1

Artificial Intelligence

The Promise of Disruption

Larry A. DiMatteo

1.1 INTRODUCTION

Disruption – societal and economic – has been a part of humankind from the beginning of time. One example is the transition from a mostly agrarian economy to the industrial age toward the end of the nineteenth century. Over time stalwart career paths were made obsolete. The need for blacksmiths gradually diminished, while the demand for welders arose. The technological age is just another example although the disruption of employable skills and ways of doing business has been amplified due to the recent acceleration of technological development. Advanced artificial intelligence (AI) or superintelligence promises much greater disruption. Although disruption of the status quo is viewed in a derogatory sense, mainly by those anchored in the status quo, taken from the broader view of the betterment of humankind disruption has been a positive force in the macro sense. However, advancement or disruption does not come without costs – the industrial age put the world on the path toward the existential threat of climate change. Nevertheless, this was not an inevitable pathway. If the wealthier nations had the political will, aided by technological breakthroughs, then the environment crisis could have been avoided or diminished.

The world is close to reaching another inflection point: the so-called existential threat of superintelligence with the potential of replacing human control and decision-making with its creation. Before that point, AI and other technologies have caused major disruption in the economy and employment, and this disruption will only accelerate in the future. The pivotal issue is not whether advanced AI is preventable. It is not. Even if it can be delayed there are strong utilitarian and deontological arguments that favor the encouragement of AI development. The focus should be on mitigating the negative effects of disruption and using smart design to prevent AI from ever becoming an existential threat to humankind. Professor John O. McGinnis makes an eloquent argument along these lines in his chapter entitled: "The Folly of Regulating against AI's Existential Threat" (Chapter 27).

Artificial intelligence is currently used in many areas of society – economic, big data, and government activities. It has shown its many benefits in medicine, industry, consumer marketing, and so forth. The acceleration of technology has made it obvious that AI will continue to get smarter with greater abilities to make decisions previously made by humans. The AI of the future will be characterized by greater degrees of autonomy in searching big data,

https://doi.org/10.1017/9781009072168.004 Published online by Cambridge University Press

accelerated machine learning, and physical robots. This brings up the issue of – despite the benefits of AI – what threats does AI pose to society and democratic institutions?[1]

Movers in society, law, design, and technology need to work together to avoid the pitfalls of future advanced AI or superintelligence. Automation has been a core factor in creating more efficient and wealthier economies. However, automation coupled with AI has led to the fear that AI decision-making, which lacks human involvement, will threaten our way of life. A truly autonomous AI system may make decisions not wanted or expected. Some have suggested that the fear of AI as an existential threat is misplaced because it will never be able to replicate the human mind. Nobel Laureate Daniel Kahneman notes that AI is only able to master System 2 thinking, which entails deliberative, rational thinking, but not System 1 thinking, consisting of intuitive and creative thinking, which will remain the domain of human beings.[2] This may be true but there is still the threat that unregulated, unmonitored autonomous systems will sometime in the future coopt the ability of humans to intervene with System 1 thinking.

This chapter sketches out the many issues related to advanced AI and its governance. It reviews the legal and ethical dimensions of that governance. Due to the acceleration of technology, the legal response will lag behind, but at some point, law will have to become forward-thinking in order to anticipate and prevent the dangers that future AI presents. For now, a few rules of thumb can be noted. AI decision-making must not be totally autonomous and must allow for human intervention. Current legal and ethical concepts, such as agency, autonomy, fairness, contract, property, intellectual property (IP), and trust, will be flexible enough to regulate abuse in the early stages of development. At some point in time specialized regulation will need to be created. Specialized regulation will include the use of technological design to ensure compliance with the law. At the ethical level, society should make normative assessments of when the pure efficiency of technological advancement is outweighed by the quality of human life and community. In the end, any solutions to future dangers presented by superintelligence must be an interdisciplinarity process that includes the input of policy makers, technologists, ethicists, and lawyers.[3]

It is clear that human beings will have an ongoing role to play in supervising AI from ethical and legal perspectives. AI will need to be monitored to prevent it from overshadowing the human element. AI will know the world differently than the way humans do; it does not possess first-person experiences that humans have developed over millennia through evolutionary processes. To lose that element would be to lose a vital piece of our humanity.

This book takes a broad view of the current and future uses of AI. It is structured along a number of topical areas. The areas chosen for study show the broad impact of AI, but remain only a selective sampling of the areas, now and in the future, that AI will impact. This introductory chapter discusses AI from a broader societal view through the perspectives of law, ethics, and public policy.

[1] This future threat was symbolized by the computer HAL in Stanley Kubrick's 1968 movie *2001: A Space Odyssey*. HAL represents humans' greatest achievement, which evolves to threaten the destruction of its human overlords.

[2] Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus & Giroux, 2011). For a fuller discussion of the application of Kahneman's theory of thinking see Joshua Davis, "AI, Ethics and Law: A Possible Way Forward," Chapter 21, pp. 306–307.

[3] "Interdisciplinary teams are necessary for AI and application design to bring together technical developers with experts who can account for the societal, ethical and economic impacts of the AI system under design." HUB4NGI, "Responsible AI – Key Themes, Concerns & Recommendations for European Research and Innovation" (June 2018), www.ngi.eu. Permission of Steve Taylor (S.J.Taylor@soton.ac.uk).

The most positive understanding of the value of AI is to see it as a public good: "One area of great optimism about AI and machine learning is their potential to improve people's lives by helping to solve some of the world's greatest challenges and inefficiencies."[4] However, many things introduced to advance the public good work for and against that good. As such, if good is used as the basis of power the illicit use of that power is likely to follow. Something promoted for the public good will still need to be regulated. The key is that regulation must be focused on the overreach of the power that is AI, while not retarding its development. The right kind of regulation works hand in hand with the expansive use of AI. The more any misuse of AI can be prevented (or punished) the greater will be the trust in its development as a safe means to cure the many problems the world faces (e.g., climate change, poverty, equal opportunity, pandemics, and scarcity of resources).

Despite the many benefits that AI promises its development must be placed within a broader landscape of public policy. Automation, for example, is disruptive of the current employment needs of companies by making current skills sets obsolete. A forward-looking industrial policy, which includes the transitioning of workers into the skill sets created by AI, will need to be created. Foreseeable disruption without planning and mitigation will unleash the demons of human nature. The consequences of disruption caused by AI can be lessened by public policies and programs that sit outside of AI innovation.

The ethical implications of replacing human decision-making with AI decision-making must also be forward thinking. There are elements that need to be incorporated into the process of AI development – transparency, human intervention, and skills training. In the area of transparency, understanding the process of AI decision-making is vital to any interests that are impacted by such decisions: "Transparency concerns focus not only on the data and algorithms involved, but also on the potential to have some form of explanation for any AI-based determination." However, the ability to understand advanced AI systems, as well as predicting their behavior, is problematic. Therefore, the second element of ethical AI requires the design of AI that allows for human intervention to monitor and overturn AI decision-making. Finally, technical skills must be accompanied by ethical and legal skills in the design and use of AI. Technological progress of what is possible needs to be done in a framework of "putting good intentions into practice by doing the technical work needed to prevent unacceptable outcomes."[5]

There are many questions being debated about the use of advanced AI, both specific and broad in nature. More specifically, what are the implications that AI systems pose for privacy, security, and protection of personal and sensitive data? What are the ethical implications of AI's use in dealing with consumers? Should AI be granted personhood? More broadly, how should ethical standards and guidelines be developed for AI? How should public policy be constructed relating to AI? What are the benefits and threats of the future development of superintelligence?

This chapter will analyze current soft law instruments and literature to determine the ways society may minimize the risks of superintelligence becoming an existential threat to humanity.[6] Types of regulations to be considered include targeted statutory law, self-regulation, and

[4] National Science and Technology Council, "Preparing for the Future of Artificial Intelligence" (October 2016), 2.
[5] National Science and Technology Council, "Preparing for the Future," 3.
[6] The idea of a robot takeover has been the grist for many sci-fi movies. Some of the doomsday scenarios pose such questions as, "What if one day machines decided that humans were just a waste of resources and started a robocalypse? Or, will artificial general intelligence be humanity's last invention?" See Doomsday Now, "Robot Takeover," https://doomsdaynow.com/robot-takeover/.

https://doi.org/10.1017/9781009072168.004 Published online by Cambridge University Press

standardization.[7] In some ways an analogy can be drawn between the development of advanced AI and the cloning of human beings. Despite the perceived benefits of cloning, such as growing human organ replacements, cloning of human beings has been universally condemned for medical, safety, and ethical reasons. The biggest concern is that it would lead to the creation of "better human beings" violating principles of equality and human dignity. In the same way, the growth of superintelligence threatens the autonomy and dignity of human beings. This chapter will explore the ways that can be used to prevent this from happening.

Section 1.2 explores the nature of law – how it evolves and how this evolution lags behind real-world developments. This lag is generally beneficial because it allows new things to develop more quickly and incentivizes innovation. However, in the age of the acceleration of technology[8] legal gradualism poses a problem in managing the development of advanced AI and fortuitously mapping out impermissible AI systems and applications. Section 1.3 will review the types of principles offered by soft law instruments to encourage the responsible development and use of AI. Section 1.4 discusses the genesis of a European approach to the development of trustworthy AI. Section 1.5 reviews the coverage of the book.

1.2 NATURE OF LAW

The relationship between law and society can be framed a binary one. On the one hand, law needs to respond to developments in society or face becoming obsolete. In this way law is purely reactive in nature. On the other hand, law can be a positive force in the development of society by placing normative limits on social development or what Karl Llewellyn called the "marking off of the impermissible."[9] In this way, law plays a proactive role in shaping how society evolves.

The evolution of law is reactive in nature resulting in a lag between real-world developments and their regulation. The virtue of "lag" is that premature regulation of something new may stifle its development and discourage innovation. The history of the Internet is an example of determining when the best time is to regulate a new technology. There were two points of view regarding the regulation of the Internet – the libertarian view that it should not be regulated in order to allow for it to continue to develop unimpeded and the traditionalist view that the newness of the Internet and unknown dangers posed by such technology required targeted law to prevent abuse. In this case, the libertarian view won out leading to the central role the Internet now plays in daily life. In recent years serious consideration has been given to enact laws, such as the EU General Data Protection Regulation (GDPR), to manage the threat that social media companies and big data, enabled by the Internet, present to human autonomy and dignity. This is the issue now presented by AI: Should it be regulated or not and, if so, when

[7] Standards provide requirements, specifications, and guidelines to be applied to ensure that AI meets its technical and ethical objectives. Standards can address issues in the areas of software engineering (security, monitoring), performance (accuracy, reliability, scalability), safety (control systems, regulatory compliance), interoperability (data, interfaces), security (confidentiality, cybersecurity), privacy (control of information, transmission), traceability (testing, curation of data), and so forth. See National Science and Technology Council, "The National Artificial Intelligence Research and Development Strategic Plan" (October 2016), 32–33.

[8] Thomas Friedman states that: "Technology is now accelerating at a pace the average human cannot keep up with." MIT News, "Thomas Friedman Examines Impact of Global Accelerations" (October 2, 2018), https://news.mit.edu/2018/thomas-friedman-impact-global-accelerations-1003#:~:text=%E2%80%9CTechnology%20is%20now%20accelerating%20at%20a%20pace%20the,added%2C%20emphasizing%20a%20key%20theme%20of%20his%20talk. See also Thomas Friedman, *Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations* (New York: Farrar, Straus & Giroux, 2016).

[9] Karl N. Llewellyn, "Book Review," *Harvard Law Review* 52 (1939): 700, 704.

https://doi.org/10.1017/9781009072168.004 Published online by Cambridge University Press

should it be regulated and how? In this case, the threats to society are on a larger scale in that the creation of better autonomous systems that will lead to a major shift from human decision-making to machine decision-making. The lure of autonomous AI decision-making is that it is more accurate and efficient. The era of big data makes automated processes a necessity.

The reactive nature of law through the ages has generally been a positive feature mainly because of the gradual nature of change. Today, the acceleration and complexity of technology renders law enfeebled in the face of modernity. This presents the problem that if law continues to lag behind the technological advancement of superintelligence (independent decision-making with the ability of human intervention) and super-superintelligence (the loss of the ability of humans to intervene) any regulation will prove to be futile. This is seen in what has been called the alignment problem[10] of advanced AI in which the autonomous system something that it "believes" is in the best interest of its human benefactor and instead the decision is not the one that the human being would have made. Stated differently, AI makes a value judgment that is not aligned with the values or expectations of the human parties. This is akin to the agency problem found in corporate law where director–officer–employee interests may diverge with the interests of the corporation and its shareholders.[11]

Law to be effective will have to be proactive. An analogy can be seen in evolutionary biology where Stephen Jay Gould contested evolutionary theory as a slow, gradual process by showing that the fossil records indicate times of evolutionary "jumps" or "punctuated equilibria."[12] In the area of AI, the law will at some point of development of AI need to jump ahead in order to prevent future dangers from occurring. Hopefully, the "angst of futuristic surrender to an AI and robotically controlled world" will provide the motivation for proactive regulation.[13]

The regulation of future, unknown technological developments seems to be an impossible task, but in fact a plausible regulatory framework can be envisioned to regulate perceived future threats. The precautionary principle used in environmental protection is a case in point. Principle 15 of the United Nations' Rio Declaration on Environment and Development states: "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."[14] Stated more simply, the precautionary principle means that an action should not be taken if the consequences are uncertain and potentially dangerous. Thus, if the consequences of a technological advancement pose potential dangers to human dignity, human rights, or democratic processes then it should be prohibited despite its perceived benefits. This is a rejection of the notion that the benefit of AI is an unassailable truth.

[10] Peter McBurney and Simon Parsons, "Talking about Doing," in Katie Atkinson, Henry Prakken, and Adam Wyner (eds.), *From Knowledge Representation to Argumentation in AI: Law and Policy Making* (London: College Publications, 2013), 151–166.

[11] See Patrick McColgan, "Agency Theory and Corporate Governance: A Review of the Literature from a UK Perspective" (May 22, 2001), https://pdfs.semanticscholar.org/79c5/2954af851c95a27cb1fb702c23feaae86ca1.pdf.

[12] Stephen Jay Gould and Niles Eldredge, "Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered," *Paleobiology* 3 (1977): 115–151.

[13] This danger is known as known as "'singularity,' whereby superintelligent machines take over and permanently alter human existence through enslavement or eradication." Mike Thomas, "The Future of Artificial Intelligence," https://builtin.com/artificial-intelligence/artificial-intelligence-future (updated April 20, 2020).

[14] Report of the United Nations Conference on Environment and Development, A/CONF.151/26 (Vol. I) August 12, 1992, www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.151_26_Vol.I_Declaration.pdf.

Big data and analytics have shown that the harvesting of personal data can produce high profits. This type of monetary incentive will result in the amoral exploitation of AI. An analogy is seen in blockchain technology, which provides a secure, efficient, and anonymous vehicle for transferring information, but in the wrong hands it can be used to illegally launder money. Just as clandestine laboratories may seek to illegally clone a human being, incredulous enterprises may seek to develop types of AI and applications prohibited by future law. The pervasiveness and depth of regulation and monitoring will be pivotal in stemming such illicit activities. The rest of the chapter will analyze a basket of regulations that can be used to prevent the exploitation of AI.

1.3 RESPONSIBLE AI

Dianna Wallis sees the speed of technological development and the complexity of the issues it presents as a call to arms. She asserts that "the sooner we start as a society discussing the issues now presented by advanced technologies, such as AI and superintelligence, the more it is likely that national and international legal systems can develop a holistic approach to the appropriate use and ethical safeguards related to such technologies."[15] This approach requires law and policy makers to be proactive. Instead of waiting until AI poses a threat to democracy and endangers society, "AI needs to be guided by a deliberative political process, to determine how fast and how far such technology should go."[16] The impact of AI programs on democratic processes has been seen in recent elections and noted by the Council of Europe: "AI-based technologies used in online media can contribute to advancing misinformation and hate speech, create 'echo chambers' and 'filter bubbles' which lead individuals into a state of intellectual isolation."[17]

1.3.1 *Landscape of AI, Society, Law, and Ethics*

AI is already in use in many sectors of society touching large companies, government operations, and the consumer marketplace. Figure 1.1 shows one view of an increasingly complex relationship between AI and its stakeholders. It focuses on six themes that need to be considered in creating responsible AI: (1) regulation and control, (2) transparency, (3) responsibility, (4) design, (5) ethics, and (6) socioeconomic impact.[18]

In the area of ethics, society must develop a framework of applied ethics for AI. This includes the selection of existing ethical norms and new norms that adhere to the use of AI. Transparency relates directly to information and education. AI or the humans implementing AI systems must be required to disclose the nature of the processes being used, the personal information being processed, and how decisions are made. Regulation and control require that humans remain in control of autonomous systems including the ability to intervene to change an AI decision. The

[15] Diana Wallis, "Visions of the Future," in Larry DiMatteo, Michel Cannarsa, and Cristina Poncibò (eds.), *Cambridge Handbook on Smart Contracts, Blockchain Technology and Digital Platforms* (New York: Cambridge University Press, 2020), 363–364.

[16] Wallis, "Visions of the Future," 368.

[17] Council of Europe, Report of Committee on Political Affairs and Democracy, "Need for Democratic Governance of Artificial Intelligence," Doc. 15150 (September 24, 2020), 9–10.

[18] See also, Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Berkman Klein Center Research Paper No. 2020-1 (January 15, 2020), https://cyber.harvard.edu/publication/2020/principled-ai. This study provides a meta-analysis of thirty-six AI principles documents and finds eight prominent themes in order of emphasis: privacy, accountability, safety and security, transparency and explainability, fairness and nondiscrimination, human control of technology, professional responsibility, and promotion of human values.

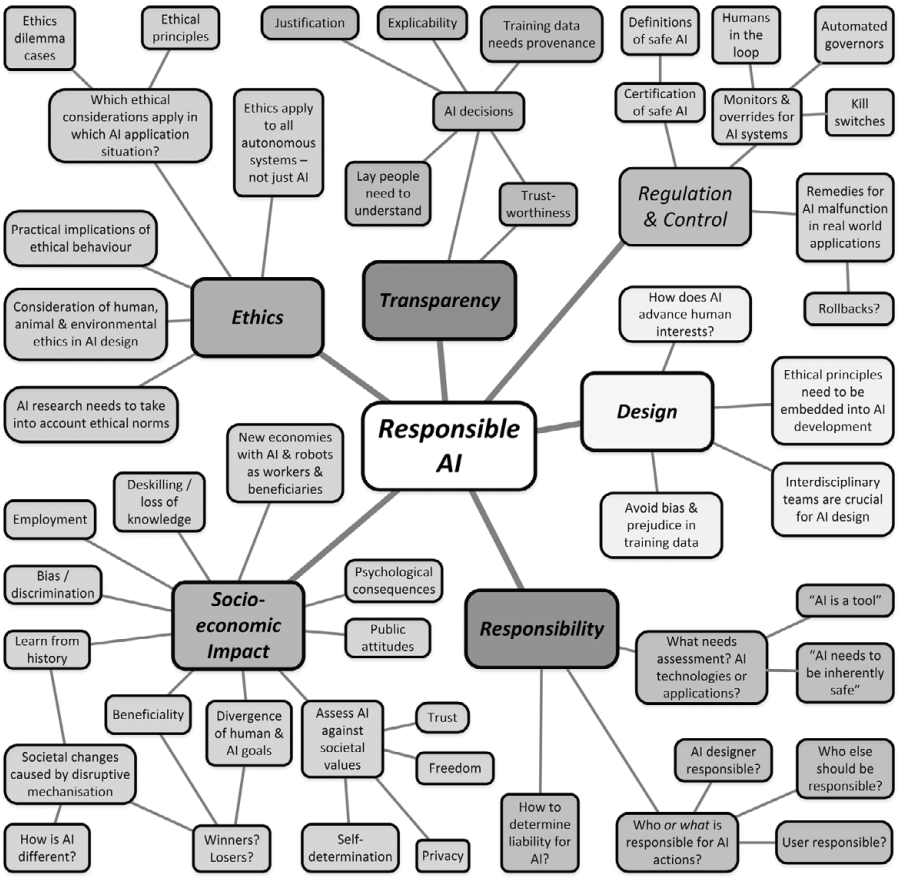

FIGURE 1.1 "Responsible AI" (six main themes)[19]

larger issue is a determination of what is safe AI and the areas where the use of AI is considered inappropriate. This type of assessment must be done through interdisciplinary dialogue as the issues involved cut across the fields of law, computer science, ethics, and technology. It manifests the urgency to turn future conversations on these questions into a "liquid network,"[20] an interdisciplinary space expanding and generating a reliable flow of knowledge.[21]

Design may consist of the use of technology to manage and monitor itself. AI systems must be designed to act ethically and ensure personal information is protected by design. Also, great precaution must be taken so that AI does not replicate the biases of its human programmers. Responsibility is the determination of which of the stakeholders – programmers, creators, owner-users – should be allocated liability if the AI system fails or causes harm.[22] Finally, due diligence

[19] HUB4NGI, "Responsible AI."

[20] Steven Johnson, *Where Good Ideas Come From* (New York: Penguin Group, 2010), 45.

[21] Barbara Pasa and Larry A. DiMatteo, "Observations on the Impact of Technology on Contract Law," in Larry A. DiMatteo, Michel Cannarsa, and Cristina Poncibò (eds.), *Cambridge Handbook on Smart Contracts, Blockchain Technology and Digital Platforms* (New York: Cambridge University Press, 2020), 338, 347.

[22] An example would be how does current product liability law apply to AI systems? See Irina Carnat, "The Notion of Defectiveness Applied to Autonomous Vehicles: The Need for New Liability Bases for Artificial Intelligence," *Trento Student Law Review* 2 (2020): 15 (notes five levels of vehicle autonomy; concludes that the American risk-utility and European consumer expectation approaches to product liability is ill-suited to AI; a better approach would include the development of harmonized technical standards to be applied in the development of autonomous vehicles).

https://doi.org/10.1017/9781009072168.004 Published online by Cambridge University Press

on the socioeconomic impact of AI should be undertaken before its creation and throughout its life cycle. The question to be asked is whether the benefits of AI outweigh its socioeconomic disruptive impact and negative effects on human well-being, which range over concerns of trust, privacy, democratic values, psychological impact in and outside the workplace, and human rights. In sum, just because AI can do something doesn't mean it should be allowed to.

1.3.2 *What Is Wrong with Existing Law?*

The decision not to overly regulate the Internet with specialized bodies of rules proved to be the right decision at least in the beginning of the era of information. Existing legal constructs proved flexible enough to deal with the issues presented. Traditional contract, tort, and intellectual property concepts proved amazingly malleable in controlling internet abuses. For example, the oldest of common law causes of action – trespass – has been used to litigate the improper encroachment on a party's bandwidth. Roger Brownsword refers this ability to fit novel real-world change to existing legal frameworks as the coherentist approach where the fit is a product of manipulation:

> Faced with new technologies, the coherentist tendency is to apply the existing legal framework (the traditional template) to innovations that bear on transactions, or to try to accommodate novel forms of contracting within the existing categories. We need only recall Lord Wilberforce's much-cited catalogue of the heroic efforts made by the courts – confronted by modern forms of transport, various kinds of automation, and novel business practices – to force "the facts to fit uneasily into the marked slots of offer, acceptance and consideration" or whatever other traditional categories of the law of contract might be applicable.[23]

In the words of Brownsword, the regulatory-instrumentalist approach provides an alternative approach. It looks at policy not doctrine as they pertain to particular communities and fundamental values. The difference in approaches is shown in this question: "Even if transactions are largely automated, are there not still Rule of Law concerns implying that there will be some limits on the permitted use and characteristics of [the technology]?"[24] In the end, a combination of both approaches may be needed. Existing legal constructs should be retained and applied to new technologies, but that application should be based on an overt discussion of how best and for what purposes should those constructs be applied. Regulatory instrumentalism will be needed when the peripheral use of existing constructs meets their limitations or borders and more specialized new laws will be needed to align a new technology, such as advanced AI, to community values. This is the point when the benefits of technocracy and efficiency must yield to core values of democracy and human dignity.

1.3.3 *Escaping the Law*

With the advancement of AI and machine learning, a paradigmatic shift, sometime in the future, may be in store, where code will be seen as having the effect of law ("code

[23] Roger Brownsword, "Smart Transactional Technologies, Legal Disruption and the Case for Network Contracts," in Larry A. DiMatteo, Michel Cannarsa, and Cristina Poncibò (eds.), *Cambridge Handbook on Smart Contracts, Blockchain Technology and Digital Platforms* (New York: Cambridge University Press, 2020), 313, 322, quoting Lord Wilberforce in *New Zealand Shipping Co Ltd.* v *A. M. Satterthwaite and Co Ltd.*: The Eurymedon [1975] AC 154, 167.

[24] Brownsword, "Smart Transactional Technologies," 332.

is law").[25] Lawrence Lessig has argued that coders and software programmers, by making a choice about the working and structure of IT networks and the applications that run on them, create the rules under which the systems are governed. The coders therefore act as quasi-legislators. In other words, "code is law" is a form of private sector regulation whereby technology is used to enforce the governing rules.[26] This may be true as a technological fait accompli, but it may not be legal or ethically just. For example, an illegal term cannot be made legal simply by placing a contract on a blockchain. Even though the term will be self-executing, and the contracting parties may have little recourse, these characteristics do not magically make the term legal.

The above example is seen as an attempt to escape the law and the court system. The future of AI will provide a similar scenario but in a more potent form. Will democratic and communal values be lost to technological decision-making? Democracy is not the most efficient of governing systems often infected by waste and corruption. In order to prevent such infections, it may be tempting to turn over governmental activities to AI that will be able to make incorruptible and efficient decisions. In this scenario, AI will rise above the law. This would lead to a diminishment in human value and dignity. AI systems lack the human empathy and judgment so vital to human governance. As stated earlier, the threat is that the advancement of AI may proceed to a point where human intervention is no longer possible. In the short term, responsible AI must be developed and monitored in order to protect basic human values. In the long term, further development of truly autonomous AI systems may have to be prohibited.[27]

1.4 REGULATING AI: AREAS OF CONCERN

The depth of the legal, ethical, and policy literature on AI is enormous. For this reason, this section will focus on the initiatives undertaken by the European Union and the Council of Europe. Many of the issues discussed above are recognized in these documents and some solutions are offered. In many cases, however, no specific solution is offered but a pathway to future regulation is given. In 2017, the European Economic and Social Committee (EESC) identified that the most important AI "societal impact domains include: safety; ethics; laws and regulation; democracy; transparency; privacy; work; education and equality."[28] The European Union and the Council of Europe have recognized the need to work toward a regulatory-ethical scheme to deal with future advances in AI. The European Commission, in June 2018, established the High-Level Expert Group on Artificial

[25] Jia Wang and Lei Chen, "Regulating Smart Contracts and Digital Perspectives: A Chinese Perspective," in Larry A. DiMatteo, Michel Cannarsa, and Cristina Poncibò (eds.), *Cambridge Handbook on Smart Contracts, Blockchain Technology and Digital Platforms* (New York: Cambridge University Press, 2020), 183, 194.

[26] Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999).

[27] See Dirk Helbing et al., "Will Democracy Survive Big Data and Artificial Intelligence?" in Dirk Helbing (ed.), *Towards Digital Enlightenment* (London: Springer, 2019), 73–98; Steven Livingston and Matthias Risse, "The Future Impact of Artificial Intelligence on Humans and Human Rights," *Ethics & International Affairs* 33 (2019): 141–158.

[28] EESC Opinion on AI and society (INT/806, 2017). It should be noted that the EU's major concern in the beginning was to encourage the development of AI. In communications of April 25, 2018 and December 7, 2018, the European Commission set out its vision for AI, which supports "ethical, secure and cutting-edge AI made in Europe." "Three pillars underpin the Commission's vision: (i) increasing public and private investments in AI to boost its uptake, (ii) preparing for socio-economic changes, and (iii) ensuring an appropriate ethical and legal framework to strengthen European values." COM(2018)237 and COM(2018)795.

Intelligence,[29] which began work on "Ethics Guidelines on Trustworthy AI" (*Trustworthy AI*).[30] Subsequently, in 2020, the Council of Europe (COE) Ad hoc Committee on Artificial Intelligence (CAHAI) published "Towards a Regulation of AI Systems" (*Towards Regulation*).[31] These documents discuss the "impact of AI on human rights, democracy and rule of law; development of soft law documents and other ethical-legal frameworks; and drafting of principles and providing key regulatory guidelines for a future legal framework."[32] These documents will be discussed below. The following three sections explore the principles needed to guide the future regulation of AI, AI's threats to human rights, and the elements of trustworthy AI.

1.4.1 *Future Regulation of AI*

Towards Regulation ferrets out a number of ethical themes and related issues:

1. Justice is mainly expressed in terms of fairness and prevention (or mitigation) of algorithmic biases that can lead to discrimination; fair access to the benefits of AI (designing AI systems especially when compiling the training datasets).
2. Nonmaleficence and privacy: misuse via cyberwarfare and malicious hacking (privacy by design frameworks).
3. Responsibility and accountability: includes AI developers, designers, and the entire industry sector.
4. Beneficence: AI should benefit "everyone," "humanity," and "society at large."
5. Freedom and autonomy: freedom from technological experimentation, manipulation, or surveillance (pursuing transparent and explainable AI, raising AI literacy, ensuring informed consent).[33]
6. Trustworthiness: control should not be delegated to AI (processes to monitor and evaluate the integrity of AI systems).
7. Dignity: prerogative of humans but not of robots; protection and promotion of human rights; not just data subjects but human subjects.[34]

Towards Regulation incorporates an Israeli Study[35] and Ethics Report.[36] The Study notes that due to the increasing complexity of AI systems "it is difficult to anticipate and validate their behavior in advance."[37] The Ethics Report enunciates six ethical principles central to creating public policy relating to AI:

[29] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

[30] European Commission, "Ethics Guidelines on Trustworthy AI" (First Draft, December 2018), April 8, 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[31] COE CAHAI, "Towards Regulation of AI Systems," DGI (2020), 16.

[32] COE CAHAI, "Towards Regulation," 7.

[33] An analogy can be drawn to the GDPR's "right to be forgotten."

[34] COE CAHAI, "Towards Regulation," 53–55.

[35] Isaac Ben-Israel, Eviatar Matania, and Leehe Friedman, "Harnessing Innovation: Israeli Perspectives on AI Ethics and Governance," Report for CAHAI, https://sectech.tau.ac.il/sites/sectech.tau.ac.il/files/CAHAI%20-%20Israeli%20Chapter.pdf; COE CAHAI, "Towards Regulation," 120.

[36] *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel*, Special Report to the Prime Minister, eds. Isaac Ben-Israel, Eviatar Matania, and Leehe Friedman (in Hebrew) (September 2020), 32.

[37] COE CAHAI, "Towards Regulation," 130.

https://doi.org/10.1017/9781009072168.004 Published online by Cambridge University Press

1. Fairness: striving for substantial equality, prevention of biases and discrimination (in information, in the process, and in the product), and avoidance of widening socioeconomic and educational gaps.
2. Accountability: incorporates the principles of transparency (information about the process and related decision-making); Explainability: being able to explain on the level of individual users, as well as on a collective level if the system affects groups; Ethical and legal responsibility: determining the responsibilities for setting reasonable measures to prevent the risk of harm.
3. Protecting human rights: preventing harm to life; Privacy: preventing damage to privacy due to collecting, analyzing, and processing information; Autonomy: maintaining the individual's ability to make intelligent decisions; Civil and political rights: right to elect, freedom of speech, and freedom of religion.
4. Cyber and information security: maintaining the systems in working order, protecting the information, and preventing misuse by a malicious actor.
5. Safety: preventing danger to individuals and to society.
6. Maintaining a competitive market and rules of conduct that facilitate competition.

The Ethics Report[38] then proposes the following model, to match different regulatory approaches based on the risk level associated with a particular activity. Thus, for example, high-risk activities are better addressed by legislation and self-regulation ex ante, than by post hoc judicial intervention. At the other end, low-risk activities do not necessarily require dedicated legislation, and can be addressed through standards and self-regulation. This model, of course, is not meant to apply in a rigid fashion. Rather, it presents a framework that enables policy makers and regulators to gauge the appropriate means of regulating an activity, factoring in a multitude of variables. The Report notes that the question of "who regulates" is no less important: regulation by a central AI body enables the development of consistent policies; however, there is a risk of overregulation and chilling innovation if a regulation is adopted across the board. Conversely, regulation could be left to different sector-based bodies, which would allow for greater experimentation, at the expense of uniformity of rules.[39]

1.4.2 *Impact of AI on Human Rights*

The Council of Europe and Commissioner for Human Rights issued a Recommendation involving steps to protect human rights from AI.[40] The Recommendation notes that the threat of AI to human rights is the central concern going forward. It suggests that public authorities perform human rights impact assessments "prior to the acquisition and/or development of [an AI] system" and that assessment must determine "whether an AI system remains under meaningful human control throughout the AI system's lifecycle."[41] The Recommendation also urges that "AI actors take effective action to prevent and/or mitigate the harms posed by their AI systems."[42] Furthermore, AI systems must not be "complex to the degree it does not allow for

[38] K. Nahon, A. Ashkenazi, R. Gilad Bachrach, D. Ken-Dror Feldman, A. Keren and T. Shwartz Altshuler, "Working Group on Artificial Intelligence Ethics & Regulation Report," in *The National Initiative for Secured Intelligent Systems*, 172.

[39] COE CAHAI, "Towards Regulation," 139.

[40] Council of Europe (COE) and Commissioner for Human Rights (CHR), "Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights" (Recommendation) (May 2019).

[41] COE and CHR, "Recommendation," 7.

[42] COE and CHR, "Recommendation," 9.

human review and scrutiny"[43] and independent oversight should be required at the "administrative, judicial, quasi-judicial and/or parliamentary levels."[44] The Recommendation also prohibits the use of "AI systems that discriminate or lead to discriminatory outcomes," which includes "transparency and accessibility of information of the training data used in the development of an AI system."[45] In the area of data protection and privacy, it states that the "use of facial recognition technology should be strictly regulated."[46] The most meaningful protection, which would largely mitigate the fears of an AI takeover, is that "AI systems must always remain under human control, even in circumstances where machine learning or similar techniques allow for the AI system to make decisions independently."[47] Finally, at a societal level, governments should promote AI literacy through "robust awareness raising, training, and education efforts" and developers and appliers of AI should be required to gain knowledge of human rights law.[48]

1.4.3 *Trustworthy AI and Regulation by Design*

The High-Level Expert Group on Artificial Intelligence guide for creating *Trustworthy AI* focuses on three general areas of responsible AI – lawful AI, such as conformity with the GDPR; ethical AI, which is especially important when hard law rules are inexistent; and robust AI. Robust AI is a relatively vague concept that requires AI to "perform in a safe, secure and reliable manner, and safeguards should be foreseen to prevent any unintended adverse impacts."[49] The idea of designing safeguards to prevent any unintended adverse impacts is vital in principle, but as a general statement it is specious and without meaningful content. Finally, the study lists seven requirements for trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, nondiscrimination, and fairness, (6) environmental and societal well-being, and (7) accountability.[50]

The High-Level Expert Group on Artificial Intelligence, discussed above, recognizes the technical component in the development of trustworthy or responsible AI. One element is the development of "whitelist" rules (acceptable or required behaviors or states) that the system should always follow and "blacklist" rules (restrictions on behaviors or states that the system should never transgress). This would be given to AI developers a priori in order to design systems that do not violate the prohibitions and incorporate the required safeguards.[51] An important component of any regulation of AI will be ex ante. Instead of waiting for problems to surface, regulation by design attempts to prevent the problems from occurring in the first place. Early examples of this coopting of technology for regulatory purposes are privacy-by-design and security-by-design. For example, the requirements of the GDPR should be incorporated into the design of an AI system. Thus, law and ethical principles would be used to standardize AI development.

[43] COE and CHR, "Recommendation," 10.
[44] COE and CHR, "Recommendation," 10.
[45] COE and CHR, "Recommendation," 11.
[46] COE and CHR, "Recommendation," 13.
[47] COE and CHR, "Recommendation," 13–14.
[48] COE and CHR, "Recommendation," 14.
[49] European Commission's High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI" (2019), 6–7.
[50] European Commission, "Trustworthy AI," 2.
[51] European Commission, "Trustworthy AI," 14 and 21.

The US Government has advanced a strategic AI development plan.[52] It is characterized as a three-level structure. At the bottom are foundational values or principles that cut across all areas of AI innovation and applications, and include: (1) ethical, legal, and societal implications; (2) safety and security; (3) standards and benchmarks; (4) datasets and environments; and (5) capable AI workforce. Based on these foundational values, basic research and development focuses on two general areas: long-term investments and human–AI collaborations. Under the former category research is focused on data analytics, perception, theoretical limitations, general AI, scalable AI, human-like AI, robotics, and hardware. In the area of human–AI collaboration, the focus is targeted at human-aware AI, human augmentation, natural language processing, interfaces, and visualizations. Finally, in the application of AI, the plan recognizes the fields or sectors of agriculture, communications, education, finance, government services, law, logistics, manufacturing, marketing, medicine, science and engineering, transportation, and security. This strategy is well thought out but implementation of it will be problematic since it is based on collaboration across public-private entities, industries, and academic disciplines.

Regulation by design would not only consist of implementing existing law and ethical principles but would include a development process that anticipates future problems that may have negative ethical and socioeconomic impact. A component of regulation by design includes the allocation of responsibility. Currently, responsibility and potential liability can be easily allocated to the humans who develop and apply AI systems. This is because AI systems today are "closer to intelligent tools than sentient artificial beings."[53] However, "should the current predictions of superintelligence become realistic prospects, human responsibility alone" may not be sufficient. Instead, interdisciplinary assessments will be needed to determine where moral and legal responsibility lies when "AI participates in human-machine networks."[54]

1.4.4 *Glance into the Future*

The current and short-term progeny of AI in commerce and government has reduced costs to businesses and consumers and has effectuated more egalitarian benefits such as greater access to justice. Thus, the fear of the dangers attributed to AI have been inflated. There remains a large chasm between today's AI and that in the foreseeable future and the creation of artificial general intelligence – "a notional future AI system that exhibits apparently intelligent behavior at least as advanced as a person across the full range of cognitive tasks."[55] Even though superintelligence may be decades away, it is important that humankind begin forming an "environment (human) protection" impact study on how best to ensure that superintelligence works to enhance human existence and preserve human dignity.

The above notion of due diligence begins with the current understanding of AI and its applications. Today's regulators need to use the new technologies of today – machine learning, autonomous vehicles and systems, and AI decision-making – to develop frameworks and human capital necessary to deal with the AI of the future. Such frameworks will need to account for numerous factors, such as the quality and costs of certain technologies, as well as security, weaponization, privacy, safety and control, workforce, and fairness and justice concerns. The future regulator will be a technologist knowledgeable of the workings of AI, as well as being

[52] National Science and Technology Council, "AI Strategic Plan," 16.
[53] HUB4NGI, "Responsible AI."
[54] HUB4NGI, "Responsible AI."
[55] National Science and Technology Council, "Preparing for the Future," 7.

steeped in the understanding of the primary importance of democratic institutions and human dignity.

1.5 SCOPE OF COVERAGE

This section describes the book's coverage of a broad selection of topical areas with their own unique issues and problems. It provides a truly interdisciplinary and global perspective of the law and ethics of AI. The author group is a cosmopolitan mix of legal scholars, legal practitioners, and technologists in a variety of countries including Austria, China, Estonia, France, Germany, Italy, Japan, Netherlands, Spain, Switzerland, Turkey, United Kingdom, and United States.

The book is unique due to its breadth of coverage. It is divided into seven parts: Development and Trends; Contracting and Corporate Law; AI and Liability; AI and Physical Manifestations; AI and Intellectual Property Law; Ethical Framework for AI; and Future of AI. Part I introduces the key elements of AI and lays the foundation for the understanding of subsequent chapters. It includes an examination of the potential of AI to make law more efficient and less biased. It also examines the dangers of AI relating to its regulation, liability of entities that use AI, the replication of bias, and threats to democratic institutions. Chapter 2 is written by law and political scientists, as well as technologists who explain the various types of AI from machine learning to AI decision-making. Finally, the impact of AI and technology on the practice of law will be explored.

Part II consists of a series of chapters covering the application of AI to contracting and company law. In the area of contracting, the impact of AI on the negotiation, drafting, and formation of contracts, as well as in the performance of contracts, will be discussed. The final chapter of the part examines the role of AI in corporate decision-making and the board directors' duty of disclosure to shareholders.

Part III examines the issues of liability related to the creation and implementation of AI, including: a comparative analysis of the application of existing tort theories and potential liabilities from the European and American perspectives; an analysis of the question of liability relating to AI decision-making, data protection, and privacy; and an analysis of the application of agency law to AI systems.

Part IV focuses on the physical manifestations of AI, such as self-driving cars, other types of autonomous systems including robots, and the interconnectivity of AI to the Internet of Things. The scholars ask what happens if there are algorithmic errors that cause harm and who is liable for damages? The conclusion is that a new liability regime will be needed to allocate liability between the creators of the AI-controlled manifestations and those who sell or implement the AI system.

Part V examines the intersection of AI and intellectual property law. Key issues to be discussed include the patentability of AI from European and American perspectives; whether AI should be recognized as the creator of intellectual property; and whether AI-generated artistic works should be recognized under copyright law.

Part VI distinguishes between the ethical and unethical uses of AI. Given that regulation often lags behind technological developments, ethics will play an important part in setting limits for AI applications. The focus is on the relationship of AI to consumers in the areas of data privacy and security and the implications of AI for consumer law in general. The topics analyzed include whether AI should be recognized under the law as an artificial being, much like corporations; that is, should advanced AI be given legal status? Also studied are the implications of AI for legal and judicial ethics. How do current ethical standards apply to the lawyer's use of AI? The final

chapter theorizes that the best approach is moving beyond traditional approaches to ethics to a model of standardizing ethical AI.

The final part, Part VII, anticipates the future of AI as a disruptive force in such areas as the role of AI in the judicial system, public policy, legality and regulation of AI, and the ability of competition law to prevent AI collusion. The penultimate chapter ("The Folly of Regulating against AI's Existential Threat") ponders the future of AI. It takes a costs-benefits approach to the potential existential danger of advanced AI and suggests the appropriate government policy toward this accelerating technology. The final chapter summarizes the major findings and recommendations of the book.

Some of the more specific perspectives captured in the analysis include small and large business, government officials and regulators, legal practitioners and educators, ethicists, consumers, and citizens. Cross-chapter analyses cover the use of AI in government decision-making; legal practice (negotiation, drafting, and performance of contracts, as well as company law); ethical use of AI; and legal liability for AI including in tort law, data protection and privacy, and in agency law. Part VII also examines the issue of the liability for AI decision-making, liability for physical manifestations of AI, such as self-driving cars, other autonomous systems, robotics, and harm related to interconnectivity. The symbiotic relationship between AI and intellectual property law is explored including AI as inventor, patentability of AI, and protection of AI-generated works under copyright law.

From a broader perspective, there is the issue of "just because something can be done or achieved does that mean it should be done?" The normative element of autonomous systems and advanced AI are discussed in relationship to consumers, ethical frameworks, AI as a legal person, and control of AI through standardization. In Chapter 27, Professor John O. McGinnis leaves the reader with a positive and hopeful view of the development of advanced AI. He notes that AI as an existential threat to democracy and humanity is mostly speculative, but not certain. In the end, rationality warrants the encouragement of AI research because of the benefits it holds for humankind. AI is not the case for the application of a precautionary principle to prevent unexpected harm. It is simply a product of human creativity that can be harnessed for the greater good. Governments through funding and facilitative regulations or standardization should play a key role in this harnessing.