Pathways to Artificial General Intelligence: A Brief Overview of Developments and Ethical Issues via Artificial Intelligence, Machine Learning, Deep Learning, and Data Science

Mohammadreza Iman, Hamid R. Arabnia, and Robert Maribe Branchinst

1 Introduction

It is nearly impossible to move around modern society without encountering a device or application powered by artificial intelligence (AI). Weather forecasts, traffic signals, airplanes, factory lines, home appliances, and mobile applications are just a few examples of areas likely to encounter elements controlled by AI. Yet, there is even more happening under the surface with AI managing countless applications including internet traffic, gene-related research, and medical image and history analyzation. For most people today, deep learning, machine learning, and AI are all terms for which they are at least familiar.

Another body of work that most people will have heard of is data science and data analytics. Technological advances over the past few decades have transferred the possibility of generating, storing, sharing, and analyzing data to nearly everyone. With data now being a true commodity, some have said that data is the new oil or gold. For example, retailers are now able to gather information about their sales as well as their customers habits and preferences to greatly benefit both parties. Retailers can then use this information to intelligently predict customer shopping habits during other times of the year as well as control their supplies based on projected demands, thus, not wasting time and money on unnecessary storage or

M. Iman (✉) · H. R. Arabnia
Department of Computer Science, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, USA
e-mail: hra@uga.edu

R. M. Branchinst
Learning, Design, and Technology Mary Frances Early College of Education, University of Georgia, Athens, GA, USA
e-mail: rbranch@uga.edu

H. R. Arabnia et al. (eds.), *Advances in Artificial Intelligence and Applied Cognitive Computing*, Transactions on Computational Science and Computational Intelligence, https://doi.org/10.1007/978-3-030-70296-0_6

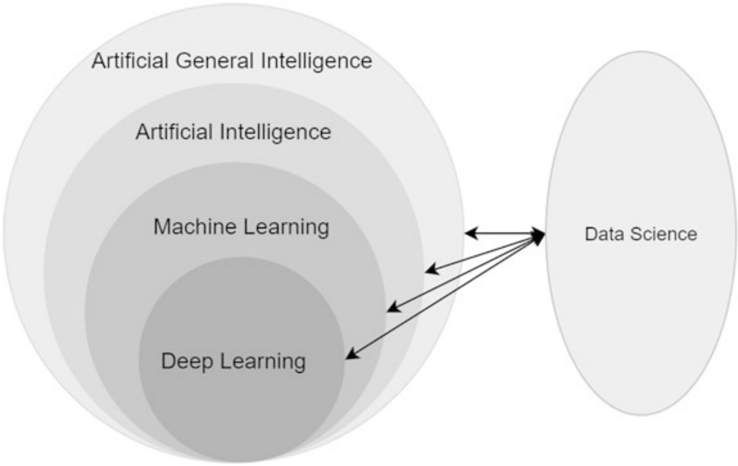

Fig. 1 Domain hierarchy

creating shortages. This is just one example of the great advances made possible by data science and its varying applications. With advances such as autonomous driving now available, there is no telling where data science and AI might take us.

In this article, we briefly review the history of these developments in artificial general intelligence, artificial intelligence, machine learning, deep learning, and data science (see Fig. 1), tracing the history from the first mechanical computer in 1850 to the current state of deep learning in 2020. We overview the many evolutions in AI and discuss possible future directions as well as some of the ethical dilemmas posed by such advances. Ultimately, our goal is to overview these processes for a lay audience who may not have intimate knowledge of AI and data science at large.

2 Artificial Intelligence

The first mechanical computer was invented in the 1850s by Charles Babbage [1]. In 1950, Alan Turing, renown for advancing the general-purpose programmable computer, asked the big question for the first time: "Can machines think?" [2]. Alan Turing proposed an operational test for machine intelligence. A machine "passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a human or a machine." [2].

In 1956, the term "Artificial Intelligence" (AI) was used for the first time in a proposal for a summer research workshop at Dartmouth College in New Hampshire. The goal of AI was to "[make] a machine behave in ways that would be called intelligent if a human were so behaving" [3]. The aim of the workshop was to develop AI such that it might pass the Turing test. To pass the Turing test the AI needs to [1]:

- Understand speech; *natural language processing (NLP)*
- Store the information and data; *knowledge representation*
- Use the stored data to draw conclusions; *automated reasoning/decision-making*
- Detect new patterns and adapt to new circumstances; *machine learning (ML)*

To fully pass the Turing test, two additional capabilities are needed [4, 5]:

- Extract knowledge and/or comprehension from images or videos (e.g., face recognition); *computer vision*
- Mimic human physical behaviors corresponding with the senses to interact with the environment (e.g., touch, motor functioning); *physical interaction*

Overall, the AI could be divided into two categories of **Artificial General Intelligence** (AGI) or strong AI (actual thinking), and narrow or weak AI (simulated thinking) [1, 6]. Some scholars argue that achieving AGI may be decades into the future, and that the emergence of AGI will bring with it an "intelligence explosion" leading to "profound changes in human civilization" [6]. Yet, the field of computer science has already begun to develop the narrower form of AI. In fact, the ability to have devices such as sensors and robots, and intelligent Decision Support Systems (DSS), such as the autocorrect software analyzing the words on this page are the result of already existing forms of AI [6]. These technologies have already evolved beyond what many people could have imagined, and yet the future of AGI has the potential to transform the experience of generations to come in ways we cannot yet predict.

Early (late twentieth century) AI approaches were rule-based and focused on attending to all possible solutions for a specific and identifiable problem [7]. Some board games and various types of robots used in factory lines are just two examples of this type of rule-based AI. Going forward, decision making systems began to advance these types of approaches [8]. Specifically, decision making attended to the fact that real-life problems are rarely contained within such specific and rule-based features. Even board games must contend with players who make unpredictable decisions. Although decision making may follow some of the same patterns of rule-based prediction systems, it began to extend the bounds of these rules by accounting for uncertainty [9]. The Boltzmann machine research line during the 1990s through the early 2000s delivered a well-known example of this type of AI, which utilizes probability and statistics predicting behavior patterns in various settings [10–12]. Sum-Products Networks (SPNs) are another advancement in AI that began to incorporate networks able to compete with deep learning models in many applications by taking into account the probability distributions of features [13].

The boom of digital storage development in the 1990s and 2000s—delivering cloud storage and advanced data collection methods—brought with it a new era of "big data." Big data refers to the vast amount of easily accessible consumer data including images, texts, audio, transactions, and human and environmental sensing data from electronic devices. This surge in available data required new methods for analyzing it, translating to Data Science (DS) and a new chapter in machine learning (ML).

3 Data Science

Data Science is divided into three main areas including collecting, storing, and analyzing data (structured or unstructured) [14]. Data collection methods were advanced through the spread of high-speed internet world wide—including less costly wireless connections—as well as increased variety of cheaper electronic connections and sensors, such as smartwatches, exercise trackers, and cameras [15]. Data storage was advanced through cloud storage, which further influenced big data collection by offering these services at a reduced cost and to an increasing proportion of the population.

Data analysis consists of two major components: preprocessing and processing. Preprocessing refers to various aspects of raw data management including unbalanced data, imputation techniques for missing data, detecting and addressing outliers, and data labeling procedures. Processing refers to extracting information and knowledge from preprocessed data to identify patterns, make predictions, and/or classify data [14]. One of the promising methodological categories for processing big data is Machine Learning (ML), a subdivision of AI.

4 Machine Learning (ML)

Machine learning (ML) is a subdivision of AI that consists of statistics, mathematics, and logical techniques to extract patterns (i.e., information) from a set of training data and apply the inferences to unseen data. Again, these recent advances in ML were made possible by the new era of big data and the vast advancement in computational capacity. Importantly, ML differs from other forms of AI in that it does not require extensive and complicated programming, but rather, has the ability to learn patterns and later apply them. Thus, ML does not need to consider every possible solution (i.e., be deterministic) and can manage noise and uncertainty [16].

Innovation in ML brought with its exponential advancement in earlier techniques—some of them developed before the 1970s—such as Linear Regressions, decision trees, Random forest, K-nearest neighbor (KNN), Support Vector Machine (SVM), Artificial Neural Networks (ANNs). For example, early ML ANN models for autonomous driving [17] and facial recognition [18] were developed in the 1980s but lacked access to the data and computation capacity needed to apply them [16].

Like any method, ML brings with it its own unique techniques and challenges. Common types of ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Each of these is discussed in brief below, followed by some of the challenges associated with ML such as overfitting and dealing with extraneous features.

Supervised learning refers to the use of an ML training data set that has been labeled, typically by humans, and the goal of which is to categorize or label the unseen data [19]. The process of categorization or labeling often occurs through **classification** or **regression** techniques, which has value for making predictions—using regression—such as predicting stock market values or classifying objects in an image, such as identifying tumors in a medical x-ray.

Unsupervised learning refers to categorizing data by analyzing patterns and shared features without utilizing a pre-labeled training dataset [19]. In unsupervised learning, **clustering** is often used to detect patterns and anomalies, such as in grouping customers for marketing strategies or marking emails from unknown sources as "spam."

In addition, a smaller (i.e., limited) labeled ML dataset may be used to improve the categorization of a larger, unlabeled dataset. This is known as ***semi-supervised learning***. Semi-supervised learning may be a more cost-effective option of labeling large datasets, in addition to allowing for greater accuracy by limiting human error [19]. For example, speech recognition errors may be reduced by 22% when human-labeled data are combined with machine-labeled data using semi-supervised learning [20].

Reinforcement learning (RL) operates using a reward-based system. Reinforcement learning attempts to select the best possible action that would maximize the final reward (or conversely, minimize the punishment), all while keeping track of these actions to improve the choice-selection of the following round. Thus, it is a trial-and-error process that works through the system's ability to learn improvement strategies and decisions through the success or failure of previous attempts. There are many different types of RL algorithms, each designed to address a specific problem [16]. Examples of RL applications include some types of board games (e.g., chess and Go), robots, and various elements of autonomous driving systems.

Although Machine Learning algorithms demonstrate immense accuracy in identifying training dataset patterns, a common problem in these models is **overfitting** the data [21, 22]. Overfitting occurs when the ML network has been trained using all labeled (i.e., training) data and cannot deal with the noise (i.e., uncertainty) in the unseen data. It also occurs when patterns observed in the limited training data are not accurate of the existing patterns in the larger data. Overfitting may occur when using unbalanced or biased datasets, indicating that the training set does not include all possible samples within the domain[16, 21, 22].

There are several ways of correcting for the risk of overfitting. One of these is to divide the training dataset into two parts: training and **validation** [6, 16]. The size of the validation set will depend on the size of the overall training set, but typically ranges from about 10 to 30% of the full set. The validation set is not used for training purposes, but is instead verified against the final dataset to ensure accuracy. In this procedure, **cross-validation** is used to correct for the risk of selecting a biased validation set [23]. For example, a 10-fold cross-validation procedure would involve

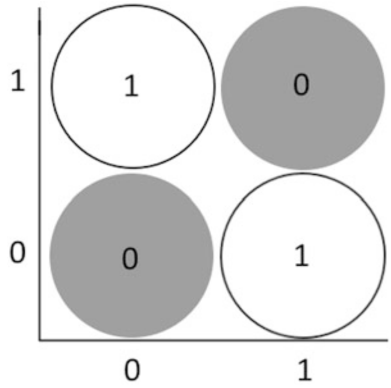

Fig. 2 Not linearly separable pattern, known as XOR problem

dividing a training set into 10 separate sets and then training the ML model 10 times using only nine of those sets each time. The final model would then be validated against the remaining set (1/10th of the original), with the accuracy being equal to the average of the 10 validation runs.

Another challenge that may arise in ML models is the issue of extraneous features, such as the vast number of potentially uncorrelated features present in some big data sets. In many cases, not all of the features present in a dataset will be related to the objective of the ML model and, thus, are not useful. For example, to predict the seasonal sales of an online store, customers' employment status and income may be related to the outcome, but their specific job title may not be. There are several known processes for responding to unrelated features in a dataset including **feature selection, combination, and extraction** [24]. These are performed through techniques like correlation analyzation, principal component analyzation (PCA), and dimensionality reduction techniques. These techniques work mainly by validating the correlation of each feature to the target [24].

Machine Learning techniques and data sets can be categorized into two groups: linear and non-linear. A linear data pattern is the simplest data pattern and can be categorized using a linear function to perform regression or classification. Many algorithms had been developed to fit linear models such as linear regression, logistic regression, classification and regression trees, K-nearest neighbors, and support vector machine [16]. Non-linear functions are those that cannot be classified using linear methods. Like other models of data analysis and management, non-linear data associations may pose additional challenges to ML [16]. The non-linear problem in ML is known as the **XOR (i.e., "exclusive or") problem**, which refers to a mixed pattern of data that cannot be categorized using linear functions, Fig. 2 [25].

Although many algorithms have been developed to manage linear data (as mentioned above), the non-linear nature of many data sets remained a challenge for ML. For example, the decision trees, k-nearest neighbors, and support vector machine mentioned above are functions that can manage some non-linear data problems; yet, they do this imperfectly, and issues remain. Artificial Neural Networks (described in the next section) began to address these issues [16].

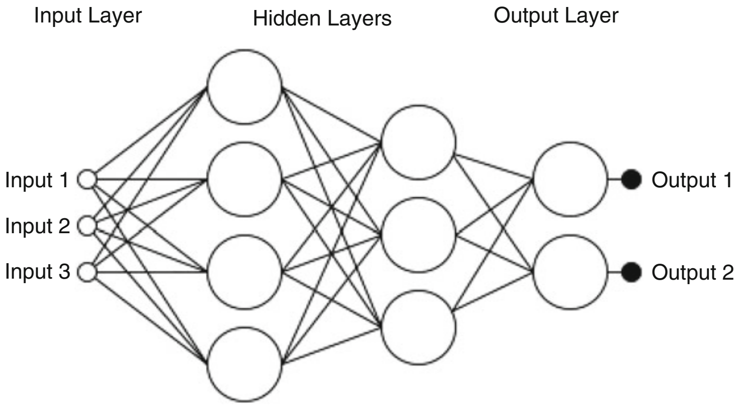

Fig. 3 An example of MLP network architecture

5 Artificial Neural Network

In 1958, the first artificial neuron was introduced—attempting to mimic the neural pathways of the human brain. Named Perceptron, it used a sigmoid function and performed linear functions with great success [26]. To advance this then new technology, several Perceptrons were later clustered into a layer, allowing for linear patterns to be detected through the use of input data connecting into the Perceptron layer. Training happens by feedforwarding the data while backpropagating the labels to tune the weights of each node. Thus, the first artificial neural network (ANN) was born [27]. Perceptron remained at the height of ANN mechanisms until 1969, when rigorous reviews demonstrated its shortcomings—namely, that Perceptron could not address the issue of non-linearity; it had hit a dead-end [28].

By the 1980s, scientists again attempted to address the issue of non-linearity (i.e., the XOR problem) by using hidden layer(s) of Perceptron, known as Multilayer Perceptron (MLP). MLP is a type of ANN consisting of one or more layers of varying nodes—the network architecture (see Fig. 3) [27]. Using an activation function, such as sigmoid, on the front end of the nodes, again combined with backpropagation techniques, allowed for increasingly advanced classification and regression models—including those for non-linear patterns [27]. These advances greatly improved the accuracy of some of the advanced technologies we enjoy today, such as autonomous driving and facial recognition.

The early ANN designs were fully connected, with each node tied to the next, and each connection having a weight. Each node contained an activation function and uses the value of prior nodes multiplied by the weight of the connection to calculate the next node in a recursive loop. The simultaneous backpropagation by means of the training dataset serves to update and fine-tune the node weights and thresholds. Similar to the reinforcement learning process described above, cost/loss functions are used as additional metrics by which to measure the compatibility between the training data (i.e., ground truth) and the network predictions [27].

Despite vast advances in theorizing, the ANNs (MLP) of the 1980s faced several challenges. Specifically, the limited number of available nodes in each MLP layer, combined with the limited number of layers, produced a heavy burden for the computers of the day. In short, advanced theorizing was limited to the computational capacity of the 1980s machines. However, by the year 2000, significant advances were made in computational capacity. These advances, paired with the ability to replace the nodes' sigmoid activation function with more efficient functions such as sign, linear, tanh [29, 30], and more recently ReLU and leaky-ReLU [31], allowed for the creation of a larger network of nodes, including more hidden layers. This led to the creation and advancement of deep neural networks (DNN), also known as deep learning (DL).

6 Deep Learning (DL)

As mentioned, the vast improvements in the computational capacity of the 2000s helped shape the development of deep neural networks (DNN) or deep learning (DL). Another shaping factor in the development of DL was the arrival of big data sets, which offered the opportunity to improve the training process and thus the performance of DNN.

Similar to ANNs, the learning in DNNs occurs through the optimization of the weights throughout the entire network. One of the well-known algorithms for handling this type of optimization problem is Stochastic Gradient Descent (SGD) [32]. There are several other methods based on the SGD algorithm such as Momentum, Nesterov Momentum, and Adam [32]. Each of these methods works by tracing the error surface of the error calculation function (known as loss function) with the goal of finding the global minima, as shown in Fig. 4 [32, 33]. The loss function is based on the adjustments of the weights of each of the connections in the network [32].

Other parameters that need to be taken into consideration in order to maximize DNNs' accuracy include data preprocessing, hyperparameter adjusting such as learning rate adjustments, weight initialization, initializing biases, and batch normalization [34, 35].

Several modifications of DNNs have vastly improved the implementation of these models. The modifications aim to reduce the models' generalization error by regularizing the weights. There are several methods to do such regularizations including considering the noise robustness, stop learning point (i.e., early stopping), parameter sharing, and dropout [34, 35]. The following section overviews some of the main events and advancements in determining the current state of DLs.

In 2007, Fei Fei Li and colleagues introduced ImageNet, the largest database of labeled images with over 14 million images categorized into nearly 22,000 indexed synsets (categories) as of 2020 [image-net.org]. These images can be used

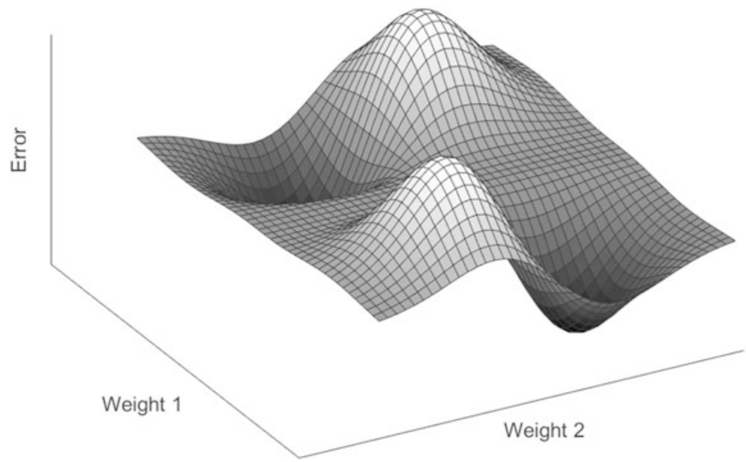

Fig. 4 An example of error surface in an ANN/DNN

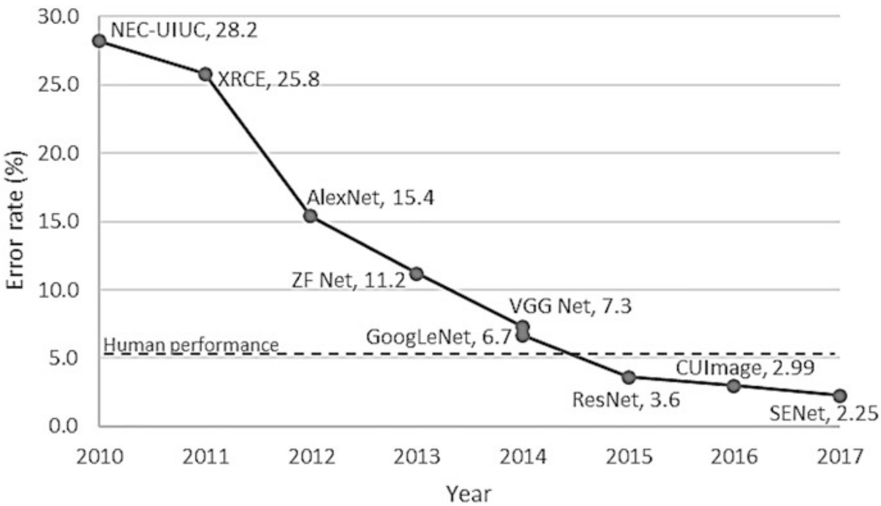

Fig. 5 ILSVRC winners

for technologies such as object location, detection, and classification in videos and other image-related media [image-net.org]. Since 2010, ImageNet has led an annual challenge—the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The challenge brings bright minds from around the globe together to explore new ideas in DL by allowing them to use a large collection of image-data they would not otherwise have access to. This challenge has brought great success in minimizing error as demonstrated in Fig. 5. Remarkably, the classification error of 28% in 2010 was reduced to less than 3% by 2017 [36].

By 2011, the Convolutional Neural Network (CNN) was beginning to grow in popularity. CNN was able to outperform humans in recognition of traffic signs with an accuracy of 99.46% (compared to humans at 99.22%) [37]. First introduced in 1997, CNN was inspired by the visual cortex of animals and attempted to regularize input data to find hierarchical patterns within image data—called self-organized map [38]. By 2020, nearly all DLs utilized CNN layer(s) for visual-related tasks. Going back to the ILSVRC, CNN was utilized by the majority of champions, but it has been used for many other applications as well [36]. Interesting research conducted in 2015 demonstrates how a DNN network functions using CNN layers (i.e., https://youtu.be/AgkfIQ4IGaM) [39].

Another network architecture from the year 1997, Long Short-Term Memory (LSTM) [40], also saw vast improvements in the 2010 decade [41]. LSTM, also known as Recurrent Neural Network (RNN), is a modified neural network that utilized feedback connections. LSTM allowed for the subsequent advancements in DL such as speech and handwriting recognition applications, as well as anomaly detection in data series (e.g., network traffic) [41].

In 2014, Ian Goodfellow and colleagues invented the Generative Adversarial Network (GAN) [42]. GAN is comprised of two neural networks competing against each other; the first is a generative network which generates new data while the second is a discriminative network that evaluates the generated data. This advanced network can generate new data based on the characteristics of the inputted training data. For example, if the training data were to be of a human face, the network would generate a new face, looking entirely human but never having previously existed [43]. A vast amount of applications benefit from GANs, such as imaginary fashion models and scientific simulations [43]. Despite their many advances, GANs raise some concerns, specifically regarding the production of falsified voice or video records [44].

By 2016, Google announced the Tensor Processing Unit (TPU), followed by the Google TensorFlow framework of open source libraries [45]. TensorFlow touts well-tailored hardware and software to be used for neural network computations and applications [45]. Using this technology, Google's DeepMind AlphaGo defeated the Go champion in 2016 by combining DL and RL in a new mechanism named Deep Reinforcement Learning [46].

Another contemporary topic in ML that was also initiated in the 1990s is transfer learning or domain adaptation, first published by Lorien Pratt [47]. Transfer learning works by using knowledge garnered through the ML model during the training phase and literally transferring the learning to another task in a similar domain [48]. For example, a DL model trained to classify flowers can be used to also classify leaves by modifying the trained model using a transfer learning technique (e.g., fine-tuning some of the layers). Since 2014, transfer learning has been used to adapt deep learning models, such as in domains like medical imaging. This is known as deep transfer learning and has been used to reduce the often-long training time as well as to handle the small training samples of some deep learning [48]. Progressive learning introduced by Google's DeepMind Project in 2016 is another specific type of deep transfer learning that is attempting to build on previous, related knowledge, similar to human learning capabilities [49].

In summary, this overview summarized a few notable types of DL that are on the rise. It is important to note that the aforementioned advancements in DL are vast topics in and of themselves, each carrying with them a research line with hundreds or even thousands of relevant articles that could not be overviewed here. In addition, there are many more DL advances not discussed here, including the autoencoders used for image segmentation models such as U-Net and new types of data compressors [50], among many others.

7 Discussion

As mentioned, the era of big data, spurred by drastic advancements in computational capacity, has brought a new chapter to machine learning (ML) and artificial intelligence (AI) since the turn of the century. In the past decade alone, the movement toward artificial general intelligence (AGI) has grown exponentially, and there is no telling where it might take us.

A common example of the great innovations of AGI is IBM's Watson, first introduced in 2011. Watson is a natural language processing (NLP) platform, whose architecture benefits from a variety of developments in AI and ML [51]. In 2011, Watson defeated the champions of the popular quiz show Jeopardy—a feat spurred by its ability to "process 500 gigabytes, the equivalent of a million books, per second" [52].

The use of even narrow AIs to mimic human cognition is opening the pathway to AGI, and AGI is a force that future humans will have to contend with. The competition is likely to be intense given that computer programs do not suffer from fatigue, boredom, or other common human ailments—and impediments to work and/or output. For example, Google's well-known program, AlphaGo, managed to train itself in one night to rise from an amateur to a champion player by the next morning [46].

There are countless other examples of the ways in which AGI is looming closer. Present-day Artificial Neural Networks (ANN) are already a simple mimic of human brain cells, and Convolutional Neural Networks (CNN) mime the human visual cortex. Generative Adversarial Networks (GAN) work like the human imagination—generating new data from observed data—which can be used for better understanding facts by imaging-related data. All this is done without needing to access all or even a vast amount of data. Long-Short Term Memory (LSTM) works similarly to the human memory and is able to solve problems related to sequential data analyzing. Transfer learning, followed by progressive learning, is attempting to mimic human skill-learning abilities, a task that is endless for human beings. However, human beings must rely on previous knowledge and skills—oftentimes garnered over a lifetime—that AI programs can learn in a matter of hours. All of this evidence suggests that AGI is close to becoming reality, and the implications of this have yet to be explored.

8 Ethics

With any scientific advance—particularly one that travels so quickly—ethical issues and considerations are unavoidable. Recent years have seen an increase in considerations of the potential ethical pitfalls of AI and the use of personal data raised by data scientists and other scholars (including social scientists, historians, and others) [53–55]. Unfortunately, the ethical consequences of many advances are difficult to assess in real-time. For example, although it is relatively obvious to see the issues with falsifying evidence through GANs [44], the impact of wrongly classifying a disease through X-ray images is more difficult to project, not to mention the social implications of these advanced changes. This section discusses some primary areas of concern in the ethics debate surrounding AI and offers some additional points for consideration.

One primary area of ethics currently involves **data privacy** surrounding sensitive data and personal information such as credit card transactions and medical data. Issues related to data privacy are complicated by the need for access to personal information in order to move many fields forward. For example, not using medical information means that patients miss out on the opportunity to have their diagnoses made by more accurate AI programs. On the other hand, there are also consequences to this data being made available. Doctors risk being sued if later AI advances point to something they missed, and patients are at risk of having their private information shared elsewhere. Like most ethical issues, it is imperative to consider both sides of this debate and to seek solutions that maximize benefits while limiting risks. Data anonymization mechanisms begin to address these issues by making data unidentifiable, which allow for the positive usage of private information without risking patient or physician privacy [56].

A second ethical implication includes the social impact of **human job loss** if AI automates such jobs. For example, the rise of autonomous driving semis and other transport vehicles has the potential to contribute to the unemployment of a large proportion of middle-class workers in the USA. This process is similar to the transition of farming to factory jobs across Europe and other parts of the world during the industrial era as well as the subsequent automation of factory lines. During these times, many workers lost their jobs, thereby moving from middle-class into poverty. Although these changes are unavoidable, the social impact on families and communities must be considered. However, existing data and insight from the industrial and automation booms can help data scientists and social science researchers better predict and prepare for the implications of AI-automation for employment. Planning for these potential consequences may help to ease the transition for society and future generations.

A final ethical implication worth noting here is the broader **impact on society**. Such impacts are often difficult to observe in real-time. One poignant example of the effects of AI is political polarization. With the rise of social media and worldwide connectivity via cell phones, tablets, smartwatches, and other devices, the implications of AI automating what information people have access to is more evident than ever. Some have suggested that the targeted marketing and information brought by AI has contributed to political and social polarization, with many people having access only to the information they already agree with. Opinions are constantly being validated, and the ability to be objective in any topic is becoming limited. This process has also occurred with consumer branding, as AI approaches target consumers with ads for products they are more likely to buy based on their previous purchasing behaviors. In fact, it has been said that these mechanisms know people better than they know themselves. The implications of the human mind becoming inundated with certain types of data remain to be seen, and the ethical considerations have yet to be examined. However, such implications must be on our radar as they have the potential to change society for generations.

Ethical dilemmas are just that: dilemmas. The vast majority are not easily solvable or even identifiable. However, their elusiveness cannot be the reason that data scientists fail to consider these issues—potential or currently a reality. Rather, it is the responsibility of the data scientist community to partner with other disciplines (e.g., social and behavioral sciences) to consider the effects of their creations on society, no matter how far into the future they may reach.

References

1. S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach* (Pearson, Kuala Lumpur, 2016)
2. I.B.Y.A.M. Turing, Computing machinery and intelligence-AM Turing. Mind **59**(236), 433 (1950)
3. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Mag. **27**(4), 12 (2006)
4. S. Harnad, Other bodies, other minds: a machine incarnation of an old philosophical problem. Minds Mach. **1**(1), 43–54 (1991)
5. P. Schweizer, The truly total Turing test. Minds Mach. **8**(2), 263 (1998)
6. J. Howard, Artificial intelligence: implications for the future of work. Am. J. Ind. Med. **62**, 917–926 (2019)
7. F. Hayes-Roth, Rule-based systems. Commun. ACM **28**(9), 921 (1985)
8. E.J. Horvitz, J.S. Breese, M. Henrion, Decision theory in expert systems and artificial intelligence. Int. J. Approx. Reason. **2**(3), 247–302 (1988)
9. P.L. Jones, I. Graham, *Expert Systems: Knowledge, Uncertainty and Decision* (Chapman and Hall, London, 1988)
10. M.C. Moed, G.N. Saridis, A Boltzmann machine for the organization of intelligent machines. IEEE Trans. Syst. Man. Cybern. **20**(5), 1094 (1990)
11. S. Ruslan, and G. Hinton, Deep Boltzmann machines., In *Artificial intelligence and statistics*, PMLR (2009), pp. 448–455

12. K.H. Cho, T. Raiko, A. Ilin, Gaussian-Bernoulli deep Boltzmann machine, in *The 2013 International Joint Conference on Neural Networks (IJCNN)* (2013), pp. 1–7
13. H. Poon, P. Domingos, Sum-product networks: a new deep architecture, in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (2011), pp. 689–690
14. W. Van Der Aalst, Data science in action, in *Process Mining* (Springer, Berlin, 2016), pp. 3–23
15. M. Iman, F.C. Delicato, C.M. De Farias, L. Pirmez, I.L. Dos Santos, P.F. Pires, THESEUS: a routing system for shared sensor networks, in *Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015* (2015), pp. 108–115
16. E. Alpaydin, *Introduction to Machine Learning* (MIT Press, Cambridge, 2014)
17. D.A. Pomerleau, Alvinn: an autonomous land vehicle in a neural network, in *Advances in Neural Information Processing Systems* (1989), pp. 305–313
18. E.L. Hines, R.A. Hutchinson, Application of multi-layer perceptrons to facial feature location, in *Third International Conference on Image Processing and its Applications, 1989* (1989), pp. 39–43
19. O. Chapelle, B. Scholkopf, A. Zien, Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Trans. Neural Netw. **20**(3), 542 (2009)
20. A. Sapru, and Sri Garimella, Leveraging Unlabeled Speech for Sequence Discriminative Training of Acoustic Models., In INTERSPEECH (2020), pp. 3585–3589
21. T. Dietterich, Overfitting and undercomputing in machine learning. ACM Comput. Surv. **27**(3), 326–327 (1995)
22. T. Poggio, K. Kawaguchi, Q. Liao, B. Miranda, L. Rosasco, X. Boix, J. Hidary, H. Mhaskar, Theory of deep learning III: the non-overfitting puzzle. CBMM Memo. 073 (2018)
23. J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 569–575 (2009)
24. S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in *2014 Science and Information Conference* (2014), pp. 372–378
25. T.L. Clarke, T.M. Ronayne, Categorial approach to machine learning, in *Conference Proceedings 1991 IEEE International Conference on Systems, Man, and Cybernetics* (1991), pp. 1563–1568
26. F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. **65**(6), 386 (1958)
27. D. Anderson, G. McNeill, Artificial neural networks technology. Kaman Sci. Corp. **258**(6), 1–83 (1992)
28. M. Minsky, S. Papert, An introduction to computational geometry, in Cambridge tiass., HIT (1969)
29. B. Karlik, A.V. Olgac, Performance analysis of various activation functions in generalized MLP architectures of neural networks. Int. J. Artif. Intell. Exp. Syst. **1**(4), 111–122 (2011)
30. P. Sibi, S.A. Jones, P. Siddarth, Analysis of different activation functions using back propagation neural networks. J. Theor. Appl. Inf. Technol. **47**(3), 1264–1268 (2013)
31. B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network (2015, preprint). arXiv1505.00853
32. S. Ruder, An overview of gradient descent optimization algorithms (2016, preprint). arXiv1609.04747
33. M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. **32**(14–15), 2627–2636 (1998)
34. J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)
35. D. Maclaurin, D. Duvenaud, R. Adams, Gradient-based hyperparameter optimization through reversible learning, in *International Conference on Machine Learning*, (2015), pp. 2113–2122
36. J.W. Tweedale, An application of transfer learning for maritime vision processing using machine learning, in *International Conference on Intelligent Decision Technologies* (2018), pp. 87–97

37. D. Cireşan, U. Meier, J. Masci, J. Schmidhuber, A committee of neural networks for traffic sign classification, in *The 2011 International Joint Conference on Neural Networks* (2011), pp. 1918–1921
38. S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. **8**(1), 98–113 (1997)
39. J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, Understanding neural networks through deep visualization (2015, preprint). arXiv1506.06579
40. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
41. Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. **31**(7), 1235–1270 (2019)
42. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems* (2014), pp. 2672–2680
43. T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 4401–4410
44. T. Zubair, A. Raquib, J. Qadir, Combating fake news, misinformation, and machine learning generated fakes: insight's from the Islamic ethical tradition. Islam Civilisational Renew. **10**(2), 189–212 (2019)
45. K. Sato, C. Young, D. Patterson, An in-depth look at google's first tensor processing unit (TPU). *Google Cloud Big Data Machine Learning Blog*, vol. 12 (2017)
46. H.S. Chang, M.C. Fu, J. Hu, S.I. Marcus, Google deep mind's alphago. OR/MS Today **43**(5), 24–29 (2016)
47. L.Y. Pratt, *Transferring Previously Learned Back-Propagation Neural Networks to New Learning Tasks* (Rutgers University, New Brunswick, 1993)
48. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning (2019, preprint). arXiv:1911.02685
49. A.A. Rusu, N.C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, R. Hadsell, Progressive neural networks (2016, preprint). arXiv:1606.04671
50. S. Zebang, K. Sei-ichiro, Densely connected AutoEncoders for image compression, in *Proceedings of the 2nd International Conference on Image and Graphics Processing* (2019), pp. 78–83
51. D.A. Ferrucci, Introduction to 'this is Watson'. IBM J. Res. Dev. **56**(3–4), 1 (2012)
52. W.-D.J. Zhu, B. Foyle, D. Gagné, V. Gupta, J. Magdalen, A.S. Mundi, T. Nasukawa, M. Paulis, J. Singer, M. Triska, *IBM Watson Content Analytics: Discovering Actionable Insight from Your Content*. (IBM Redbooks, Armonk, 2014)
53. N. Bostrom, E. Yudkowsky, The ethics of artificial intelligence. Cambridge Handb. Artif. Intell. **1**, 316–334 (2014)
54. A. Etzioni, O. Etzioni, Incorporating ethics into artificial intelligence. J. Ethics **21**(4), 403–418 (2017)
55. A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1(9), 389–399 (2019)
56. A. Gkoulalas-Divanis, G. Loukides, J. Sun, Publishing data from electronic health records while preserving privacy: a survey of algorithms. J. Biomed. Inform. **50**, 4–19 (2014)