



Derek C.
Schuurman

DOI: <https://doi.org/10.56315/PSCF12-23Schuurman>

Virtue and Artificial Intelligence

Derek C. Schuurman

Two questions concerning artificial intelligence (AI) and virtue are explored. The first question is whether AI is capable of virtue, and the second question explores if AI can assist humans in the acquisition of virtue. It is argued that AI cannot be a moral agent and therefore cannot genuinely be virtuous. However, AI can perform behaviors that are in accordance with virtuous behavior. The notion of “virtue-by-proxy” is introduced in which virtuous programmers can strive to design AI programs that are trained to mimic certain virtues or behave in accordance with virtue. Next, it is argued that since AI systems can nudge humans toward repeated practices and habits, they will inevitably shape and form users. Thus, AI systems might be designed to assist humans with virtue formation; likewise, they could be misdirected to encourage certain vices. Finally, the concept of virtue in the Christian tradition is contrasted with secular notions of virtue and is used to inform limits on the role of AI in virtue formation.

Keywords: Artificial Intelligence (AI), machine learning, virtue, vice, design, habits, ethics

As I write this article, my streaming music service is offering new songs that match my current selections. Recommendation engines can make use of AI (artificial intelligence) to recognize my listening patterns to make future song suggestions. In fact, I receive daily nudges from AI systems through internet search suggestions, purchase recommendations, video streaming prompts, auto-completion of texts and emails, and personalization of my social media feeds. These daily prompts and nudges can have major impacts on our habits and practices.¹

Aristotle, the ancient Greek philosopher, observed that “moral virtue comes about as a result of habit.”² Likewise, certain vices can be encouraged through poor habits. Aristotle concludes,

It makes no small difference, then, whether we form habits of one kind or another from our very youth; it makes a very great difference, or rather all the difference.³

If AI can be used to influence habits, and habits shape the kind of person we

become, then it follows that “it makes a very great difference” how we will design these new tools.

The topic of AI and virtue pairs a computer science term with a philosophical term. This topic is intrinsically interdisciplinary and requires drawing upon technical, theological, social, and philosophical resources. In fact, any attempt to address this topic strictly from a technical perspective will necessarily involve philosophical and religious presuppositions. As such, these presuppositions are best laid out on the table right from the beginning. Likewise, a strictly philosophical approach to this topic without technical grounding will treat AI as a “black box” (that is, the inner workings are unknown), and consequently, AI will be susceptible to popular myths and assumptions about its capabilities, limits, and features. One

Derek C. Schuurman (PhD in Electrical Engineering, McMaster University) is professor of computer science at Calvin University and a Fellow of the ASA. He is author of *Shaping a Digital World* and co-author of *A Christian Field Guide to Technology for Engineers and Designers*, both published by IVP Academic.

Article

Virtue and Artificial Intelligence

benefit of joining conversations about AI and virtue is that it brings into dialogue “the two cultures” of technology and humanities.⁴

Philosopher Rebecca Konyndyk DeYoung defines virtue as “habits or dispositions of character” that help us “to live and to act well.”⁵ The question is, can an AI have virtue? In other words, can we take AI and “train it up in the way it should go” to show virtue?⁶ A related question is, might AI serve to help humans in the acquisition of virtue? In this article, I will argue that although AI is not capable of virtue itself, it can display a certain degree of *virtue-by-proxy*. I conclude with some thoughts about how AI might assist humans with virtue formation, along with insights from the Christian tradition on virtue.

Is AI Capable of Virtue?

The first question to be addressed is whether AI is capable of virtue. This article will concur with the conclusions of prior works that have claimed, “AI systems cannot genuinely *be* virtuous but can only *behave* in a virtuous way.”⁷ In this section, I will explore how AI and virtue may be connected through a concept that will be referred to as “virtue-by-proxy.”

If virtue helps us to live and to act well, this presupposes a moral agent exercising moral responsibility. Aristotle reflects on moral responsibility in *Nicomachean Ethics* and suggests that moral responsibility hinges on two conditions. The first is a “control condition” which requires that an agent must have a choice over whether to perform an action. The second is an “epistemic condition” that requires the agent to be aware of what they are doing.⁸

In the classic text, *Artificial Intelligence: A Modern Approach* by Stuart Russell and Peter Norvig, AI is defined as “the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment.”⁹ Although an AI system receives inputs and produces output, it meets neither the control condition nor the epistemic condition. While an AI can produce outputs which have ethical implications, it does not meet the control condition since its outputs are predetermined by the computations within its neural network.

Neural networks are “trained” using an algorithm (such as backpropagation) to adjust the weights within a network to minimize or maximize some mathematical goal function. Once the weights are set, the future outputs for a given set of inputs are predetermined, and hence, the AI system does not directly *control* its output.¹⁰ Even AI systems with stochastic elements rely on pseudorandom algorithms, which are also deterministic. Secondly, an AI system does not have *awareness* since it is simply performing calculations. Even impressive large language models (LLMs) are “simply a system for haphazardly stitching together sequences of linguistic forms ... without any reference to meaning: a stochastic parrot.”¹¹ An AI system has no more awareness than a spreadsheet and therefore does not meet the epistemic condition. In a nutshell,

To be responsible, you need to know what you are doing and bringing about, and, in retrospect, know what you have done ... Responsibility then means answerability and explainability.¹²

Since AI systems do not meet these two conditions for moral responsibility, neither can they be capable of virtue.

To be clear, the lack of moral responsibility does not imply that AI is neutral, nor does it preclude the responsibility of those who design and deploy AI systems. Moral responsibility is distinct from the area of AI ethics, which is the application of ethical principles to ensure that machines are designed in ways to protect people and the environment. A helpful document titled “Moral Responsibility for Computing Artifacts,” developed by an interdisciplinary group of philosophers, computer scientists, practitioners, and lawyers, states this clearly: “The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact.”¹³

The Possibility of Virtue-by-Proxy

Since AI systems cannot have moral responsibility, it follows that they cannot display virtue and any appearance of virtues are, in fact, *ersatz* virtues. However, some have speculated that autonomous software systems might conceivably serve as a *proxy* for human responsibility. Computer scientist Nick Breems proposes the notion of “subject-by-proxy”

by which “responsibility could be inherited by programs from the programmer.”¹⁴ Breems suggests that a developer “exercises her responsibility by creating a system that will behave normatively *in the real world*, after the developer’s participation is no longer active” and can do so “by encoding normativity.”¹⁵ However, Breems is careful to qualify his proposal, acknowledging the challenge of encoding the “nuanced, intuitively grasped concepts of diverse normativity ... into a form that could be actualized as ‘goals’ for the artificial agent.”¹⁶ Breems relies on the philosophical framework of the Dutch philosopher Herman Dooyeweerd which rejects the notion that everything can be reduced to algorithms. Dooyeweerd’s philosophy contends that only humans can function as subjects in normative areas such as justice, ethics, and faith.

The notion of “subject-by-proxy” could be extended to a similar notion of “virtue-by-proxy.” Using this approach, one might maintain that AI systems are not capable of virtue but, nevertheless, serve as a proxy to the virtue of the programmers. Virtuous programmers can strive to create AI programs that are trained to mimic virtue-like behaviors. Such virtues might include humility; for example, by anticipating the need for extensive error detection and exception handling. AI systems might also echo the virtue of civility through friendly and hospitable user interfaces, or autonomous vehicles could mimic the virtues of courteous drivers. Furthermore, AI programs could create conditions where users are afforded opportunities to practice habits that accord with humility.

Recent research has uncovered an effect called “latent persuasion” in which large language models (LLMs) can provide nudges to change human behavior “by making some choices more convenient than others.”¹⁷ Whereas this could be exploited for ill, it could also be directed toward virtue by nudging people toward a “disposition to live well with one’s fellow citizens” in their online interactions.¹⁸

AI system design might exercise virtue-by-proxy by being attentive to justice and fairness and addressing bias in machine learning. Author Cathy O’Neil provides insightful suggestions for working toward justice in machine learning in her book *Weapons of Math Destruction*.¹⁹ The virtue of empathy may be implemented by proxy by creating software that

responds to the emotional state of the user. One researcher, Rosalind Picard, has explored “affective computing” by designing “computers that interact with people.” She writes that these computers “recognize emotions and how to intelligently respond to them, including when to show empathy.”²⁰ To say a computer can “show empathy” is problematic language since it implies agency, but the notion of virtue-by-proxy shifts the agency to a virtuous programmer who designs an AI system to *mimic* virtuous behaviors, such as empathy.

Since machine learning requires a mathematical goal function to optimize, the question immediately arises as to how behaviors that accord with virtue might be implemented as goal functions. One recent approach that has been explored is reinforcement learning from human feedback (RLHF) in which human feedback is used to further tune an AI model.²¹ In the case of virtue-by-proxy, human feedback could be used to nudge a machine learning model to exhibit behavioral outputs that accord with virtue. In this case, the virtues that are implemented by proxy are not those of the programmers, but rather of the humans providing the reinforcement learning feedback. One example might be to train an LLM to mimic the virtue of civility. However, recent work with LLMs has demonstrated that RLHF tuning has many challenges and tamping down unwanted behavior remains challenging.²² Some of the issues include the vast amount of feedback needed to tune a large model, variance in feedback among multiple human trainers, and the fact that feedback is typically limited to simple preference ordering of outputs.²³

Aside from the limits intrinsic to reinforcement learning from human feedback, there are additional limits to virtue-by-proxy. For example, while an AI system may be able to mimic empathy, it is entirely incapable of *feeling* empathy. The social scientist Sherry Turkle suggests that “children need to be with other people to develop mutuality and empathy; interacting with a robot cannot teach these.”²⁴ Likewise, it should be noted that there are many challenges in implementing justice and fairness in computers. For example, individual and group fairness can sometimes form competing requirements in machine learning.²⁵ Other justice challenges can arise in datasets due to effects such as Simpson’s Paradox which underscores “the importance of human experts in

Article

Virtue and Artificial Intelligence

the loop to examine and query Big datasets.”²⁶ In fact, there will be times when justice may demand that certain things *not* be automated. Frankly, it is difficult to imagine how many of the “technomoral virtues” suggested by philosopher Shannon Vallor might even be approximated in software—virtues like courage or magnanimity.²⁷ This presents further complications if one holds to the “unity of the virtues” (as Aristotle did) in which one virtue depends on all the others.²⁸ Furthermore, if computers can manipulate only quantifiable values, and if virtue includes factors that are not easily quantifiable, then virtuous behavior can only be approximated at best. The adage is true that “not everything that counts can be counted,” and thus, virtue cannot be reduced to mathematical computations or an algorithm. Hence, one should be quick to acknowledge the many limitations to the notion of virtue-by-proxy.

The wider challenge of steering AI toward human goals and ethical behavior is an open area of research referred to as the “value alignment” problem.²⁹ Already in 1960, the AI pioneer Norbert Wiener anticipated this problem when he wrote,

If we use, to achieve our purposes, a mechanical agency with whose operation we cannot interfere effectively ... we had better be quite sure that the purpose put into the machine is the purpose which we *really desire*.³⁰

Cautionary tales include “The Sorcerer’s Apprentice” in the Disney film *Fantasia*, in which Mickey Mouse instructs a broom to fill a cauldron, only to have it multiply and run amok. Similarly, philosopher Nick Bostrom’s thought experiment imagines an AI whose goal function is to maximize the production of paper clips, and then it proceeds to convert the earth and large portions of the observable universe into paper clips.³¹

In the end, it requires wisdom to discern the extent to which virtue-by-proxy is appropriate or even possible. For this reason, while the notion of virtue-by-proxy may be philosophically intriguing, its practicality will be extremely limited. In his “subject-by-proxy” proposal, Breems is careful to remind us that we should “avoid attempts to imbue software with emergent moral agency.”³² He concludes that it is ultimately “involved human beings,” both users and developers, “that must bear the responsibility,”

and they must “*delegate* [their] responsibility to the computer with great care.”³³

Finally, it should be noted that attempts to build machines using the notion of virtue-by-proxy should *not* be conflated with creating machines that pretend to be human. Creating machines that look and sound like human persons can lead to a kind of ontological confusion. Machines are machines—they are not human—and the virtue of honesty should oblige us not to create machines that pretend to be human persons. In fact, one could make the case that building a machine that looks and responds like a human is essentially a form of deception.³⁴ In the words of theologian Craig Bartholomew,

We should start with ontology—this is our Father’s world, and we are creatures made in his image—and then move on to epistemology—as his creatures, how do we go about knowing this world truly?³⁵

A common pitfall is to anthropomorphize our machines and, in so doing, to elevate machines and reduce the distinctiveness of human beings.³⁶ Once we have established the ontological distinction of who we are and what machines are, we can then begin addressing questions about the appropriate use of AI.³⁷

Can AI Assist Humans in the Acquisition of Virtue?

If virtuous AI is not possible, could it still be used to assist humans in the acquisition of virtue? In a recent paper by Boyoung Kim et al., an experiment was performed in which a robot verbally provided advice to “guide humans to comply with the norm of honesty.”³⁸ Their experiment “indicated that robots may not be suitable for serving in the role of a moral advisor.”³⁹ While verbal advice from a robot may have limited impact, the ability for AI systems to nudge humans toward repeated practices and habits will inevitably shape and form users toward virtue—or vice.⁴⁰

Some current examples of software that can nudge us toward virtues of self-control are apps which remind users to exercise, or even gamify exercise to entice users toward improved fitness. Dieting apps can help users manage their appetites and food

intake and digital well-being apps can help users limit screen time and social media usage. There are also apps that can help users cultivate spiritual disciplines, such as prayer, personal devotions, as well as scripture reading and memorization. A focus on virtue formation could stimulate further innovative ideas that leverage the capabilities of AI.

In a similar manner, AI can be crafted to encourage vice. In book 2 of his *Republic*, Plato describes the “Ring of Gyges.” The ring is a kind of technology that allows the user to become invisible at will. Plato uses this thought experiment to consider whether such a technology might encourage a rational person to act unjustly since they could perform actions without being seen and therefore avoid any consequences. Plato observes,

If you could imagine anyone obtaining this power of becoming invisible, and never doing any wrong or touching what was another’s, he would be thought by the lookers-on to be a most wretched idiot ...⁴¹

A modern equivalent could ask the question, “Would a decent person act differently when they are able to view and post anonymously online?”

There are many examples of how AI-driven algorithms can encourage certain types of vices. Consider how video streaming services entice you to binge-watch by automatically playing the next episode or recommending other things to view. Likewise, consider the dopamine effects of video games and social media that keep their users playing or scrolling for long periods of time. Such systems can encourage the vice of sloth. Social media can also encourage the vice of envy as we absorb the highlights of other people’s curated lives. Moreover, social media can “foster and feed on vainglory impulses.”⁴² Online pornography inflames lust, and online conversations driven by social media algorithms optimized for engagement can often spiral into outrage and wrath. AI can be easily misdirected to encourage each one of the seven vices.⁴³

Virtue in the Christian Tradition

AI may plausibly assist in a limited way with virtue formation through nudging us toward good habits and practices. But virtue formation in the Christian tradition is not just about “what to do and what not to

do,” it also involves the “larger category of the divine purpose for the entire human life.”⁴⁴ Philosopher Alasdair MacIntyre observes, “I can only answer the question ‘What am I to do?’ if I can answer the prior question ‘Of what story or stories do I find myself a part?’”⁴⁵ For the Christian, virtue involves living into the biblical story. Aristotle’s vision of virtue was that of the “moral giant striding through the world doing great deeds and gaining applause.”⁴⁶ In contrast, “Christian virtue isn’t about you ... It’s about God and God’s kingdom.”⁴⁷

The word for virtue does not occur in the New Testament, but there is an emphasis on “the careful development and cultivation of Christian character.”⁴⁸ In fact, the goal of the Christian life is to become more like Christ—something we cannot do on our own. Saint Augustine recognized this when he hears God ask, “Why are you relying on yourself, only to find yourself unreliable?” Rebecca Konyndyk DeYoung observes, “You won’t practice the spiritual disciplines long, however, before you confront the sober truth: We can’t make ourselves Christlike, no matter how hard we try.”⁴⁹ She continues, “Practice, discipline, and all the things we do can’t be the whole story, because human agency is not the whole story.”⁵⁰ Theologian N. T. Wright observes that the Christian virtues “remain both the work of the Spirit and the result of conscious choice and work on the part of the person concerned.”⁵¹

In addition to the four “cardinal” virtues described by the ancient Greeks (wisdom, justice, courage, and temperance), the Christian tradition recognizes the three theological virtues of faith, hope, and love. While ancient Greek virtues were aimed at cultivating the individual, Christian virtues “point away from ourselves and outward: faith, toward God and his action in Jesus Christ; hope, toward God’s future; love, toward both God and our neighbor.”⁵² If love is the primary virtue, then it is one that needs to be practiced in the context of community.⁵³

In fact, modern notions of virtue are often humanistic versions of what were once distinctively Christian concepts, what MacIntyre calls “fragmented survivals from an older past” and “ghosts of conceptions of divine law.”⁵⁴ Many modern conceptions of virtue are operationally defined and are very different outside their original theological frame. For example, a Christian view of humility is not just a view of self

Article

Virtue and Artificial Intelligence

or others but is also grounded in “a trust that one’s well-being is entirely secured by the care of God.”⁵⁵ In some modern definitions, humility might be connected to a “wonder at the universe’s retained power to surprise and confound us”; that is quite different from trusting in the care of a personal God.⁵⁶

Since virtue is not just operationally defined in the Christian tradition, the notion of virtue-by-proxy is a limited concept. Likewise, the potential role for AI in virtue formation is more modest. But Christians should nevertheless recognize the contribution of habits and rituals in their spiritual formation, including the nudges that may come from the AI systems they encounter. Christian philosopher, James K.A. Smith, refers to habits and practices as kinds of *liturgies* that “take hold of our gut and aim our hearts toward certain ends.”⁵⁷ It is for this reason that Smith recommends that we perform a “liturgical audit” of our lives.⁵⁸ A prudent extension to this advice would be to include an audit of the liturgies that may come with AI technology, for both discerning users and responsible designers.

Conclusion

In conclusion, I have argued that AI is not capable of virtue, but there might be an argument for a very limited form of virtue-by-proxy. While virtue-by-proxy is an intriguing philosophical notion, ultimately, it has many limitations and shortcomings. At the very least, the notion of virtue-by-proxy is a reminder that AI systems should be designed with care and responsibility since they operate far from the programmer in both time and space. Of course, virtue-by-proxy presupposes a virtuous system designer. For this reason, it is essential that the education of engineers and computer scientists address virtue formation alongside the development of technical skills.⁵⁹

Although AI is not capable of virtue, AI systems are capable of nudging users in a variety of ways, and thus may have some limited role in virtue formation (or alternately, in encouraging vices). In the case of the Christian tradition, the role of AI in virtue formation will be even more limited, since the Christian notion of virtue is situated within the context of the biblical story and is not just operationally defined.

The Christian computer scientist, Frederick Brooks, has suggested that rather than striving for AI (arti-

ficial intelligence), a better approach would be IA (intelligence amplification). Rather than striving to build “giant brains” with AI, Brooks suggests that IA is the better approach—using a machine alongside a human mind.⁶⁰ This sentiment might inform AI and our approach to virtue as well: instead of trying to build “AV” (artificial virtue), a wiser approach will be to build machines for “VA” (virtue amplification)—machines that can assist humans in exercising virtue. But first we need to practice virtue ourselves—cultivating habits and liturgies that help shape us into the kind of people God calls us to be. Only then can we begin to develop AI with the wisdom needed to direct it in responsible and obedient ways.

Notes

¹Richard H. Thaler and Cass R. Sunstein, *Nudge: The Final Edition* (New York: Penguin Books, 2021), 4.

²Aristotle, *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis, IN: Hackett Publishing, 1985).

³Ibid.

⁴C.P. Snow, *The Two Cultures: And a Second Look* (Cambridge, UK: Cambridge University Press, 1964).

⁵Rebecca Konyndyk DeYoung, *Glittering Vices: A New Look at the Seven Deadly Sins and Their Remedies*, second edition (Grand Rapids, MI: Brazos Press, 2020), 7–8.

⁶This is a similar phrase to that found in Proverbs 22:6 where it refers to training up children in the way they should go.

⁷Mihaela Constantinescu and Roger Crisp, “Can Robotic AI Systems Be Virtuous and Why Does This Matter?,” *International Journal of Social Robotics* 14, no. 6 (August 2022): 1548, <https://doi.org/10.1007/s12369-022-00887-w>.

⁸Matthew Talbert, “Moral Responsibility,” in *The Stanford Encyclopedia of Philosophy* (Fall 2023 edition), ed. Edward N. Zalta and Uri Nodelman, <https://plato.stanford.edu/archives/fall2022/entries/moral-responsibility/>.

⁹Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, third edition, Prentice Hall Series in Artificial Intelligence (Upper Saddle River, NJ: Prentice Hall, 2010), viii.

¹⁰There are systems called hypernetworks which use a network to generate the weights for another network, but the weights of the hypernetwork themselves are fixed. For more information on hypernetworks, see David Ha, Andrew Dai, and Quoc V. Le, “HyperNetworks,” version 4, December 1, 2016, <https://arXiv.org/abs/1609.09106>.

¹¹Emily M. Bender et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?,” in *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (March 2021), 617, <https://doi.org/10.1145/3442188.3445922>.

¹²Mark Coeckelbergh, *AI Ethics*, The MIT Press Essential Knowledge Series (Cambridge, MA: The MIT Press, 2020), 115.

¹³Keith W. Miller, “Moral Responsibility for Computing Artifacts: ‘The Rules,’” *IT Professional* 13, no. 3 (May 2011): 57–59, <https://doi.org/10.1109/MITP.2011.46>.

- ¹⁴Nick Breems, "Subject-by-Proxy: A Tool for Reasoning about Programmer Responsibility in Artificial Agents," *ACM SIGCAS Computers and Society* 47, no. 3 (September 25, 2017): 65–71, <https://doi.org/10.1145/3144592.3144599>.
- ¹⁵Since AI systems are not moral agents, it might be more philosophically precise to state that they can "behave in accordance with normativity in the real world." See Breems, "Subject-by-Proxy."
- ¹⁶Breems, "Subject-by-Proxy," 71.
- ¹⁷Maurice Jakesch et al., "Co-Writing with Opinionated Language Models Affects Users' Views," in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (April 2023), 1, <https://doi.org/10.1145/3544548.3581196>.
- ¹⁸Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (New York: Oxford University Press, 2018), 140–45.
- ¹⁹Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).
- ²⁰Rosalind W. Picard, *Affective Computing* (Cambridge, MA: MIT Press, 2000), 77.
- ²¹Gabrielle Kaili-May Liu, "Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback," March 6, 2023, <https://doi.org/10.48550/arXiv.2303.02891>.
- ²²Yotam Wolf et al., "Fundamental Limitations of Alignment in Large Language Models," 2023, <https://doi.org/10.48550/arXiv.2304.11082>.
- ²³Liu, "Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback," 2.
- ²⁴Sherry Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other*, Revised and expanded edition (New York: Basic Books, 2017), 56.
- ²⁵Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian, "The (Im)Possibility of Fairness: Different Value Systems Require Different Mechanisms for Fair Decision Making," *Communications of the ACM* 64, no. 4 (April 2021): 136–43, <https://doi.org/10.1145/3433949>.
- ²⁶Rahul Sharma et al., "Why Not to Trust Big Data: Discussing Statistical Paradoxes," in *Database Systems for Advanced Applications: DASFAA 2022 International Workshops: BDMS, BDQM, GDMA, IWBT, MAQTDS, and PMBD*, ed. Uday Kiran Rage, Vikram Goyal, and P. Krishna Reddy. DASFAA 2022, Lecture Notes in Computer Science, vol. 13248 (Cham, Switzerland: Springer-Verlag, 2022), 61, https://doi.org/10.1007/978-3-031-11217-1_4.
- ²⁷Vallor, *Technology and the Virtues*, 120–54.
- ²⁸Alasdair C. MacIntyre, *After Virtue: A Study in Moral Theory*, third edition (Notre Dame, IN: University of Notre Dame Press, 2007), 157.
- ²⁹Stuart J. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (New York: Penguin Books, 2020), 137.
- ³⁰Norbert Wiener, "Some Moral and Technical Consequences of Automation: As Machines Learn They May Develop Unforeseen Strategies at Rates That Baffle Their Programmers," *Science* 131, no. 3410 (May 6, 1960): 1355–58, <http://www.jstor.org/stable/1705998>.
- ³¹Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, reprinted with corrections (Oxford, UK: Oxford University Press, 2017), 150.
- ³²Breems, "Subject-by-Proxy," 71.
- ³³Ibid.
- ³⁴Amanda Sharkey and Noel Sharkey, "We Need to Talk about Deception in Social Robotics!" *Ethics and Information Technology* 23, no. 3 (September 2021): 309–16, <https://doi.org/10.1007/s10676-020-09573-9>.
- ³⁵Craig G. Bartholomew, *Contours of the Kuyperian Tradition: A Systematic Introduction* (Downers Grove, IL: IVP Academic, 2021), 103.
- ³⁶Some forms of transhumanism can also blur the lines between humans and machines.
- ³⁷Derek C. Schuurman, "Artificial Intelligence: Discerning a Christian Response," *Perspectives on Science and Christian Faith* 71, no. 2 (June 2019): 79, <https://www.asa3.org/ASA/PSCF/2019/PSCF6-19Schuurman.pdf>.
- ³⁸Boyoung Kim et al., "Robots as Moral Advisors: The Effects of Deontological, Virtue, and Confucian Role Ethics on Encouraging Honest Behavior," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21: ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO: ACM, 2021)*, 10, <https://doi.org/10.1145/3434074.3446908>.
- ³⁹Ibid., 17.
- ⁴⁰Kate Lucky, "Imago AI," *Christianity Today*, October 2023.
- ⁴¹Plato, *Republic*, trans. G.M.A. Grube and rev. C.D.C. Reeve (Indianapolis, IN: Hackett Publishing, 1992), 360 b–d.
- ⁴²DeYoung, *Glittering Vices*, 41.
- ⁴³Robinson Meyer, "The Seven Deadly Social Networks," *The Atlantic*, May 9, 2016, <https://www.theatlantic.com/technology/archive/2016/05/the-seven-deadly-social-networks/480897/>.
- ⁴⁴N.T. Wright, *After You Believe: Why Christian Character Matters* (New York: HarperCollins, 2012), 69.
- ⁴⁵MacIntyre, *After Virtue*, 216.
- ⁴⁶Wright, *After You Believe*, 70.
- ⁴⁷Ibid., 70.
- ⁴⁸Ibid., 60.
- ⁴⁹DeYoung, *Glittering Vices*, 224.
- ⁵⁰Ibid., 224.
- ⁵¹Wright, *After You Believe*, 197.
- ⁵²Ibid., 204–5.
- ⁵³Ibid., 144.
- ⁵⁴MacIntyre, *After Virtue*, 111.
- ⁵⁵Kent Dunnington, *Humility, Pride, and Christian Virtue Theory*, Oxford Studies in Analytic Theology (New York: Oxford University Press, 2018), 88.
- ⁵⁶Vallor, *Technology and the Virtues*, 126–27.
- ⁵⁷James K.A. Smith, *Desiring the Kingdom: Worship, Worldview, and Cultural Formation*, volume 1 of Cultural Liturgies (Grand Rapids, MI: Baker Academic, 2009), 40.
- ⁵⁸James K.A. Smith, *You Are What You Love: The Spiritual Power of Habit* (Grand Rapids, MI: Brazos Press, 2016), 53–55.
- ⁵⁹William Jordan, "A Virtue Ethics Approach to Engineering Ethics," in *2006 Annual Conference & Exposition Proceedings* (2006 Annual Conference & Exposition, Chicago, IL: ASEE Conferences, 2006), 11.142.1-11.142.11, <https://doi.org/10.18260/1-2--595>.
- ⁶⁰Frederick P. Brooks, "The Computer Scientist as Toolsmith II," *Communications of the ACM* 39, no. 3 (March 1996): 64, <https://doi.org/10.1145/227234.227243>.