AI and the Future of Work: A Pro-active Strategy[1]

Thomas A. Kochan
George M. Bunker Professor, MIT Sloan School of Management
and
Co-Director, Institute for Work and Employment Research

December 2019

Support for this research is provided by the MIT Good Companies Good Jobs Initiative and the MIT Task Force on Work of the Future. The views expressed are solely those of the author.

Abstract

One of the greatest challenges and opportunities of our time lies in harnessing the innovative potential of emerging technologies to help achieve a more prosperous and just society. Discussions of how to do so are at the center of debates over how artificial intelligence (AI), machine learning, and associated tools might affect the future of work. In this paper I will outline a proactive strategy for addressing these issues.

Background

It is clear that advances in Artificial Intelligence (AI) and related tools have enormous potential for driving innovation and productivity and will have significant effects on the quality and quantity of work in the future. Yet there is considerably uncertainty over whether the dominant effect will be to eliminate jobs and replace human judgement with algorithms or to inform and support human judgments with more, better, and more timely information. Moreover, the ability of AI tools to obtain and use personal data about how people work and live adds another challenging dimension to the governance of this particular technology. The purpose of this paper is to lay out a framework and set of concrete actions for how AI tools could be developed and implemented to ensure they contribute to creating more productive and fair workplaces and more broadly shared prosperity. The centerpiece of the action plan lies in bringing the multiple stakeholders that are, or will be, affected by AI into the processes of defining the problem(s) AI is asked to address and into the design, implementation, and use of these tools in practice.

In 2018 the President and Provost of MIT charged a university-wide Task Force on Work of the Future to shift discussions regarding the future of technology and work from debates over how many jobs advances in technology will destroy or create to a more proactive exploration of

how technologies can by used to create a more inclusive society and broadly shared prosperity (MIT Work of the Future, 2018). The decision to organize this Task Force was motivated by the recognition that our societies cannot afford to perpetuate or worsen existing income inequalities and divisions between "winners" and "losers" from deployment of AI, robotics, advanced manufacturing, and other technological breakthroughs that institutions like MIT help to invent and move into practice. MIT is not alone in taking up this question. Organizations as diverse as the World Economic Forum, the Organization for Economic Cooperation and Development, global labor union federations as well as those in Europe and the U.S. and many consulting firms are now focused on the same questions. There is a growing recognition (see for example, the Institute for Electric and Electronic Engineers, 2017) that multiple stakeholders need to discuss the ethical issues associated with AI and that there are big risks to leaving the development and application of AI tools to the sole discretion of the companies, researchers, and vendors competing to advance their use in industry.

In this paper I will outline my own views on how to harness the use of AI in constructive ways and deal with the disruptive effects it will have on jobs and those who will be displaced. This approach recognizes the obvious: AI can be developed in ways that augment how humans do their jobs and will change how work is organized *and* it will displace some unpredictable number of workers. A comprehensive strategy has to deal with both sets of effects. Doing so (See Figure 1) will require changes in who defines the problems AI is asked to address, how AI systems are designed and deployed in organizations, how the workforce is trained and engaged in the design and deployment process, how data privacy and ownership is managed, and how public and private resources are allocated to address the impacts of AI on workers who are displaced.

Defining the Problems AI is asked to Solve

Russell (2019) provides a clear articulation of the starting point for using AI constructively.

> "Because machines, unlike humans, have no objectives of their own, we give them objectives to achieve. In other words, we build optimizing machines, we feed objectives into them, and off they go… Machines are beneficial to the extent that their actions can be expected to achieve our objectives." (Russell, 2019; 10-11).

Given these statements, the key question is who is the "we" in determining the objectives specified for AI/Machine Learning tools? The most frequent answer to this question is the inventor or developer of the tool and/or the organization paying for it. Yet the effect of leaving the definition of the problem to the private interests of entrepreneurs or developers will inevitably bias AI development toward narrow economic objectives that don't take account of the full costs, or potential benefits, of how the technology is deployed or to the full range of effects it has on the individuals who either supply data or are targets/users of data.

Consider, for example, the race of so many entrepreneurs and firms to develop self-driving vehicles. Why is the objective of this particular suite of technologies defined as achieving driverless trucks, cars, and other vehicles, i.e., the so-called "level 5" (Car and Driver, 2017) of autonomous vehicles? Doesn't society have an interest in how these technologies are designed and developed? Why isn't the overall goal to develop safer, more efficient, and more accessible transportation systems for all? Clearly, the specific technologies in semi-or fully autonomous vehicles will need to interact and be supported by a range of infrastructure changes paid for by public authorities. If these broader goals, stakeholders, and needed investments were embedded in defining the problem and choosing the design parameters, perhaps the ultimate

solutions would be a better integration of technical support for the humans in the drivers' seats and the many users of the next generation transportation systems.

Moreover, if the problem definition stays focused narrowly on achieving driverless vehicles, others—highway and urban planning authorities, those who drive for a living, and other drivers and pedestrians who interact with driverless vehicles will be left to pay the costs of adjusting to the new technology. These stakeholders should have a voice in shaping how the initial problem and objectives are defined and how these and other consequences are considered and addressed in the technology development process.

In the context of work, leaving it to vendors and employers to define the objectives of new technology is likely to translate into defining the primary objective of the technology as a labor- saving strategy, i.e., replacing human labor and judgement with technologies aimed at increasing productivity or profitability by driving down labor costs and/or limiting human judgments about how work is carried out. While driving down labor costs is a legitimate objective to ask of any technology, it has three negative consequences if it is the sole or dominant objective for AI.

First, it leads to inefficient and over-investment in automation largely because inventors lack either the incentives or the knowledge to consider how workers and human judgment might complement what machines can do (Thomas, 1996; Mindell, 2015). Elon Musk, founder and CEO of Tesla Motors provided the most recent example of this longstanding tendency. In 2016 he announced he was building the most automated auto assembly operation in the world. Two years later he had to admit that he had "underrated" workers (Geggell, 2018). He had to bring in large numbers of workers because his automated operations couldn't get Tesla's Model 3 out the door in time to satisfy either his customers or investors. As will be noted below, he apparently

didn't learn this lesson from General Motors' experience in trying the same strategy in the auto industry in the 1980s.

Second, much of the basic research that AI system developers build on is, or has been, funded with public resources. Thus, the public has a legitimate right to insist that AI systems address the public's interest as well as the private economic interests of those purchasing or selling AI tools. The good news is some small steps are being taken to support research and development of AI tools that address big social problems. The U.S. Air Force and MIT recently announced agreement to create an Artificial Intelligence Accelerator Program that will support projects using AI to address critical social problems. And, more broadly, the Organization for Economic Cooperation and Development (OECD) reports that a number of countries such as Finland, Japan, Korea, Saudi Arabia, China, and others, as well as the European Union are engaging multiple stakeholders in developing national AI governance guidelines (OECD, 2017). Thus, there is growing recognition of the public's stake in AI development and use.

Third, labor saving benefits that accrue to an individual firm are associated with corresponding costs, some of which are borne by the individuals displaced and some by society; only a fraction is borne by the individuals and organizations that realize the benefits of the increased productivity. The effect is to increase inequality between the investors who benefit from the productivity improvements and those who lose jobs and income (MIT Work of the Future, 2019). At a societal level this difference between winners and losers from economic changes has been shown to contribute to social divisions and political polarization (Autor, Dorn, and Hanson, 2013).

In the specific context of work, the workforce is one of the key stakeholders that needs to be better represented in defining the problems AI is asked to address. Yet, in many countries

workers and their representatives lack the legal rights or power to gain a voice in problem definition.

The need for this early stage involvement in the uses of AI in employment relationships is especially critical as the number of vendors of human resource (HR) information systems proliferate. A growing number of external vendors and internal HR data analysts are developing and offering AI tools for recruitment, selection, training, performance assessment, monitoring the pace of work, and other aspects of employment relationships (Bersin, 2018; Tambe, Cappelli, and Yakubovich, 2019). At a recent meeting of high-level HR executives, concerns were raised over the lack of transparency of some of these systems. HR executives were purchasing and using some of these systems to support their decision making but lacked an understanding of how the system's algorithms generated the decision outcomes. This violates a basic requirement of "explanation" the European Union built into its regulation of AI systems (Goodman and Flaxman, 2017). It is not surprising that several lawsuits have been filed against leading vendors of these systems. If the users (HR staff) of these systems cannot explain and defend the reasons the system produces a given set of results (e.g., why one person is hired while another is screened out), then perhaps the system shouldn't be used.

Sequential versus Integrated Technology and Work Design Processes

Significant changes will be needed in the standard way organizations manage the introduction of AI and related technologies for AI to both increase productivity and improve work processes. Figure 2 illustrates two alternative processes for introducing technological innovations into organizations (Christidis, Miller, and Kochan, 2018). The typical process is sequential. A vendor or other inventor develops a generic tool for digitizing or automating some set of work processes and sells this generic tool to organizational executives. Then a process of

negotiations begins between the vendor/inventors and the managers of the operations involved over how to tailor the tool to fit into the organization's specific business and work processes. This can take considerable time and often involves considerable negotiations and conflicts with engineers, IT specialists, and managers who are invested in legacy systems and ways of organizing work systems in their domain. At some point HR executives are brought into the discussion to assess whether the workforce has the skills needed to work with the new technology and/or how many workers either lack these skills or will not be needed because the technology promises to be labor saving. The Tesla example noted above and evidence from as far back as the introduction of robotics into auto manufacturing in the 1980s (MacDuffie and Krafcik, 1992), investments in information technologies (IT) in the 1990s (Bresnahan, Sadun, Brynjolfsson, and Hitt, 1999), and electronic records technologies in health care (Draca, Sadun, and Van Reenen, 2007; Litwin, 2011; Hitt and Tambe, 2016) all demonstrate that technology strategies conceived and implemented separate from consideration of management practices, work systems, and workforce changes prove to be less productive than systems that are informed by those who understand these issues. AI developers need to learn the lesson from these prior examples.

The alternative to the sequential strategy is to integrate technology design and decision-making processes with decisions about how work systems might be changed to put the technology to best use. This idea is as old as the socio-technical studies in coal mining and other industries conducted by the Tavistock Institute in the 1940s and 1950s (Trist and Bamfort, 1952). More recently, information technology (IT) specialists describe this as user-based or agile development processes (Sacolick, 2018). Specifically, these processes call for engaging the end users of the new tools at each stage of the development, implementation, and use of the new

technology. In workplaces this may mean managers and front-line workers and, if the workforce is represented by a union and/or a works council, with worker representatives. To get the best returns from technology, AI included, technology and work design should be integrated, not developed in a sequential process where technology design is determined prior to and apart from considering its effects on work and the workforce.

An integrated strategy requires significant organizational, institutional, and public policy changes. In the US, workers have a very weak and at best secondary right to participate in negotiations over the impacts of technological change but no rights to participate directly in the design process. To make an integrated strategy viable labor law needs to change in two respects. First, the scope of collective bargaining would need to be expanded to give workers the right to negotiate over the design of technology and to have advance notice of technological changes that will affect work processes. A bill to do so has been introduced in the U.S. Congress but is unlikely to get serious attention for some time to come (Brown, 2019). Second, regardless of whether a union is present, workers would need to have some type of consultative body available to obtain information and consult on the design and implementation of AI and other technologies (Tambe, Cappelli, and Yakubovich, 2019).

In Europe unions tend to be stronger and works councils provide more opportunities for engaging the workforce in the implementation process. Indeed, some union leaders are calling for the expansion of work council responsibilities so workers, managers, engineers, and others have access to information on the plans to introduce new technologies before they are finalized, and can discuss and advise developers on issues of data privacy, use, and other relevant issues (Colclough, 2019).

Training before the new Technologies Appear

If workers and their representatives are to participate effectively in technology decision making they need to have knowledge and skills required to add value to these processes. This will require a massive investment in continuous ("life-long-learning") education and training to prepare the workforce to engage new technologies, including AI, before the technologies are brought into their workplaces (World Economic Forum, 2019). Training should include the technical and social skills needed to add value to technology strategies and decisions. Waiting to "retrain" workers when new technologies are being introduced is too late to prepare the workforce to adapt to, and work with, new technologies when they appear at the workplace.

Labor organizations also need to develop the capacity to participate in technology decision making. Workers and their representatives should have access to educational programs and resources to develop expertise on AI, robotics and other emerging technologies. Our group at MIT plans to produce an online teaching module that addresses what workers and their representatives need to know about emerging technologies, including but not limited to AI. The goal is to develop a cadre of labor experts who in turn can represent and educate/train others in what they need to know to effectively represent workers in the technology decision-making processes.

Labor organizations around the world are taking similar steps by recruiting data scientists to advise their local officers and to develop applications for using data provided by their members to track the effects of technology on the host of issues of concern to the modern workforce. These professionals can provide an important service in "training the trainers" to empower the workforce to engage in deploying technologies in productive and equitable ways, and to suggest ways of improving and finding new applications of these tools after they are implemented. This is especially important because a critical source of productivity improvement

from technological innovations comes from the tacit knowledge individuals and teams of workers draw on to drive incremental continuous improvement efforts after technological changes are put in use (Polanyi, 1966; Orlikowski, 2000; Piore, 2019).

Strengthening Adjustment and Compensation Policies

Since some unknown numbers of workers will be displaced by AI and related technologies, employers, labor organizations, and public policy makers all need to work together to provide effective adjustment assistance and fair compensation to those whose jobs and careers are adversely affected. The US ranks near the bottom of OECD countries in the percentage of GDP invested in retraining and labor market adjustment programs (OECD, 2019). Many European countries have active labor market adjustment programs that are funded through combinations of public and private investments and often administered by labor organizations or via collaborative business, labor, and government institutions (World Economic Forum, 2019). Singapore has taken a very proactive approach to this task (Skills Future, 2019). As part of its national digitization and innovation strategy, the government has constructed twenty industry transformation maps that project changes in technologies relevant to each industry. The government also provides training vouchers for workers to update their skills in anticipation of the changes that lie ahead.

Data Ownership and Privacy

AI technologies pose significant data ownership and privacy issues. These issues are now emerging as a highly contested area of concern and debate. For example, Figure 3 lists a set of demands that the UNI Global Union has put forward regarding ownership and protection of worker data (Colclough, 2017). Figure 4 summarizes the recommendations of the OECD

Council on Artificial Intelligence which were drafted by a multi stakeholder expert group consisting of employer representatives, governments, academics experts, and labor unions.. The European Union has recently adopted a directive governing AI that, among other provisions, requires AI system designs to protect individuals against discrimination and to be transparent, i.e., to able to provide individuals with an explanation of the causal reasoning or logic embedded in the system. As Goodman and Flaxman (2017) note, these EU regulations may be difficult to apply or enforce in practice. By definition AI system predictions discriminate among factors that are correlated with a target outcome and it is extremely difficult to identify the relative importance of a specific variable or set of variables in producing the predictions of a complex AI/machine learning system. Thus, how these regulations will be applied and enforced is an open question. They may, however, be a harbinger of similar regulatory debates to come in other government bodies.

These are just several examples of the escalating interest in and discussions of AI system design and use that are underway across the world. Together they signal the determination of a growing number of societal groups to develop a more inclusive and comprehensive set of principles, guidelines, and perhaps regulations governing use of personal data in AI systems.

Conclusion

The central point of this paper is that how AI is received by and affects society will depend on who participates in the key decisions influencing its design and use. While there is no single deterministic path for AI, I offer the following predictions about the future of AI if no changes are made in who defines the problems it is asked to address and in the design and implementation processes that currently dominate how it is introduced and used in workplace settings:

1. Vendors and developers will more frequently use AI systems to replace rather than augment workers and human judgment.

2. AI deployment will result in over-automating tasks and under-achieving its potential for increasing productivity and innovation.

3. Use of AI will increase inequality in the distribution of benefits and costs and leave significant social and economic costs of its effects to be borne by society.

4. There will be significant opposition and backlash from the workforce and others in society over the use of their personal data in AI systems.

These effects could be mitigated (and the predictions therefore proven wrong) if a) the AI community engages the stakeholders who share an interest in AI development and use, b) organizations that invest in AI take an integrated approach to its design and implementation, c) the workforce and its representatives become more proactive and influential in AI and related technology decision-making and workforce development, and d) societies expand and deepen labor market adjustment policies for displaced workers. In short, a new governance system could help AI achieve its full innovative and productivity-enhancing potential while ensuring the benefits it generates are widely shared.

Figure 1

Elements of an Integrated Technology and Work Strategy

1. **Involve key Stakeholders in Defining the Problem(s) Technology is being asked to Solve**

2. **Engage Workers early and directly in the Design and Implementation of an Integrated Technology and Work Design Process**

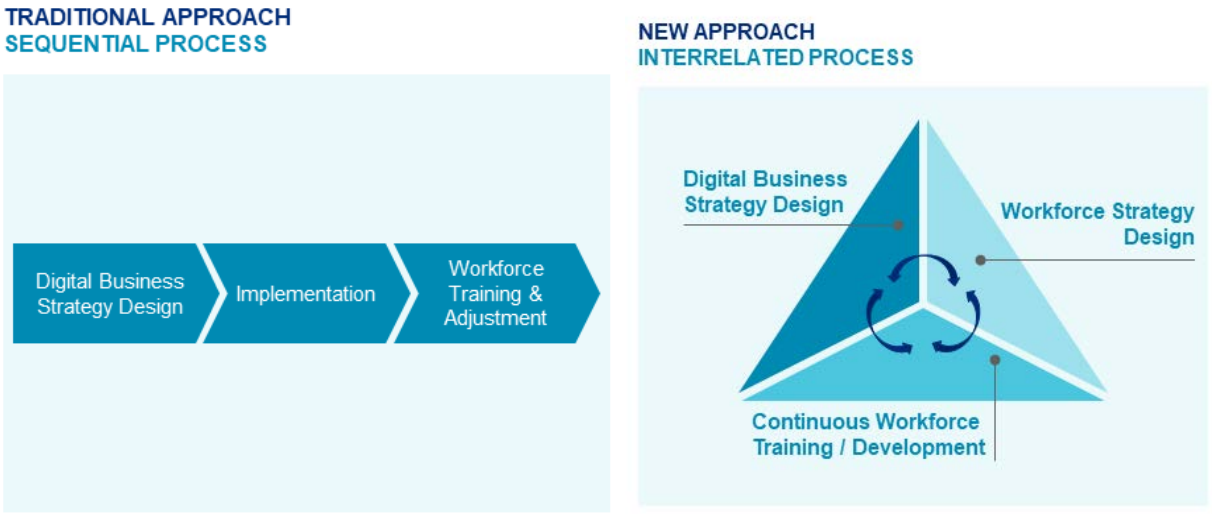
3. **Train the Workforce before the Technology is Deployed**

4. **Protect the Ownership and Privacy of Personal Data**

5. **Provide Adjustment Assistance and Fair Compensation to those Adversely Affected**

Figure 2

Sequential and Integrated Technology and Work Design Processes



Source: Antonis Christidis, Miller, Axel, and Kochan, Thomas. 2018. "Engaging the Workforce in Digital Transformations." Mercer/Oliver Wyman. https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2018/april/Engaging-The-Workforce-In-Digital-Transformation.pdf

Figure 3

Key Demands: UNI Future of Work

1. Put Data Rights and Privacy at the top of the Bargaining Agenda: Management must be accountable to workers/shop stewards on their use of data throughout the data life cycle.

2. When investing in new technologies, the organization should be obliged to create at the same time a 'people plan.'

3. Negotiate for an expansion of the Works Council responsibilities or establish an AI and Data Governance Council.

4. Follow the OECD principles calling for attention to people and plant issues, social dialogue, skill development and just transitions to new technologies. (http://bit.ly/OECDPrinciplesFWOW.)

Source: UNI:FWOW.
file:///C:/Users/Tkochan/AppData/Local/Microsoft/Windows/INetCache/Content.Outlook/UHF9IZJD/key%20demands%20UNI%20FWOW.pdf.

Figure 4

OECD Council Recommendations of the OECD's Council on Artificial Intelligence

1.1. Inclusive growth, sustainable development and well-being: Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.

1.2. Human-centred values and fairness: AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle.

1.3. Transparency and explainability: AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information…to foster a general understanding of AI systems, ii. to make stakeholders aware of their interactions with AI systems, …understand the outcome, and, enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

1.4. Robustness, security and safety: a) AI systems should be robust, secure and safe throughout their entire lifecycle so that they… do not pose unreasonable safety risk.

1.5. Accountability: AI actors should be accountable for the proper functioning of AI systems.

Source: OECD Council "Recommendations of the Council on Artificial Intelligence, May 22, 2019. https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf

References

Antonis, Christidis, Axel Miller, and Thomas Kochan. 2018. "Engaging the Workforce in Digital Transformations." Mercer/Oliver Wyman. https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2018/april/Engaging-The-Workforce-In-Digital-Transformation.pdf.

Autor, David H., David Dorn and Gordon H. Hanson. 2013. "The China Syndrome: The effects of Import Competition in the United States." *American Economic Review* 103, 6, 2121-2168.

Bersin, Josh. 2018. AI in HR. *Forbes*, June 18, 2018. https://www.forbes.com/sites/joshbersin/2018/06/18/ai-in-hr-a-real-killer-app/#51d903f348f1.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012), "Americans Do I.T Better: US multinationals and the productivity miracle" *American Economic Review* 102 (1), 167-201

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 1999. "Technology organization and the demand for skilled labor," in Margaret Blair and Thomas Kochan (eds.) The New Relationship, Washington, D.C.: The Brookings Institution, 145-84.

Colclough, Christina. 2019. "Top Ten principles for Ethical Artificial Intelligence". UNI Global Union. http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf.

Draca, Mirko, Raffaella Sadun and John Van Reenen (2007), "ICT and Productivity" Handbook of Information of Information and Communication Technologies" R. Mansell, C. Avgerou, D. Quah and R. Silverstone (eds) Oxford Handbook on ICTs, Oxford University Press.

Hitt, Lorin M. and Prasanna Tambe. 2016. "Health care, Information technology, work organization and nursing home performance, *Industrial and Labor Relations Review*. 69, 4, 824-859.

Institute for Electric and Electronic Engineers. 2017. "The IEEC Global Initiative on Ethics of Autonomous And Intelligent Systems." https://ethicsinaction.ieee.org/.

Litwin, Adam Seth. 2011. "Technological change at work: The impact of employee involvement on the effectiveness of health information technology." *Industrial & Labor Relations Review, 64*(5), 863-888, 2011.

MacDuffie, John Paul and John F. Krafcik, "Integrating technology and human resources for High-Performance Manufacturing: Evidence from the international auto industry. In Thomas Kochan and Michael Useem (Eds.) Transforming *Organizations*. MIT Press, 1992; 209-226.

Mindell, David A. 2015. *Our Robots, Ourselves: Robotics and the Myths of Autonomy*. New York: Viking Press.

MIT Work of the Future. 2018. https://workofthefuture.mit.edu/

Organization for Economic Cooperation and Development (2017). "AI Policies and Initiatives." https://www.oecd-ilibrary.org/sites/cf3f3be0-

en/index.html?itemId=/content/component/cf3f3be0-en&mimeType=text/html. Downloaded December 8, 2019.

Organization for Economic Cooperation and Development. 2019. "Active Labor Market Policies Connecting People with Jobs. https://www.oecd.org/employment/activation.htm.

Piore, Michael J. 2019. "Tacit Knowledge and the Future of Work Debate." Paper presented at the Annual Meetings of the Society for Advancement of Socio-Economics. New York, June.

Polanyi, Michael (1966) *The Tacit Dimension*. Garden City, N.Y.: Doubleday.

Russell, Stuart. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking/Random House.

Sacolick, Isaac. 2018. "What is Agile Methodology: Modern Software Development Explained. *Inforworld*, March 15, 2018. https://www.infoworld.com/article/3237508/what-is-agile-methodology-modern-software-development-explained.html

Skills Future. 2019. https://www.skillsfuture.sg/.

Tambe, Prasanna, Peter Cappelli, and Valery Yakubovic. 2019. "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward." *California Management Review*. 6, 14, 15-42.

Thomas, Robert J. 1996. *What Machines Can't Do.* Berkeley, CA: University of California Press.

Trist, Eric and Ken Bamforth. 2019. "Some social and psychological consequences of the long wall method of coal getting." *Human Relations. 4, 1, 3-36.*

World Economic Forum. 2019. "A Global Standard for Life Long Learning and Worker Engagement to Support Advanced Manufacturing. https://www.weforum.org/whitepapers/a-global-standard-for-lifelong-learning-and-worker-engagement-to-support-advanced-manufacturing.