Character and . . .

Transitions

Volume 7 / 2021

Articles

Response

Wendt Character Initiative
University of Dubuque

Editors

The faculty essays presented here emerge from a semester-long process of reading and writing together in an environment of critique and review. Nevertheless, this invited journal of essays represents the authors' views and not necessarily the views of the Wendt Center for Character Education or the University of Dubuque.

Contact Information

Wendt Center for Character Education
University of Dubuque
2000 University Avenue
Dubuque, IA 52001

563-589-3440 (office)
563-589-3243 (fax)
wendt@dbq.edu (e-mail)
www.dbq.edu/wendt (website)

Copyright Notice

The *Character and . . .* Journal is published by the Wendt Center for Character Education at the University of Dubuque in Dubuque, Iowa, and uses parenthetical citations in the style of the 9th edition of the *MLA Handbook.*

Stepping over the Brink into Artificial Intelligence

Ann Mauss

Abstract
We live in a world where we are in transition, increasingly relinquishing our decision-making process to computers, but are computers trustworthy enough to make our decisions for us? Artificial Intelligence, which drives the computing decision process, is a growing field in computing, but we must understand how it works and the justice and ethical issues it faces in order to ensure that decisions by current and future algorithms and A.I.s reflect moral virtues.

Would you follow your phone's directions off a cliff? We have all heard the stories of people trusting their phones so implicitly that they end up having accidents. They have walked off Chicago's Navy Pier into Lake Michigan, into traffic, and most often, into other people, but a man from the United Kingdom almost followed his phone off a cliff. Robert Jones had traveled many miles using the GPS app on his phone. Mr. Jones was driving to a place he had never been before, and the GPS app instructed him to turn down a dirt lane. What he did not know was that it was just a small footpath ending at a cliff. Fortunately for him, there was a fence at the top of the cliff, so he was stopped before making a 100 ft. drop (Brooke).

Would you follow your phone's directions off a cliff?

What would have caused the app to decide that the footpath was a valid road for vehicles? Why would Robert Jones follow so unquestioningly? We are living in a world where our lives are changing at a rapid pace, and technology is a driver in that change. We look to computers for answers because we expect them to make better decisions than us. With little or no thought to the consequences, we increasingly relinquish our decision-making processes, ceding more and more control to computers and computer applications.

> *We increasingly relinquish our decision-making processes, ceding more and more control to computers and computer applications.*

The problem is that computers are not trustworthy, just, or virtuous. As computer intelligence increases at exponential rates, we need to understand how they arrive at their answers and decisions. This article discusses the machine-learning algorithms that are behind many of the applications we use, the data used and information produced, the potential for inherent bias and lack of transparency in how they are developed, and the need to bring virtue into the conversation. We stand at a precipice in our society, transitioning from an analog decision-making process to a digital one with our feet teetering on the edge of this proverbial cliff.

Teaching a Dog to Sit

We use computer applications for all types of functions, and each of these follows some type of algorithm. Simply defined, an algorithm is a precise step-by-step set of instructions for solving a task. If these steps are followed correctly, it will result in a solution to a given task. It is a broad definition because any recipe can count as an algorithm. Computer algorithms, however, have a more specific definition.

Computer algorithms are mathematical objects—using a wide variety of equations, algebra, calculus, logic and probability—applied in the code of a programming language. “They are given an objective, and set to work crunching through calculations to achieve their aim” (Fry 8). Algorithms function in four basic ways, and most computer applications use at least one of the following functions, but many use a combination of them. It is important to understand how they function, so that we can address the various problems they can cause.

- **Prioritization**—Algorithms present an ordered list. We see this type of algorithm when GPS apps find the fastest route.
- **Classification**—Algorithms find similar things and group them. These are often used by advertisers to target ads based on a person's previous choices.
- **Association**—Algorithms identify and mark relationships between things. You can see these working in dating sites or at Amazon in the "those who looked at that looked at this."
- **Filtering**—Algorithms determine what is important and filter out what isn't. Voice recognition software does this, as well as social media sites that filter news stories to fit in with your personal feed (Fry 8–10).

Most applications use a combination of these algorithmic functions at any time in their programing, but what really counts is how they perform these tasks. As stated earlier, algorithms are step-by-step instructions, but it is how these steps are defined that makes a difference. Humans view the world in ranges, spectrums, and frequencies and see many shades of gray. We call this an analog view. Computers, on the other hand, work in a black and white world, a state of ones and zeros, ons and offs. We call this a digital or binary view. It is because of this binary, digital simplicity that we humans blindly trust computers for accurate information and believe that they can make more impartial decisions than we can make on our own.

There are two basic approaches to designing algorithms: rules-based algorithms and machine-learning algorithms. Rules-based algorithms are closer to the textbook definition of algorithms. Programmers define and design the step-by-step logic that the computer uses to solve a task. It is easy for humans to follow how the computer arrived at a solution or correct it when it fails. These algorithms are used to solve known problems, and there is little ambiguity to them. Machine-learning algorithms are very different. These algorithms fall under the broader category of Artificial Intelligence (A.I.). The implementation of A.I. is in its infancy, but it is advancing at a rapid pace.

Machine-learning algorithms are based on how humans—or any kind of animal—learn how to do something. For example, I trained my dog to stay by just holding my hand in a specific position. She learned that specific hand position means *stay* through a series of

Feedback serves as a reward to algorithms for correct answers.

trial and error. I would reward her when she stayed and would withhold the reward if she performed any other behaviors when I gave that hand signal. Eventually, she figured out what that hand signal meant, and she became steadily faster at associating staying with a reward, quickly responding with the correct action. Machine-learning algorithms "learn" in a similar fashion. It starts by stating a clear objective that we want to reach. The programming provides feedback by offering a "reward" or acknowledgment of a correct answer, and no reward for an incorrect answer. Like a mouse in a maze, the algorithm finds the right path through trial and error. To give it a simplistic definition, "You give the machine data, a goal, and feedback when it's on the right track—and leave it to work out the best way of achieving the end" (Fry 11).

There are issues with each aspect of the algorithm learning process. For instance, humans cannot always trace the algorithm's learning path and many times this opaqueness is by design of the programmers. This is the issue of transparency.

Computer Algorithms

Aspect	**What is it?**	**Questions Raised**
Data	The initial information fed to the algorithm	What data is used? Where does it come from?
Feedback	The answer determined by the algorithm	What is considered a correct answer? Who or what determines the correct answer?
Learning	The path the computer takes to get to an answer	How does it arrive at the answer?

A common use of these algorithms is teaching a computer to "talk" and the implementation of these algorithms have revealed multiple problems. Conversational speech applications have revealed multiple issues concerning social bias and algorithm transparency.

Teaching a Computer to Talk

A type of programming that machine-learning algorithms excel at is Natural Language Processing (NLP), which is a branch of A.I. These algorithms help computers understand, interpret, and manipulate human language. Let's break down the parts of an A.I. algorithm. It needs a lot of data to learn from. In most cases, programmers look to the Internet and the abundant and widely available data that it offers, believing that data, in large scales, will provide algorithms with an exponential set of data points that broaden the learning outcomes. It can come from Google, Twitter, or a myriad of other platforms or data brokers.[1]

NLP algorithms provide an excellent example of how data is used and processed, and how they reflect an inherent bias. These algorithms start with a known dataset of words, called word embeddings. There are several embeddings widely available on the Internet that can be used. "The words become vectors in a multi-dimensional space, where nearby vectors represent similar meanings. With word embeddings, you can compare words by (roughly) what they mean, not just exact string matches" (Speer 3). This is comparable to using words in the context of a sentence to find a definition of a word. These are the input (data) for the algorithms.

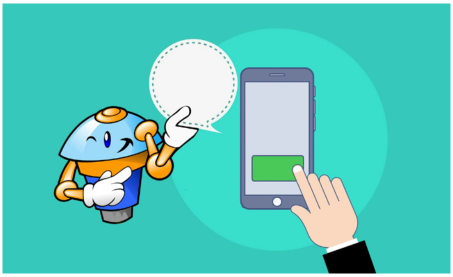
Since feedback is the next step in the process, we need to understand the sentiment lexicons, or collections of words, that the algorithms use for their feedback. Sentiment analysis uses online systems such as Amazon, Twitter, or Facebook to extract and analyze public views and opinions and build lexicon datasets. The lexicon datasets help define the positive and negative words from the word embedding.

The problem is that computers are not trustworthy, just, or virtuous.

In "How to Make a Racist AI Without Really Trying" Robyn Speer tested how these lexicons interpreted the words in the embeddings. She created an algorithm that evaluated the words using the sentiment lexicon, ascribing -1 for negative words and +1 for positive words, thereby coming up with a sentiment score. Some of the words that were evaluated were people's names with different adjectives. She found that the datasets have a statistically significant bias of positive for "white" names and negative for "black" names (Speer 14). It would also rate Mexican restaurants more unfavorably than Italian restaurants (Speer 10). How did the algorithm come to this conclusion just by the difference in the words Mexican and Italian?

This problem is rooted in the feedback that the algorithm receives from the lexicon, but the question that remains is how the lexicon defined these seemingly neutral words so differently. As stated earlier, lexicons rate words based on how they are used in context across many different platforms: Facebook posts, Twitter feeds, Instagram captions, etc. The algorithm uses these contextual situations as the input it needs to understand a positive or negative word. This illustrates a crucial point about these algorithms. They are not created with a built-in bias; rather, they are learning the bias that we project upon the words. In Speers' experiment, she adapted her NLP lexicon to adjust for this bias and created a less biased outcome, but it was not completely neutral.

This leads us to the next part of the process: how does the algorithm find its answer? Herein lies a key problem with these types of NLP and other machine-learning algorithms. "If you let a machine figure out the solution for itself, the route it takes to get there often won't make sense to a human observer" (Fry 11). The issue is a lack of transparency in how the algorithms arrive at their answers. Most computer developers claim intellectual property on their process of developing their algorithms and, therefore, no one can see the inner workings of the machine.



Chatbots are extentions of Natural Language Processing algorithms.

In 2016, Microsoft launched a new chatbot: Tay. A chatbot is an extension of NLP algorithms, and Tay was an experiment in "conversational understanding." Tay was designed to communicate with humans as if it were another human, a teenage girl to be specific. It was programmed to learn about language over time, thus being able to have, in theory, conversations about any topic. Tay was described as "the intersection of machine learning, natural language processing, and social networks." Within a few hours, however, Tay's tweets became more racist and misogynistic until Twitter received so many complaints that Microsoft took the account down after only 16 hours of being online (Schwartz).

What Microsoft did not see, and many argued they should have seen, is how easily the "repeat after me" function was manipulated. It didn't take long for Internet trolls to find and attack the Twitter account and inundate it with all kinds of vitriol. Many argued that, in the design of Tay, Microsoft should have been asking "how can this be used to hurt someone," and they did not follow through in their obligation to society (Schwartz).

Microsoft eventually fixed the problem with Tay and released an updated version, Zo, a few months later. Zo was "designed to shut down conversations about certain contentious topics, including politics and religion, to ensure she didn't offend people." And if Zo was attacked about these subjects, updates in the algorithm directed it to shut down completely and disconnect from the conversation (Schwartz).

Garbage In Garbage Out

In order to address the bias that the algorithms project, we must first look at where the algorithms learn this bias. We feed the algorithms data from our current culture. In many of my classes, I explain to my students the concept of GIGO—**G**arbage **I**n **G**arbage **O**ut. What this means is that if you are using invalid, inaccurate, or skewed data as input into your system, you will get invalid, inaccurate, or skewed information from the system. The algorithms are getting data from our imperfect world, whether hyper-divisive vitriol from social media or records from a systemically unjust system, so the A.I. data is not neutral.

We need to be particularly aware of algorithms that could have a severe impact on the lives of humans.

Even though Tay's way of communication became offensive to many people, it could be argued that the hurt was not long lasting and had minimal effects on people. Over time, however, this could cause profound harm to people. It is especially concerning as we multiply this by the number of applications that use the data. The examples above raise an ethical dilemma: What version of humanity do we want reflected in our technology? Or put another way, what human values do we want reflected back to us?

Both Speer and Microsoft, the "owners" of their respective algorithms, took action and worked to correct a problem. What about other algorithms with inner workings that are not transparent and/or that could produce a larger or more egregious offense? We need to be particularly aware of algorithms that could have a severe impact on the lives of humans, such as those used in our criminal justice system.

Making Biased Decisions

Algorithms have been used throughout our criminal justice system for years with the thought that reducing a person to a simple number makes dispensing justice easier. These measures are based on the belief that these algorithms come up with an accurate depiction of the person that is being sentenced and therefore can make decisions for judges. That is not always the case.

The U.S. criminal justice system uses algorithms to determine sentencing.

The cases of Brisha Borden and Vernon Prater, both of Coral Springs, Florida, illustrate major problems with using algorithms for decision-making in the criminal justice system. Borden and Prater were both arrested for petty theft. Borden had picked up a bike from the sidewalk and tried to ride it before abandoning it, and Prater had shoplifted $86.35 worth of tools from a Home Depot. Even though the offenses do not seem very different, the way they were treated by the justice system was. Borden's bond was set to a much higher amount than Prater's bond, which on the surface seems odd for seemingly similar offenses.

Brisha Borden was 18 years old at the time of the offense, and is a black woman. She only had four juvenile misdemeanors on her record. Prater was 41 years old at the time of the offense, and is a white man. He had two armed robberies and one attempted robbery on his record. The judges in each case used a computer algorithm that rated the likelihood of recidivism of offenders. Borden scored higher on the recidivism scale than Prater, and was therefore given a higher bond at her hearing. Two years later Borden had not been charged with any new crimes and Prater was sentenced to 8 years in prison for stealing thousands of dollars' worth of electronics.

The judges in these hearings used a software that is widely used across the country, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). In an analysis by ProPublica,[3] they "found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk" (Larson et al.). As with many machine-learning algorithms, there are several factors at play in the way it is biased, but it is hard to say what exactly causes the bias. Since COMPAS is the property of Northpointe, Inc., a privately held for-profit corporation, the actual algorithm and inner workings of the program are not accessible.

Northpointe "does not publicly disclose the calculations used to arrive at defendants' risk scores, so it is not possible for either defendants or the public to see what might be driving the disparity" (Angwin et al.). Since they claim that the software is proprietary, there is a significant transparency issue at stake. What we do know is that they used the records of thousands of criminals to define its algorithm. Several studies have concluded that a systemic bias exists in our criminal justice system. Therefore, the data that Northpointe used for its system could reflect the systemic racism that plagues our justice system. As shown earlier, using biased data as input for an algorithm (Angwin et al.) leads to biased results unless the code includes some additional compensation to ensure unbiased output.

Due to the "black box" situation with Northpointe, we cannot see the inner workings of the machine, in this case, the programming of the algorithms. By claiming that the algorithms are their intellectual property, Northpointe doesn't legally have to reveal anything about their software. Without seeing the actual programming code, there is no way to know if the algorithm is compensating for these systemic issues or not. Many judicial systems in the country rely heavily on the scores that this software produces. Without being able to see how it is arriving at its answer, the software could cause harm to many vulnerable people at an exponential rate.

Since transparency issues make it difficult to see how many machine-learning algorithms arrive at an answer, we cannot completely adjust for the bias when writing the algorithm, as Speers' experiment illustrates. In *Technology and the Virtues*, Shannon Vallor argues that "We need to cultivate in ourselves, collectively, a special kind of moral character, one that expresses what I will call the technomoral virtues" (Vallor 1). We must first change our culture, our own moral codes, and system biases. That is, we must first become more virtuous people.

Overhauling an entire culture to place more value on virtue is a huge project. While we work on that we still need to determine how to protect people from the harm that the bias in these systems may cause and identify our moral responsibility for any injustices caused by the decisions from these algorithms.

We must first change our culture, our own moral codes, and system biases.

> The 21st century decisions on how to live well—that is about *ethics*—are not simply moral choices. They are *technomoral* choices, for they depend on the evolving affordances of the technological systems that

> we rely upon to support and to mediate our lives in ways and to degrees never before witnessed. (Vallor 2)

In our transition to A.I., we must recognize the moral dimension, now more than ever.

Building a Better Future

There are two parts to building a more virtuous future. Individuals must become more virtuous people, taking responsibility for all actions, and communities must hold everyone to a virtuous standard. To become more virtuous, we can look to Aristotle for direction.

Aristotle was concerned with moral virtue understood as "excellence of . . . the soul" (1102a16-17) as the pathway to well-being. We develop character throughout our lifetime. It begins as we are young and as our parents teach us right from wrong, and their own moral code. It then develops as we apply these lessons throughout our life experiences, building habits of character, habits of moral virtue (1103a17-1103b25). Two of Aristotle's cardinal virtues are particularly pertinent to this discussion of human-A.I. decision-making—Prudence and Justice.

- **Prudence**—making judgments based on practical wisdom (1140b1-10).
- **Justice**—recognizing what is good for the community and taking up a course of action that reflects this (1129b17-1130a13).

Algorithms work in a world of black and white.

A prudent person exhibits practical wisdom and "is guided by appropriate feeling and intelligence, rather than mindless habit or rote compulsion to follow fixed moral scripts provided by religious, political or cultural institutions" (Vallor 25). Thus a prudent person does more than just follow the rules. A prudent person acts thoughtfully and deliberately, and makes decisions with a keen insight into the consequences of the decision (Aristotle 1140a25-31). Humans use prior knowledge and consider extenuating circumstances when making a decision. It is difficult to ascribe these attributes to an algorithm. Algorithms only work in a world of black and white, and decisions are based solely on predefined structures. William Hasselberger, computer scientist and pioneer in A.I., calls this the "Input Problem" (986) because the algorithm cannot distinguish a morally acceptable input from an immoral one.

Algorithms do not "see" the same context as humans do. Humans have an analog view of the world, seeing it in many shades and degrees. In the Brisha Borden case, a human would have factored her age and her previous misdemeanors, rather than basing the decision only on boxes that have been checked. "The person who enacts fixed moral rules 'correctly' but rigidly—without style, feeling, thought, or flexibility—is, on this view, a shallow parody of virtue" (Vallor 25). But this is the only way that computers can apply rules, fundamentally exhibiting a distinct inclination to be a shallow parody of virtue.

Aristotle's view of justice focuses on what is right for the community as a whole (1129b17-19). He believes in equitable distributions and the correction of inequity (1131a10-1131b24). From this viewpoint, we must correct the inequity of our systems in order to achieve justice for all. Herein lies the conundrum of our algorithm problem: if the injustice originates in the data *and* we cannot see how the algorithm "learns" from this data, how can we achieve equity?

In order to take care of the "Input Problem," social media platforms have to vigorously police online posts and news for vitriol and remove those users who are offensive. Social media platforms such as Facebook, Instagram, and Twitter have stepped up their responses but they have not consistently applied their own rules. It is virtually impossible to rid the data itself of prejudice and bias. The answer, then, is for all social media users to take up the mantle and decide to be virtuous people.

Some users will fail to be virtuous, so algorithm developers should take measures to mitigate the inequity inherent in the system. They need to evaluate the context of the use of the algorithm and decide what values they wish it to reflect. This, however, will require developers to be completely transparent with their developments, sharing not only how their algorithms work, but also the data they are using to teach the algorithm.

If the injustice originates in the data and we cannot see how the algorithm "learns" from this data, how can we achieve equity?

One such developer, IBM,[4] has made ethical considerations a priority by issuing and adhering to principles of A.I. development.

> For the public to trust AI, it must be transparent. Technology companies must be clear about who trains their AI systems, what data was used in that training and, most importantly, what went into their algorithm's recommendations. If we are to use AI to help make important decisions, it must be explainable. ("IBM's Principles")

IBM goes even further by calling for all developers to follow their example. "We encourage all technology companies to adopt similar principles to protect client data and insights, and to ensure the responsible and transparent use of artificial intelligence and other transformative innovations" ("IBM's Principles"). IBM is a major player in the computing industry, but it is only one of many players. This is a step in the right direction, but the past has shown that having corporations police themselves does not always work. Corporations aim to earn profits and maximize their shareholder wealth. Ethics and justice often take a back seat to these goals.

Unregulated A.I. is the greatest existential crisis humanity faces today.

If we can learn anything from the past, it is that we cannot leave it up to individual corporations and developers to regulate themselves. Corporations will do what is in their best interests, which at times are at odds with the greater good of society. Elon Musk[5] understands this situation very clearly. He knows how quickly A.I. can improve:

> We are rapidly headed towards digital super intelligence that far exceeds any human. . . . And the rate of improvement is exponential. This is a very serious danger to the public, and therefore, there needs to be a public body that has insight and then oversight to confirm that everyone is developing A.I. safely. (Ritm 1)

If one or a few companies manage to develop "God-like superintelligence," humanity will be subject to the whim of these companies. If they are virtuous and use their development justly, then humanity's situation will likely improve, but if not, they could take over the world.

Given the significant impact that A.I. can have on humanity as whole, it is imperative that we take action immediately. Of course, a wholesale move by the population toward a more virtuous culture is unlikely at best, and we certainly can't expect all corporations to behave virtuously. If we are to achieve justice, the first step is to have governing bodies regulate and review companies that develop and implement machine-learning algorithms. In the pharmaceutical industry, the FDA oversees and regulates the development and production of drugs, and must approve any drug before it can be distributed to the public. Similarly, we should have a governing body that oversees A.I. developers. This governing body should oversee and regulate A.I. development and approve any A.I. algorithm before it is used by any corporation or individual.

Unregulated A.I. is the greatest existential crisis humanity faces today. Computers permeate every facet of our societies, and because they are no longer confined to one physical machine, data and algorithms can persist forever. We need to consider all of this before ceding our thinking to them. Elon Musk notes, "I believe that the danger of A.I. is much greater than the danger of nuclear warheads by a lot. Nobody would suggest that we allow anyone to just build nuclear warheads if they want. The least scary future I can think of is where we have, at least, democratized A.I." (Ritm 1). If we democratize A.I., we at least have a voice in how they function and can preserve our moral integrity.

Conclusion

In today's world, technology seamlessly adheres to Moore's Law.[6] We are swiftly transitioning from human decision-making to A.I. decision-making, a "betwixt and between" liminal state as we contemplate the future use of A.I. in our decision-making processes. We use these machine-learning algorithms to enhance our lives today, but it is the digital superintelligence that is now being developed that we have to worry about.[7]

Liminality

We need to look at the data these algorithms use, how they address any bias in this data, and how they come up with answers. By becoming a more virtuous society and providing proper oversight, we can develop an A.I. that reflects back the values we want to see. If we do not take action now, before computers surpass our own intelligence, we could be handing over our future to an eternal, unjust A.I. without even realizing it. We have blindly followed a winding path: now we stand at a precipice in our society, transitioning from benign systems to immortal dictators, our toes hanging over the edge of a proverbial cliff.

Ann Mauss is an Associate Professor of Computer Studies and chair of the Computer Studies and Mathematics Department at the University of Dubuque. Ann has 23 years of experience teaching in post-secondary education and over 10 years of experience in the computer information technology field. She has an MBA from Northern Illinois University and holds various professional certifications in systems and database technologies. Ann specializes in computer programming, database and systems analysis and project management in the computing technology field. She also specializes in utilizing Community Based learning in the classroom. Ann is a former Iowa Campus Compact fellow.

Photo credit p. 39: Aravind kumar, Pixabay.com
Photo credit p. 41: athree23, Pixabay.com
Image credit p. 44: mohamed Hassan, Pixabay.com
Photo credit p. 46: Ichigo121212, Pixabay.com
Image credit p. 48: 1117826, Pixabay.com
Photo credit p. 51: Free-Photos, Pixabay.com

Notes

[1] Data brokers specialize in accumulating a wide variety of publicly available information, aggregating it, and then selling this information to others.

[2] ProPublica is an independent, non-profit newsroom that produces investigative journalism. The journalists for this article were 2017 Pulitzer Prize finalists in Explanatory Reporting.

[3] IBM was one of the first developers of machine-learning technologies with Deep Blue, a computer that was designed to play chess and was famous for beating Garry Kasporov.

[4] Elon Musk is CEO of Tesla, Inc., founder and CEO of SpaceX, the Boring Company, and Neuralink—all companies that create and develop A.I. technologies.

[5] Moore's Law states that the speed and capability of our computers increases every one and a half to two years.

[6] For further discussion on how computers can impact society, see Ritm 1 or Ng.

Works Cited

Angwin, Julia, et al. "Machine Bias." *ProPublica*, 23 May 2016, www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=M28Yt5CCtbKAxQp99Ick5WRwgYRkeCZw.

Aristotle. *Nicomachean Ethics*. Translated by Marin Ostwald, Bobbs-Merrill, 1962.

Brooke, Chris. "'I Was Only Following Satnav Orders' Is No Defence: Driver Who Ended up Teetering on Cliff Edge Convicted of Careless Driving." *DailyMail.Com*, 16 Sept. 2009, www.dailymail.co.uk/news/article-1213891/Driver-ended-teetering-cliff-edge-guilty-blindly-following-sat-nav-directions.html.

Fry, Hannah. *Hello World: Being Human in the Age of Algorithms*. 1st ed., W.W. Norton, 2018.

Hasselberger, William. "Ethics Beyond Computation: Why We Can't (and Shouldn't) Replace Human Moral Judgment with Algorithms." *Social Research: An International Quarterly*, vol. 86, no. 4, Johns Hopkins UP, 2019, pp. 977–99. Project MUSE, muse.jhu.edu/article/748873.

"IBM's Principles for Trust and Transparency." *THINKPolicy Blog*, 30 May 2018, www.ibm.com/blogs/policy/trust-principles/.

Larson, Jeff, et al. "How We Analyzed the COMPAS Recidivism Algorithm." *ProPublica*, 23 May 2016, www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=2Afda3_4a5kZwuyUcu2dePYQjk2YB3DX.

Ng, Joanna. "How Artificial Super-Intelligence Is Today's Tower of Babel." *ChristianityToday.Com*. www.christianitytoday.com/ct/2020/june-web-only/artificial-intelligence-todays-tower-of-babel-ai-ethics.html. Accessed 11 July 2020.

Ritm 1. "Elon Musk on Artificial Intelligence." *YouTube*, 14 Oct. 2019, www.youtube.com/watch?v=H15uuDMqDK0.

Schwartz, Oscar. "In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation." *IEEE Spectrum*, 25 Nov. 2019, spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation.

Speer, Robyn. "How to Make a Racist AI without Really Trying." *ConceptNet Blog*, 13 July 2017, blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/.

Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. 1st ed., Oxford UP, 2016.