Theoretical Foundations Paper
Artificial Intelligence and Ethics

Introduction

Artificial Intelligence (AI) stands to impact all areas of the global society. Dr. John McCarthy was involved in AI research since 1948 and was the first to use the term AI. Dr. McCarthy's 1955 proposed that "… every aspect of learning and other properties of intelligence can be described so precisely that a machine can simulate it" (Lischka, 2011). The first industrial revolution began around 1800. During this revolution, for the first time, goods and services were produced by machines. "Hard" capital (e.g., railways, steam engines, etc.) were essential inventions during this phase. The second revolution began at the end of the 19th century with electrification and the "soft" process innovation of the assembly line. The third industrial revolution started in the 1970s and was distinguished by further automation through electronics (e.g., personal computers, the internet). AI is being called the fourth industrial revolution. This newest revolution is the technical integration of electronics and physical systems. Now production is becoming controlled by decentralized self-organized machines being optimized in real-time (Lischka, 2011). AI is a General Purpose Technology (GPT) and will be pervasive, improve over time, and create even newer innovations (Brynjolfsson & McAfee, 2014).

AI has a wide range of applications, including natural language processing, allowing computers to understand language, machine vision enabling computers to see objects, and automated model building allowing machines to operate autonomously (Chen, Christensen, Gallagher, Mate, & Rafert, 2016). In the longer term, Kurzweil (2005) sees nonbiological intelligence matching the range and subtlety of human intelligence by 2029 and what he calls the "Singularity" – the point at which AI surpasses human intelligence - to occur by 2045 (Pratt, 2015).

The impact AI will have on labor economics (see Frey & Osborne, 2017, Smith, 2019, Wisskirchen, et al. 2017), financial markets (Pisani, 2015), military and security systems, AI itself, and, ultimately, the distinction between humans and machines, is of great concern to many. These potentially far-reaching impacts raise serious ethical issues. To address the ethical problems from AI institutions have developed to both in the private and public sectors. The purpose of this paper will be to explore how those institutions are forming within social-economic institutional theories.

Research Question

Are current ethical rules and norms governing the development of AI evolving under the social-economic institutional principles of efficient solution hypothesis and social institutional construction?

Roadmap

This paper will proceed first by providing a brief review of ethical standards in general and then review what ethical concerns arise from the development and implementation of AI technology. Second, the paper will review the current literature relating to the establishment of AI policies and institutions. The literature review will demonstrate that little attention is directed toward social-economic institutional principles in AI technology development. Third, the paper will describe two important and applicable social-economics institutional theories: the efficient solution hypothesis and social institutional construction theory. Fourth, the paper will briefly review the status of ethical AI standards within the institutional framework. Finally, the paper will apply the two social-economic institutional theories to two specific United States (US) institutions: one public and one private. The paper will finish with a few closing conclusions.

Importance of AI Ethics

During the axial age, along with one's duty to the tribe, developed the idea of a duty to do good. From that point on, the concept of ethics has continued to evolve. Today ethics can be divided into three categories: 1) Consequential, 2) Deontological, and 3) Virtue ethics. Consequentialists claim that the right action is the one that brings about good consequences. The most common form of consequential ethics is utilitarianism. Utilitarianism holds that any action should bring about the greatest happiness for the largest number of people. Deontological theories (derives from the Greek words for duty or deon and science or logos) claim that right actions should be per a principle, such as 'do not take innocent life' or 'do not lie.' Finally, virtue ethics focuses on what a virtuous person would do in each situation (Boddington, 2017). Under these precepts, today's ethics has increasingly been concerned with obligations of specialized disciplines along professional lines. In this sense governing bodies are typically formed to develop ethical codes for their members (Honderich, 2005). Regarding AI ethics, no formal consensus has yet emerged. However, the ethical risks have been developing and have essentially coalesced around certain topics such as justice and fairness, privacy, military security, and transparency and responsibility.

Regarding justice and fairness, Eric Sammer, a distinguished engineer at Splunk, a NASDAQ traded machine learning software company, says, "A lot of these [AI] algorithms are being trained on existing human practices that are inherently biased …" (Splunk, 2020). AI is being used to screen job applicants, grade college entrance exams, assess creditworthiness, and provide information to police and the justice systems. These algorithms "learn" from existing datasets. Unfortunately, these datasets can contain existing demographic group biases (Boddington, 2017). This fact concerns policymakers and developers. Some have argued that privacy is a thing of past generations; others disagree. There is an ever-increasing level of personal data being generated by commercial enterprises and by individuals sua sponte. Technology is developing the ability to process recognizable private data and the

enormous amount of dark data (i.e., data not previously reviewable such as emails and text messaging) that society does not yet perceive as private data. AI can penetrate this data and develop detailed personal individual profiles raising significant societal privacy concerns. AI will affect many issues of military security, such as the ability to analyze governmental communications and extract important strategic information.

Additionally, AI can give weak states and non-state actors access to a type of autonomous weaponry previously unavailable. This could allow them to punch far above their weight. Because AI code can be written and developed so that the details of exactly how the algorithms are performing may be lost, transparency and responsibility issues have risen as ethical and legal concerns (Jobin, Lenca, & Vayena, 2019). AI compounds all the above risks and others because of its potential to become self-developing, resulting in exponential growth. This means that an AI algorithm can grow and have far-reaching impacts beyond the ability for societal control (Makridakis, 2017).

Literature Review

AI catches the imagination of everyone. Most people have read science fiction tales or seen popular movies regarding intelligent machines. Isaac Asimov's novel, "I, Robot," Steven Spielberg's film, "A.I.," or the character Data in television's Star Trek the Next Generation. These popular books and shows often address ethical issues, but there has also been serious real-world theoretical work into AI ethics. A great number of articles address specific issues. For example, one study reviewed AI's pretrial in the media and found that while the media's presentation "has a fairly realistic and practical focus in its coverage of the ethics of AI [its] … coverage is … shallow" (Ouchchy & Dubljević, 2020, p. 1). Another study evaluated AI-equipped machines' ethical decision-making, such as driverless cars (Etzioni & Etzioni, 2017). Bias and discrimination from AI algorithms are an important specific topic often addressed in the literature (Polonski, 2018 and Boddington, 2017). Unfortunately, war and military

defense continue to be a human activity, and much literature has been devoted to the implications that AI brings to this human endeavor. AI's control of lethal autonomous weapons systems, namely systems that destroy military targets absent human management, has earned much debate (Russell, Hauert, Altman, & Veloso, 2015). Similarly, concerns over malicious exploits from bad actors have been raised. Dangerous AI can spontaneously emerge from intelligent algorithms, or it can be deliberately created. These programs can do enormous damage to the cyber-ecosystem and our physical world (Pistono & Yampolskiy, 2016).

These particularized concerns have driven a class of survey AI ethical literature. For example, a survey was conducted on how technologists, innovators, developers, and business leaders feel about AI and its impact on humanity. Not surprisingly, the responses were patterned after the specific issues noted in the previous paragraph (Anderson, Rainie & Luchsinger, 2018). Another survey polled experts to answer the question, "what is AI?" This review concluded that "while AI researchers favor definitions of AI that emphasize technical functionality, policymakers instead use definitions that compare systems to human thinking and behavior" (Krafft, Young, Katell, Huang, & Bugingo, 2020, p. 1). Like this polling research, other research has analyzed the available published materials to help understand the current state of AI ethical visions. Some surveys have been narrowly focused on such areas as the pharmaceutical industry (Herxheimer, 2003), but others have a wider scope. One such wide scope study looked at European policy documents and found that European states have similar AI ethical visions, but none have addressed the issues comprehensively (Vesnic-Alujevic, Nascimento, & Pólvora, 2020). Another global AI ethics survey discovered five ethical principles (i.e., transparency, justice and fairness, non-maleficence, responsibility, and privacy) appear to be emerging within published materials. However, that same survey found substantive divergence in how these principles should be interpreted and implemented (Herxheimer, 2003).

The empirical studies above have also been supported by more normative literature. Some articles have looked to the past and examined analogous inventions of previous revolutionary developments (i.e., industrial and digital ages) to help guide AI's future ethical development (Makridakis, 2017). Other articles examine how various ethical codes, such as Utilitarianism, might be programmed into AI software (Anderson, Anderson & Armen, 2004, Wiegel, 2010). Still, other articles examine how different types of moral machines might function. For example, is there a difference between algorithms that operate independently of humans (e.g., autonomous weapons systems) and those that work as merely advisors to humans (Anderson & Anderson, 2007)? The question of whether ethics is computable or if some moral decisions just cannot be programmed at all has also been considered (Anderson & Anderson, 2007 and Etzioni & Etzioni, 2017).

The advances in AI and all the descriptive and normative studies have been cumulating in the need to plan for AI's ethical implementation directly. One theme focuses on safety and who the law is to hold accountable when something goes wrong (Carrillo, 2020 and Pavaloiu & Kose, 2017). Questions like, "how much power and authority these intelligent machines ought to have-and just who, if anyone, will control them?" (Waldrop, 1987 p. 35) are being addressed. Practical solutions to these problems are being considered. One article seeking to promote safety and security proposes that third-party auditing of software, red team exercises, and bias and safety bounties be developed (Brundage et al., 2020). Other works present entire comprehensive frameworks. One such comprehensive framework examines each of the major governmental entities' competencies, namely the legislature, government agencies, and courts. The framework proposes that the legislature establish guiding principles for AI regulation and assign the administration of those principles to the expertise of agencies. The framework concludes that the courts are best equipped to allocate responsibility for harm after the legislative and agency processes are completed (Scherer, 2015). Closer to this current paper's work, another comprehensive discussion of AI ethics is given in the book "Towards a code of ethics for

artificial intelligence" by Paula Boddington (Boddington, 2017). Here, Boddington discusses how to identify new ethical problems associated with AI, how to solve those problems, and discusses the social nature of ethical code development. For example, she discusses how codes of ethics can arise from catastrophes (e.g., Tuskegee Syphilis trials, thalidomide tragedy, etc.) and how often social overreaction can occur in these situations. She also notes that codes of ethics are often modeled on other codes (e.g., social science ethics modeled on medical ethics) or are influenced by powerful groups (e.g., European vs. US approaches to privacy) or individuals (powerful and wealthy innovators).

Other than the last book mentioned immediately above, of all the materials reviewed in this literature review, little work has been performed on the social processes that are forming ethical AI models. None have addressed the social-economic institutional development of AI ethics. This will be the topic we turn to next in this paper.

Social-economics Institutional Theories

Two social-economics institutional theories will be described in this section. First, the efficient solution hypothesis argues that economic institutions develop in response to market failures and that the institution's emergence can be explained by examining the market failure that it solves. Second, social institutional construction, which alternatively argues that economic institutions originate from a source then develop along a trajectory.

Efficient Solution Hypothesis

Neoclassical economists argue that economic institutions emerge as efficient solutions to market failures (e.g., Bates 1994). One way to view the efficient solution hypothesis is that it is an "after-the-fact" analysis. That is, it generally takes the institution as a given. The theory then looks backwards and seeks to determine how the established institution was an efficient solution to its problem. Unlike Adam Smith, who in a 1755 lecture said, "Little else is requisite to carry a state to the highest degree of

opulence … but peace, easy taxes, and a tolerable administration of justice …" (Stewart, 1858, p. 68), authors like North and Thomas claim that the "efficient economic organization is the key to growth; the development of an efficient economic organization in Western Europe accounts for the rise of the West" (North & Thomas, 1973, p. 1). Others have echoed such sentiments (e.g., Reyerson, 2006). Oliver Williamson viewed the efficiency gains in terms of transaction cost economizing. For Williamson, whichever organizational processes minimized transaction cost, the greatest would be the ultimate institutional organization (Williamson, 1985). The efficient solution hypothesis has also been applied to explain the emergence of substructures within large American corporations. Chandler, Granovetter, and Swedberg's analysis suggest that corporations' multidivisional structure is an efficient result of increasingly complicated technological advances and the emergence of a national market (Chandler, Granovetter, & Swedberg, 1992). However, as useful as the efficient solutions model might be, it does suffer one serious flaw. It uses the results of the process in its explanation of its occurrence (Roy, 1990). There is also an empirical issue because not all outcomes are efficient, so efficiency cannot automatically be assumed. Finally, there is potentially a range of possible efficient solutions, and the hypothesis does not evaluate why one solution developed versus another.

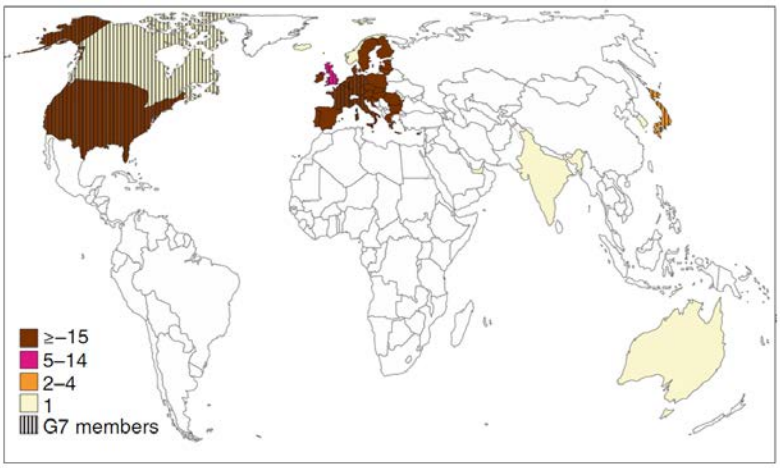
Social Institutional Construction

As noted, the efficient solution hypothesis suffers from limitations. Most notable that it assumes that actual history can be determined by what economic behaviorists call "revealed preferences." Under this criticism, the approach pays insufficient attention to social structures and the beliefs of the actors within those structures that existed before the new institutions develop. That is, it substitutes a hypothetical history for an actual history (Swedberg & Granovetter, 2018 and Schotter, 1981). In doing so, the efficient solution hypothesis will tend to underestimate the roles of potential counterfactuals. Additionally, it ignores the possibility of error in historical development. As Douglass North contends, there is a distinction between "allocative efficiency" and "adaptive efficiency." Allocative efficiency

assures that goods are properly priced, while adaptive efficiency has a trial-and-error component (North, 1990). North goes as far as to argue that "inefficient . . . institutions are the rule, not the exception" (North, 1993, p. 252). According to Peter Berger and Thomas Luckmann, in The Social Construction of Reality (1966), institutions are generally the result of a slow social maturation process that finally becomes "the way things are done." For example, Marie-France Garcia-Parpet describes a situation where a new institutional strawberry market developed over time from a struggle between strawberry producers and distributors within a pre-existing organization, ultimately evolving into a new alternative institution. Here Garcia-Parpet not only demonstrated the maturation process but also highlighted that unless the pre-existing organization was accounted for, the new institution could not be understood (Garcia-Parpet, 2007). The social maturation process will likely have constraints, however, because of these pre-existing organizations or technologies. The term "path-dependent development" is associated with this concept (David, 2003 and Arthur, 1989). The idea is exemplified by the development of the typewriter keyboard, as described by Paul David (2003). The QWERTY keyboard layout made sense from a technical and efficiency viewpoint early in the typewriters' development. As technology changed, the design became less efficient; however, as many people became accustomed to the QWERTY layout, it became more costly to retrain people than to maintain the inefficient design. Granovetter has generalized the idea of path-dependent development to institutions, suggesting that they too are constructed against a background of social, political, market, and technological constraints (Swedberg & Granovetter, 2018). Therefore, under a social institutional construction approach, to understand an institution, the history of existing institutions and actors, the links between those institutions and actors, the motivations of the actors, exiting conflicts, and the limits to development due to path-dependencies all must be considered. Also, the possibility for missteps and errors must be considered.

Social-economics Institutional Theories and AI Institutional Ethics Development

A 2019 paper conducted an internet analysis of AI literature containing principles and guidelines for ethical AI. Below is a map indicating the document count and country of origin from that search.



Source: Jobin, A., Ienca, M., & Vayena, E. (2019), "The global landscape of AI ethics guidelines. Nature Machine Intelligence"

Most ethics guidelines were released in the US (21) and within the European Union (19), followed by the UK (13) and Japan (4). Sources of these statements include government organizations, research institutions, and private companies (Jobin, Ienca, & Vayena, 2019). A breakdown of sources is presented below:

Entity	Document Count
Private Company	19
Government	18
Research Institution	9
Intergovernmental / Supernational Organization	8
Non-profit, professional assoication, scientific society	7
Private sector alliance	4
All others	8
Total	73

Institutions developing AI ethics are far too numerous and varied to analyze in total. Therefore, this paper will focus on two US organizations. One governmental, namely the US White House's Office of Science and Technology Policy, and the other private, namely OpenAI.

White House's Office of Science and Technology Policy

In 2020 the White House's Office of Science and Technology Policy (OSTP) released a draft of a document entitled "Guidance for Regulation of Artificial Intelligence Applications." This document listed ten principles that US agencies should consider when evaluating AI regulations (AI Update, 2020). The OSTP is the principal agency that directly advises the US President regarding science and technology policy matters. Congress established the OSTP through the National Science and Technology Policy, Organization, and Priorities Act of 1976 (P.L. 94-282). The act states, "The primary function of the OSTP Director is to provide, within the Executive Office of the President, advice on the scientific … issues." Further, "The Office shall serve as a source of scientific and technological analysis and judgment for the President concerning major policies, plans, and programs of the Federal Government." (P.L. 94-282).

Evaluating the OSTP from an efficient solution hypothesis would first ask, what efficiency problem the institution solves? The answer would be that as science and technology have become more complex and integral to a nation's functioning, top-level policy development becomes more important and the coordination of that policy at the top level is an efficient solution. Absent coordinated efforts, science programs are likely to be badly designed and prone to costly mistakes. Additionally, policymakers are likely to fall prey to special interest groups (Kelly, 2004). Using taxpayer resources to centralize the function, rather than developed through multiple uncoordinated agencies and departments, minimizes transaction costs, decreases errors, and enhances efficiency.

However, using a social institutional construction evaluation paints a much richer image of the OSTP. In the past, science and technology policy issues reach the Presidential only if they were highly controversial or were a matter of great public concern. When they did reach the President, the President would generally seek out trusted advisors. Starting in the 1930s, Presidents began creating

boards and committees to advise on science matters. However, these boards and committees lack permanency. World War II demonstrated the importance of science to national economics and military strength. As such, President Franklin D. Roosevelt wrote a letter to his science director, Vannevar Bush, seeking recommendations on how to support science during peacetime. Bush's response laid the foundation for today's US science policies, demonstrating a degree of path-dependency. Subsequent Presidents continued to use various councils and committees, but none of those organizations had statutory authority and would often be disbanded, restructured, or assembled anew with each successor President. Due to political conflicts, in 1973, President Nixon abolished the then existing science office and let part of the organization return to the private sector, leaving only the National Security Council in a public role. With this backdrop, President Ford sought legislation to create a permanent office, and with the National Science and Technology Policy, Organization, and Priorities Act of 1976 (P.L. 94-282), the OSTP was created (Sargent & Shea, n.d. and Kelly, 2004). With this history born mostly out of World War II and subject to politicization, it becomes apparent why many Presidents have used the OSTP, not for overriding national policy initiatives, but rather for military concerns and pet projects. For example, under President Reagan, the OSTP was subjugated into being a cheerleader for the administration's Strategic Defense Initiative (SDI). Again, under President George W. Bush, the OSTP devoted much of its time to homeland security issues (Kelly, 2004). President Trump had a recognized desire to undermine and outdo all things President Obama. Under this analysis, the (OSTP) document "Guidance for Regulation of Artificial Intelligence Applications" could be seen as a continuation to President Trump's Executive Order Exec. Order No. 13859, "Maintaining American leadership in artificial intelligence" issued on January 7, 2019 (Exec. Order No. 13859, 2019) which in turn could be seen as a "not to be outdone" reaction to the President Obama administration's The National Science and Technology Council report titled, "Preparing For the Future of Artificial Intelligence." (White House, n.d.). Thus, perhaps the "Guidance for Regulation of Artificial Intelligence

Applications" policy does not represent a new direction for the US but rather a political reaction to a past administration.

OpenAI

OpenAI is a private non-profit with only a five-year history. At its founding in 2015 by Sam Altman and Elon Musk, two major wealthy technology innovators, the organization's blog posted that their "aim is to build value for everyone rather than shareholders," that "researchers will be strongly encouraged to publish their work," and that "our patents (if any) will be shared with the world." The organization stated that it would "freely collaborate" with other organizations. The founders committed $1 billion to the organization (Introducing OpenAI, 2015). OpenAI's contribution to AI ethics lay in its Charter. OpenAI's charter states, "OpenAI's mission is to ensure that artificial general intelligence (AGI) … benefits all of humanity … [with a] "primary fiduciary duty … to humanity." OpenAI asserts that its goal is "Long-Term Safety" of AGI use (OpenAI, 2020). On February 21, 2018, Musk resigned, citing "a potential future conflict (of interest)" with his ownership of Tesla and Tesla's AI development for self-driving cars (Openai, 2021). However, others claim it was because Mr. Musk disagreed with the organization's plans (Bass, 2019). Later, OpenAI started a for-profit arm to help it raise more money. It then partnered with Microsoft as a solution to OpenAI's funding needs (Bass, 2019). Some "worry that this move — or rather the way they made it — may result in making the innovative company no different from the other AI startups out there." (Coldewey, 2019). Still, others have signaled concerns about the growing secrecy at OpenAI, with employees having to sign nondisclosure agreements and being restricted from communicating with the press (Hao, 2020).

Evaluating OpenAI from an efficient solution hypothesis again, the first question is, what efficiency problem the institution solves? The answer is difficult to cognize. It is possible that Altman and Musk saw that regulatory oversight of AI technology was too slow and ineffective. Perhaps OpenAI would

help create a more efficient ethical and legal development process. Maybe OpenAI might create a better sharing of AI technologies, thus benefiting all companies and organizations equally and potentially accelerating economic returns. Certainly, the parties funding the organization are the correct entities; however, only a subset of the potentially interested entities.

Applying a social institutional construction evaluation suggests that sometimes powerful actors and their beliefs can organize entire industries (e.g., consider Edison and Insull's impact on the American electric power generation industry). This might be true of OpenAI. However, given OpenAI's current more normal venture capitalistic direction, perhaps Douglass North's idea that trial-and-error (North, 1990) plays an important part in institutional development might be a correct analysis for OpenAI. Future developments will be telling.

Conclusion

This paper sought to answer whether current ethical rules and norms governing the development of AI are evolving under the social-economic institutional principles of efficient solution hypothesis and social institutional construction? To accomplish that goal, the paper reviewed the outstanding literature and discovered that little consideration had been given to the topic of AI ethics and social-economic institutional principles. To fill this gap, the paper identified two US institutions: one public and one private. It then applied two social-economic institutional theories, namely the efficient solution hypothesis and the social institutional construction theory, to each of those organizations. The paper did find that the institutions could be interpreted under each of those theories. However, the latter rather than the former idea provided better illumination of the implicit processes. However, given the limited scope of this review, how illustrious the general approach might be must remain tentative. Further examination of more AI ethics institutions would likely better support whether social-economic institutional principles can help gain insight into AI ethics institution activities.

References

AI Update (2020): White House Issues 10 Principles for Artificial Intelligence Regulation (https://www.insidetechmedia.com/2020/01/14/ai-update-white-house-issues-10-principles-for-artificial-intelligence-regulation/). Inside Tech Media . 2020-01-14. Retrieved 2021-04-03.

Anderson, J., Rainie, L., & Luchsinger, A. (2018). Artificial intelligence and the future of humans. Pew Research Center, 10, 12.

Anderson, M., & Anderson, S. L. (2007). The status of machine ethics: a report from the AAAI Symposium. Minds and Machines, 17(1), 1-10.

Anderson, M., Anderson, S. L., & Armen, C. (2004). Towards machine ethics. In AAAI-04 workshop on agent organizations: theory and practice, San Jose, CA.

Bass, D. (2019). Microsoft to invest $1 billion in OpenAI. Retrieved April 04, 2021, from https://www.latimes.com/business/story/2019-07-22/microsoft-openai

Bates, Robert H. (1994). “Social Dilemmas and Rational Individuals: An Essay on the New Institutionalism.” Pp. 43–66 in Anthropology and Institutional Economics, edited by James M. Acheson. Lanham, Md.: University Press of America.

Berger, Peter L., and Thomas Luckmann. (1966). The Social Construction of Reality: A Treatise in the Sociology of Knowledge. New York: Anchor Books.

Boddington, P. (2017). Towards a code of ethics for artificial intelligence. Springer International Publishing.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... & Anderljung, M. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. arXiv preprint arXiv:2004.07213.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. Telecommunications Policy, 44(6), 101937.

Chandler, A. D., Granovetter, M., & Swedberg, R. (1992). The emergence of managerial capitalism (pp. 131-158). Boulder, Colo.: Westview Press.

Chen, N., Christensen, L., Gallagher, K., Mate, R., & Rafert, G. (2016). Global economic impacts associated with artificial intelligence. Analysis Group.

Coldewey, D. (2019). Openai shifts from nonprofit to 'capped-profit' to attract capital. Retrieved April 04, 2021, from https://techcrunch.com/2019/03/11/openai-shifts-from-nonprofit-to-capped-profit-to-attract-capital/

David, P. A. (2003). Clio and the Economics of QWERTY. International Library of Critical Writings in Economics, 163, 607-612.

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

Exec. Order No. 13859, (2019) Maintaining American leadership in artificial intelligence, 84 C.F.R. 3967-3972.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. Technological forecasting and social change, 114, 254-280.

Garcia-Parpet, M. F. (2007). The social construction of a perfect market. The Strawberry Auction at Fontaines-en-Sologne, D. MacKenzie, F. Muniesa, L. Siu (eds.), Do Economists Make Markets, 20-53.

Hao, Karen (2020). "The messy, secretive reality behind OpenAI's bid to save the world" (https://www.technologyreview.com/s/615181/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/). MIT Technology Review. Retrieved 2021-04-02.

Herxheimer A (2003) Relationships between the pharmaceutical industry and patients' organisations. BMJ 326(7400):1208

Honderich, T. (Ed.). (2005). The Oxford companion to philosophy. OUP Oxford.

Introducing OpenAI (2015), (https://blog.openai.com/introducing-openai/). OpenAI Blog. 12 December 2015.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Kelly, H. (2004). Flying blind: The rise, fall, and possible resurrection of science policy advice in the United States. Washington, D.C.: The Federation.

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020). Defining AI in policy versus practice. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 72-78).

Kurzweil, R. (2005). The singularity is near: When humans transcend biology. Viking Press.

Lischka, K. (2011). John McCarthy: Der Vater der Rechner-Cloud ist tot - DER SPIEGEL - Netzwelt. Retrieved May 5, 2020, from https://www.spiegel.de/netzwelt/web/john-mccarthy-der-vater-der-rechner-cloud-ist-tot-a-793795.html

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures, 90, 46-60.

North, D. C. (1990). Institutions, institutional change and economic performance. Cambridge university press.

North, D. C., & Thomas, R. P. (1973). The rise of the western world: A new economic history. Cambridge University Press.

North, Douglass C. (1993). "Institutions and Economic Performance." Pp. 242–261 in Rationality, Institutions and Economic Methodology, edited by Uskali Mäki, Bo Gustafsson, and Christian Knudsen. London: Routledge.

OpenAI. (2020). OpenAI charter. Retrieved April 02, 2021, from https://openai.com/charter/

Openai. (2021). Retrieved April 04, 2021, from https://en.wikipedia.org/wiki/OpenAI

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. AI & SOCIETY, 35(4), 927-936.

Pavaloiu, A., & Kose, U. (2017). Ethical artificial intelligence-an open question. arXiv preprint arXiv:1706.03021.

Pisani, Bob. (2015). What caused the flash crash? CFTC, DOJ weigh in. Retrieved from https://www.cnbc.com/2015/04/21/what-caused-the-flash-crash-cftc-doj-weigh-in.html

Pistono, F., & Yampolskiy, R. V. (2016). Unethical research: how to create a malevolent artificial intelligence.

Polonski, V. (2018). AI is convicting criminals and determining jail time, but is it fair. In World Economic Forum.

Pratt, Gill A. (2015) “Is a Cambrian Explosion Coming for Robotics?” Journal of Economic Perspectives 29,(3): 51-60.

Reyerson, K. (2006). Institutions and the Path to the Modern Economy: Lessons from Medieval Trade. By Avner Greif. Cambridge: Cambridge University Press, 2006. Pp. xix, 503. $34.99, paper. The Journal of Economic History, 66(4), 1080-1081.

Roy, W. G. (1990). Functional and historical logics in explaining the rise of the American industrial corporation. Comparative Social Research, 12(20), 19-44.

Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Ethics of artificial intelligence. Nature, 521(7553), 415-416.

Sargent, J. F., & Shea, D. A. (n.d.). Office of Science and Technology Policy (OSTP): History and overview (Vol. R43935) (U.S.A., Congressional Research Service). Congress. Updated March 3, 2020

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harv. JL & Tech., 29, 353.

Schotter, Andrew. (1981). The Economic Theory of Social Institutions. Cambridge: Cambridge University Press.

Smith, A. (2019). Future of Workforce Automation: America's Predictions. Retrieved May 5, 2020, from https://www.pewresearch.org/internet/2016/03/10/public-predictions-for-the-future-of-workforce-automation/

Splunk. (2020). Emerging Tech Predictions 2020. Retrieved September 29, 2020, from https://www.splunk.com/pdfs/ebooks/emerging-tech-predictions-2020.pdf.

Stewart, D. (1858). The Collected Works: Biographical memoirs of Adam Smith, William Robertson, Thomas Reid: to which is prefixed a memoir of Dugald Stewart with selections from his correspondence (Vol. 10). Constable.

Swedberg, R., & Granovetter, M. (2018). Introduction to the third edition. The Sociology of Economic Life.

Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. Telecommunications Policy, 44(6), 101961

Waldrop, M. M. (1987). A question of responsibility. AI Magazine, 8(1), 28-28.

Wiegel, V. (2010). Wendell Wallach and Colin Allen: moral machines: teaching robots right from wrong: Oxford University Press, 2009, 273 pp, ISBN: 978-0-19-537404-9. Ethics and Information Technology, 12(4), 359–361. https://doi.org/10.1007/s10676-010-9239-1

Williamson, O.E., (1985), The Economic Institutions of Capitalism. (New York, New York Free Press).

Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Soler, G. J., & von Brauchitsch, B. (2017). Artificial intelligence and robotics and their impact on the workplace. IBA Global Employment Institute, 2012-2017.