

# It's not the algorithm, it's the ethics

Received (in revised form): 26th August, 2022

**Gary M. Shiffman\***  
Giant Oak, Consilient, USA

**Christopher Wall\*\***



Gary M. Shiffman



Christopher Wall

**Gary M. Shiffman** is an economist working to counter coercion and organised violence and to support others who do the same. After earning an undergraduate degree in Psychology, his career began in the US Navy with two tours in the Gulf War, and several positions in the national security community in Washington, DC. Dr Shiffman earned his PhD in Economics and joined the faculty of Georgetown University's School of Foreign Service in 2002. He published 'Economic Instruments of Security Policy' in 2006. He started incorporating machine learning into his research while a principal investigator on Department of Defense funded R&D projects related to insurgency, terrorism, and human trafficking. He created two technology companies founded upon behavioural science-based machine learning software. In 2020, he published 'The Economics of Violence: How Behavioral Science Can Transform Our View of Crime, Insurgency, and Terrorism' (Cambridge University Press). He has published essays in *The Hill*, *the Wall Street Journal*, *USA Today*, and other outlets.

**Christopher Wall** is pursuing his PhD at King's College, London, researching and writing about political violence and the use of ML/AI for national security. Previously he was involved with DARPA research on ML/AI and he frequently lectures at several military commands throughout the Department of Defense, to include SOCOM's Strategic Leadership International School. He also holds an appointment as an Adjunct Professor at Georgetown University, where he teaches a course titled 'The Science of National Security', designed to help future policymakers become more intelligent consumers of data. In 2018, he co-authored 'The Future

of Terrorism' with the late Georgetown historian, Walter Laqueur, and his writing has appeared in outlets such as *The Hill*.

## ABSTRACT

Machine learning and artificial intelligence (ML/AI) technologies have transformed nearly every industry, helping to realise unprecedented efficiency and effectiveness in a variety of tasks once thought the exclusive domain of humans. The financial compliance industry, however, lags its peers in adopting ML/AI tools in spite it being readily available and promising to reduce costs for financial institutions. This paper argues that the reason for delay in adoption is not ignorance of the technology but the lack of a moral consensus around its use in financial compliance. The ethics and morality behind the adoption of ML/AI tools and why compliance professionals are discouraged from adopting it in their compliance programmes are explored. The paper introduces the trolley car problem and how this explains the lack of a moral consensus of the use of ML/AI in compliance. It then explores why, even though machines today can pass the Turing test, machines are not capable of making moral judgments, meaning humans remain responsible for the actions taken by ML/AI. This creates an unprecedented burden about making moral decisions without any real benefit to compliance officials who want to do good. The argument is that if regulators change the incentive structure away from conformity to saving lives, and making this the moral regime guiding the use of ML/AI, technology adoption would increase and allow the compliance industry to change the world for the better.

**Keywords:** machine learning, artificial intelligence, ethics, moral philosophy, Turing test, financial compliance

\*Giant Oak,  
2300 Wilson Boulevard,  
Suite 700 PMB 142,  
Arlington, VA 22201,  
USA  
E-mail: gary.shiffman@  
giantoak.com

\*\*E-mail: christopher.s.wall@  
kcl.ac.uk

The financial compliance industry uses automation — the ‘use of machines to execute linear functions once thought only possible by humans’ — to improve productivity, yet compliance costs increase each year at a massive rate, with one study reporting an 18 per cent increase year on year.<sup>1</sup> Financial institutions (FIs) now automate procedural guardrails, moving humans away from laborious tasks to supervising technology that flags suspicious transactions, alerts when names on watchlists appear, and inspects credit, marketing and other departments to minimise inappropriate and illegal biases. These uses of automation are good, in the sense of providing a public benefit to the world, and, therefore, perhaps attract talented people to compliance as a career. But automation reflects the technological advances of the 20th century, where machines could be fed strict rules and process information in a world where rules accounted for all but the most sophisticated crimes. With time, automation became the bare minimum an FI needed to accomplish to remain compliant.

In the 21st century, automation no longer suffices. Criminals today operate on transnational scales, with resources that outmatch many law enforcement bodies, and that mutate faster than the pace at which regulators set standards. Fortunately, machine learning and artificial intelligence (ML/AI) technologies have made massive advances in the past decade in detecting threats more dynamically and robustly while reducing the labour-intensive nature of compliance. Indeed, the technology available for regulatory compliance has become its own industry and more readily available.<sup>2</sup> The emergence of this technology would suggest that compliance costs should fall, but instead, the opposite has occurred. If talented people have access to historic technologies to perform ethically good and important missions, then costs increase for one of two reasons: FIs

voluntarily spend more money on compliance to harness increased effectiveness of technology to catch more criminals, or FIs are not adopting innovations in ML/AI and are paying additional compliance costs in the form of labour on existing technologies to satisfy regulatory requirements.

For the purposes of this paper, the first explanation is dismissed, although arguments in support of it are welcomed, and the second is explored: FIs lag behind many industries in adopting available innovation in ML/AI to increase efficiency and effectiveness, continuing to rely instead on brute automation. Instead of providing lengthy data in support of this claim, it is stipulated here in order to get at a potentially much more interesting topic. If financial compliance, as an industry, lags behind other industries in technology adoption, how can this behaviour be explained and justified?

Talented and well-intentioned people enter into a career with an important mission — to combat injustice and crime — and yet these same professionals fail to make the obvious choices people in other industries have made to improve effectiveness and efficiency. For many years, it was believed that teaching algorithms and measurement would lead to technology adoption among leading compliance professionals. Believing that people fear what they do not understand, enabling understanding was embarked upon with the aim of correcting the mistaken notion that ML/AI could replace humans by explaining that it instead augments and empowers humans to make better informed decisions. Through the American Bankers Association (ABA) and many invitations to roundtable discussions with government officials, five years was spent in delivering content. The motivation for this paper is to share what was learned: understanding how ML/AI works was not precluding technology adoption in compliance. It is not the algorithm; it is the ethics.

## THE TROLLEY PROBLEM AND THE IMITATION GAME

You are driving a trolley and the brakes fail, sending your trolley speeding down the tracks, requiring you to make the choice: do nothing and the trolley will kill five distracted track workers, or pull a lever to change tracks to save those five workers but kill one worker on the other track. The famous ethical thought experiment known as the ‘Trolley Problem’ was developed in 1967 by philosopher Philippa Foot, and it has become a vehicle for understanding the moral universe unfolding with the emergence of ML/AI.<sup>3</sup>

The ‘trolley problem’ provides the ethical framework to solve the technology adoption puzzle in financial compliance. In the thought experiment, the individual driving the trolley must either kill one person or let five people die. What to do? When this question was posed to a room of banking compliance professionals this summer, about 70 per cent of the room indicated they would choose to pull the lever. This scenario is straightforward and suggests a strong consensus around the tradeoff — better to kill one instead of letting five people die. But change the scenario slightly, and the tradeoff becomes complicated. Would it be acceptable to push a big person on to the tracks to stop the trolley, killing one to save five? What if the five track workers are receiving hazardous duty pay because of extreme risks, and the lone person on the second track is a child? What if that child ignored posted ‘Danger’ signs? What if the trolley driver died of a heart attack when the brakes failed, but there was a passenger on the trolley, would the same 70 per cent pull the lever, or would it be possible to justify remaining in the passenger seat away from the lever, conferring the power of deciding who dies? This last scenario seems most instructive for the purposes of this paper — what circumstances incentivise the bystander to watch the tragedy and not step up to the lever?

To use this thought experiment to solve the compliance technology puzzle, we need one additional concept: thinking versus imitation. Unlike automation, machine learning learns inductively from human-generated data.<sup>4</sup> Therefore, machines mimic human activity with no evidence of actual consciousness of the ethical complexities. If 70 per cent of human trolley drivers pull the lever to change tracks, then machines mimicking the physical or cognitive processes of humans will also pull the lever. But making moral judgements remains an entirely human activity. We argue this distinction — thinking versus imitation — through Alan Turing’s ‘imitation game’.<sup>5</sup>

In 1950, when pioneering digital computing machines, Alan Turing expressed interest in the question: can computers think? Given the state of technology of the day, he then proposed a falsifiable theory: in 50 years, will machines have the ability to imitate humans? In his paper,<sup>6</sup> he describes the imitation game describing how one might create data to test related hypotheses. By the time of the writing of this paper in 2022, countless machines have been deployed that use ML/AI to mimic humans. Among the authors’ favourite demonstrations of machines fooling humans is Google’s CEO’s introduction of the Google assistant in May of 2018.<sup>7</sup> In his demonstration at Google’s IO conference, Sundar Pichai played clips of an AI called Duplex calling a hair salon and restaurant to make reservations. In both instances, the AI understood the human, and neither human showed any awareness that they were talking with a non-human agent — a machine.

A computer has been able to beat the best human chess player for more than 20 years.<sup>8</sup> Machines can now write poetry and make visual and musical art, imitating a particular artist’s style to the extent that most people are none the wiser.<sup>9</sup> Machines perform tasks in automobile design and manufacturing.<sup>10</sup> In hospitality, machines

automate call centres and perform customer service tasks.<sup>11</sup> But does this impressive and growing list of ML/AI accomplishments lead us to conclude that machines think? No, because morality and ethics are based on the actions necessary for humans to live fulfilling and virtuous lives within society.<sup>12</sup> No ML/AI platform is capable of making value judgements because these are inherently human activities.

This claim is grounded in moral philosophers going back to Aristotle, but present-day scenarios are easily supportive of the premise. Consider self-driving cars (SDCs). If you were told that a few moments ago, a coding error in a self-driving car just killed a family with three children in a major accident, how would you feel about them then? What if this was not a coding error but, rather, part of the design to avert killing another family with three children that accidentally crossed in front of the car?

Humans are uncomfortable with cars making these moral decisions, but this is in part because they represent subjectivity, such as when one outcome causes more damage than another — when one life exceeds the value of another. A lack of consensus about what moral regime should govern SDCs complicates human perspectives on ML/AI engaged in processes with significant consequences. For instance, researchers noted that individuals from Western Europe and North America prefer SDCs to make decisions that would harm the elderly if it meant saving the young, a preference not widely shared in East Asia.<sup>13</sup> So companies making SDCs construct fractured moral regimes,<sup>14</sup> segmented by local institutional norms.<sup>15</sup>

The compliance world faces a similar challenge when it comes to ML/AI because there is no overarching moral regime to guide its use. ML/AI for compliance at present prioritises flagging content for human adjudicators to make value judgements. In practice, this means humans use machines to sift through cases that have the highest

probability of being a true positive with indicators of illicitness and make decisions. But humans become fatigued and there are only a finite number of cases humans can realistically process in a day without suffering burnout. If the value proposition of ML/AI is to find more bad elements without increasing the number of human staff, what type of behaviours should compliance departments prioritise? For instance, they could focus on uprooting all human traffickers from their systems, thereby helping law enforcement to save an untold number of victims, but this would come at the cost of humans not identifying cases of money laundering or terrorism financing. Or it could direct an ML/AI platform to find sanctions busters wishing to use safe financial institutions to finance a deadly war where thousands are killed, and many more suffer. In either case, there would be trade-offs that banks have to make that would save some lives at the cost of helping others. There is no moral consensus on these choices, so compliance departments continue directing resources to existing tools and methods that satisfied their regulators and kept pace with their peers in the past; this passes for ethical decision making today.

For those who read the preceding paragraphs and maintain that this is incorrect and that computers possess moral consciousness, the premise is that even if computers can indeed think, such an error would not alter the arguments of this paper. If machines only imitate, then humans teach machines to mimic ethical decision making in those circumstances that have occurred in the past and for which there is training data. In other words, in spite of the distance in time and space, humans own the consequences of machine decisions of a trolley-like tragedy. If machines think, in the sense of a conscious awareness of moral right and wrong, because they learned from human furnished data, is the human off the hook? Creating a 'conscious' AI — if one ever comes into



Figure 1 *The imitation game*

existence — will be a result of human agency driving technology adoption. Regardless of thinking or imitating machines, a human does not avoid ethical accountability.

### BECOME THE JUDGE

If computers successfully mimic more human behaviours, even if this is not really thinking in a true human sense, then perhaps the appearance of consciousness is sufficient. To explore this, the imitation game, updated to the circumstance of humanity in 2022, is re-examined.

The imitation game, today often called the Turing test, features two people and a machine.<sup>16</sup> The initial scenario starts with a man in Room A and a woman in Room B, and they must answer questions from an interrogator in Room C. In the game, the interrogator must determine which room contains the man and which contains the woman. The interrogator knows absolutely

nothing about either individual, and the interrogator must determine their identity by asking questions. In this scenario, the man in Room A tries to trick the interrogator while the woman in Room B tries to help the interrogator. For example, if the interrogator asks about hair length, the man can lie, anticipating how the woman will answer.

The follow-on scenario is similar, except that the man in Room A is replaced by a machine. Again, the interrogator is tasked with correctly identifying the machine, and the machine's goal is to confuse the interrogator into thinking it is human (Figure 1).

The best strategy for the machine is to 'provide answers that would naturally be given by a [person]'.<sup>17</sup> For example, in response to maths and chess problems, the machine should not get every answer correct, and instead should mimic human performance. For example, Turing suggests the following



exchanges between the judge and Rooms A and B.

Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34,957 to 70,764

A: (Pause about 30 seconds and then give as answer) 105,621.

Q: Do you play chess?

A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R–R8 mate.<sup>18</sup>

Even Turing anticipated the machine would work faster and more reliably: notice the 30 second pause followed by a wrong answer to imitate a human. Humans working in 2022 face machines that successfully imitate human behaviours to such a high degree that they even appear to ‘think’.<sup>19</sup> In an age when machines not only perform human functions — automation — but also seem to think, the lesson of the imitation game is for humans not to compete, but to judge.

When a machine searches orders of magnitude faster than a human with higher accuracy, the human knows to stop competing at the act of searching and instead to judge what the machine returns. The same applies to compliance-related tasks. Humans no longer manually inspect customer names against lists of sanctioned entities. Instead, a machine compares the two lists of names and identifies matches. The human in 2022 investigates those results. Humans can now push for improved performance through innovation; increases in efficiency enable increases in effectiveness at lower costs. Searches for sanctioned entities, for example, can now include contextual information, improving risk identification at lower costs.

The financial industry recognises the importance of using ML/AI, and more companies now pride themselves on deploying ML/AI for RPA, screening and risk management.<sup>20</sup> These tools outperform incumbent automation tools and increase human productivity by detecting greater amounts of risk with less effort.<sup>21</sup> But trepidation persists because ML/AI remains unknown, and examples of ML/AI harming people are legion.

For instance, in 2018 Amazon discovered the AI recruiting tool it was developing showed bias against women.<sup>22</sup> In 2021, an investigation found that AI bias was causing 80 per cent of Black mortgage applicants to be denied.<sup>23</sup> A tool built to enable decreasing prison populations based on predicting recidivism probability was found to compute biased results.<sup>24</sup>

In all these examples, the algorithms worked in a technical sense, but failed in the more important human sense because the training data was poor. ML algorithms do not think — they simply learn from the data. An algorithm trained on data created by biased human systems will reflect those biases. The machine is literal and lacks a consciousness of morality. Without humans reviewing outcomes, recognising bias and validating the training data, these AI systems would have been given free rein to continue causing inappropriately discriminatory outcomes. Humans must judge the ethical consequences of ML/AI systems because the algorithm just reflects data. Great algorithms can, unfortunately, lead to morally wrong outcomes and do so at extremely efficient and effective levels of performance.

With the advanced state of ML/AI today and the rapid growth in its capabilities expected to continue, compliance professionals must view themselves not as competitors with technology, but as the judges of the systems’ performance. In the terms of Turing’s imitation game, the compliance professional’s position

description must move the person from Room B to Room C. But these talented and well-intentioned people in financial compliance are hesitating to make this move. Hesitancy about judging the ethics of technology output is causing the lag in technology adoption, which is impeding improving efficiency and effectiveness at fighting crime and injustice. This is not occurring from any lack of familiarity with how ML algorithms work.

### THE ETHICS OF TECHNOLOGY ADOPTION

Nobody wants to be the driver of a runaway trolley. Transformative ideas often come from simple insights, and, in retrospect, appear obvious. After years of exploring the depths and complexities of human acts of violence and coercion, and then using this knowledge to ‘encourage bank compliance professionals to see ML/AI as the enabler’ in fighting crime and injustice, this simple idea has come up.<sup>25</sup> Compliance officers’ hesitancy arises from an ethical challenge. Understanding the ‘why’ of the hesitancy clears a path to change the way all compliance officers react to innovations, so that they move toward increasing effectiveness while decreasing costs.

Returning to the trolley problem thought experiment presented in an earlier section of this paper, where the question was asked as to what circumstances would incentivise the bystander to watch the tragedy and not step up to the lever. But now the thought experiment is framed in such a way as to explain failures to innovate in financial compliance.

You are a passenger on a trolley whose driver has just shouted that the trolley’s brakes have failed, and who then died of the shock. On the track ahead are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and you can turn the trolley

onto it. Unfortunately, there is one person on the right-hand track. You are unaware of the identities of the six people — track workers, children, elderly, friends or family members of yours. You can turn the trolley, killing the one, or you can refrain from turning the trolley, letting the five people die.<sup>26</sup>

Humans have not agreed upon a universal ethical framework for the variations of who is the trolley driver and who are the victims. Therefore, when the brakes fail, nobody wants to be driving that trolley, because nobody wants to face the choice: kill or let die. Without consensus among one’s community or professional peers, the driver fears second guessing and disapprobation, or legal, criminal and civil consequences.

The very nature of a compliance professional’s job revolves around ethics and consensus. Examiners and auditors measure the performance of others based upon well-known regulatory standards — the standard measure of compliance. The questions asked during a review, audit, or inspection relate to achieving a standard, or measuring distance below the standard. Regulators do not yet give out extra credit.<sup>27</sup> For example, compliance professionals do not get measured on lives not murdered, victims rescued from human traffickers, elder abuse scammers arrested, or loans provided to deserving people today who were unjustly denied last week. Because the compliance officer’s metric of performance is negative measurement, the position might feel as if they must choose between the lesser of two evils, such as killing or letting die.

Here is an example of the type of challenge facing compliance officers, regulators and examiners. A bank runs an incumbent versus challenger test on a stratified random sample of 100,000 customers. The incumbent system alerts on 1,500, yielding 25 cases requiring action. The challenger ML/AI system alerts on 1,000 of that same 100,000

random sample, one third less work for the human reviewers, and 75 of those 1,000 alerts yield cases for action, 300 per cent more. The quality of the alerts is equal across both systems, and the challenger increases efficiency by 33 per cent and effectiveness by 300 per cent. But the Chief Compliance Officer does not adopt the innovation. Why? Because the challenger system identified a different set of specific threats that did not overlap entirely with those yielded by the incumbent system.

Here is the thought experiment. You are a compliance officer. If you take no action, your team will human review 1,500 of every 100,000 customers in order to yield 25 compliance risks to the FI. Or, if you 'pull the lever' and switch technologies, your team will review only 1,000 of every 100,000 customers to yield 75 risks. But the new system will miss 10 of the 15 identified by the incumbent one. Do nothing and let 50 threats continue undiscovered, or switch, and intentionally miss 10 threats to discover those 50. Without consensus that your regulators and peers approve of the ethical framework, do you change? Who wants to be put in this position?

Eight years after Turing created the imitation game in the journal *Mind*, in 1958, Philippa Foot asked, 'Why do moral arguments break down while other arguments do not?'<sup>28</sup> Is it possible to draw a strict line between 'statements of fact and statements of value'?

When people argue about what is right, good, or obligatory, or whether a certain character trait is or is not a virtue, they do not confine their remarks to the adducing of facts that can be established by simple observation, or by some clear-cut technique.<sup>29</sup>

Efforts to promote technology adoption that rely upon facts do not lead directly to statements of value. Technologists can bring ML/AI to the industry, but they do not bring consensus on newly emerging ethical

trade-offs. The argument is not mathematical, but relies upon a consensus of moral right and wrong.

If presented with all the facts about the earth being round, someone who did not accept this truth would be criticised. The idea of earth being round is not a moral question. But when presented with all the facts on the efficiency and effectiveness of an innovative technology able to discover drug and human traffickers, people may ethically choose not to adopt the new technology and that is accepted. Foot asked why we accept the 'breakdown' on ethical questions but not others.

People making moral judgements seek invulnerability to criticism. An automobile executive, for example, who knows that fewer deaths will occur with widespread SDC deployments, but who will face criticism and legal penalties for the many instances when SDCs do kill people, is like a financial compliance executive who knows that more crime and injustice will be identified with ML/AI systems, but fears criticism and legal penalties for the many instances of missed cases. The net benefit from innovation adoptions is significant, but in some sectors, such as financial compliance, the lack of consensus on the ethical framework causes those charged with making the decisions to fear criticism on moral grounds, and, therefore, they delay adoption.

Delayed adoption of innovation harms the public welfare. The illicit actors that compliance professionals seek to deter do not await consensus to go out and execute new scams. Terrorists and war financiers, criminals and crooks of all types, and all those seeking to corrupt governments, take advantage of opportunities afforded them, whether it is some technological edge or some loophole created by regulatory stickiness. Inaction has material consequences for the well-being of each member of society, perhaps especially the most vulnerable.



## THE PATH FORWARD

It is now clear to these authors that the financial compliance industry will adopt innovations only after a consensus emerges, regardless of the scale of improved accuracy. For ethical choices, conformity is easier than change. Compliance officers' decisions have the following attribute: the duty to avoid a negative (exam) exceeds the duty to provide aid (to victims of crime and injustice).<sup>30</sup> Compliance is a negative duty: do not fail an exam.

Where will this consensus come from? Through instead measuring the aid and assistance provided by the talented and well-intentioned people choosing compliance careers. In 1976, ethics philosopher Judith Jarvis Thomson presented an alternative to the trolley problem where she turned 'evils to goods'.<sup>31</sup> You are driving a trolley from Point A to Point B to deliver a life-saving drug to an unsuspecting patient. However, en route you learn that if you divert the trolley to Point C, the drug can save five lives. You cannot get to both Point B and C in time to save everyone. It seems permissible to divert, saving five lives versus one life. The moral clarity seems to come from positive facts more directly than negative. Approbation for this action will result when measured on the aid provided, choosing between two positives, rather than on the action not taken, which would be choosing between two negatives.

In compliance scenarios, the 'good' being accomplished is significant: loans successfully given to the underserved, victims rescued, terrorist financiers thwarted. If nobody wants to drive a runaway trolley, others will volunteer to deliver life-saving medicines. In a world where compliance officers are measured on goodness — aid provided, victims saved — they would seek out the innovation to improve effectiveness in combating crime and injustice. This world then seems like a better place.

## REFERENCES AND NOTES

- (1) Parasuraman, R., Sheridan, T. B. and Wickens, C. D. (May 2000) 'A Model for Types and Levels of Human Interaction with Automation', *IEEE Transactions On Systems, Man, And Cybernetics – Part A: System and Humans*, Vol. 30, no. 3, 286. See also: 'True Cost of Financial Crime Compliance Study: Global Report', *Lexis Nexis Risk Solutions* (June 2021), available at <https://risk.lexisnexis.com/insights-resources/research/true-cost-of-financial-crime-compliance-study-global-report> (accessed on 15th July, 2022).
- (2) Institute of International Finance (1st October, 2015) 'RegTech: Exploring Solutions for Regulatory Challenges', available at [www.iif.com/Publications/ID/4229/Regtech-Exploring-Solutions-for-Regulatory-Challenges](http://www.iif.com/Publications/ID/4229/Regtech-Exploring-Solutions-for-Regulatory-Challenges) (accessed on 15th July, 2022).
- (3) Foot, P. (1967) 'The Problem of Abortion and the Doctrine of the Double Effect', *Oxford Review*, Vol. 5, 5–15.
- (4) Murphy, K. P. (2012) 'Machine Learning: A Probabilistic Perspective' MIT Press, Cambridge, MA, p. 1.
- (5) Turing, A. M. (October 1950) 'Computing Machinery and Intelligence', *Mind: A Quarterly Review of Psychology and Philosophy*, Vol. LIX, no. 236, 433–60.
- (6) *Ibid.*
- (7) Bergen, M. (8th May, 2018) 'Google's Assistant Will Now Book You an Appointment on Its Own', *Bloomberg*, available at [www.bloomberg.com/news/articles/2018-05-08/google-s-assistant-will-now-book-you-an-appointment-on-its-own#xj4y7vzkg](http://www.bloomberg.com/news/articles/2018-05-08/google-s-assistant-will-now-book-you-an-appointment-on-its-own#xj4y7vzkg) (accessed on 15th July, 2022).
- (8) Greenemeier, L. (2nd June, 2017) '20 Years after Deep Blue: How AI Has Advanced Since Conquering Chess', *Scientific American*, available at [www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/](http://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/) (accessed on 15th July, 2022).
- (9) The *Economist* (11th June, 2022) 'Huge "Foundational Models" Are Turbo-charging AI Progress: They Can Have Abilities Their Creators Did Not Foresee', available at [www.economist.com/interactive/briefing/2022/06/11/huge-foundational-models-are-turbo-charging-ai-progress](http://www.economist.com/interactive/briefing/2022/06/11/huge-foundational-models-are-turbo-charging-ai-progress) (accessed on 15th July, 2022).
- (10) BMW (30th October, 2018) 'Computer-Assisted Art – The Fascination of AI Design', available at [www.bmw.com/en/design/ai-design-and-digital-art.html](http://www.bmw.com/en/design/ai-design-and-digital-art.html) (accessed on 17th July, 2022).
- (11) Fettes, J. (12th July, 2022) 'Attended AI Is the Future of Customer Service. Here's What That Means for Brands', *Fast Company*, available at [www.fastcompany.com/90766613/attended-ai-is-the-future-of-customer-service-heres-what-that-means-for-brands](http://www.fastcompany.com/90766613/attended-ai-is-the-future-of-customer-service-heres-what-that-means-for-brands) (accessed on 17th July, 2022).
- (12) This idea finds its genesis in early moral philosophy, starting with Aristotle's 'Ethics'. Aristotle, Ross, D. (Trans.) (2009) 'The Nicomachean Ethics', Oxford University Press, New York, pp. 198–203; see also

- Adkins, A. W. H. (1984) 'The Connection Between Aristotle's Ethics and Politics', *Political Theory*, Vol. 12, no. 1, 29–49.
- (13) Maxmen, A. (24th October, 2018) 'Self-driving Car Dilemmas Reveal that Moral Choices Are Not Universal: Survey Maps Global Variations in Ethics for Programming Autonomous Vehicles', *Nature*, available at [www.nature.com/articles/d41586-018-07135-0](http://www.nature.com/articles/d41586-018-07135-0) (accessed on 17th July, 2022).
- (14) Davis, W. (26th March, 2022) 'Very Significant: The Mercedes-Benz Decision that Could Fast Track Autonomous Cars', *Drive*, available at [www.drive.com.au/news/very-significant-the-mercedes-benz-decision-that-could-fast-track-autonomous-cars/](http://www.drive.com.au/news/very-significant-the-mercedes-benz-decision-that-could-fast-track-autonomous-cars/) (accessed on 17th July, 2022).
- (15) It is worth noting that how each company frames ethical decisions can only be inferred from public statements and technical white papers. The previous footnote mentioned how Mercedes Benz is assuming responsibility for fatalities associated with self-driving cars, something that Tesla is not willing to do at the time of writing. However, there is some empirical evidence based on public statements that suggest that companies in the self-driving car industry rely on the trolley problem to think about ethics and all have arrived at different conclusions, and this is probably affected by the legal regime in which they operate. See: Martinho, A., Herber, H., Krosen, M. and Chorus, C. (2021) 'Ethical Issues in Focus by the Autonomous Vehicle Industry', *Transport Reviews*, Vol. 41, no. 5, 556–77; Nash, J. (26th March, 2018) 'Around the World, Driverless Car Rules in Flux: Current Self-driving Car Rules in Germany, Japan, and Singapore Show a Mix of Motivations and Approaches to Regulation', *Robotics Business Review*, available at [www.roboticsbusinessreview.com/unmanned/around-the-world-driverless-car-rules-in-flux/](http://www.roboticsbusinessreview.com/unmanned/around-the-world-driverless-car-rules-in-flux/) (accessed on 17th July, 2022); Taihagh, A. and Si Min Lim, H. (2019) 'Governing Autonomous Vehicles: Emerging Responses for Safety, Liability, Privacy, Cybersecurity, and Industry Risk', *Transport Reviews*, Vol. 39, no. 1, 103–28; 360 Business Law (14th September, 2021) 'Mapping the Global Legal Landscape for Self-Driving Cars', available at [www.360businesslaw.com/blog/self-driving-cars-laws/](http://www.360businesslaw.com/blog/self-driving-cars-laws/) (accessed on 17th July, 2022).
- (16) Turing, ref 5 above.
- (17) *Ibid.*, p. 435.
- (18) *Ibid.*
- (19) Tiku, N. (11th July, 2022) 'The Google Engineer Who Thinks the Company's AI Has Come to Life', *Washington Post*, available at [www.washingtonpost.com/technology/2022/06/11/google-ai-lambda-blake-lemoine/](http://www.washingtonpost.com/technology/2022/06/11/google-ai-lambda-blake-lemoine/) (accessed on 18th July, 2022).
- (20) Swabey, P. (24th March, 2022) 'Intelligent Automation in Financial Services: Leading the Way', *Tech Monitor*, available at <https://techmonitor.ai/technology/ai-and-automation/intelligent-automation-financial-services-leading-the-way> (accessed on 18th July, 2022).
- (21) Barefoot, J. A. (24th May, 2022) 'The Case for Placing AI at the Heart of Digitally Robust Financial Regulation', Brookings Center on Regulation and Markets Policy Brief, available at [www.brookings.edu/research/the-case-for-placing-ai-at-the-heart-of-digitally-robust-financial-regulation/](http://www.brookings.edu/research/the-case-for-placing-ai-at-the-heart-of-digitally-robust-financial-regulation/) (accessed on 18th July, 2022).
- (22) Dastin, J. (10th October, 2018) 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women', Reuters, available at [www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G](http://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G) (accessed on 18th July, 2022).
- (23) Hale, K. (2nd September, 2021) 'A.I. Bias Caused 80% of Black Mortgage Applicants to Be Denied', *Forbes*, available at [www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=c9a597436feb](http://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=c9a597436feb) (accessed on 18th July, 2022).
- (24) Hao, K. (21st January, 2019) 'AI is Sending People to Jail – and Getting It Wrong', *MIT Technology Review*, available at [www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/](http://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/) (accessed on 18th July, 2022).
- (25) Shiffan, G. (2020) 'Economics of Violence: How Behavioral Science Can Transform Our View of Crime, Insurgency, and Terrorism', Cambridge University Press, New York; Laqueur, W. and Wall, C. (2018) 'The Future of Terrorism: ISIS, al-Qaeda, and the Far-Right', St. Martin's Press, New York; Shiffan, G. (May/June 2021) 'Artificial Intelligence and the Future of Bank Compliance: How Do You Know If It Is Working?' *ABA Bank Compliance Magazine*, available at [https://magazines.aba.com/bcmag/may\\_june\\_2021/MobilePagedArticle.action?articleId=1677907#articleId1677907](https://magazines.aba.com/bcmag/may_june_2021/MobilePagedArticle.action?articleId=1677907#articleId1677907) (accessed on 18th July, 2022).
- (26) Thomson, J. J. (April 1976) 'Killing, Letting Die, and the Trolley Problem', *Monist*, Vol. 59, No. 2 – Philosophical Problems of Death, 204–217, p. 207.
- (27) The authors are optimistic that the Anti-Money Laundering Act of 2020 may change this.
- (28) Foot, P. (October 1958) 'Moral Arguments', *Mind*, Vol. LXVII, No. 268, 502–13.
- (29) *Ibid.*, p. 513.
- (30) Foot, ref 4 above.
- (31) Thomson, ref 26 above, p. 209.