Business Horizons (2020) 63, 37—50

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY
www.journals.elsevier.com/business-horizons

Rulers of the world, unite! The challenges and opportunities of artificial intelligence

Andreas Kaplan [a,*], Michael Haenlein [b]

[a] ESCP Europe, Heubnerweg 8-10, D-14059, Berlin, Germany
[b] ESCP Europe, 79 Avenue de la République, F-75011, Paris, France

KEYWORDS
Artificial intelligence;
Artificial superintelligence;
Human-machine symbiosis;
Machine learning;
Robotics;
Work displacement

Abstract A decade ago, we published an article in *Business Horizons* about the challenges and opportunities of social media with a call to action: "Users of the world, unite!" To celebrate its anniversary, we look at artificial intelligence and the need to create the rules necessary for peaceful coexistence between humanity and AI. Hence, we now are urging: "Rulers of the world, unite!" In this article, we outline six debates surrounding AI in areas like artificial superintelligence, geographical progress, and robotics; in doing so, we shed light on what is fact and what is utopia. Then, using the PESTEL framework, we talk about the six dilemmas of AI and its potential threat and use. Finally, we provide six directions on the future of AI regarding its requirements and expectations, looking at enforcement, employment, ethics, education, entente, and evolution. Understanding AI's potential future will enable governments, corporations, and societies at large (i.e., the rulers of this world) to prepare for its challenges and opportunities. This way, we can avoid a scenario in which we return in 10 years to write the article: "Dreamers of the world, unite!"

1. From users to rulers of the world, unite!

A decade ago, we wrote an article on the challenges and opportunities of social media (Kaplan & Haenlein, 2010). In it, we explained what social media is and what it is not, how social media applications can be classified and how firms can make use of them. Since then, social media has exploded in terms of both impact and relevance. While the tools themselves may have changed-think less Myspace, hi5, and Friendster and more Instagram, Snapchat, and WhatsApp—the number of users on social networks now amounts to an impressive 3 billion, roughly one-third of the world's population. In parallel, the amount of research on social media has exploded; Google Scholar now includes 4 million articles on the topic. Our article (Kaplan & Haenlein, 2010) alone has been cited over 16,000 times, becoming one

* Corresponding author
E-mail addresses: kaplan@escpeurope.eu (A. Kaplan), haenlein@escpeurope.eu (M. Haenlein)

https://doi.org/10.1016/j.bushor.2019.09.003
0007-6813/

of the most cited articles published in *Business Horizons*. Neither we nor anyone else would have been able to predict the sheer size of the phenomenon 10 years later.

Along with the positives, with growth we saw an increase of the negatives of social media that we did not anticipate or, at least, did not anticipate to such an extent. The disadvantages of social media include:

- Fake news (Allcott & Gentzkow, 2017), which we hinted at when looking at Wikipedia (see also Kaplan & Haenlein, 2014) but we underestimated its importance;
- Cyberbullying, which today results in emotional damage and even suicide among many young people (Smith et al., 2008);
- Online trolling and online harassment, which have led some celebrities to shut down their entire social media presence (Philipps, 2015); and
- A rise in addictions, anxiety, and depression attributed to excessive online gaming or a perception gap between what people see on social media and their own lives (Banjanin, Banjanin, Dimitrijevic, & Pantic, 2015).

Since we published our article on social media, we have seen artificial intelligence evolve. In 2011, IBM Watson defeated two human champions on the game show *Jeopardy*. In 2015, Stephen Hawking and 3,000 other researchers signed an open letter calling for a ban on autonomous weapons and, in 2018, Google presented Duplex, an AI service with the ability to book appointments by phone. Today, AI can generate artificial faces[1] and make those faces talk and say whatever you want (Zakharov, Shysheya, Burkov, & Lempitsky, 2019). We all know those emails in which a long lost prince or head of state is in desperate need to give you $20 million and we all have learned to ignore them. But what if tomorrow you get an email from a friend, written in his usual style, or a call in his voice, or even a video message with his face talking, asking you for $100? How to know what is real and what is illusion?

While social media was the hot topic in every boardroom in 2010, today, everyone is talking about AI. We think it is time to explain what AI is and what it is not, how the delusions and dilemmas of AI can be classified, and what directions we need to take to make the best use of this new technology. We first talk about the main debates in the field of AI: why there are so many misconceptions, how AI can be defined, how the future of AI may look, and what risks it may bring. We then turn toward the dilemmas of AI and look at the good and bad in the areas of politics, economics, society, technology, environment, and law. Our article finishes with a discussion of future directions for AI in terms of legislation, employment, ethics, education, international collaboration, and evolution with an emphasis on implications for firms.

So let us dive into this world in which robots may turn evil, we all will be out of a job soon because AI can do everything better and faster than us, and inequality will reach new levels. The same world in which AI will help us to combat isolation, employees will no longer have to deal with routine tasks, and we may have a shot of finding a solution to the problems we are unable to address today. In the world of AI, the rulers—not the users—of the world have to unite in order to create a sustainable future for humanity. Oh, brave new world, that has such systems in it!

2. Six debates: Fact and utopia

When people talk about AI, confusion, forecasts, and storytelling are usually just around the corner. Three experts in a room discussing AI may use four different definitions for the term. Some believe that evil robots will, at some point, destroy us all while others think that AI may never reach a higher level than what you can expect from a more advanced version of your smartphone. We want to shed some light on this discussion by focusing on six delusions and separating current facts from what seem—based on today's knowledge at least—to be utopian ideas of the future (see Figure 1).

2.1. Why all those misconceptions?

Why do so many people have such strong ideas about the potential and threat of AI? The blame lies, as it often does, with Hollywood. Movies like *The Matrix* (Do you take the blue pill or the red pill?) have shown us that the entire world is probably just an elaborate simulation created by machines powered through AI. *The Terminator* (Hasta la vista, baby!) provided convincing proof that evil robots may actually be out to kill all of us. And the *Avengers*, *Power Rangers*, and *Iron Man*

[1] https://thispersondoesnotexist.com/

Figure 1. The six Ws of artificial intelligence

WHY	Misconceptions exist because media tend to emphasize the dangers and threats of AI more than its potential and opportunities
WHAT	AI is "a system's ability to interpret external data correctly, to learn from such data and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019b, p. 17)
WHO	As with any tool, AI-driven robots can be used for good or evil. It is unclear if robots can turn evil by themselves (wireheading), but likely that they do not and will not think like humans
WHEN	It is impossible to say when artificial superintelligence will arrive—maybe tomorrow, in our lifetime, or never
WHERE	The current epicenters of AI development are China and the U.S., but other world regions (especially Europe) will play important (niche) roles in the fugure
HOW	The extent of AI risk is hard to predict but likely to be substantial. Even if the chance of occurrence is small, the discounted risk is sufficiently large to warrant preventive measures today

illustrate the combination of humans, machines, and AI (e.g., J.A.R.V.I.S. - Just A Very Intelligent System).

These movies are mirrored in press articles that naturally have an interest in positioning AI as more controversial than it actually is. Sex sells, and so do fear and doomsday scenarios. Interestingly, this does not stop at tabloid journalism but also applies to respected news outlets. In November 2018, *The Guardian* published an article titled "The truth about killer robots," and a month later, *The* Economist (2019) wrote: "There are no killer robots yet—but regulators must respond to AI in 2019." Such reporting is not a new phenomenon. In 1946, when ENIAC, the first general-purpose computer, was presented at a price of $6.3 million in today's currency, the media fallout ranged from calling it a "magic brain" to a "mathematical Frankenstein." It is problematic though as inaccurate and sensational stories can create unrealistic expectations, a problem which Zachary Lipton from Carnegie Mellon has named the "AI misinformation epidemic."

The question then becomes how to separate the wheat from the chaff and how to identify which of these conceptions are probable, which ones possible and which ones unlikely. So let us look next into what the term AI actually means.

2.2. What, again, is artificial intelligence?

Broadly speaking, there are three reasons why it is so difficult to define AI. First, it is already difficult to define what human intelligence is and, hence, applying this fuzzy concept to machines is a complicated endeavor. Second, once we grow used to a machine performing a complex task, we stop considering the ability to do this task a sign of intelligence. This is generally referred to as the *AI effect* (Haenlein & Kaplan, 2019; McCorduck, 2004). The AI effect makes the definition of AI a moving target in the sense that AI always seems to be out of reach. Third, AI has different evolutionary stages—from narrow to general to superintelligence— and can be classified into analytical, human-inspired, and humanized AI depending on its cognitive, emotional, and social competences (Kaplan & Haenlein, 2019b). It is easy to mix up those different stages and types leading to confusion over the term itself.

A decade ago, we faced a similar situation with respect to the term social media and, in response, we developed a definition to identify what exactly should be included in this fuzzy concept and how it differs from related ones such as Web 2.0 and user-generated content. At the time, we defined social media as: "a group of internet-based applications that build on the ideological and technological

foundations of Web 2.0, and that allow the creation and exchange of user-generated content"[2] (Kaplan & Haenlein, 2010, p. 61).

Today, we are facing the same issue with respect to AI and its relationship to concepts such as big data and Internet of Things (IoT). We therefore define AI as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019b, p. 17). IoT (Krotov, 2017; Saarikko, Westergren, & Blomquist, 2017) can be seen as one specific way of obtaining the external data required as an input for AI. It is one input toward big data (Lee, 2017), which describes data sets characterized by huge amounts (i.e., volume) of frequently updated data (i.e., velocity) in various formats such as numeric, textual, or images/videos (i.e., variety). AI uses external information obtained through IoT or other big data sources as an input for identifying underlying rules and patterns by relying on approaches from machine learning, which describes methods that help computers learn without being explicitly programmed. We refer readers to Kaplan and Haenlein (2019b) for more details.

2.3. Who will those robots of the future be?

One idea that has inspired dozens of blockbusters is that of the evil robot that starts turning against its human creator—think of the Sentinels in the *X-Men* series or Bender in *Futurama*. It is obvious that robots can be programmed to perform tasks that can be considered evil; military robots and drones are examples of that. But the question is whether a robot that was initially conceived to help humans, say as a service robot in a restaurant or an eldercare robot, can turn evil autonomously. The intuitive answer to this question is no; no robot can, by itself, change the goal for which it was programmed. And while this may be true, although not certain, a robot may very well use whatever it takes to achieve its goal.

In a similar vein, as humans may commit crimes to obtain drugs that generate pleasure, an AI system may decide to cheat or even kill humans to achieve the goal it has been programmed to maximize. This phenomenon is called *wireheading* and makes the definition of the goal function of an AI system particularly important (Yampolskiy, 2014). Yet again, such logic assumes that a super intelligent AI system actually thinks and behaves as humans do, which is far from certain (Mosser, 2018). Since artificial intelligence is not human intelligence, it is not clear what artificial superintelligence actually looks like. When AlphaGo, an AI system developed by Google (Silver et al., 2016), was able to beat Lee Sedol in the highly complex board game Go, it was interesting to observe how nonhumanly the system played. Moves that were considered to be a mistake by human observers later turned out to be the winning strategy. AI is more than the formalization or mirror image of human intelligence. It is an entirely different thing altogether. One point where this becomes particularly obvious is creativity. While AI is fundamentally based on pattern recognition and curve fitting, "creativity is intelligence having fun" as Albert Einstein said, and it is the foundation for all aspects of art, be it in the form of painting, literature, music, or food. At this stage, it seems unlikely that AI systems will be able to solve truly creative tasks.

2.4. When will we see systems with artificial superintelligence?

The holy grail of AI is *artificial superintelligence*: systems that are self-aware, and capable of scientific creativity, possess social skills and general wisdom, and thus make humans redundant (Kaplan & Haenlein, 2019b). When can we expect to see such a jack-of-all-trades? When confronted with this and other questions on the future of AI, a common fallacy is to make a linear projection of progress done in the past. Although such projections are often misleading, this happens to the best of us. Look at Herman Simon, who predicted in 1957 that computers would be able to beat the world champion in chess within a decade. Yet, it took four times as long for IBM's Deep Blue to defeat Gary Kasparov in 1997.

There are numerous reasons such projections are likely to fail. Computing power, for example, has increased continuously in the past, a phenomenon known as Moore's Law. Yet, this is unlikely to continue in the future since we are approaching physical limitations for how dense transistors can be packed on a chip. A significant driver of past AI progress is the increasing availability of data but this is likely to slow down as well. And even if the performance of AI systems continues to increase, they can never generate more information than is available in the input data in the first place— a concept referred to as the *data processing inequality*.

[2] See Kaplan (2012) for the extension of social media to mobile devices.

Will we ever be able to artificially replicate a human brain with its 200 billion neurons connected by 10,000 synapses that each can have about 1,000 states? Given current technology, probably not. But then again, many achievements that seemed impossible a decade ago have become mainstream today. The simple answer to the question of when we will see artificial superintelligence is: nobody knows. It could be around 2050 as predicted by Vincent Müller and Nick Bostrom (2014), it could be never, or it could be tomorrow.

2.5. Where in the world will progress in AI originate?

If we cannot answer when, then we can at least say where progress in AI is most likely to take place. Currently, the centers of AI research are in China and the U.S. Who will win this cold tech war is still uncertain and so are the implications AI will have on people's lives in the future. While China is betting heavily on facial recognition—currently using 170 million cameras powered by AI to fight crime and increase general security—more and more places in the U.S. are getting increasingly concerned about the creation of a surveillance state. San Francisco recently banned facial recognition and other cities are likely to follow suit. It is most likely such differences in governmental power and attitudes toward privacy that will give China an advantage in the race for global AI leadership (Zhu, Huang, Chen, & Gao, 2018).

The remaining world regions are far behind China and the U.S., yet they may shape the future in different ways. The strict privacy requirements implemented in Europe through the General Data Protection Rule (GDPR) in 2018 will impact the way firms are required to handle and store personal data way beyond European borders. To what extent this will slow down the evolution of AI overall, since powerful AI needs big data, is unclear. On one hand, it can be argued that less data means less powerful AI because AI is only as powerful as the data it can learn from. On the other hand, children learn to recognize an elephant after only a handful of images, while modern AI systems usually need millions of pictures. If advances in AI will at some point make it possible to train systems based on much fewer input data than current requirements, then data volume would become a much less critical resource. In such a world, privacy concerns are likely to be valued higher, giving Europe a head start compared to other world regions. Future considerations will also include the need for regulation and the desire to generate decentralized systems independent from GAFA (i.e., Google, Amazon, Facebook, and Apple).

2.6. How serious are the risks of AI?

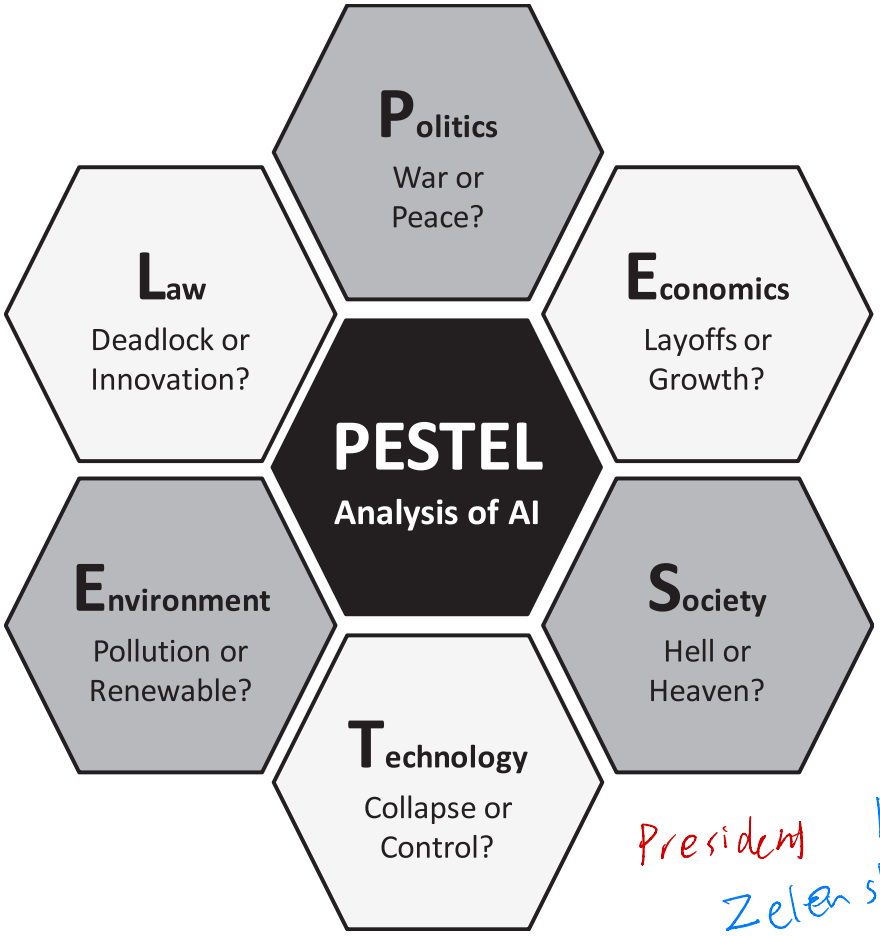
Putting all of this together—the misconceptions, the rise of evil robots, the emergence of artificial superintelligence, and the cold tech war—makes one wonder how serious the risks of AI actually are. Is AI a wolf in sheep's clothing? According to the Future of Humanity Institute at the University of Oxford, which specializes in identifying risks that could threaten humanity, it might well be. For many risks we commonly fear (e.g., meteor strikes, pandemics, nuclear war), the chances that they will lead to the total extinction of humanity are actually pretty slim. While these events and their direct consequences may lead to the disappearance of 90%, 95%, or even 99% of humanity, there will still be 10%, 5% or 1% who will survive them. If AI turns bad—really bad—the risk could be full extinction with 0% survival. What would happen to the proverbial cockroaches remains a matter of debate.

In response, many ask for more AI regulation even if we may still be decades away from the emergence of artificial superintelligence. Still, it never seems too early to prepare for what may come in the distant future. In 1933, physicist Ernest Rutherford said that nuclear energy was moonshine; 24 hours later, the nuclear chain reaction was invented. And it was 78 years before Japan's Fukushima reactor exploded in 2011. Even if we try to balance the potentially serious consequences of AI going wrong with the remote chance of this actually happening, the discounted risk is enough to warrant preventive measures today. This raises a question: Should not every citizen, company, or government be advocating for research and control in this area? To find an answer, one can draw a comparison to the protection of the environment for which many say that severe actions are needed to combat climate change but only too little is actually being done. So let us now look at the main dilemmas of AI and the key domains for which good and bad often go hand in hand with each other.

3. Six dilemmas: Threat and use

The Russian historian Aleksandr Solzhenitsyn once famously said that the "battle line between good and evil runs through the heart of every man." The same can be said for AI. Given the recency of the phenomenon, it is uncertain how good and evil will

Figure 2. PESTEL analysis of artificial intelligence



play out in the world of AI. We now look at six key dilemmas of AI, illustrating this tradeoff using the PESTEL framework as a basis (Aguilar, 1967). Figure 2 applies this framework to AI.

3.1. Politics: War or peace?

There are numerous ways in which AI and robotics could be used for war and military purposes (Leveringhaus, 2018). This not only includes military robots and autonomous drones but also exoskeletons, which increase the strength and endurance of human soldiers. In the future, we may see evolutions such as insectoid robots[3] and the use of AI to support decision making and plan maintenance of planes and ships. But an opponent may be able to manipulate your devices via hacking or they may deviate from the right path themselves due to the aforementioned issue of wireheading. Your enemies may even decide to rely on subtler methods and influence the democratic voting process, as the example of Cambridge Analytica and Facebook in the 2016 U.S. presidential election have shown.

These examples should not give a false impression that AI is all bad. Would it not be better to use a military robot to map a minefield or explore an unknown building than rely on human soldiers who may die in the process? And in the same way AI can be used to influence, it can also be relied on to help humans make better decisions. It has long been known that humans are terrible in decision making (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981). Tools such as RoboVote[4] already exist to help citizens identify the best party to vote for given their preferences. And on

[3] A prototype has already been developed by Harvard's microbiotics laboratory.

[4] http://robovote.org/

the topic of outsourcing decision making: Might it possibly be better to be governed by machines than human politicians altogether? A recent survey conducted by the Center for the Governance of Change at Spain's IE University showed that 25% of Europeans would prefer if policy decisions were made by AI instead of politicians, who may turn out to be corrupt or ideologically extreme.

3.2. Economics: Layoffs or growth?

It seems that, at some point, every discussion of AI turns to the issue of layoffs and the question of whether human workers are still needed if machines can do everything. The underlying argument goes back nearly a century when, in 1930, John Keynes introduced the concept of technological unemployment (Keynes, 1930). In their famous study, "The Future of Employment", Frey and Osborne (2017) noted that 47% of total U.S. employment may be subject to automation. While some argue that technological change has always destroyed some jobs and created others in exchange (e.g., the introduction of the car transformed many carriage drivers into chauffeurs), others say that a sufficiently advanced AI system can do everything better and cheaper than humans ever could (Atack, Margo, & Rhode, 2019). In such a world, why would one want to employ humans for anything? When cars became more prominent, horses, which used to be a key form of transportation at the time, essentially vanished from our streets. Sure, some still are around for racing but for the average Tennessee Walking Horse or Appaloosa there really was no job anymore and their prospects looked relatively grim.

Yet such a negative outlook is not necessarily likely. An average employee performs dozens if not hundreds of different tasks in a given day and only a few of them can actually be taken over by machines. Given the high cost of purchasing an AI system as well as customizing and maintaining it, it seems unlikely that firms will replace humans by machines if those machines can only do part of their job. Other types of employment, including those relying on feeling tasks, will probably become even more important in future (Huang, Rust, & Maksimovic, 2019). The World Economic Forum (2018) predicted that over 50 million new jobs will be created through AI in the next 5 years. It seems likely that the future of automation lies in job enhancement instead of job displacement (Metcalf, Askay, & Rosenberg, 2019) with AI systems taking over routine and boring tasks that human employees preferred not to do in the first place, at least in the short to medium term.

3.3. Society: Hell or heaven?

Two problems that put a strain on most societies today are the rise in inequality and loneliness. AI will most likely have an impact on both of them. While the technological change that will be triggered by AI may not lead to the disappearance of jobs, it will most likely disconnect firm productivity from labor productivity. Hence the increase in productivity brought by AI may not benefit everyone equally, leading to rising inequality. It has been shown that the rate of return on capital tends to be larger than the rate of economic growth (Piketty, 2014) and AI is likely to widen this gap even further. According to the World Economic Forum, this rise in inequality is the greatest societal concern of the robotics revolution. In addition, AI may lead to a rise in isolation if, for example, nurses are replaced by care robots to respond to a labor shortage in hospitals, social care stations or elderly homes.

Yet the pendulum could also swing in the other direction. In an increasingly aging society, the problem of lonely citizens, especially elderly ones, will become more and more severe. Robots such as Pepper, a semihumanoid robot manufactured by Softbank Robotics, are already equipped with AI that allows them to detect emotions and react accordingly. While not perfect, they may turn out to be of great help for people who would have limited social contact otherwise. Staying in the medical domain, AI can also help to predict serious health risks such as skin cancer and strokes, frequently outperforming human experts. The question then becomes: If an AI system could tell you with near certainty which disease you will get in the next decade and which one may kill you, would you want to know? Or would rather remain oblivious to when death may hit you? Even if there is no work anymore, since all is done by AI, humans could focus on their spiritual and physical development. Is such a life the Elysium of enlightenment or the evilness of ennui?

3.4. Technology: Collapse or control?

In a world in which an increasing share of activities will be controlled by AI, how can humans ensure they still keep the upper hand? The case of the evil robot is clearly an extreme scenario but there are many other ways AI could get out of control.

First, an AI system might simply misinterpret a user request or take it too literally. If you ask your self-driving car to take you to the hospital as fast as possible, you may want to specify that you are

planning to arrive alive and ideally without running over anyone else. Or, if there will ever be an AI system that one could ask to save the world or bring world peace, a perfectly efficient move could be to eradicate all humans.[5] As intelligent as AI may seem, it can still be made stupid by having to interpret fuzzy human demands.

Second, AI systems may be biased from the beginning. As mentioned above, AI is all about interpreting external data and learning from it. If the external data used to train AI is biased (e.g., because it has been generated by humans using biased heuristics), then such bias will carry over or even be amplified in the AI system. Examples of this abound, and include image recognition systems that perform better recognizing white people versus people of color due to the biased input data.

Third, once AI systems get increasingly performant—and, hence, more complex—they will be more and more difficult for humans to understand. Artificial neural networks and deep learning, some of the most common machine learning tools used in AI, are inherently a black box. This opens, in principle, the possibility for AI to outsmart us. Humans were able to control the planet not because we are the fastest or strongest creatures but because we are the smartest. If we no longer are the smartest, how can we ensure we remain in control?

3.5. Environment: Pollution or renewable?

Every major change in the human economy has put a significant strain on the environment. The agricultural revolution, which helped humans to settle and combat famine, resulted in the disappearance of entire forests to make space for farming and the emission of huge amounts of carbon dioxide (e.g., from cattle ranching). The industrial revolution led to dependence on fossil fuels and associated climate change. Unfortunately, AI is not exempt from this rule. The servers, which run computations in the cloud or store big data, require vast amounts of energy for cooling. Producing them requires raw materials such as cobalt, nickel, and lithium in such high amounts that Earth may soon no longer be able to support them in sufficient quantities. And once they are outdated, they generate electronic waste, the processing of which affects human health and damages the environment even further.

[5] The Marvel movie *Avengers: Age of Ultron* is an apt illustration of this.

Yet, it is conceivable that the benefits created by AI will outweigh those costs. Humanity is facing serious issues, climate change being most prominent among them, which seem difficult to address using the approaches at hand. AI can be a major game changer in this context (VoPham, Hart, Laden, & Chiang, 2018). Using its Deep Mind AI system, Google was able to reduce the cooling bill of its data centers by 40%—a performance that would have been difficult to achieve using more conventional approaches. In Norway, AI helped to create a flexible energy grid that integrates more renewable energy than before. One could argue that firms should not have to care about environmental goals but should focus their attention on profit maximization. Still, in a framework of creating shared value (Porter, 2011), both may be achievable in one go. In addition, more and more firms are getting certified as B-Corporations, which pushes them to take social and environmental performance as well as shareholder value into account. It seems difficult to say which other advances may be in store for the future but it seems likely that our best shot of combating climate change, either by reducing carbon emissions or by filtering existing carbon out of the atmosphere, lies in the use of AI.

3.6. Law: Deadlock or innovation?

Coming back to our definition of AI mentioned above, it is easy to see that external data and AI go hand in hand. This brings at least two major legal challenges. The first one is the issue of privacy. Much of the data used to train AI systems have been generated in one way or another by humans. Hence privacy is a major concern in this context. Many rumors—some of them true—deal with this issue. You may have heard that your smart speaker may secretly be listening to your conversations or that the free game app you just downloaded can access the location data of your phone. Governments will find themselves in the complicated position of regulating privacy; too little regulation may inevitably lead to the violation of civil rights, while too much may motivate firms to move their AI investments to another jurisdiction.

The second challenge is liability. Who is to blame if an AI system makes a decision that generates some harm? The mathematician who developed the underlying algorithm? The manufacturer who produced the software? The database that provided the external data for the system to learn from, or the customer who purchased and used the system? Given the huge penalties that may be associated with liability lawsuits, this issue

is probably one of the most pressing ones in need of legal clarification. Some countries suggest to treat an AI system as an animal and make the holder or owner responsible for any harm caused. Alternatively, AI systems may receive their own rights, an approach that some jurisdictions have used with respect to natural phenomena. New Zealand, for example, recognized the status of the Whanganui River as a legal entity in 2017. And let us not forget Saudi Arabia, which granted citizenship to Sophia, an AI robot.

4. Six directions: Requirements and expectations

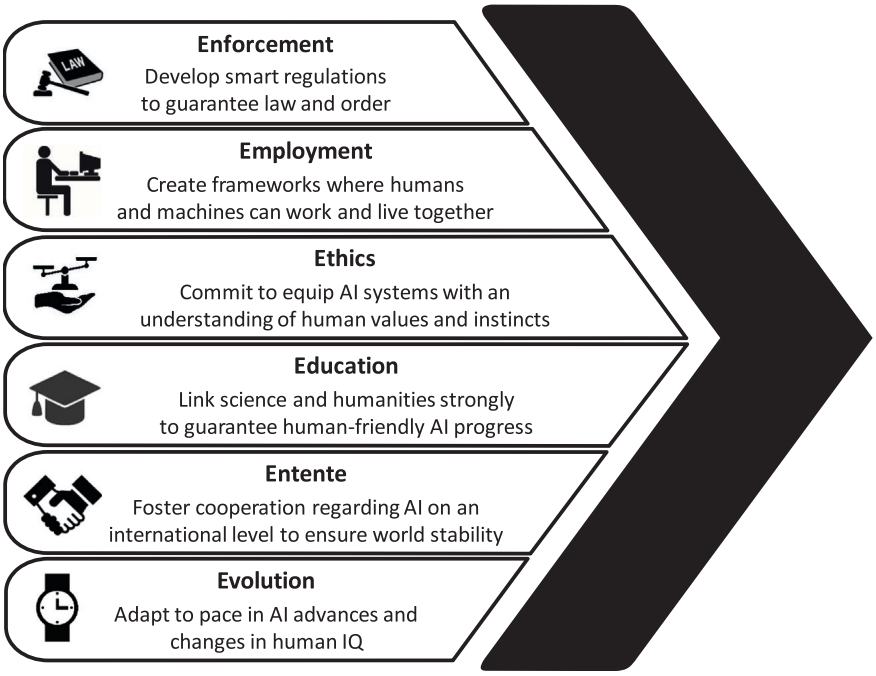
Let us end this article by giving six general directions on how the world, regulators, and society might prepare for the AI revolution. We also show how being aware of these upcoming changes will help firms to know what to expect and prepare accordingly (see Figure 3).

4.1. Enforcement: Law and order

All of the above makes it clear that law, order, and regulation are necessary conditions for AI to develop to its full potential. But this is easier said than done. Given that AI is still in its infancy, it is difficult to predict what will happen in the next years or even months. What amplifies this further is the AI effect (Haenlein & Kaplan, 2019). It therefore seems likely that regulation will not focus on AI itself but on the process used to generate AI. This can include common norms regarding the collection, processing, and storage of personal data as well as guidelines and procedures for testing algorithms and ensuring their transparency. For example, self-driving cars might be required to have black boxes, similar to flight recorders, which can be used to provide objective information in case of an accident.

Regulation will also be needed to prevent excessive concentration in the AI space. Already today GAFAM (Google, Amazon, Facebook, Apple, and Microsoft) in the U.S. and BAT (Baidu, Alibaba, and Tencent) in China are dominating the market and this is likely to become more pronounced in the future. The power of AI is driven by the amount of input data present and the performance of algorithms and hardware to learn from such data. Both these dimensions have considerable network effects, making it likely that larger firms will become even bigger in the future. Many have described big data as the new oil and, in that logic, AI companies will probably be the utility providers of the future, transforming data into information similar to how classical utilities transform oil into energy. In the future, firms may be required to share their data or give details about their

Figure 3. Six dirEctions for AI to unite the rulers of the world



algorithms, probably after some period following the logic of patent protection that is the main mechanism in the pharmaceutical industry.

This specter of regulation will have at least three implications for firms. First, companies need to be aware that regulation will come eventually and prepare for it sooner rather than later. They also should avoid pushing the abilities of AI too far since technically feasible but ethically unacceptable solutions may just be the straw that breaks the camel's back and may trigger more severe regulation in response. Second, AI should not be left in the hands of IT since doing things wrong or with a limited scope in mind can have substantial and costly implications. And finally, firms need to be aware of the national context in which they are active (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018) since standards and current as well as future regulation vary considerably from one country to another.

4.2. Employment: Humans and machines

Although massive job displacement due to AI is not necessarily likely, some industries will clearly see a very significant change. Restaurants of the future will more likely be staffed by service robots than by human waiters, and call center agents are likely to be replaced by chatbots. Similar to how domestic appliances reduced the need for household staff and mechanical looms put weavers out of a job, AI will touch some pink- and white-collar jobs in the same way as blue-collar workers were affected by automation on the shop floor decades ago. At first, governments may decide to keep people in the job by forcing companies to reduce the use for automation, train employees for new tasks, or distribute the total work hours differently, or they may restrict the use of automation altogether. In France, for example, a law already requires that electronic platforms are only usable during normal office hours under certain conditions.

Yet it seems unrealistic to expect that all employees will have the skills to train for a new occupation or the ability to develop such skills. AI may increase unemployment rates to a certain extent, at least for some groups of society. This brings us to the idea of a universal basic income. Whether such an idea should apply to every citizen or be targeted toward certain groups is a matter of ongoing debate. A recent study in Finland provides empirical evidence that a broad universal basic income may be less beneficial than initially anticipated. There is also the question of who should pay for such an idea. One obvious solution may be to collect additional taxes on automation, similar to how VAT works today; however, that could distort competition in a severe way. Connected to this are philosophical and ethical questions about the value of work and the importance of meaningful work to personal happiness.

What does all of this mean for managers? First, managers need to be aware that many employees will be scared of being replaced by AI, independent of whether this fear is justified or not. This requires strong skills in leading an open dialogue, resolving conflict, and—broadly speaking—a human, ethical, open, and transparent leadership style. Second, managers need to identify the skills of their human employees and find a place for them in an ecosystem in which humans and machines will work hand in hand. This will include a stronger focus on emotional or feeling tasks for humans, for which they have an inherent advantage over machines (Huang et al., 2019). Third, all of this needs to be done in a bottom-up versus top-down approach. Involving employees in the process of developing and implementing AI systems makes such systems more successful (Tambe, Cappelli, & Yacubovich, 2019). In short, managers will need to act as empathetic mentors and data-driven decision makers (Kaplan & Haenlein, 2019b).

4.3. Ethics: Values and instincts

At the end of the day, an AI system is a tool that will try to achieve whatever task it has been ordered to do. The problem comes when those orders are given by humans who are rarely as precise as they should be. AI systems, therefore, need to have some understanding of human values to interpret fuzzy commands correctly. For example, an AI system that receives the order of protecting a person from harm should know that locking said person into a room 24/7 is not the ideal course of action. This implies that AI systems will need to be equipped with rules that go beyond their main task, including ethical guidelines (Etzioni & Etzioni, 2017). One example that received public interest is the Moral Machine Experiment (Awad et al., 2018) in which humans were asked to make tradeoffs in the context of self-driving cars. If your car had to choose between running over one child or one pensioner, which one do you pick? What about one child and two pensioners?

Besides being important in itself, providing AI systems with values and ethical guidelines will also be crucial since an increasing reliance on AI will make existing company rules and actions more transparent to the outside world. An AI system that

has been trained on biased or erroneous data will formalize and amplify such errors. If before your firm had a small bias toward hiring some group of applicants, then the AI system will go all the way. This is what happened to Amazon; it had to scrap a recruiting tool that was systematically biased against women. Also, if an AI system gave you the ability to track your employees in real time, finding out who is sending emails to whom, editing files and meeting, should a firm use such a tool? What if this tool helps you to identify the key influencers and change makers in your organization? The AI tool Isaak developed by Status Today, which obtained the award for the Best AI Startup in 2017, can do just that.

For managers, trust will be a crucial metric to monitor closely. Consumers need to be able to trust that firms make good use of their data (Kaplan & Haenlein, 2019b; Rossi, 2019). If firms fail to build such trust, the consequences can be substantial, ranging from customers actively distorting information to boycotting firms altogether. Given these possibilities for substantial adverse reactions, AI systems should be launched only after extensive trial phases, similar to the extensive testing conducted by pharmaceutical companies before new product launches. The more transparent an AI system is, the better. If needed, rules that can be communicated to consumers and other stakeholders may need to be extracted from AI systems post-hoc to make the inner workings of such systems less opaque.

4.4. Education: Science and humanities

In a world of AI, educating humans in the right way will be central. First and foremost, human workers will need to accept that machines will, in one part or another, make up a section of their colleagues in the future. Thus, each person should have at least a basic understanding of programming and should learn a programming language such as Python. The idea is not to be able to program in Python at an expert level but, just as speaking some basic Chinese is helpful when working with colleagues in China, knowing the basics of programming will allow employees to better understand their AI counterparts. Combining the need for new skills with the increasing lifespan of humans gives the concepts of lifelong learning a whole new dimension (Kaplan, 2018; Pucciarelli & Kaplan, 2016). Given the rapid pace of change, it seems very likely that most human workers will have multiple phases of studying in their life.

In the medium term, however, humans probably need to accept that machines will outperform them in most analytical tasks. This means they need to be educated in more subtle skills—previously named feeling tasks (Huang et al., 2019)—as those will probably be the domain where humans will continue to outperform AI. Broadly speaking, the human side and an ethical approach to its use will be of rising importance to avoid businesses and society becoming too technology-focused, a la science fiction disaster scenarios. In response, some have argued that universities should broadly introduce courses on AI and humanity to answer questions around equity, ethics, and privacy (Keating & Nourbakhsh, 2018).

Such training should not be limited to universities only but also practiced in firms. Companies have a tendency to see such training as a burden since there is no guarantee that trained employees will stick with the company in the future. But this is the wrong perspective. Firms need employees able to work in an AI-enabled context, and providing extensive training is not only an investment in building such skills but also in attracting new talent and retaining existing ones. Employees who have a natural talent for AI should be spotted and leveraged as AI evangelists to inspire the workforce. A culture of trial and error can help employees experiment with AI without the fear of making mistakes or looking dumb. Waiting until the government forces firms to implement such training on a company level is both inefficient and short sighted. The sooner companies can ensure that there are AI-literate staff on all levels of the organization, the better.

4.5. Entente[6]: Cooperation and stability

Regulation in the space of AI is complex and will require at least some form of international collaboration and diplomacy to avoid the use of AI in war, terrorism, or tax evasion. Several international bodies have already developed recommendations in this regard. The UN and Human Rights Watch, for example, have advocated a treaty banning AI-driven weapons. Looking at tax evasion, if firms rely increasingly on AI, they will be able to shift the center of value creation to any jurisdiction they wish by putting their AI-powered servers offshore. Since this reduces corporate profits to arbitrarily low levels (e.g., due to the payment of fictitious royalty fees), the EU is considering the introduction of a tax of 3% on

[6] Entente is a term primarily used in international diplomacy to describe "an international understanding providing for a common course of action" (Merriam-Webster Dictionary)

revenues generated within its borders. The OECD has also recommended developing international guidelines regarding AI to avoid de facto rules being made by the two AI superpowers: China and the U.S.

Looking at a broader scale, it is clear that the main world regions have differing approaches to how AI should be managed. In China, AI is largely controlled by the government and used to increase the economic standing of the country and to support, for example, general security through facial recognition systems on streets. In the U.S., the evolution is driven by a handful of for-profit tech companies that mainly have corporate profits in mind. In the EU, the focus is on the citizens themselves and their need for data protection and privacy. The coordination of these partly contradictory goals will be a key challenge. Some even speak of the concept of *AI boxing*: the creation of three non-connected networks that serve those three regions separately using their own rules and guidelines.

It could be argued that such macroeconomic factors do little to influence firms in their day-to-day operations. Yet, as the 2019 trade disputes between China and the U.S. have shown, this is clearly not the case. If different viewpoints on the use of AI between the U.S., China, and Europe result in different technological infrastructures and standards, then this has direct implications for the operations of most firms. The same applies when political forces result in export or import restrictions of AI technology. Look at the business restrictions the U.S. Department of Defense imposed on Huawei Technologies as an illustration of this point. For firms, this implies that they may need to develop infrastructure consistent with multiple regional/national rules and standards. It may also suggest that countries who fight for having the big AI champions within their borders may offer tax advantages and subsidies as incentives to companies.

4.6. Evolution: Pace and IQ

Last but not least, we should not forget that all of the above will not happen from one day to another. Laws will likely focus on supporting innovation at the beginning and then gradually, over time, place a higher value on privacy concerns and issues of market concentration. Jobs will not disappear from one paycheck to another. Instead, some tasks will see more and more automation, giving employees sufficient time to adapt their skills and move into new positions. In parallel, we will most likely see a premium placed on solid journalistic work. In a world in which AI systems can generate videos from only a few pictures and make anyone appear to have said anything (Suwajanakorn, Seitz, & Kemelmacher-Shlizerman, 2017), and social media newsfeeds can be biased and excessively customized, many people will value journalists who provide the least biased truth available. In addition, there will probably be a general move back to more authenticity, similar to how social media brought the internet back to its roots (Kaplan & Haenlein, 2012). Instead of looking at edited versions of selfies on Instagram, people may want to see natural and unfiltered beauty (Kaplan & Haenlein, 2019a).

At the end, let us come back to artificial superintelligence and the question of when machines will be as smart as humans. We mentioned that what AI is and what it is not is considered a moving target, but what everyone should remember is that the same applies to human intelligence. Humans are likely to adapt to the increasing prominence of AI and develop new and different types of intelligence. They also tend to get smarter over time. People suffering from torsion dystonia have been shown to be 10 IQ points more intelligent than the average person (Eldridge, Harlan, Cooper, & Riklan, 1970). Our inherent drive to be lazy may also be an inherent advantage as it allows us to find clever and efficient solutions. Have you ever tried and failed to assemble a piece of IKEA furniture by yourself? By placing the package in the hallway, you may motivate your spouse to do the job on your behalf. For robots, such thinking is alien and assembling a chair from dozens of pieces can be a near-impossible task—for now at least.

The conceptualization of AI as a moving target implies that managers need to adapt constantly. Organizational structures may need to be flexible in order to account for different shares of humans vs. machines and a changing distribution of tasks among them. Competitors need to be monitored regularly for their access to superior data or hardware (Kaplan & Haenlein, 2019b). In the beginning, AI may mainly help to make existing processes faster and more efficient—a concept Brock and von Wangenheim (2019) call *realistic AI*—but at some point processes may have to be redesigned entirely. Not missing this tipping point, while still keeping the human aspect of AI and teams in mind, is a key challenge for which every manager will need to adjust.

5. More to lose than their chains

In 2010, we ended our analysis of social media by concluding that users had "nothing to lose but

their chains" (Kaplan & Haenlein, 2010, p. 68). Today more seems to be at stake. If we assume for a moment that AI systems will actually take over a substantial share of our jobs and more and more humans will live on some form of universal basic income (Van Parijs, 2004), and if we combine this scenario with the substantial advances made in virtual reality technology, how far are we from a world like the one illustrated in Neal Stephenson's *Snow Crash*? We already described this world in 2009 when we discussed Second Life (Kaplan & Haenlein, 2009, pp. 563–564):

> In this book, Stephenson tells the story of a protagonist named Hiroaki Protagonist, who physically lives in Los Angeles during the early 21st century but who mentally spends most of his time in a three-dimensional virtual world called the Metaverse. He, as well as other people, access this Metaverse using personal computer terminals that project pictures of a virtual urban environment situated on a virtual artificial planet onto goggles. Within the Metaverse, everyone appears in the form of personalized avatars; that is, pieces of software that are the audiovisual bodies that people use to represent themselves and communicate with other people in the Metaverse. These avatars, which may have any appearance the user desires (except for limitations of height ''to prevent people from walking around a mile high''), can perform any activities familiar from their real life, such as visiting night clubs, making friends, or consuming virtual drugs, like the pseudo-narcotic snow crash. In the 21st century the Metaverse is so popular and attractive that some people even decide to remain continuously connected to it by spending their real life in storage units, surrounded only by the technical equipment necessary to enter the virtual world.

As unrealistic as such a scenario may be, research has shown that mice (and humans) implanted with electrodes that can stimulate the brain's pleasure center will become obsessed by stimulating those as much as possible. Remember the concept of wireheading mentioned above? The term actually stems from those wires which enter a human's brain. Now, this all sounds like crazy talk and science fiction, right? But it is not; Neuralink, a startup founded by Elon Musk, is currently working on developing implantable brain-machine interfaces to achieve a symbiosis between human and AI. Let us hope that, 10 years from now in 2030, this world will still be a fiction. And, if not, we already have a great title for an article on the topic: "Dreamers of the world, unite!"

References

Aguilar, F. J. (1967). *Scanning the business environment*. New York, NY: Macmillan.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives, 31*(2), 211–236.

Atack, J., Margo, R., & Rhode, P. (2019). "Automation" of manufacturing in the late nineteenth century: The hand and machine labor study. *The Journal of Economic Perspectives, 33*(2), 51–70.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., et al. (2018). The moral machine experiment. *Nature, 563*, 59–64.

Banjanin, N., Banjanin, N., Dimitrijevic, I., & Pantic, I. (2015). Relationship between internet use and depression: Focus on physiological mood oscillations, social networking, and online addictive behavior. *Computers in Human Behavior, 43*(2), 308–312.

Brock, J. K.-U., & von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review, 61*(4), 110–134.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528.

Economist. (2019). *The world in 2019*. Available at: https://www.economist.com/the-world-in/2018/12/17/there-are-no-killer-robots-yet-but-regulators-must-respond-to-ai-in-2019

Eldridge, R., Harlan, A., Cooper, I., & Riklan, M. (1970). Superior intelligence in recessively inherited torsion dystonia. *The Lancet, 295*(7637), 65–67.

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics, 21*(4), 403–418.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280.

Haenlein, M., & Kaplan, A. M. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14.

Huang, M.-H., Rust, R. T., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of AI. *California Management Review, 61*(4), 43–65.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Kaplan, A. M. (2012). If you love something, let it go mobile: Mobile marketing and mobile social media 4x4. *Business Horizons, 55*(2), 129–139.

Kaplan, A. M. (2018). A school is a building that has 4 walls – with tomorrow inside": Toward the reinvention of the business school. *Business Horizons, 61*(4), 599–608.

Kaplan, A. M., & Haenlein, M. (2009). The fairyland of Second Life: About virtual social worlds and how to use them. *Business Horizons, 52*(6), 563–572.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68.

Kaplan, A. M., & Haenlein, M. (2012). Social media: Back to the roots and back to the future. *Journal of Systems and Information Technology, 14*(2), 101—104.

Kaplan, A. M., & Haenlein, M. (2014). Collaborative projects (social media application): About Wikipedia, the free encyclopedia. *Business Horizons, 57*(5), 617—626.

Kaplan, A. M., & Haenlein, M. (2019a). Digital transformation and disruption: On big data, blockchain, artificial intelligence, and other things. *Business Horizons, 62*(6), 679—681.

Kaplan, A. M., & Haenlein, M. (2019b). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15—25.

Keating, J., & Nourbakhsh, I. (2018). Teaching artificial intelligence and humanity. *Communications of the ACM, 61*(2), 29—32.

Keynes, J. M. (1930). Economic possibilities for our grandchildren. In J. M. Keynes (Ed.), *Essays in persuasion*. London, UK: Macmillan.

Krotov, V. (2017). The internet of things and new business opportunities. *Business Horizons, 60*(6), 831—841.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293—303.

Leveringhaus, A. (2018). What's so bad about killer robots? *Journal of Applied Philosophy, 35*(2), 341—358.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. Natick, MA: AK Peters Ltd.

Metcalf, L., Askay, D., & Rosenberg, L. (2019). Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. *California Management Review, 61*(4), 84—109.

Mosser, C. (2018). Mechanical dreams of humanity: When machines long for humanity. *Interdisciplinary Humanities, 35*(1), 78—90.

Müller, V. C., & Bostrom, N. (2014). Future progress in artificial intelligence: A survey of expert opinion. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence*. Berlin, Germany: Springer.

Philipps, W. (2015). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. Cambridge, MA: The MIT Press.

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: Harvard University Press.

Porter, M. (2011). Creating shared value. *Harvard Business Review, 89*(1/2), 62—77.

Pucciarelli, F., & Kaplan, A. (2016). Competition and strategy in higher education: Managing complexity and uncertainty. *Business Horizons, 59*(3), 311—320.

Rossi, F. (2019). Building trust in artificial intelligence. *Journal of International Affairs, 72*(1), 127—133.

Saarikko, T., Westergren, U. H., & Blomquist, T. (2017). The internet of things: Are you ready for what's coming? *Business Horizons, 60*(5), 667—676.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*, 484—489.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, 49*(4), 376—385.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics, 36*(4), 95:91, 95:13.

Tambe, P., Cappelli, P., & Yacubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15—42.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453—458.

Van Parijs, P. (2004). Basic income: A simple and powerful idea for the twenty-first century. *Politics & Society, 32*(1), 7—39.

VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.-Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology. *Environmental Health, 17*(40), 1—6.

World Economic Forum. (2018). *Machines will do more tasks than humans by 2025 but robot revolution will still create 58 million net new jobs in next five years*. Available at: https://www.weforum.org/press/2018/09/machines-will-do-more-tasks-than-humans-by-2025-but-robot-revolution-will-still-create-58-million-net-new-jobs-in-next-five-years/

Yampolskiy, R. V. (2014). Utility function security in artificially intelligent agents. *Journal of Experimental & Theoretical Artificial Intelligence, 26*(3), 373—389.

Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). *Few-shot adversarial learning of realistic neural talking head models*. Cornell University. Available at: https://arxiv.org/abs/1905.08233

Zhu, J., Huang, T., Chen, W., & Gao, W. (2018). The future of artificial intelligence in China. *Communications of the ACM, 61*(11), 44—45.