Ethical Implications of the 4th Industrial Revolution for Business and Society

John Hooker and Tae Wan Kim
Tepper School of Business
Carnegie Mellon University

A book chapter of Business & Society 360 (Emerald)

Abstract
Businesses are rapidly automating workplaces with new technologies (e.g., driverless cargo trucks, artificially intelligent mortgage approvals, machine-learning based paralegals, algorithmic managers). Such technological advancement raises a host of questions for business and society. As Thomas Donaldson recently remarked, “It’s instance of a problem that more sophisticated engineering cannot solve, and that requires a more sophisticated approach to values” (Ufberg, 2017). In this chapter, we explore the value questions as follows: What is the purpose of business in the machine age? What model for business will best serve society in coming decades: profit maximization, stakeholder theory, or another conception entirely? Is it time for a new social contract between business and society? Do firms have a natural duty to offer employment? Are existing concepts of responsibility/liability adequate for an age in which companies use autonomous robots as scapegoats? How can we protect our humanity and dignity in an algorithm-based society? Do we need to teach ethics to robots?

Robots are coming. Computational systems are becoming increasingly autonomous and displacing current capacities at an exponentially increasing rate. Businesses are rapidly gearing up to automate workplaces by utilizing state-of-the-art technologies, including but not limited to autonomous customer assistance systems, driverless cargo trucks, artificially intelligent hedge funds, autonomous mortgage approvals, robotic paralegals, and autonomous killing machines. Such technological advancement raises a host of questions about value and humanity: If artificial intelligence harms someone without meaningful human control, who must be held responsible? If robots take over human jobs on an unprecedented scale in the near future—in the era Brynjolfsson and McAfee (2014) dub “the Second Machine Age,” and Schwab (2016) calls “the Fourth Industrial Revolution”—should the government offer a basic income to everyone?
Business Ethics and the 4th Industrial Revolution

In this chapter, we explore what role and responsibility business enterprises can and should take for the future of society. The chapter is comprised of three parts. The first part explores various topics in employment in the coming society: the gig economy, technological unemployment, meaning of work, basic income, and a duty to hire. The second section explores three ethical issues in artificial intelligence: machine ethics, explainable AI and AI as scapegoat. The final section introduces and defends business’ accountability of augmentation.

**The Gig Economy**

Entertainers often depend on gigs, or temporary engagements, for income. This practice is rapidly spreading across the economy at large, as organizations increasingly rely on temporary positions and contractual workers rather than employees. Technology is an increasingly important driver of this trend. It is said that the gig economy is the economy of the future, or even the present. This raises ethically urgent questions, because gigs may not provide the economic security or legal protections of employment. In particular, one can ask if companies shirk their ethical duty when they rely on temporary workers.

A gig economy is often associated with a sharing economy, perhaps because both use online services such as TaskRabbit, Airbnb or Uber. Yet the two are quite different. A sharing economy is one in which individuals share their possessions with others for a fee, such as tools or recreational vehicles. TaskRabbit supports a gig economy because it recruits people to do small jobs, while Airbnb is part of a sharing economy because it allows users to share their homes. Uber happens to participate in both kinds of economy because it connects drivers with riders as well as allowing them to share their cars.

The connection with online apps is misleading in another way, because they have so far accounted for only a small portion of the rising gig economy. The last two decades have seen a
Business Ethics and the 4th Industrial Revolution

quiet shift from full-time employees to workers who contract for short-term projects or obtain work through temporary employment agencies. This results in a large “contingent” or “freelance” workforce. Construction workers have long been part of it, but its numbers are growing rapidly. Hotels and restaurants, for example, frequently outsource tasks to temporary workers rather than hiring employees to do them, a practice that is largely invisible to customers. According to one study, 36% of U.S. workers were freelancers in 2017 (Edelman Intelligence, 2017).

Technology is an important factor in this shift. The ability to work online rather than in the office makes workers more independent of the firm. Technological change requires every-changing skills, and it is often easier to find these skills in the gig economy than to teach them to long-term employees. Online apps come in on top of this trend and will doubtless accelerate it.

Purely economic incentives play a major role as well. It is often cheaper to outsource work than to assign it to employees. This is sometimes because outsourcing the work is inherently more efficient. For example, it may be more efficient for a single trucking firm to deliver goods for several bakeries than for each bakery to employ its own drivers. Yet the contract workers themselves are typically cheaper than employees, aside from any efficiency advantage. They are unprotected by overtime and labor laws and typically receive no health coverage or other benefits. When demand for a product drops, there is no need to pay the same number of employees to do less work or to undertake expensive layoffs and rehires. Gig workers are also less likely to unionize, although there are movements underway to change this (Davidson, 2016).

Compounding these incentives is the brute fact that base wage rates for contract workers equilibrate at a lower level than for employees doing the same job. In one striking example, a
transcriptionist employed at University of Pittsburgh Medical Center saw her hourly wages drop from $19 to $6 per hour when the hospital required her to work for an outside transcription service—even though she did the same work at the same desk! (Vinik, 2018) Economic theory may or may not predict this outcome, but it is the current reality, and possibly the future reality. Perhaps the social norms that historically accompany employment simply evaporate when the worker is no longer “part of the family.” The gig economy may represent cultural change as well as an economic or technological phenomenon.

Whatever the case, the future is rapidly approaching, and it is crucial to understand, at a fundamental level, what it means to abandon the employment model. In principle, the state can provide health coverage and other benefits in a gig economy. Labor laws can regulate compensation for a certain kind or quantity of piecework, much as price controls regulate the cost of consumer goods or apartments. This suggests that the essence of employment is job security. Employment-at-will, a peculiarly U.S. practice, provides little or no job security, but perhaps this kind of employment is employment in name only. Employees-at-will receive some legal protections, but again, this is incidental to the concept. It therefore seems reasonable to define employment as an engagement that provides some degree of job security—either through contract, as in the U.S., or by virtue of labor law, as in Europe.

Defenders of the gig economy point out that one can have economic security without job security. If an independent worker can go online and find suitable work at any time, a steady job is unnecessary. Maybe pickup work will be abundant once the once the gig economy is dominant, much as Uber customers are abundant in some cities. Yet it is hard to believe that gig workers will be immune to the business cycle. The very characteristics of the gig economy that
Business Ethics and the 4th Industrial Revolution

protect employers from fluctuations in demand require that workers bear the brunt of these fluctuations.

Even if economic security does not require job security, it presupposes a living wage. As already noted, compensation is often significantly less for gig workers than employees doing the same work, thus exacerbating inequality. Even a given level of income is less valuable when it uncertain, because funds must be kept in reserve for lean periods.

As online apps assume a larger role in the gig economy, they may create another source of inequality. Social networking sites, sharing apps, and brokerage services compete in a largely winner-take-all arena. The winners establish near-monopolies and accumulate wealth that is far out of proportion to the value they create. It is relatively easy to set up networking and brokerage functions online. In exchange for this limited contribution, these sites extract substantial wealth from the economy, often to the benefit of a few at the expense of many (Rushkoff, 2016).

The coming gig economy demands an ethical assessment. As for the individual firm’s responsibility, one might make a case that switching from employees to contingent workers is ethically problematic in the early stages of a labor market transition. Yet this issue will quickly become moot as the freelance economy sets in. Companies will be unable to compete unless they exploit the cost advantages of freelance workers. It is necessary to examine the ethical pros and cons of an established gig economy to determine whether it should be regulated or reshaped by society at large.

The benefits of the gig economy to companies have already been outlined. Obvious benefits to workers include flexibility and freedom. Flexibility allows workers to take time off for child care, illness, and other family issues without asking anyone’s permission. Freedom
Business Ethics and the 4th Industrial Revolution

allows them to switch more easily between different kinds or work, and focus at times on avocations, pro bono activities, or artistic creation. This arguably promotes self-realization and life satisfaction, as well as releasing volunteer labor that benefits society in ways that paid labor cannot.

The liabilities of the gig economy for workers add up to unpredictability, economic insecurity, and the stress that results from lack of control over one’s life. The psychological and health effects of excessive stress are well documented. Indeed, they incur economic costs that tend to offset the claimed efficiencies of a gig economy. One might argue that society should at least insulate workers from such major shocks as serious illness, long-term work shortages, and exploitative wages.

Inequality raises further ethical issues that are already the subject of much discussion. Two basic philosophical positions have evolved. One regards inequality as intrinsically evil or incompatible with a broader concept of social justice (Parfit, 1997; Temkin, 1993). The other sees the evil of inequality as reducible to that of poverty or deprivation that results from it (Frankfurt, 2015). In any event, the inequalities we see in the real world almost always marginalize lower classes, sometimes in extreme ways. This only exacerbates the insecurity and stress that accompany an unpredictable labor market.

Perhaps the fundamental ethical question for the gig economy stems from its uncertainty and the resulting existential stress. Government regulation can conceivably regulate wage rates offered by a gig economy, but it is unclear how the government can tame its inherent volatility. An independent worker’s struggle to stay afloat in a huge and unpredictable labor market may be incompatible with human flourishing.

**Technological Unemployment**
The AI revolution could lead to a more radical outcome than the gig economy. It could displace workers on a scale that has not been seen in history. A large fraction of the population could lose the opportunity to work for money. The reality of this threat has been hotly debated in recent years, with no consensus in sight. Everyone agrees that AI will eliminate a large fraction of the existing labor market, but many insist that new opportunities for labor will arise, much as they did during and after the Industrial Revolution. Others say it will be different this time.

AI has already taken over much of manufacturing. On many a factory floor, there are no human beings in sight. Next to go will be business services. Warehousing and retailing are shedding workers by the day, and the automation of freight transport and delivery is just around the corner. AI software based on “deep learning” in neural networks is targeting a wide array of white-collar services. These include loan application processing, investment advice, credit rating, online ad placement, college admissions, job applicant screening, and even parole decisions for prisoners (O’Neil, 2017). It is often assumed that personal services and the professions are immune for the time being, but even they are under attack. Robots are replacing aides in some nursing home and home care roles, and there is anecdotal evidence that their charges bond with the robots almost as though they are humans (or at least pets). Machines are rapidly overtaking medical technologists, and even doctors, in their ability to interpret scans and diagnose illness. Education is rather easily automated, at least where basic skills are concerned.

During the Industrial Revolution, textile workers feared that weaving machinery would threaten their jobs. The workers attacked machines under the leadership of a mystical figure called Ned Ludd. Those who followed the figure were called “Luddites.” The economic history of the 19th century shows that Luddites were wrong. Capitalism is dynamic. Technological
Business Ethics and the 4th Industrial Revolution

innovation in aggregate created more new jobs than were lost. To deny this is the “Luddite fallacy” (Autor 2015).

We may indeed get through the AI revolution much as we got through the Industrial Revolution in the 19th century, but “getting through” the latter entailed 150 years of political upheaval, violence, and economic suffering. It was not until the 1940s or 1950s that developed economies began to deal seriously with the dislocation caused by industrial employment, mainly through labor laws and welfare policies. While the Industrial Revolution created legions of new blue and white-collar jobs, it is difficult to imagine new job categories that the AI revolution might create on a similar scale.

The Luddite fallacy may therefore not be a fallacy in the near future. Although technological innovation creates more new jobs, it is likely that workers will compete with robots for them (Brynjolfsson and McAfee 2014). Those who lose jobs will need to be reeducated, which takes months or years, whereas smart systems with the power of advanced algorithms and big data will learn the necessary skills far faster than human workers. The upshot, then, is that robots will take many of the new jobs that the dynamic nature of capitalism will create through technological innovation.

The above argument is corroborated by a growing number of data-driven reports. The *Oxford Martin Programme on the Impacts of Future Technology* (Frey & Osbornem, 2013) predicts that 47 percent of total U.S. employment will likely be automated over the coming twenty years. Using the Oxford research and others, the *Economic Report of the President to the Congress* (2016) predicts that the probability of robots taking over the lowest-paid jobs (hourly wage less than 20 dollars) in the coming decades is 0.83, the middle-paid jobs, 0.31, and the highest-paid jobs (more than 40 dollars), 0.04. The consulting firm *McKinsey* similarly predicts
that “as many as 45 percent of the activities individuals are paid to perform can be automated by adapting currently demonstrated technologies” (Chui, Manyika, Miremadi, 2015). Finally, the World Bank predicts that about 65 percent of jobs in developing countries are at serious risk of automation (World Bank, 2016).

Inequality

We have no reason to completely accept the predictions above. At any rate, it is difficult to predict the future. Nonetheless, the real possibility of technological unemployment in the coming machine age is sufficient to lead us to several questions about the role and responsibility of business in society. Remarks by W. Brian Arthur well summarize one of the questions. Arthur (2011) states,

“The second economy [the coming machine age] will certainly be the engine of growth and the provider of prosperity for the rest of this century and beyond, but it may not provide jobs, so there may be prosperity without full access for many. […] [W]e face a problem.”

One might say, at this juncture, “Why is business responsible for this? Isn’t this why we have the government?” There are at least two philosophical arguments why business should be held responsible. The first is the so-called “responsibility principle,” according to which you are responsible for minimizing any negative impact and cleaning up the negative impact you made. Unemployment through automation, especially, at a massive scale, can be understood as a negative externality, as is pollution created by fossil fuels, for instance. Just as responsible firms should minimize and remedy pollution by sustainability initiatives, responsible firms have a duty to minimize and remedy the negative impact of technological unemployment upon society. Second, from a Rawlsian perspective, technological unemployment is justified only if the worst-off group---the unemployed or those who systematically lack opportunities for gainful
Business Ethics and the 4th Industrial Revolution

employment---has reason to accept the societal structure in which massive technological unemployment exists.

**Market-Based Redistribution**

Societal remedies for technological unemployment are often identified with government redistribution programs, such as the universal basic income to be discussed below. Yet the economy itself already performs a massive wealth redistribution function. The task before us is not so much to introduce redistribution as to reform and strengthen redistribution mechanisms that already exist, whether through government or by other means.

The primary role of redistribution in advanced economies is to spread the benefits of manufacturing, which enjoys extraordinary labor productivity. Greater labor productivity means a higher standard of living, because the economy generates more wealth in exchange for the same amount of human effort. Yet for society as a whole to enjoy this wealth, it must somehow be transferred from factory workers to the population at large.

There are several reasons that an hour of human labor in a modern factory generates such enormous benefit. One is that a manufactured good is a “gift that keeps on giving.” A refrigerator assembled in a matter of minutes can provide service for decades. A second is that technology allows the few workers who run a factory to generate prodigious output. A third factor, less often acknowledged, is the series of innovations in operations management introduced by Japanese manufacturing in the 1970s and 1980s. They led to a revolution in manufacturing efficiency worldwide.

The market value of manufactured goods does not reflect their true value, because their prices are low due to the very efficiency of manufacturing. The low market prices are, in fact, one way the economy redistributes the value created by factory workers. Another mechanism is
Business Ethics and the 4th Industrial Revolution

the distribution chain, which rewards a wide of array of occupations for supplying and servicing the fruits of industry. Other mechanisms are more controversial. Owners of the means of production appropriate part of the value created by labor. The financial sector collects interest due to its control of money and credit. Companies and individuals that control scarce resources collect economic rent. Persons who hold powerful positions in society leverage those positions to accumulate wealth, and so forth.

Technological unemployment radically intensifies the need to redistribute wealth, not only from manufacturing but also from other sectors, as they become more labor productive. One does not have to be a leftist to acknowledge flaws in the haphazard way a market economy accomplishes this. Even staunch libertarians F. A. Hayek and Milton Friedman advocated a universal basic income (Hayek, 1960; Friedman, 1962) as a non-market distribution mechanism.

**Universal Basic Income**

A universal basic income is perhaps the most straightforward non-market corrective for structural inequality. It is usually defined as “an income paid by a political community to all its members on an individual basis, without means test or work requirement” (Parijs, 2004, 8). A number of countries have experimented with the idea on a small scale, and most report salutary effects despite the perceived risk that it will encourage laziness. A few programs have provincial or national scope, such as the Alaska Permanent Fund, which pays out $800-2000 annually in oil revenues to every resident of the state. The largest national program is Brazil’s *Bolsa Familia*, which makes small cash payments to poor families who send children to school and have them vaccinated. Similar programs have been adopted in Mexico (the first), Chile, Honduras, Jamaica, Malawi, and Zambia. None of the national programs are truly universal, however, since they are conditional on prior income level and recipient behavior.
Business Ethics and the 4th Industrial Revolution

An obvious barrier to a successful transfer program is political resistance to extracting enough money from productive sectors to finance an adequate subsidy. Deeper objections relate to the lack of autonomy of recipients. A government-sponsored transfer system leaves many people with limited control over their lives. Many of those who wish to better their lot beyond the basic income are unable to do so, because the economy provides them no opportunity to earn money. A consistent critic of basic income programs has been the Roman Catholic Church, whose objections are based on the Thomist idea that humans have a natural right to own or control property necessary for their sustenance, including the means of production (Aquinas, Part II, Question 66, ca. 1265-1274).

Meaning of Life

A universal basic income, if properly executed, would offer basic welfare to those who systematically lack work opportunities. However, the goal of basic income is not just to offer material comfort but also to give a broader spectrum of work choices. Left-libertarians have defended a basic income on the ground that it provides more freedom to choose the kind of work and working conditions one pursues. A premise of this argument is that people have a better chance to live a meaningful life through meaningful work. Yet in the coming machine age, in which much of the population systematically will lack work opportunities, this philosophical justification for basic income will not be fully realizable. To the extent that the unemployed wish to find meaning through work, they will be frustrated. They can in principle contribute through volunteer labor, but contributions that make a real difference will often require capital or wealth beyond their means. Thus, the problem of materialistic inequality can be mitigated by basic income, but the problem of eudemonistic or axiological inequality perhaps cannot.
Of course, work may not be not necessary for a meaningful life. Rethinking the connection between work and a meaningful life will, in fact, become an important element of the coming age. People may re-educate themselves to mitigate the construct of a work-oriented civilization. The business sector can and should promote such education. At the same time, work has long played a central role in human civilization, and it is unrealistic to suppose that society can fundamentally reorient its stance toward work within the coming two decades.

**Duty to Hire?**

The problem of how a meaningful life relates to work is directly relevant to business ethics, in particular to the literature about corporate purpose, including but not limited to shareholder theory and the stakeholder perspective. Whether the problem can be mitigated or not depends largely upon society’s choice of model for corporate purpose in the coming decades. Corporations exist because society allows them to exist for some purpose. If the current model of corporate purpose, e.g., maximization of profit through maximized automation, is not adequate for future society, we have reason to rethink the model. It is an open question whether the capital-owner-profit-maximization model will be a legitimate corporate purpose in the coming decades, even if the model has been workable for previous decades (Kim and Wolf, forthcoming).

Shareholder profit maximization is not a natural duty; it is a human artifact, at best a social contract. Before the rise of the Chicago School in the 1970s, most corporate executives believed that corporations had pluralistic purposes. “Fifty years ago, if you had asked the directors or CEO of a large public company what the company’s purpose was, you might have been told the corporation has multiple purposes: to provide equity investors with solid returns, but also build great products, to provide decent livelihoods for employees, and to contribute to
Business Ethics and the 4th Industrial Revolution

the community and the nation.” (Stout 2012, p. 3) A recent Aspen Institute survey shows that even contemporary corporate executives do not believe that serving shareholders’ interests is the most important purpose of a corporation. Most important is to “serve customers’ interests,” second is to “deploy large amounts of capital for projects that strengthen society,” while “serve employees’ interests” and “serve shareholders’ interests” tie for third place.¹

Then, do corporations or firms more generally have a natural duty to hire human workers? This question deserves deep analysis, but a short answer may be that corporations are legal entities, and so it may be a stretch to say that corporations have natural duties. Yet it is also plausible that having options to create a meaningful life is a natural right for free and equal citizens in a society. A meaningful life is strongly connected to the option to belong to an organization in which individuals are encouraged, even pressured, to develop virtue and excellence on a regular basis, and by doing so contribute to the common good. In contemporary society, corporations/firms frequently serve this purpose. Then the business sector’s duty to hire could be recognized as an extended natural duty, to the extent that it serves citizens’ natural right to have options for a meaningful life.

Third-Way Solutions

“Third way” proposals address the causes of unequal distribution without making a commitment to either socialism or laissez-faire capitalism. One much-discussed source of inequality is the concentration of productive capital in relatively few hands. A possible remedy is to facilitate widespread ownership of the means of production, perhaps in the form of stock holdings in high-tech companies (Albus, 1976; Rotman, 2015; Smith, 2013). We are already seeing a growing

¹Aspen Institute’s Survey is found here: https://assets.aspeninstitute.org/content/uploads/files/content/upload/Unpacking%20Corporate%20Purpose%20May%202014.pdf
Business Ethics and the 4th Industrial Revolution

share of income from capital rather than labor (Piketty, 2015), and universal ownership of capital would harness this trend rather than try to resist it. The question, of course, is how to bring about universal ownership.

One strategy is to move toward worker ownership of firms. Worker ownership is already a well-established practice in many countries, with perhaps the most famous example being the vast Mondragon conglomerate in the Basque region of Spain. Yet worker-owned companies must obtain startup capital from somewhere, the most common source being loans from banks, governments, or NGOs. This tends to concentrate power in lending institutions, defeating one of the purposes of universal capital ownership.

Another strategy is to call on the government to provide everyone a capital inheritance, analogous to the inheritance some receive from their families. It is also known as a “capital homestead,” in reference to a 19th century U.S. land grant program. Grants of this kind do not centralize power as loans do, because individuals can do what they will with their inherited capital, without having to repay the state. They can invest the money in education and any enterprise they please, thus exercising control over their lives. A hefty capital gains tax seems an inevitable part of such a plan, because the next generation of capital grants must somehow be financed. There are also a number of practical issues. For example, people who squander their inheritance can presumably be supported by a modest basic income, but it is unclear what to do about those who go into bankruptcy through bad luck.

A second source of inequality is centralized control of the money supply. Many third-way advocates support the use of alternative currencies that are independent of banks and unregulated by governments. These currencies have attracted enormous attention in recent years (Lietaer & Dunne, 2013; Ammous, 2018). The best known is Bitcoin, a digital cryptocurrency in
Business Ethics and the 4th Industrial Revolution

which transactions are anonymously recorded online by blockchain technology. Bitcoins are “minted” by awarding them to individuals who help supply the massive computing power needed to maintain blockchains. There are hundreds of other, lesser-known currencies, many intended for local use.

Alternative currencies are created for many purposes, but most relevant to technological unemployment is their ability to establish an informal market apart from the main economy. For example, a group of individuals may wish to exchange goods and services but lack enough of the national currency to support such an exchange. They may wish to sell home-grown food, crafts, or artistic performances, when their productive efficiency is too low to compete in the main economy or attract financial capital. Yet productivity is not essential, since the purpose of the exchange is to supplement income from the main economy while providing the satisfaction of meaningful work. An alternative currency can enable this sort of activity, and in so doing help provide the kind of autonomy that third-way proposals seek.

A third source of inequality is economic rent collection. Those who control scarce resources such as land, minerals, energy, and intellectual property often exploit this control to extract wealth from the economy. One possible strategy to minimize rent collecting is to implement a “post-scarcity economy,” in which scarcity is converted into abundance (Aguilar-Millan, Fenney, Oberg, & Rudd, 2010). The AI revolution plays a major role in this conception, because it produces goods and services with very little human labor and therefore makes them more abundant. Perhaps this can be extended to energy resources through the efficient production of solar energy systems, given the inherent abundance of energy from the sun. Energy is the most basic resource because it can create other resources, such as fresh water through desalinization or minerals through recycling. It was the primary driver behind the first
Business Ethics and the 4th Industrial Revolution

Industrial Revolution and can perhaps similarly power the next one, in part by eliminating rent collection on once-scarce resources.

As for intellectual property, we are already seeing a movement away from scarcity. The GNU project was an early harbinger of this trend in its encouragement of open-source software (Stallman, 1985), and enormous volumes of freely available content are now posted online. Copyright and patent protections remain, but they create artificial scarcity because ideas are infinitely reproducible. Partly for this reason, there is arguably no Lockean natural right to intellectual property, and the very term “intellectual property” may be a misnomer (Steidlmeier, 1993). In any case, the growing “creative commons” may point to an altruism economy in which many goods and services are freely donated.

The altruism economy is an obvious challenge to traditional business practice, which for centuries has been based on the profit model. Yet even now, altruistic motives play a major role in the world of work. A nurse’s compassion, a teacher’s dedication to learning, and a firefighter’s determination to save lives are not motivated by their modest salaries. Good business managers know that an opportunity to make a positive contribution to society is a strong motivator, and countless successful organizations are not based on a profit model in any case. The human tendency to altruism is well documented (Ricard, 2015), and anthropologists tell us that human culture began with gift economies in which goods and services were given without an explicit agreement for compensation (Cheal, 2015). Perhaps, after a long detour into scarcity economics, human society is beginning to return to its original model of exchange.

**Three Issues in AI ethics**

As technology moves in the direction of artificial intelligence, we will face AI systems that make decisions in ways that humans have no feasible way to foresee. Consider the following scenario:
Business Ethics and the 4th Industrial Revolution

You are a person from a racially underrepresented group, say black, and you recently applied to an online mortgage approval system and were rejected. The bank that hosts the online application system has recently started using AI to recommend mortgage applications for approval. You happened to know that the bank’s approval rate for clients of the same race as yours has recently abnormally decreased, for no good reason. You meet with a representative of the bank and claim that the bank has racially discriminated against you, and that the bank should be held liable for the discrimination (e.g., remedying the problem, compensating for any physical or psychological losses, if any, and taking action to prevent similar problems in the future). The bank representative says that it is impossible for the autonomous artificial agent to discriminate racially against applicants, because the algorithms it uses were designed to be indifferent to the race of applicants. To prove that, in front of you the representative submits ten fake applications equally qualified as yours (as judged by independent human evaluators) that consist of 5 whites and 5 blacks. The AI accepts all white applicants but only 2 black. The representative looks puzzled.²

This kind of scenario raises at least three different ethical issues a) a duty to design AI to behave ethically, b) a duty to develop explainable AI, and c) a duty to voluntarily “take” responsibility. We discuss the three issues in what follows.

**Machine Ethics**

How can businesses that use autonomous artificial agents effectively prevent the agents from incurring losses to humans, e.g., racially underrepresented mortgage applicants? To address this issue, there has been a rapidly growing scholarship under the name of “machine ethics” (e.g., Anderson & Anderson, 2011; Wallach & Allen, 2009). At some point of our near future, we will

² We borrow the example from Bostrom and Yudkowsky (2014) with some added details.
see remarkably advanced autonomous artificial intelligence that has the capability to render itself, on some abstract level, as an intentional agent who can set principles or reasons upon which it makes decisions and acts. A primary purpose of the scholarship of machine ethics is to develop algorithms that such intentional artificial agents can learn as environmental data and use for their decision-making. For instance, if the artificially intelligent robot in the scenario above can learn about moral principles underlying ideas of equal respect, equal citizenship, etc., the machine would know whether the approvals she makes are consistent with principles of equal respect and fair treatment. In this sense, the computer scientist Stuart Russell (2015) expects that in the near future ethics will be a lucrative business in tech industries, because autonomous robot companies will probably collaborate with “values companies” that provide ethics algorithms to autonomous robots.

A dominating model of machine ethics is human intuition-based, in which training data for AI is based on opinions made by humans. MIT Media Lab’s research project, “The Moral Machine Experiment: 40 Million Decisions and the Path to Universal Machine Ethics,”\(^3\) gathered 40 million opinions from 3 million people in 200 countries about how people intuitively make choices about the trolley dilemma-like scenarios that involve autonomous vehicles. Using the data, they aim to offer “global, harmonious, and socially acceptable principles of machine ethics.” But the intuition-based approach is a poor standard for moral decision making by autonomous machines. Just because 40 million people think that X is ethical does not mean that X is ethical. Just because bribery is a social norm in a certain culture does not mean that bribery is ethical in that culture. There is sufficient evidence that humans’ intuitive decisions are susceptible to situational cues. A prime motivation for developing autonomous vehicles is that

\(^3\) The research was presented at AAAI Conference on Artificial Intelligence, Ethics and Society 2018 (http://www.aies-conference.com/invited-talks/#invited_talk1).
human error is a leading cause of accidents, and autonomous vehicles minimize human involvement. Thus, if human moral intuitions are not reliable, machine ethics should be developed so as to avoid human moral intuition. We develop a principle-based Kantian machine ethics elsewhere (Hooker & Kim, 2018). Our deontological — or principle-based — approach would remove the human biases inherent in intuition-based models. In addition, the processes and outcomes can be rigorously expressed via computational language, which is more consistent and transparent.

**Explainable AI and a Right to Explanation**

The thinking machine’s algorithms are based on complicated neural networks, and, in the mortgage application scenario, it is almost impossible to investigate why and how the autonomous artificial agent rejected black candidates, including you. The autonomous thinking machine in the scenario might have learned from already biased existing data that reflected preexisting patterns of inequality in society (Crawford, 2013). If the machine had learned from the environmental data generated by past human brokers who had implicit racial prejudices and had denied black candidates’ applications based on these biases, the appraisal machine’s decisions might have followed the past prejudices and accordingly rejected qualified applicants who fell into historically discriminated groups. And this kind of process is in most cases an emergent property of autonomous artificial agents, so it is in many cases inherently difficult for human users (the manager of the bank, in our case) to foresee what specific problems may occur or pinpoint why they occur (Selbst & Barocas, 2016).

Due to the un-explainable nature of some kinds of AI, e.g., deep learning, there is growing demand for developing explainable/interpretable AI. There are, largely, two different kinds of explanations in terms of content, as follows:
• **System functionality**, i.e., the logic, significance, envisaged consequences, and general functionality of an automated decision-making system, e.g., the system’s requirements specifications, decision trees, pre-defined models, criteria, and classification structures;

• **Specific decisions**, i.e., the rationale, reasons, and individual circumstances of a specific automated decision, e.g., the weighting of features, machine-defined case-specific decision rules, information about references or profile groups (Wachter, Mittelstadt, & Floridi, 2017, p. 78).

Additionally, there can be two different kinds of explanations, in terms of timing, as follows:

• An *ex ante* explanation occurs prior to an automated decision taking place. Note that an *ex ante* explanation can logically address only *system functionality*, as the rationale of a specific decision cannot be known before the decision is made;

• An *ex post* explanation occurs after an automated decision has taken place. Note that an *ex post* explanation can address both *system functionality* and the rationale of a *specific decision* (ibid.).

Hence, there can be three different categories of possible explanations as follows:

• An *ex ante* explanation about system functionality (or an *ex ante* generic explanation)

• An *ex post* explanation about system functionality (or an *ex post* generic explanation)

• An *ex post* explanation about a specific decision (or an *ex post* specific explanation)

The first kind, an *ex ante* explanation, is similar to the kind of explanation given to users under the name of informed consent. So, there is not much controversy about the importance of *ex ante* explanation. But a right to *ex post* explanation has been underexplored.

A trust-based argument can be made to defend a right to *ex post* explanations. Informed consent can be understood as a promissory transaction that inherently involves invitation of trust. When data subjects (users, employees, or applicants) consent to a certain company’s collecting and processing their personal data, which inherently involves errors, risks, and uncertainty that cannot be reasonably well explained in an *ex ante* manner, data subjects cannot but take the company’s action of obtaining informed consent as assurance. That is, within the context of informed consent, the data subjects have reason to take the company’s act of obtaining informed consent, to assert, therewith, to invite data subjects to *trust* the company. Thus, similar to the promise in algorithmic contexts that we discussed, companies, by obtaining informed consent from data subjects, commit themselves to a certain moral demand—that is, an obligation not to breach the trust that data subjects are invited to place in them. But data subjects should not
blindly place trust in companies. Under what conditions are data subjects justified in placing trust in companies that use automated algorithmic decision systems? First of all, no companies can assure that there will be no risks or uncertainties. However, it is reasonable for data subjects to expect companies to assure them that once harms or wrongs occur, the company will respond in a fair and responsible manner. To meet this demand, some *ex post* explanation is essential (Kim and Routledge, 2018).

**AI as Scapegoat and the Principle of Fair Reciprocity**

Even if AI learns the best interpretations of machine ethics and explainable AI is feasible, there may be the risk that AI *accidentally* harms humans in ways that human users cannot specifically foresee or effectively avoid. More importantly, it is almost pointless to hold autonomous robots responsible for harms that they create. For instance, as in our mortgage scenario above, autonomous artificial agents would not have the managerial authority to fix problems that they created or any pecuniary resources to compensate for losses. Companies might want to use robots as a *scapegoat*, which would facilitate serious moral hazard in the coming AI age.

Imagine that the discriminated applicant above asks the representative of the bank, “Why did you do that?” The human representative answers, “We are so sorry that this happened to you. But no one in the bank did it; it is just misfortune. There is no good way for us to predict or control this kind of accident.” Intuitively, the representative’s moral defense sounds terribly irresponsible, but the defense could be philosophically further articulated and grounded as follows:

*The Scapegoat Argument*

- P1) Agent A is morally responsible for Action X, which means just that X is properly *attributable* to A in a way that renders A open to moral appraisal for performing X.
- P2) Agent A is open to moral appraisal for Action X just when X is expressive of A’s reflective or deep self or practical agency.
P3) Action X is expressive of Agent A’s self or agency only when X identifies with A’s desires, reasons, attitudes, or commitments that move A to perform X, whereas X is not expressive of A’s self or agency when X does not identify A’s desires, reasons, attitudes, or commitments, especially when A does not have volition or control over doing X or A cannot be aware of X.

P4) In the mortgage bank case above, the racial discrimination was not expressive of any humans’ desires, reasons, attitudes, or commitments, and none of the humans’ practical identities moved the thinking machine to racially discriminate. More simply, the humans did not have volition or control over the autonomous artificial mortgage appraiser’s creating the emergent property of racial discrimination and the humans in the bank were not able to be aware of the autonomous machine’s discriminative appraisal.

C) Thus, the humans in the bank are not morally responsible for the outcome action.

As premises 1 through 4 show, asking whether a certain agent is morally responsible for some action is the same thing as asking whether the action is properly attributed to the agent in the way that the connection between the agent and the action is used as a ground for moral appraisal. If the action is morally bad or wrong, the attributable connection can be used as a reason for making a negative moral appraisal (e.g., blame, reproach, or indignation) to the agent who is attributed to the action.

The bank’s reason for using the autonomous machine was in fact not to racially discriminate against underrepresented races. Thus, the outcome action was not really expressive of the people’s judgment-sensitive attitudes. In short, there is no attributable connection between the humans in the bank and the artificial agent’s ethically problematic performance. The applicant who is racially discriminated against does not have a basis for morally condemning the humans in the bank.

It is worth noting that the scapegoat argument is based upon a certain conception of moral responsibility, often called “attributability-responsibility.” It is also worth noting that the attributability-responsibility is not the only conception of responsibility that plays an important role in our moral lives. There is another important kind of responsibility, one that the
representative of the bank ignored, and one that can make sense of our belief that the bank’s scapegoat defense is irresponsible. It is the conception of “responsibility as accountability” (Watson, 1996). Most fundamentally, attributability-responsibility is about “being responsible,” while accountability is about “being held responsible” or to use David Enoch’s (2014) term, “taking responsibility.”

When a 7 year old child carelessly damages a toy store’s property, the owner of the store may find it appropriate to hold the parents responsible for compensating for the damage, because in our society it is, in most cases, parents’ expected role to properly educate their children not to behave carelessly in other’s property. Even if the child’s parents properly educated the child, the owner may find it appropriate to hold the parents responsible, although the parents are not blameworthy. As Watson (1996) says, “Holding people responsible is not just a matter of the relation of an individual to her behavior, it also involves a social setting in which we demand (require) certain conduct from one another and respond adversely to one another’s failures to comply with these demands.”(p. 229). This sense of responsibility—namely, accountability—does not need to show any attributable connection between the action and the agent who should be held responsible. If there is appropriate reason to hold a person responsible, then the person can be held responsible, regardless of the existence of attributability.

Also consider Tim Scanlon’s (1998) example of accidentally contaminated milk. Suppose that a milk producer accidentally—meaning that there was due care—sold contaminated milk to some customers, and as a result they became seriously ill. Under the strict liability law, the milk producer is legally liable for compensating for the harms the customers incurred. This is true even if the milk producer is not properly attributable to the outcome action—selling contaminated milk. That is, there is no basis for moral appraisal for the milk producer, so the
customers are not in a position to blame the milk producer. Yet, the customers can hold the milk producer to taking responsibility for compensating for the damages. This is so because the milk producer has an ex ante role obligation to remedy or compensate for accidental damage, and such a role obligation is in this case legally codified in the milk industry under the strict liability law. Namely, as Scanlon (1998) says, entering the milk industry involves companies “laying down one’s (legal) right not to be penalized in the event that the milk one sells turns out to be impure even though one has not been negligent in handling it” (p. 266).

Of course, saying that the company is legally liable does not mean that it should be held morally responsible. What the case above shows is that without attributability, one can be held responsible if there is an ex ante positional obligation to “take” responsibility and that whether or not one has such a role obligation may be determined by whether or not there are moral reasons for demanding and assigning such role obligations to some parties.

To articulate such moral reason for demanding that companies that use AI take responsibility, we find Kantian social contract thinking useful (Keating, 1997, 2001; Scanlon, 1998, Ch. 6). The goal of Kantian social contract thinking is to secure the most favorable societal structure for free and equal persons to pursue their own conceptions of the good. To actively pursue one’s conception of the good life, freedom of action is essential. But one’s freedom of action inherently imposes risk of injury and accidental damage upon others (and oneself). If accidental or unforeseeable harm is an inevitable externality of freedom of action, a just society should implement a reasonable principle to fairly allocate the cost of unforeseeable harms. In a liberal society in which equal and free persons, who have different conceptions of good, live together, reciprocity is one of the few agreed upon principles. Reciprocity here means that burdens must be borne by benefits. The cost of unforeseeable harms created by a company
that uses AI, accordingly, must be proportionately aligned with the benefit that companies and other parties gain by using AI—which we call the “principle of fair reciprocity.” One efficient way to satisfy the principle is to require companies that use AI to take the proportionate responsibility to remedy unforeseeable harms. By doing so, the burden is accordingly apportioned across companies and across customers who benefit from the companies’ AI services.

If the principle of fair reciprocity is tenable, the mortgage bank or companies in similar situations, although they are not attributably responsible, may have a duty to take responsibility—namely, they may be accountable for remedying, compensating, or taking some action to fix the problem. In such circumstances, business enterprises are not directly (or attributably) responsible for harms or losses caused by their uses of autonomous artificial agents due to lack of meaningful human controls, but they may be held responsible due to the principle of fair reciprocity. Note that in such circumstances, companies are not negatively appraised for the harms caused by the autonomous robots. But if they do not take responsibility to redress harms without overriding moral reasons, then that becomes a basis for a negative moral appraisal. Thus, the public should understand that the bank in the case above should not be regarded as blameworthy for the unforeseeable harm per se.

If society enters a new era during which using AI inherently imposes the risk of unforeseeable harm upon people, a new social contract is needed to justify the imposition of risk, and the principle of fair reciprocity informs us that it may be fair to allow such a new technology only if we as a society demand that those who benefit from the imposition of risk take a positional responsibility for the harmed party. In the coming age of increasingly autonomous
artificial agents, taking responsibility—or “accountability”—will be an important moral virtue that business enterprises should possess and cultivate.

**Automation vs. Augmentation**

Although “augmentation” has only recently received much attention, computer scientist Douglas C. Engelbart proposed it in 1962 as an alternative mindset to automation. He wrote:

> By “augmenting human intellect” we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. [...] We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life [...]..

Engelbart used the mindset of augmentation to encourage developing technologies, for instance, email and Internet, to empower workers to solve problems that they otherwise would not, due to limited human conditions. All of this may sound trivial, but the simple idea may be a powerful mindset to properly fine-tune the purpose of future business.

Recently, the public media has paid significant attention to the idea of augmentation, mainly because of its usefulness as career advice for individuals who want to survive the second machine age (Davenport & Kirby, 2015). If you choose a job that a robot cannot do alone but one in which the robot can help you significantly increase productivity and efficiency, you can keep the job—with the robot. But the philosophy of augmentation goes deeper than the employment advice. The idea underlying Engelbart’s suggestion for augmentation as an alternative mindset is an ethical vision about what justifies technological innovation and businesses’ uses: The philosophy suggests that humans and robots must *co-exist* to best develop human capabilities. More specifically, technology and computational systems must exist, be developed, and be utilized for humans by rendering humans more capable, whole, and connected, not by making them limited, redundant, isolated, and idle. IBM’s artificial intelligence Watson must be developed to coexist with humans, as Dr. Watson does not replace
Business Ethics and the 4th Industrial Revolution

but collaborates with Sherlock Homes. In the automation mindset, there would inevitably be little room left for humans; in such a mindset, the robot/technology is the future. In the world of augmentation, robots would co-exist with humans, complementing humans, and collaboration would win over robot-only systems. In the augmentation mindset, the robot/technology is for the future.

The ideal of augmentation does not simply mean that businesses charitably delay automation for the sake of humans or that businesses should provide a certain number of jobs as a corporate social responsibility. Businesses can surely do that, but unless we radically change basic incentive systems of the market economy, most delay strategies would be inevitably defeated by economic incentives toward automation. The ideal of augmentation more directly challenges the very nature of business strategy, by calling for business enterprises to explore and exploit new technologies in ways that enable humans to race with machines rather than against them. In this way, as many humans as possible can sustainably obtain opportunities to be part of collective value creation, to contribute to the good of society, and to add meaning to their lives. This is a real challenge to businesses, but one that is worth confronting for future society.

At this point, it is not easy to imagine how varied businesses can augment rather than automate. Still, augmentation, though not perfect, may be a useful overarching mindset to entice business leaders to rethink the role of business as we face technological unemployment and the challenge of finding meaning in life. In terms of their long-term impact upon business and society, the automation mindset and the augmentation mindset would make substantially different impacts on. To embrace the latter, business leaders should take steps toward augmentation sooner than later, study the future of business, and think about the axiological loss that they would likely incur with respect to meaning in life if they did nothing but recklessly
automate in the coming decades. Leaders can incentivize their researchers and employees to realize augmentation and minimize automation in their business operations, spend resources to support entrepreneurs who want to develop concrete ways to augment in economically workable ways, and sponsor university research that aims to develop workable business ideas of augmentation.

Toward the Accountability of Augmentation

Major normative theories—consequentialism and deontology—support the ethical accountability of augmentation. From a utilitarianism standpoint, broadly defined, in order to do the ethical action, businesses should best estimate the expected consequences of their ways of doing business upon all involved parties and act in ways that result in the maximum expected social utility. Compare two imagined states of affairs: the era of augmented abundance and the era of automated abundance. In both societal conditions, people, regardless of employment conditions, secure basic welfare from basic income or similar distribution schemes. A major difference, though, is that the era of augmented abundance is a better condition in which individual persons can find meaning in life by having more work opportunities to contribute to the good of society, whereas the era of automated abundance is a limited condition to that end. Note that for the utilitarian, having more opportunities to contribute to the good of society is usually superior, because the best life for the utilitarian is the one that best contributes to the good of society (Singer, 1995). Furthermore, all other things equal, being in a societal condition that does not allow people to contribute to society due to lack of work opportunities usually makes them unhappy or at least makes them less happy than they would be if they could partake in collective value creation. This is an empirical claim open to counterexamples, but most people will concede that it is basically true.
Business Ethics and the 4th Industrial Revolution

The Kantian tradition reinforces the accountability of augmentation. The formula of humanity demands that we respect all humans by not violating or compromising their autonomy, and by helping them foster their life plans (Wood, 1999). The mindset of augmentation encourages businesses to offer more options to be part of collective value creation to those who seek to contribute to the good of society, and is broadly consistent with the latter part of the formula. Like imperfect duties, the accountability of augmentation is an open-ended responsibility. But, as Allen Buchanan (1996) explains, imperfect duties do not mean “moral laxity,” which is in fact failure of imperfect duties. Still, fulfilling an imperfect duty is more difficult than not violating a perfect duty, due to its vague nature, its requirement of long-term planning, and its required efforts to consistently take steps toward goals. One solution is to perfect imperfect duties through assurance that others are doing their fair share. The business world would be able to collectively improve its chances of being more accountable by implementing some industry and global compact that promotes and incentivizes firms to fulfill the accountability of augmentation.

Finally, it is hard to imagine that those who follow the liberty tradition would not favor augmentation over automation. One might say that libertarians will not care whether businesses maximize automation or augmentation, because business people who want to automate should be ethically free to automate no matter what. Yet libertarians would not necessarily think so. Part of what justifies the system of the free market and property rights as whole in the libertarian tradition is the belief that such a system gives everyone better lives. The ideal of liberty does not guarantee good consequences, but libertarians do not necessarily deny that the ideal may impose some action-guiding norms when there exists a systematic failure to give everyone the opportunity to be a self-author (Brennan, 2012, p. 104). It is for this reason that libertarians have
Business Ethics and the 4th Industrial Revolution

supported some form of basic income guarantee for everyone. Yet, as we have discussed above, a basic income may help people secure basic welfare but would not be empowering enough to promote people to be self-authors who freely exercise economic liberty by being part of economic cooperation in the second machine age. Libertarian business leaders should consider the possibility that the mindset of automation might yield an economic structure that systematically leaves behind many of the population to write their own stories in economic life. Libertarian business leaders may have, therefore, a liberty-based reason to consider that the mindset of augmentation may be more libertarian than that of automation. The augmentation mindset may be an important libertarian virtue that a liberal citizenship in the second machine age can impose on leaders of future businesses.

Ethical theories can shed more light on the nature of the accountability of augmentation. At the same time, another relevant perspective to discuss accountability lies within frameworks about the purpose of business. In the end, a choice between automation and augmentation is really a matter of which purpose businesses choose for the future. In that vein, business ethics' contribution, in addition to providing a normative analysis, would be to provide a language to discuss the purpose of future business in ways that do not gratuitously neglect the significance of meaning of work and the accountability of augmentation.

Conclusion: Reinventing Business

In 2015, a group of leading technologists wrote the Open Letter on Digital Economy, which states:

Previous surges brought with them greatly increased demand for labor and sustained job and wage growth. This time, around, the evidence is causing some people to wonder if things are different. Or, to paraphrase many recent headlines, will robots eat our jobs? We think this is the wrong question, because it assumes that we are powerless to alter or shape the effect of technological change on labor. We reject this idea. […] The corporation is itself a powerful innovation, and one that can do far more than just
Business Ethics and the 4th Industrial Revolution

generate profits and provide a competitive return to those who supply capital and take risk. It is both a tool for transforming ideas into products and services that address society’s challenges, and the means by which many people earn their living. Along with current waves of innovation in technology, we also have an opportunity to reinvent the corporation and our business systems.4

We agree. The question is, how? We have not offered concrete and detailed suggestions of what business can do—that is beyond the capacity of our philosophical exploration. Nonetheless, in this chapter, we have preliminarily sketched what business can and ought to do to minimize the negative societal impact of automation and AI. The best way to predict the second machine age is to make it, and society’s response to the question “What is the purpose of business in the coming decades?” will have significant impact on how we make the future.

References


Aquinas, T. Summa Theologiae.

Business Ethics and the 4th Industrial Revolution


Davidson, P. (2016, April 29)). Temporary, part-time workers not taking it anymore,” *USA Today*.

Business Ethics and the 4th Industrial Revolution

Edelman Intelligence (2017, September). Freelancing in America: Retrieved from
https://www.slideshare.net/upwork/freelancing-in-america-2017/1


Business Ethics and the 4th Industrial Revolution


Business Ethics and the 4th Industrial Revolution


Business Ethics and the 4th Industrial Revolution