws and regulations, we may trust the AI behaviors, but the responsibility shall always be linked to a person.

4.2 Artificial general intelligence (AGI)

ANI systems are characterized by a limited range of well-defined intelligence-related functionalities. The design is therefore also highly specialized. In such context, the advancement in the ANI techniques is accompanied by the awareness of the significance in the correct formulation (well-posedness) of the problem and proper definition of targets ("be careful of what we wish for"). Nevertheless, it will not always be like this. The relationship between AI and humans is likely to change in the long term. As shown in Fig. 3, at stage I of AI, i.e., ANI (the current era), humans create AI systems and use them as tools. At stage II of AI, i.e., Artificial General Intelligence (AGI), humans interact with AI systems and treat them as omnipotent helpers. At stage III, named Artificial Super-Intelligence (ASI), AI may be able to create better AI systems than humans, in which sense they may be considered a novel form of life. The coming of stage III is also considered as a singularity in the development of AI (see Sect. 4.3).

As pointed out by Professor Stuart Russell, the field has operated for over 50 years on one simple assumption: the more intelligent, the better [142]. Yet, unconstrained success brings huge risks and huge benefits. Beyond the degree of intelligence that AI systems can reach, the discussion of which belongs to the technical dimension in the development of AI, more attention is needed on the socioeconomic factors. In fact, new challenges are already here today, such as trust-worthiness, privacy-preserving, cyber-security, social equality, policy and law, education, and employment, and so forth.

In our society, each occupation corresponds to a group of people and many more related to them. Today, the skills required by the train or the subway operators have changed. In the past, a considerable amount of time was devoted to training on how to manipulate vehicles, whereas nowadays the focus is on emergency events when human decisions override computer instructions. Likewise, in the aviation industry, it is a common practice for pilots to take over control during take-offs and landings because there are so many uncertainties that require human judgment. Not doing so may lead to not only safety risks but also critical ethical and law issues in case of accidents. Such risks are currently unacceptable because AI has not yet gained the trust that it can act more rationally and judiciously under all circumstances. It is also important to be able to know on what factors a decision was based upon so that a correct interpretation of it can be made. Such issues have formed a subtopic of research, called Explainable AI (XAI). The goal is to understand how AI functions, such as what factors and logic are adopted when outputting a decision. This would contribute significantly to the establishment of a formal science (compared against heuristic learning) that would enable a methodology for systematic analysis and theoretically guaranteed performance. The subject has still not matured enough in policy and

Fig. 3 Looking to the long-term future

law aspects to enable us to reach a justifiable decision on who should take the responsibility for errors and failures related to AI systems [143].

4.3 Singularity and artificial super intelligence (ASI)

It was about 15 years ago, in 2005, that a non-fiction book, titled "Singularity is Near" was published and soon became very popular [144]. Its author: the futurist Ray Kurzweil, defined technological singularity as "a time when machines will have and be able to make other machines with intelligence comparable to human beings." His prediction at that time was that this would happen by 2045, i.e., about 25 years later. In more general terms, singularity is a status that separates quantitative and qualitative changes in the process of AI development. In the context of this paper, singularity refers to the time when AI has the ability to design AGI systems better than humans. In addition to its supercomputing ability, AI outperforms mankind in learning, prediction, creativity, productivity, decision-making, organization, management, as well as survival. If the narrow understanding of learning is limited to the algorithmic/software level (which is the case for today), we need to note that post-singularity AI systems may have the ability to reprogram the firmware and even to design hardware by itself, starting from scratch. This is a domain (rebuild our brains) that humans can hardly reach at present and it is uncertain whether human can achieve it in the future. Besides, post-singularity AI can be smart enough to guarantee the supply of resources while figuring out a way for sustainable development. From then onwards, the evolutionary process of AI is dominated by itself, towards an unpredictable direction with an unimaginable speed.

Singularity is a subtle borderline, crossing which the AI systems become out of human control. It is well acknowledged that beings with the highest level of intelligence live at the top of the food chain. Nevertheless, although there is no evidence that silicon-based being can exist at all, there is neither proof that silicon cannot develop novel forms of intelligence as carbon-based creatures. With generation-by-generation iterations, there are chances that the capability of AI exceeds all humans in all aspects, which constitutes the definition of artificial super intelligence (ASI). ASI learn from experience, adapt to new situations, handle abstract concepts, and use knowledge to manipulate the surrounding environment. It can do everything (and does better) that humans can do—logic reasoning, thinking, creativity, strategy, learning a skill on its own, identifying knowledge resources on its own. Today, there is hardly any consensus reached whether such singularity exists and whether AGI is in the foreseeable future. Nevertheless, if there would be evidence of such, what is needed to prepare for it? We need to consider the means to steer the wheel in such a way that the probability of stepping into an unexpected and unfortunate future is minimized.

5 Concluding remarks

Profound technological changes have taken place around us during the last 2 decades, supported by disruptive advances both on the software and the hardware sides. The dominant feature of the changes is the integration of the virtual world with the physical world through the Internet of Things (IoT). The most recent development is the radical paradigm shift from "connected things" to "connected intelligence". Table 1 tries to emphasize the accelerating nature of technological developments by listing the important milestones in AI during the last 25 years.

Throughout the history of scientific and technological development, the emergence of any scientific and technological revolution is not only reflected in technology, but also changes in human social structure, moral constraints, laws, and education. In this article, at the critical point of AI technological revolution, a brief history of AI development is presented, the advancement and the state-of-the-art of the AI discipline are reviewed, and a perspective on the future of AI is offered. AI integrates with a variety of disciplines, plays an important role in the history and might impact the future of humanity. Today, AI is more available now than ever. Products and services driven by AI technologies have widely emerged around us and have led to a profound impact on daily life. Predictably, AI will act as the one of the backbones of new technological revolutions. Nevertheless, several troughs may once again take place if the technologies/applications cannot be grounded swiftly according to the expectations from the dominant stakeholders (inventors rather than the general public). The article is intended to help the readers to better understand AI, accept AI, and think more deeply about AI. With the long-lasting curiosity and wisdom, we are likely to witness the beauty of a world with the novel AI ecology. As Ophelia stated in Shakespeare's play Hamlet, "we know what we are, but we know not what we may become."

Declarations

Competing interests The authors declare no competing interests.

References

1. Kaynak O. The golden age of Artificial Intelligence. Discov Artif Intell. 2021;1:1. https://doi.org/10.1007/s44163-021-00009-x.
2. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. Stanford: AI Magazine; 1995.

3. Market research report, markets and markets, report code: TC 7894, May 2021. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html.
4. Hinton G, Deng L, Yu D, Dahl GE, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag. 2012;29(6):82–97.
5. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In: IEEE international conference on acoustics, speech and signal processing. IEEE: Piscataway; 2013. p. 6645–9.
6. Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE: Piscataway; 2017. p. 1251–8.
7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE: Piscataway; 2016. p. 770–8.
8. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. Comput Lang. 2018. https://doi.org/10.48550/arXiv.1810.04805.
9. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. Canada: NeurIPS; 2014. p. 3104–12.
10. Gualtieri L, Rauch E, Vidoni R. Emerging research fields in safety and ergonomics in industrial collaborative robotics: a systematic literature review. Robot Comput Integr Manuf. 2021;67:101998.
11. Palagi S, Fischer P. Bioinspired microrobots. Nat Rev Mater. 2018. https://doi.org/10.1038/s41578-018-0016-9.
12. Menouar H, Guvenc I, Akkaya K, et al. UAV-enabled intelligent transportation systems for the smart city: applications and challenges. IEEE Commun Mag. 2017;55(3):22–8.
13. Grigorescu S, Trasnea B, et al. GA survey of deep learning techniques for autonomous driving. J Field Robotics. 2019;37(3):362–86.
14. Lu R, Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Appl Energy. 2019;236:937–49.
15. Raza MQ, Khosravi A. A review on artificial intelligence-based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev. 2015;50:1352–72.
16. Yu KH, Beam AL, Kohane S. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719–31.
17. Me A, Mc A, Fkhg A, et al. Harnessing artificial intelligence for the next generation of 3D printed medicines. Adv Drug Deliv Rev. 2021. https://doi.org/10.1016/j.addr.2021.05.015.
18. Jaksic M, Marinc M. Relationship banking and information technology: the role of artificial intelligence and FinTech. Risk Manag. 2017;21(1):1–18.
19. Zhang L, Zhang B. A geometrical representation of McCulloch-Pitts neural model and its applications. IEEE Transact Neural Networks. 1999;10(4):925–9.
20. Kuriscak E, Marsalek P, Stroffek J, Toth PG. Biological context of Hebb learning in artificial neural networks, a review. Neurocomputing. 2015;152:27–35.
21. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev. 2000;44(1.2):206–26.
22. Freund Y, Schapire RE. Large margin classification using the perceptron algorithm. Mach Learn. 2000;37(3):277–96.
23. Hay JC, Martin FC, Wightman CW. The mark-1 perceptron-design and performance. Proc Inst Radio Eng. 1960;48(3):398–9.
24. Newell A, Shaw JC, Simon HA. Report on a general problem-solving program. IFIP Congr. 1959;256:64.
25. Lindsay RK, Buchanan BG, Feigenbaum EA, Lederberg J. DENDRAL: a case study of the first expert system for scientific hypothesis formation. Artif Intell. 1993;61(2):209–61.
26. Oke SA. A literature review on artificial intelligence. Int J Inf Manage Sci. 2008;19(4):535–70.
27. Hendler J. Avoiding another AI winter. IEEE Intell Syst. 2008;23(2):2–4.
28. Jones KS. Natural language processing: a historical review. Current issues in computational linguistics: in honour of Don Walker. New York: Oxford University Press; 1994. p. 3–16.
29. Marvin M, Seymour AP. Perceptrons. Cambridge: MIT Press; 1969.
30. Lighthill J. Artificial intelligence: a general survey. In: Artificial Intelligence: a paper symposium. Science Research Council: Brooklyn; 1973.
31. McCarthy J. Some expert systems need common sense. Ann NY Acad Sci. 1984. https://doi.org/10.1111/j.1749-6632.1984.tb16516.x.
32. Edwards PN. The closed world: computers and the politics of discourse in Cold War America. Cambridge: MIT Press; 1996.
33. McCorduck P, Cfe C. Machines who think: a personal inquiry into the history and prospects of artificial intelligence. Boca Raton: CRC Press; 2004.
34. Mohri M, Afshin R, Ameet T. Foundations of machine learning. Cambridge: MIT Press; 2018.
35. Zhou Z. Machine learning (In Chinese). Beijing: Tsinghua University Press; 2016.
36. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. Nature. 1986;323:533–6.
37. Zhang D, Han X, Deng C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE J Power Energy Syst. 2018;4(3):362–70.
38. Givigi S, Jardine PT. Machine learning for data-driven control of robots. IEEE Potentials. 2018;37(4):35–9.
39. Patrício DI, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput Electron Agric. 2018;153:69–81.
40. Mundher Z, El-Shafie YA, Jaafar O, et al. Artificial intelligence-based models for stream-flow forecasting: 2000–2015. J Hydrol. 2015;530:829–44.
41. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436.
42. Yin S, Kaynak O. Big data for modern industry: challenges and trends. Proc IEEE. 2015;103(2):143–6.
43. Yin S, Li X, Gao H, Kaynak O. Data-based techniques focused on modern industry: an overview. IEEE Trans Industr Electron. 2014;62(1):657–67.
44. Shen Y, Harris N, Skirlo S, et al. Deep learning with coherent nanophotonic circuits. Nat Photon. 2017;11:441–6.

45. Zaman KS, Reaz MBI, Ali SHM, et al. Custom hardware architectures for deep learning on portable devices: a review. IEEE Transact Neural Networks Learn Syst. 2021. https://doi.org/10.1109/TNNLS.2021.3082304.
46. Li X, Jiang Y, Rodriguez-Andina JJ, Luo H, Yin S, Kaynak O. When medical images meet generative adversarial network: recent development and research opportunities. Discov Artif Intell. 2021. https://doi.org/10.1007/s44163-021-00006-0.
47. Meidan Y, Lerner B, Rabinowitz G, Hassoun M. Cycle-time key factor identification and prediction in semiconductor manufacturing using machine learning and data mining. IEEE Trans Semicond Manuf. 2011;24(2):237–48.
48. Sun C, et al. Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. IEEE Trans Ind Informat. 2019;15(4):2416–25.
49. Bau D, Zhu J-Y, Strobelt H, Lapedriza A, Zhou B, Torralba A. Understanding the role of individual units in a deep neural network. Proc Nat Acad Sci USA. 2020;117(48):30071–8.
50. Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the digital twin: a systematic literature review. CIRP J Manuf Sci Technol. 2019;544:1–17.
51. Yeh Y-R, Lin T-C, Chung Y-Y, Wang Y-CF. A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection. IEEE Trans Multimed. 2012. https://doi.org/10.1109/TMM.2012.2188783.
52. Krizhevsky A, Sutskever I, Hinton GE. "ImageNet classification with deep convolutional neural networks. Canada: NIPS; 2012. p. 1106–14.
53. Moravcik M, Schmid M, et al. DeepStack: expert-level artificial intelligence in heads-up no-limit poker. Science. 2017;356(6337):508–13.
54. Johnson N, et al. Abrupt rise of new machine ecology beyond human response time. Sci Rep. 2013;3:2627.
55. Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. Nature. 2017;550:354–9.
56. Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. Nature. 2019;575:607–17.
57. Skripcak T, Tanuska P, Konrad U, Schmeisser N. Toward nonconventional human–machine interfaces for supervisory plant process monitoring. IEEE Trans Human-Mach Syst. 2013;43(5):437–50.
58. Jiang Y, Yin S, Kaynak O. Performance supervised plant-wide process monitoring in industry 4.0: a roadmap. IEEE Open J Ind Electron Soc. 2021;2:21–35.
59. Javed K, Gouriveau R, Zerhouni N. A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering. IEEE Trans Cybern. 2015;45(12):2626–39.
60. Aizpurua JI, McArthur SDJ, Stewart BG, Lambert B, Cross JG, Catterson VM. Adaptive power transformer lifetime predictions through machine learning and uncertainty modeling in nuclear power plants. IEEE Trans Ind Electron. 2019;66(6):4726–37.
61. Lee H, Kim Y, Kim CO. A deep learning model for robust wafer fault monitoring with sensor measurement noise. IEEE Trans Semicond Manuf. 2017;30(1):23–31.
62. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R. Intriguing properties of neural networks. https://arxiv.org/abs/1312.6199. 2013.
63. Zhang T, Su G, Qing C, Xu X, Cai B, Xing X. Hierarchical lifelong learning by sharing representations and integrating hypothesis. IEEE Transact Syst Man Cybern Syst. 2021;51(2):1004–14.
64. Matignon L, Laurent GJ, Le Fort-Piat N. Hysteretic Q-learning: an algorithm for decentralized reinforcement learning in cooperative multi-agent teams. IEEE RSJ Int Conf Intell Robot Syst. 2007;2007:64–9.
65. Banerjee R, Pal SK. Data-structures for multisensory information processing in an embodied machine-mind. IEEE Transact Cogn Dev Syst. 2018;10(3):726–37.
66. Yang X, Lo D, Xia X, Zhang Y, Sun J. Deep learning for just-in-time defect prediction. IEEE Int Conf Softw Qual Reliab Secur. 2015. https://doi.org/10.1109/QRS.2015.14.
67. He Y, Mendis GJ, Wei J. Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism. IEEE Transact Smart Grid. 2017;8(5):2505–16.
68. Stone P, Brooks R, Brynjolfsson E, et al. Artificial Intelligence and Life in 2030. One hundred year study on artificial intelligence: report of the 2015–2016 study panel. Stanford: Stanford University; 2016.
69. Khan M, Waseem H, et al. A novel image encryption scheme based on quantum dynamical spinning and rotations. PloS ONE. 2018;13(11):e0206460.
70. Ning H, Liu H. Cyber-physical-social-thinking space based science and technology framework for the Internet of Things. Sci China Inf Sci. 2015;58:1–19.
71. Jarrahi MH. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus Horiz. 2018;61(4):577–86.
72. Chuquicusma MJ, Hussein S, Burt J, Bagci U. How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis. Kolkata: IEEE International Symposium on Biomedical Imaging; 2018. p. 240–4.
73. Liu S, Li L, Tang J, Shuang W, Gaudiot J-L. Creating autonomous vehicle systems, second edition. Synth Lect Comput Sci. 2020;8(2):216.
74. Li Y, Zhang D, Lan Z, Tan K. Context-aware advertisement recommendation for high-speed social news feeding. In: 2016 IEEE 32nd International Conference on Data Engineering (ICDE). Piscataway: IEEE; 2016. p. 505–16.
75. Shum H, Ho PH, et al. Real-time posture reconstruction for microsoft kinect. IEEE Transact Cybern. 2013;43(5):1357–69.
76. Stanley KO, Bryant BD, Miikkulainen R. Real-time neuroevolution in the NERO video game. IEEE Trans Evol Comput. 2005;9(6):653–68.
77. Wang Y, et al. On the philosophical, cognitive and mathematical foundations of symbiotic autonomous systems. Phil Trans R Soc A. 2021;379:20200362. https://doi.org/10.1098/rsta.2020.0362.
78. Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. Nature. 2019;568:477–86.
79. Zhao Y, Li T, Zhang X, Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future. Renew Sustain Energy Rev. 2019;109:85–101.
80. Abbass H, Petraki E, Hussein A, McCall F, Elsawah S. A model of symbiomemesis: machine education and communication as pillars for human-autonomy symbiosis. Phil Trans R Soc A. 2021;379:20200364. https://doi.org/10.1098/rsta.2020.0364.
81. James K, et al. Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci. 2017;114(13):3521–6.
82. Flammini F. Digital twins as run-time predictive models for the resilience of cyber-physical systems: a conceptual framework. Phil Trans R Soc A. 2021;379:20200369.

83. Jiang Y, Yin S, Li K, Luo H, Kaynak O. Industrial applications of digital twins. Phil Trans R Soc A. 2021;379:20200360. https://doi.org/10.1098/rsta.2020.0360.

84. Dittrich BSM, Clausmeyer T, et al. Shifting value stream patterns along the product lifecycle with digital twins. Ethiopia: 7th CIRP Global Web Conference; 2019. p. 3–11.

85. Bragazzi NL, Dai H, Damiani G, et al. How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health. 2020;17(9):3176.

86. To A, Mt B, Eay C, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020. https://doi.org/10.1016/j.compbiomed.2020.103792.

87. Shi F, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE Rev Biomed Eng. 2021;14:4–15.

88. Rosen R, Fischer J, Boschert S. Next generation digital twin: an ecosystem for mechatronic systems? IFAC PapersOnLine. 2019;52(15):265–70.

89. Schalk G, Leuthardt EC. Brain-computer interfaces using electrocorticographic signals. IEEE Rev Biomed Eng. 2011;4:140–54.

90. Wang Y, et al. Brain-inspired systems: a transdisciplinary exploration on cognitive cybernetics, humanity, and systems science toward autonomous artificial intelligence. IEEE Syst Man Cybern Mag. 2020;6(1):6–13.

91. Chen X, Wang Y, Nakanishi M, et al. High-speed spelling with a noninvasive brain-computer interface. United States: Proceedings of the National Academy of Sciences of the United States of America; 2015.

92. Nagel S, Spüler M. Asynchronous non-invasive high-speed BCI speller with robust non-control state detection. Sci Rep. 2019. https://doi.org/10.1038/s41598-019-44645-x.

93. Muller-Putz GR, Pfurtscheller G. Control of an electrical prosthesis with an SSVEP-based BCI. IEEE Trans Biomed Eng. 2008;55(1):361–4.

94. Liu X, Zhang M, Subei B, Richardson AG, Lucas TH, Van der Spiegel J. The PennBMBI: design of a general-purpose wireless brain-machine-brain interface system. IEEE Trans Biomed Circuits Syst. 2015;9(2):248–58.

95. Maharbiz MM, Muller R, Alon E, Rabaey JM, Carmena JM. Reliable next-generation cortical interfaces for chronic brain-machine interfaces and neuroscience. Proc IEEE. 2017;105(1):73–82.

96. Yuan H, He B. Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. IEEE Trans Biomed Eng. 2014;61(5):1425–35.

97. Bashivan P, Kar K, Dicarlo JJ. Neural population control via deep image synthesis. Science. 2019. https://doi.org/10.1126/science.aav9436.

98. Nunes DS, Zhang P, Sá Silva J. A survey on human-in-the-loop applications towards an internet of all. IEEE Commun Surv Tutorials. 2015;17(2):944–65.

99. Munir S, Stankovic JA, Liang C-JM, Lin S. Reducing energy waste for computers by human-in-the-loop control. IEEE Trans Emerg Topics Comput. 2014;2(4):448–60.

100. Thatte N, Duan H, Geyer H. A sample- efficient black-box optimizer to train policies for human-in-the-loop systems with user preferences. IEEE Rob Autom Lett. 2017;2(2):993–1000.

101. Chiang H-H, Wu S-J, Perng J-W, Wu B-F, Lee T-T. The human-in-the-loop design approach to the longitudinal automation system for an intelligent vehicle. IEEE Trans Syst Man Cybern A Syst Humans. 2010;40(4):708–20.

102. Li Z, Liu J, Huang Z, Peng Y, Pu H, Ding L. Adaptive impedance control of human-robot cooperation using reinforcement learning. IEEE Trans Industr Electron. 2017;64(10):8013–22.

103. Sartoretti G, et al. PRIMAL: pathfinding via reinforcement and imitation multi-agent learning. IEEE Robot Autom Lett. 2019;4(3):2378–85.

104. Seeber I, Bittner E, et al. Machines as teammates: a research agenda on AI in team collaboration. Inf Manag. 2020. https://doi.org/10.1016/j.im.2019.103174.

105. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.

106. Li X, Jiang Y, Yin S. Lightweight attention convolutional neural network for retinal vessel image segmentation. IEEE Trans Industr Inf. 2020;17(3):1958–67.

107. Liu S, Li X, Jiang Y, Luo H, Gao Y, Yin S. Integrated learning approach based on fused segmentation information for skeletal fluorosis diagnosis and severity grading. IEEE Transact Ind Informatics. 2021. https://doi.org/10.1109/TII.2021.3055397.

108. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35(5):1240–51.

109. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJNL, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging. 2016;35(5):1252–61.

110. Liu S, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans Biomed Eng. 2015;62(4):1132–40.

111. Haider F, de la Fuente S, Luz S. An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. IEEE J Sel Top Signal Process. 2020;14(2):272–81.

112. Urban G, et al. Deep learning for drug discovery and cancer research: automated analysis of vascularization images. IEEE ACM Transact Comput Biol Bioinform. 2019;16(3):1029–35.

113. Liu X, Deng RH, Choo KR, Yang Y. Privacy-preserving outsourced support vector machine design for secure drug discovery. IEEE Transact Cloud Comput. 2020;8(2):610–22.

114. Biane C, Delaplace F. Causal reasoning on boolean control networks based on abduction: theory and application to cancer drug discovery. IEEE ACM Transact Comput Biol Bioinform. 2019;16(5):1574–85.

115. Saracco R, Grise K, Martinez T. The winding path towards drug symbiotic autonomous systems. Phil Trans R Soc A. 2021;379:20200361. https://doi.org/10.1098/rsta.2020.0361.

116. Liu S, Li L, Tang J, Wu S, Gaudiot J-L. Creating autonomous vehicle systems. San Rafael: Morgan & Claypool; 2017.

117. Paden B, Čáp M, Yong SZ, Yershov D, Frazzoli E. A survey of motion planning and control techniques for self-driving urban vehicles. IEEE Transact Intell Veh. 2016;1(1):33–55.

118. Rawat DB, Bajracharya C. Vehicular cyber physical systems adaptive connectivity and security. Springer Nat. 2017. https://doi.org/10.1007/978-3-319-44494-9.

119. Yin S, Rodriguez J, Jiang Y. Real-time monitoring and control of industrial cyberphysical systems with integrated plant-wide monitoring and control framework. IEEE Ind Electron Mag. 2019;13(4):38–47.

120. Zhang J, Jiang Y, Li X, et al. An adaptive remaining useful life prediction approach for single battery with unlabelled small sample data and parameter uncertainty. Reliab Eng Syst Saf. 2022. https://doi.org/10.1016/j.ress.2022.108357.

121. Jiang Y, Yin S, Kaynak O. Optimized design of parity relation based residual generator for fault detection: data-driven approaches. IEEE Trans Industr Inf. 2020;17(2):1449–58.

122. Alireza F, Biao H. Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay. J Process Contr. 2017;53:15–25.

123. Nair AM, Fanta A, Haugen FA, Ratnaweera H. Implementing an extended kalman filter for estimating nutrient composition in a sequential batch MBBR pilot plant. Water Sci Technol. 2019;80(2):317–28.

124. Romero A. GPT-3 scared you? Meet Wu Dao 2.0: a monster of 1.75 trillion parameters, towards data science. 2021. https://towardsdatascience.com/gpt-3-scared-you-meet-wu-dao-2-0-a-monster-of-1-75-trillion-parameters-832cd83db484.

125. Samek W, Montavon G, Lapuschkin S, et al. Explaining deep neural networks and beyond: a review of methods and applications. Proc IEEE. 2021;109(3):247–78.

126. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. https://arxiv.org/abs/1702.08608. 2017.

127. Li Z, Tao R, Wang J, Li F, Niu H, Yue M, Li B. Interpreting the latent space of GANs via measuring decoupling. In: IEEE transactions on artificial intelligence. IEEE: Piscataway; 2021.

128. Liu Y, Barr K, Reinitz J. Fully interpretable deep learning model of transcriptional control. Bioinformatics. 2020;36(1):i499–507.

129. Ren Z, Yang X, et al. Learnable faster kernel-PCA for nonlinear fault detection: deep autoencoder-based realization. Mach Learn. 2021. https://doi.org/10.48550/arXiv.2112.04193.

130. Ren Z, Zhang W, Zhang Z. A deep nonnegative matrix factorization approach via autoencoder for nonlinear fault detection. IEEE Trans Industr Inf. 2020;16(8):5042–52.

131. Chen Z, et al. A just-in-time-learning-aided canonical correlation analysis method for multimode process monitoring and fault detection. IEEE Trans Industr Electron. 2021;68(6):5259–70.

132. Wang G, Yin S, Kaynak O. An LWPR-based data-driven fault detection approach for nonlinear process monitoring. IEEE Trans Industr Inf. 2014;10(4):2016–23.

133. Alippi C, Roveri M. Just-in-time adaptive classifiers—part i: detecting nonstationary changes. IEEE Trans Neural Networks. 2008;19(7):1145–53.

134. Yuan X, Ge Z, Huang B, Song Z. A probabilistic just-in-time learning framework for soft sensor development with missing data. IEEE Trans Control Syst Technol. 2017;25(3):1124–32.

135. Vijaykumar S, Dsouza A, Schaal S. Incremental online learning in high dimensions. Neural Comput. 2005;17(12):2602–34.

136. Fortuin V, Hüser M, Locatello F, Strathmann H, Rätsch G. SOM-VAE: interpretable discrete representation learning on time series. Austria: ICLR; 2019.

137. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.

138. Athalye A, Engstrom L, Ilyas A, Kwok K. Synthesizing robust adversarial examples. In: International conference on machine learning. New York: MLR Press; 2018. p. 284–93.

139. Li X, Jiang Y, Liu C, Liu S, Luo H, Yin S. Playing against deep neural network-based object detectors: a novel bidirectional adversarial attack approach. IEEE Trans Artif Intell. 2021. https://doi.org/10.1109/TAI.2021.3107807.

140. Zhang J, Jiang Y, et al. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. Reliab Eng Syst Saf. 2022;221(108297):1–10.

141. Chen Z, Guo R, et al. A data-driven health monitoring method using multiobjective optimization and stacked autoencoder based health indicator. IEEE Trans Industr Inf. 2021;17(9):6379–89.

142. Russell SJ, Norvig P. artificial intelligence: a modern approach. 3rd ed. London: Pearson Education; 2010.

143. McDermid JA, Jia Y, Porter Z, Habl I. Artificial intelligence explainability: the technical and ethical dimensions. Phil Trans R Soc A. 2021;379:20200363. https://doi.org/10.1098/rsta.2020.0363.

144. Kurzweil R. The singularity is near: when humans transcend biology. London: Penguin; 2005.