Vanderbilt Journal of Entertainment & Technology Law

Volume 19
Issue 1 *Issue 1 - Fall 2016*

Article 5

2016

Keeping AI Legal

Amitai Etzioni

Oren Etzioni

Keeping AI Legal

Amitai Etzioni and Oren Etzioni***

ABSTRACT

AI programs make numerous decisions on their own, lack transparency, and may change frequently. Hence, unassisted human agents, such as auditors, accountants, inspectors, and police, cannot ensure that AI-guided instruments will abide by the law. This Article suggests that human agents need the assistance of AI oversight programs that analyze and oversee operational AI programs. This Article asks whether operational AI programs should be programmed to enable human users to override them; without that, such a move would undermine the legal order. This Article also points out that AI operational programs provide high surveillance capacities and, therefore, are essential for protecting individual rights in the cyber age. This Article closes by discussing the argument that AI-guided instruments, like robots, lead to endangering much more than the legal order—that they may turn on their makers, or even destroy humanity.

---

* Amitai Etzioni is a University Professor at The George Washington University and has previously taught at Columbia University, Harvard Business School, and University of California at Berkeley. His major books include *The Limits of Privacy, Privacy in a Cyber Age, The New Normal,* and *The Active Society.* The authors are indebted to Rory Donnelly for major comments on a previous draft.

** Dr. Oren Etzioni is Chief Executive Officer of the Allen Institute for Artificial Intelligence. He has been a Professor at the University of Washington's Computer Science department since 1991, receiving several awards including, the Robert Engelmore Memorial Award (2007), the IJCAI Distinguished Paper Award (2005), AAAI Fellow (2003), and a National Young Investigator Award (1993). He was the founder or co-founder of several companies including Farecast (sold to Microsoft in 2008) and Decide (sold to eBay in 2013), and the author of over 100 technical papers that have garnered over 27,000 citations. The goal of Oren's research is to solve fundamental problems in AI, particularly the automatic learning of knowledge from text. Oren received his Ph.D. from Carnegie Mellon University in 1991, and his B.A. from Harvard in 1986.

TABLE OF CONTENTS

I. THE UNIQUE ATTRIBUTES OF ARTIFICIAL INTELLIGENCE 136
II. AI GUARDIANS 138
III. LOCK OR OVERRIDE? 141
IV. NO FISHING 142
V. AI DOOMSAYERS 144
VI. CONCLUSION 146

Policy makers and academics are raising more and more questions about the ways the legal and moral order can accommodate a large and growing number of machines, robots, and instruments equipped with artificial intelligence (AI)—hereinafter referred to as "smart instruments." Many of these questions spring from the fact that smart instruments, such as driverless cars, have a measure of autonomy; they make many decisions on their own, well beyond the guidelines their programmers provided.[1] Moreover, these smart instruments make decisions in very opaque ways, and they are instruments capable of learning with guidance systems that change as they carry out their missions.[2]

For example, a California policeman stopped a Google self-driving car because the car impeded traffic by traveling too slowly.[3] But who could the policeman have cited? The passenger? The owner? The programmer? The car's computer? Similarly, Google faced allegations that its search engine discriminated against women by showing ads for well-paying jobs to men more frequently than to

1. Kamala Kelkar, *How Will Driverless Cars Make Life-or-Death Decisions?*, PBS (May 28, 2016, 11:34 AM), http://www.pbs.org/newshour/rundown/how-will-driverless-cars-make-life-or-death-decisions/ [https://perma.cc/X93D-4KLT].

2. Jason Tanz, *Soon We Won't Program Computers*, WIRED (May 17, 2016, 6:50 AM), https://www.wired.com/2016/05/the-end-of-code/ [https://perma.cc/UQ7F-7VYX].

3. *See* Don Melvin, *Cop Pulls over Google Self-Driving Car*, CNN (Nov. 13, 2015, 11:03 AM), http://www.cnn.com/2015/11/13/us/google-self-driving-car-pulled-over/ [https://perma.cc/8K5B-XRK5].

women,[4] and that it favored its own shops in search results.[5] The inability of mere mortals to trace how such biases come about illustrates the challenges smart machines pose to the legal and moral order. The same questions apply to findings that advertisements on websites providing arrest records were "significantly more likely to show up on searches for distinctively black names or a historically black fraternity."[6] Was there intent? Who or what should be held liable for the resulting harm? How can the government deter repeat offenses by the same instruments? This Article provides a preliminary response to these and several related questions both in cases of limited harm (e.g., a program that causes a driverless car to crash into another)[7] and with regard to greater potential harm (e.g., the fear that smart instruments may rebel against their makers and harm mankind).[8]

This Article focuses on the relationship between AI and the legal order. The relationship between AI and the moral order requires a separate analysis.[9] Although both the legal and moral orders reflect the values of one and the same society, this Article treats them separately because they choose and enforce values in different ways. In the legal

---

4. Kristen V. Brown, *Google Showed Women Ads for Lower-Paying Jobs*, FUSION (July 8, 2015, 1:33 PM), http://fusion.net/story/162685/google-ad-algorithms-gender-discrimination/ [https://perma.cc/67JU-ELN]; Julia Carpenter, *Google's Algorithm Shows Prestigious Job Ads to Men, but Not to Women.*, WASH. POST (July 6, 2015), https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/ [https://perma.cc/Z6D7-3CJT]; Claire Cain Miller, *When Algorithms Discriminate*, N.Y. TIMES (July 9, 2015), http://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html [https://perma.cc/M9HN-5QF2].

5. Issie Lapowsky, *Study Offers New Evidence That Google Skews Search Result*, WIRED (June 29, 2015, 11:23 AM), http://www.wired.com/2015/06/google-wu-study/ [https://perma.cc/TH95-FA7T]. Brian Souter has voiced concerns regarding the fairness of Google's PageRank and search results after his web sites disappeared from Google's first-page results. Brian Souter, *Disappearing Tycoon Souter Blames Google*, BBC (September 12, 2011), http://www.bbc.com/news/technology-14884717 [https://perma.cc/VA6F-ZLMZ]. In the case of MyTriggers.com, the Ohio-based shopping comparison search site accused Google of favoring its own services over others in search results (although the judge eventually ruled that the site failed to show harm to other similar businesses). Dan Levine, *Google Wins Antitrust Victory in Ohio Case*, REUTERS (Sep. 1, 2011, 4:20 PM), http://www.reuters.com/article/us-google-antitrust-ruling-idUSTRE7805O420110901 [https://perma.cc/AWQ8-UK4P].

6. Miller, *supra* note 4.

7. Alexa Liautaud, *Driverless Car Push Faces Risk of Hacker Hijacking*, BLOOMBERG (Sep. 8, 2014, 2:06 AM), http://www.bloomberg.com/news/articles/2014-09-04/driverless-car-push-faces-risk-of-hacker-hijacking [https://perma.cc/J7WE-V8UB].

8. Nick Bostrom, *When Machines Outsmart Humans*, CNN (September 10, 2014, 9:12 AM), http://www.cnn.com/2014/09/09/opinion/bostrom-machine-superintelligence/index.html [https://perma.cc/5YUW-TC5Q].

9. See generally Amitai Etzioni & Oren Etzioni, *AI Assisted Ethics*, 18 ETHICS & INFO. TECH. 149 (2016), or Amitai Etzioni & Oren Etzioni, *Designing AI Systems that Obey Our Laws and Values*, COMM. ACM, http://cacm.acm.org/magazines/2016/9/206255-designing-ai-systems-that-obey-our-laws-and-values/fulltext [https://perma.cc/6YCP-6Q9Z] (last visited Sept. 29, 2016), for a consideration of the relationship between AI and the moral order.

realm, long-established institutions like the legislature and courts sort out which values to enforce, but there are no such authoritative institutions in the social and moral realms. There is no Supreme Court for ethics—nor is one called for. Instead, the moral realm chooses values to enforce through continuous moral dialogues that often lead to new shared moral understandings over time.[10]

It cannot be stressed enough that "legal order" means not just law enforcement, but also preventive law, such as routinely auditing businesses, positioning speed cameras, and employing customs officials. This Article will reveal that maintaining the law in the cyber age requires new instruments much more than new laws.

Part I begins by specifying the unique attributes of AI programs, which can make numerous decisions on their own, lack transparency, and change frequently. Part II suggests that unassisted human agents—from auditors and accountants to inspectors and police—cannot ensure that smart instruments will abide by the law. Human agents need the assistance of AI programs (this Article call them "AI Guardians") that analyze and oversee the operational AI programs that guide smart instruments. Part III asks whether operational AI programs should be programmed to enable human users to override them. Part IV points out that smart instruments can conduct highly effective oversight and that such AI Guardians are essential for the protecting individual rights in the cyber age. The Article closes with Part V by discussing the argument that the smart instruments' autonomy may endanger much more than the legal order in that smart instruments may turn on their makers, kill them, or even destroy humanity.

I. THE UNIQUE ATTRIBUTES OF ARTIFICIAL INTELLIGENCE

Reports about the legal challenges posed by smart instruments may at first seem overblown. After all, the law has successfully regulated a wide variety of instruments; regulations govern a great range of things from the level of noise a lawn mower may legally make to the emissions a factory can legally produce.

Some argue that it would be easy to require self-driving (alternately called autonomous or driverless) cars to heed the same laws as old-fashioned cars.[11] This, however, would be akin to requiring that Model T cars obey the laws set for horse-drawn carriages. Forcing

10. Amitai Etzioni, *Moral Dialogues in Public Debates*, PUB. PERSP., Mar.–Apr. 2000, at 27.

11. Danielle Muoio, *Driverless Cars Always Obey the Law*, TECH. INSIDER (Dec. 18, 2015, 11:59 AM), http://www.techinsider.io/driverless-cars-always-obey-the-lawand-its-a-problem-2015-12 [https://perma.cc/F898-S596].

autonomous cars to abide by prevailing laws would sacrifice many of their capabilities. For example, if granted a lane of their own, driverless cars could travel safely at much greater speeds than old cars. Indeed, history shows that the invention of new technologies—from guns to DNA typing, from steam engines to unmanned aerial vehicles (UAVs)—has required some new legislation. While bolstering the legal order may require a few new laws, it is more important to develop new instruments to keep AI legal.

Both new and old laws require the help of AI because of the unique attributes of smart instruments. These devices have *considerable autonomy* in the sense that they make numerous choices "on their own."[12] That is, these instruments use complex algorithms to respond to environmental inputs independently of real-time human input; they "can figure things out for themselves."[13] Smart machines may deviate or act against the guidelines the original programmers installed.[14] For instance, self-driving cars decide when to change speed, how much distance to keep from other cars, and may decide to travel faster than the law allows—when they learn that other cars often violate the speed limits.[15] Automatic emergency braking systems,[16] which stop cars in response to perceived dangers without human input, are becoming more common.[17] Consumers complain of false alarms, sudden stops that are dangerous to other cars,[18] and that these brakes force cars to proceed in a straight line even if the driver tries to steer them elsewhere.

AI-equipped autonomous operating systems are becoming *highly opaque*—black boxes to human beings. That is, people are unable to follow the steps these machines are taking to reach whatever conclusions they reach. Viktor Schönberger and Kenneth Cukier note:

---

12. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 16–17 (2013).

13. Mass. Inst. of Tech., *New Algorithm Lets Autonomous Robots Divvy up Assembly Tasks on the Fly*, SCI. DAILY, (May 27, 2015), http://www.sciencedaily.com/releases/2015/05/150527142100.htm [https://perma.cc/K6NK-YHM5].

14. Kelkar, *supra* note 1.

15. Joe Miller, *Google's Driverless Cars Designed to Exceed Speed Limit*, BBC (Aug. 19, 2014), http://www.bbc.com/news/technology-28851996 [https://perma.cc/Q3TB-CQNZ].

16. Chris Knapnan, *Auto-Braking: A Quantum Leap for Road Safety*, TELEGRAPH, (Aug. 14, 2012), http://www.telegraph.co.uk/motoring/road-safety/9429746/Auto-braking-a-quantum-leap-for-road-safety.html [https://perma.cc/XC6L-6GMJ].

17. Mark Phelan, *Automatic Braking Coming, but Not All Systems Are Equal*, DETROIT FREE PRESS, (Jan. 1, 2016), http://www.freep.com/story/money/cars/mark-phelan/2016/01/01/automatic-braking-safety-pedestrian-detection-nhtsa-iihs/78029322/ [https://perma.cc/G2NR-X55X].

18. Eric Limer, *Automatic Brakes Are Stopping for No Good Reason*, POPULAR MECHANICS, (June 19, 2015), www.popularmechanics.com/cars/a16103/automatic-brakes-are-triggering-for-no-good-reason/ [https://perma.cc/Q4TD-HJSC].

"Today's computer code can be opened and inspected . . . . With big-data analysis, however, this traceability will become much harder. The basis of an algorithm's predictions may often be far too intricate for most people to understand."[19] They add that "the algorithms and datasets behind them, will become black boxes that offer us no accountability, traceability, or confidence."[20]

Moreover, the AI programs that guide smart instruments are *learning systems* that constantly review changing conditions and the performance of the instruments they guide—and then modify the internal guidelines accordingly.[21] Smart instruments do not stop collecting data once they have been launched; instead, further data collection enables smart instruments to keep learning from experience and improve their performance.[22] These AI programs, therefore, may stray considerably from the guidelines their programmers initially gave these programs.[23] Indeed, smart instruments may counteract their makers' and users' instructions. Hence, self-driving cars cannot be tested and certified before hitting the road and let loose under the assumption that their guidance systems will not change in response to new information collected as these cars drive about.

II. AI GUARDIANS

Smart instruments' unique attributes pose a legal challenge when these instruments cause harm. Was there intent? Who or what is responsible for the harm? And whom should the law hold liable? The following mental exercise illustrates the issue. Imagine a bank is sued for denying a disproportionate amount of loan applications made by African Americans compared to those made by Caucasian Americans. In response, the bank's officials point out that for the past three years the bank has relied on an AI program to grant or deny loans. When selecting a program, the bank stipulated that the software must refrain from using race or any surrogate variable, such as zip code, to determine creditworthiness. Still, the plaintiffs show that the program discriminated against them by presenting to the court instances in which the bank denied loans to African American applicants with credit scores as good as or better than Caucasian applicants whose loans the bank approved.

---

19. MAYER-SCHÖNBERGER & CUKIER, *supra* note 12, at 178.
20. *Id.* at 179.
21. Tanz, *supra* note 2.
22. *Id.*
23. *Id.*

A finding of discrimination does not settle the matter. The questions of intent and responsibility for discrimination stand because the law generally punishes deliberate offenses much more harshly than unintended ones—see, for instance, the difference between first-degree murder and involuntary manslaughter.[24] It is hence necessary to answer the questions of intent and responsibility in order to determine who should be held liable for harm done. To return to our mental exercise: the hypothetical court established that harm had occurred, but it still needs to determine whether the bank deliberately caused the harm by instructing programmers to use race as a variable—despite its claims. Or did the program "learn" by looking at the data that race can serve as an efficient surrogate variable for other factors such as class, education, and geography?

The court could ask an expert in computer programming to serve as a witness, but she is likely to point out that no human being can "read" an AI program to determine whether the bias it showed reflects the programmers' actions or the program's autonomous actions. Above all, no person can trace the steps a program went through to reach its autonomous decisions, as the program maintains no records of these steps.[25]

This Article suggests that what the court—and all those who need to determine intent, responsibility, and liability for the acts of smart instruments—needs are *AI programs to examine AI programs.* The law needs smart instruments to deal with smart instruments. Until now, society has treated AI largely as one field that encompasses many programs, ranging from airplane autopilots to surgical robots. From here on, AI should be divided into two categories. The first category would consist of operational AI programs—the computerized "brains" that guide smart instruments. The second category would be composed of oversight AI programs that review the first category's decision making and keep the decisions in line with the law. These oversight programs, which this Article calls "AI Guardians," would include AI programs to interrogate, discover, supervise, audit, and guarantee the compliance of operational AI programs.

Self-driving cars illustrate the role of such AI Guardians. Because these cars are programmed to learn and adapt, they need a particular kind of AI Guardian program, an AI Monitor, to come along for the ride to ensure the autonomous car's learning and decision making does not lead it to violate the law. Unlike human passengers,

---

24. *See, e.g.*, Griggs v. Duke Power, 401 U.S. 424 (1971) (further highlighting the import of establishing whether the harm was deliberately inflicted).

25. The steps are carried out by the computers involved, on their own, which do not keep a list of the very large number of complex calculations they make. *See* Tanz, *supra* note 2.

these programs would not tire of constantly checking the speed limit and distance from other cars, and they could carry out their oversight duties even in the absence of a passenger.

The AI community has not yet differentiated between operational and oversight AI programs in large part because many AI scholars shared the original ideals associated with the formation of the Internet. Those holding original ideals hoped that the Internet would be a "flat" realm, a village in which all people could cooperate.[26] They did not envision a hierarchy in which some supervise and regulate others.[27] However, over the years, the Internet has turned from a village into a jungle riddled with hackers, con artists, thieves, bullies, and free riders. It increasingly needs order-enhancing institutions. Hence, the world would benefit from the development of a slew of AI Guardians, to prevent deviations from the instructions incorporated into AI programs by their human designers.

An Interrogator AI would establish whether operational AI programs observe privacy laws by determining whether such programs use personal medical information to target consumers, make employment decisions, extend or withdraw credit, and more. Such an AI Interrogator could determine not merely whether there was an illegal use of medical information, but also whether the abuse was a deliberate act on the part of the programmers (or those who retained them) or came about as a result of the operation of the AI system. That is, an AI Interrogator could find out if the misuse of information was the result of illegally obtaining medical data or of ferreting out medical information from other personal information, the latter of which is currently legal. For instance, if an AI program at a bank called in a cancer patient's loan, the program's AI Interrogator would assess whether the program acted on information illegally obtained from a hospital or doctor's office or ferreted out the person's condition on the basis of consumption decisions (e.g., a person purchased a wig, great amounts of soap, and vitamin supplements).

Other AI Guardians could carry out a wide range of oversight roles. Auditor AI programs could determine whether financial planning software directs its users to investments or insurance plans in which those who developed the software have a financial interest. AI Auditors could also establish whether search engine results are biased in favor of the corporation that provides the search results or its advertisers.

---

26. SHANE GREENSTEIN, HOW THE INTERNET BECAME COMMERCIAL: INNOVATION, PRIVATIZATION, AND THE BIRTH OF A NEW NETWORK 33–64 (2015).

27. John Perry Barlow, *A Declaration of the Independence of Cyberspace*, https://projects.eff.org/~barlow/Declaration-Final.html [https://perma.cc/FF4F-KSQH] (last visited Sept. 19, 2016).

Meanwhile, Inspector AI programs could review AI cyber security programs, such as those that restrict access to information, and could report and follow up the response to incidents of unauthorized access. As more instruments incorporate operational AI programs, the need for various AI Guardians to carry out oversight tasks will grow. That is, there will be more need for AI programs to keep AI programs legal.

AI Guardians have two major advantages over human "guardians." First, AI Guardians are much less likely to violate the intellectual property rights and privacy of those they review because they have no motives or interests of their own. Second, AI Guardians need only a tiny fraction of the resources and time it would take for a human being to carry out the same oversight missions—if humans could carry out such reviews in the first place.

III. LOCK OR OVERRIDE?

At first blush, it may seem obvious that there should be an override device to limit smart instruments' autonomous acts. Such a device would provide humans with a sort of veto power over the acts of the smart instruments. For instance, if passengers in an autonomous car witness people trapped in a burning car on the side of the road, they would be able to stop their car in order to get out and help; the self-driving car, without such an override, would otherwise just barrel along. People should be able to slow the car down to enjoy the scenery or exceed the speed limit to rush to a hospital. Some driverless cars already have such a mechanism,[28] and several states require that self-driving cars only operate in the presence of a passenger qualified to drive.[29] New York law even requires that someone keep one hand on the steering wheel at all times.[30]

By contrast, some have argued that no override should exist because people would abuse it by speeding while intoxicated or driving recklessly out of "road rage" and, in so doing, put themselves and others in danger. As one observer put it, "[W]e often regulate and take control from individuals precisely because we cannot trust them to refrain from acting in their own interest."[31] There is also a communitarian side to

28. John Markoff, *For Now, Self-Driving Cars Still Need Humans*, N.Y. TIMES (Jan. 17, 2016), http://www.nytimes.com/2016/01/18/technology/driverless-cars-limits-include-human-nature.html [https://perma.cc/EVW9-K25M].

29. *See* Bryant Walker Smith, *Automated Vehicles Are Probably Legal in the United States*, 1 TEX. A&M L. REV. 411, 500–08 (2014).

30. N.Y. VEH. & TRAF. LAW § 1226 (McKinney 2016).

31. Joshua Gans, *Who Should Control Your Car's Software*, DIGITOPOLY (Dec. 28, 2015), https://medium.com/@joshgans/who-should-control-your-cars-software-c5ecd8c1e129#.5qkj39p8k [https://perma.cc/F8BA-9YJQ].

this argument: if self-driving cars coordinate their movements, which would greatly enhance safety, individualized overrides would undermine this benefit of autonomous cars.[32]

In response, it must be noted that the law in free societies rarely prevents people from modifying instruments they own and operate—the exception being those situations in which modifications would cause great harm (e.g., driving without a seat belt). Society deters most "bad" use of tools and instruments by punishing, after the fact, those who abuse their power. The same principle should apply to autonomous instruments.

Moreover, given that AI Guardians' oversight programs can accommodate many permutations suggests that the two viewpoints can be reconciled. An AI program could be designed to steer to the side of the road and stop if a person overrides the original program and then engages in dangerous behavior, but otherwise allow passengers to override the program at will. That is, the program could assess each override and could overrule some. AI programs should also be able to coordinate group behavior even if some members of the group are robots and some are human. None of this may be true of today's AI programs, but it seems reasonable that they will be able to do so in the future.

IV. NO FISHING

Although it is rarely phrased in this way, civil societies do not seek full law enforcement. This odd preference stems in part from the likelihood that most, if not all, citizens commit a crime at some point—many commit quite a few. If the authorities fined or arrested everyone who smoked a joint, drove faster than the speed limit after a few drinks, or who did not pay gift tax on large expenditures they made for their children, few citizens, if any, would be spared. Civil societies, therefore, often tend to look the other way and rely on sporadic enforcement. This quest for less-than-full law enforcement is one reason civil libertarians reject "fishing expeditions," that is, cases in which a law enforcement agent abuses a targeted search to try to find evidence of *any* wrongdoing, not just that covered by the warrant. Such searches are viewed as a violation of one's civil rights.[33] Indeed, this is the reason warrants include "particularity"—details about what the authorities

32. *Id.*

33. *See* Katherine M. Shelfer & Hiaohua Hu, *Making Better Sense of the Demographic Data Value in the Data Mining Procedure, in* FOUNDATIONS AND NOVEL APPROACHES IN DATA MINING 331–61 (Tsau Young Lin, Setsuo Ohsuga, Churn-Jung Liau, and Xiaohua Hu eds., 2015); Brent Skorup, *Cops Scan Social Media to Help Assess Your 'Threat Rating'*, REUTERS (Dec. 12, 2014), http://blogs.reuters.com/great-debate/2014/12/12/police-data-mining-looks-through-social-media-assigns-you-a-threat-level/ [https://perma.cc/YCB2-YXGJ].

claim to be looking for, rather than just going fishing.[34] Other reasons for limiting the scope of search warrants include preventing privacy violations, opposition to the surveillance of innocent people, and preventing the authorities from harassing civilians.[35]

Making instruments smarter has a major side effect: it makes detecting even minor crimes and misdemeanors easy, threatening the ban on fishing and all that it protects. Both the private and the public sectors are developing programs that can track an individual's Internet activity,[36] turn cell phones into microphones and tracking devices and computers into cameras,[37] implant tiny radio transmitters into clothes,[38] and much else. The development of these programs is escalating due to the advent of cloud storage and the "Internet of Things" wherein objects from refrigerators to thermostats and fitness-tracking bands have sensors that can communicate personal information to third parties and government authorities.[39]

The compilation, analysis, and extrapolation ("cybernation")[40] by AI programs of large amounts of personal information, stored or collected by these various smart instruments, further increase the effects of these new technologies, making higher levels of law enforcement much easier. For example, typical CCTVs—private surveillance cameras owned and mounted in one's business, parking lot, or residential lobby—pick up few facts about one person at one locality at one point in time, and keep the information for a short period. The opposite holds true for Microsoft's Domain Awareness System, first tested in New York City in 2012.[41] The program collects information

---

34. *See, e.g.*, State v. Retherford, 639 N.E.2d 498 (Ohio Ct. App. 1994).

35. *See id.* at 505, 510.

36. *See, e.g.*, Peter Eckersley, *How Online Tracking Companies Know Most of What You Do Online (and What Social Networks Are Doing to Help Them)*, ELECTRONIC FRONTIER FOUND. (Sept. 21, 2009), https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks [https://perma.cc/HL4J-RT9M].

37. *See, e.g.*, Mike Masnick, *Smartphone Apps Quietly Using Phone Microphones and Cameras to Gather Data*, TECH DIRT (Apr. 18, 2011, 9:41 AM), https://www.techdirt.com/blog/wireless/articles/20110417/21485513927/smartphone-apps-quietly-using-phone-microphones-cameras-to-gather-data.shtml [https://perma.cc/85AP-ZSJX].

38. Jenny Strasburg & Matthew Yi, *Clothing Will Have Transmitters*, SF GATE (Mar. 12, 2003, 4:00 AM), http://www.sfgate.com/business/article/Clothing-will-have-transmitters-Benetton-to-2628532.php [https://perma.cc/T7DC-KW9N].

39. *See* ANDREW HILTS, CHRISTOPHER PARSONS & JEFFREY KNOCKEL, EVERY STEP YOU FAKE: A COMPARATIVE ANALYSIS OF FITNESS TRACKER PRIVACY AND SECURITY 2, 76 (2016), https://openeffect.ca/reports/Every_Step_You_Fake.pdf [https://perma.cc/MRK9-TWHB].

40. AMITAI ETZIONI, PRIVACY IN A CYBER AGE: POLICY AND PRACTICE (2015).

41. Press Release, N.Y.C., Mayor Bloomberg, Police Commissioner Kelly and Microsoft Unveil New, State-of-the-Art Law Enforcement Technology, http://www.nyc.gov/portal/site/nycgov/menuitem.c0935b9a57bb4ef3daf2f1c701c789a0/index.jsp?pageID=mayor_press_release&catID=1194&doc_name=http%3A%2F%2Fwww.nyc.gov%2Fhtml%

from all over the city from various CCTV cameras, speed cameras, license plate readers, and radiation detectors.[42] While the system does not yet utilize facial recognition, it could in the future be expanded to include such data, as well as cell phone location information.[43] The Domain Awareness System stores this information for five years or more, and authorities can use it at will to draw a full profile of a person's public life.[44] This is but one example of many in which "spot" information about a person is combined with other information about that person and then those data are subjected to AI analysis that enables authorities to draw conclusions about the person, well beyond what is revealed by direct observation.[45]

To prevent such comprehensive and continuous surveillance of people in public, legislatures should pass new legislation that would require the automatic erasure of information gathered by localized instruments such as toll booths and CCTVs after a short period of time, except in special situations like following a terrorist attack or an Amber alert. Legislation should also prohibit cybernation of all information except insensitive personal information,[46] such as information about one's medical condition, and ban the use of insensitive information to divine sensitive information. To enforce these regulations, governments should pass laws mandating the use of AI Guardians to audit and monitor operational AI surveillance programs. In short, the law could use AI-assisted oversight to curb AI-enhanced surveillance.

V. AI DOOMSAYERS

A small but oft-cited group of AI mavens at highly regarded institutions like MIT, Cambridge, and Berkeley warn that smart instruments threaten to become so smart that they will surpass human intelligence, and these instruments may well rebel against their makers and take over—if not destroy—the world. Rory Cellan-Jones

---

2Fom%2Fhtml%2F2012b%2Fpr291-12.html&cc=unused1978&rc=1194&ndi=1 [https://perma.cc/H592-7M6V].

42. Joe Coscarelli, *The NYPD's Domain Awareness System Is Watching You*, N.Y. MAG. (Aug. 9, 2012, 8:50 AM), http://nymag.com/daily/intelligencer/2012/08/nypd-domain-awareness-system-microsoft-is-watching-you.html [https://perma.cc/283U-52GY].

43. *See* Neal Ungerleider, *NYPD, Microsoft Launch All-Seeing "Domain Awareness System"*, FAST COMPANY (Aug. 8, 2012, 12:07 PM), https://www.fastcompany.com/3000272/nypd-microsoft-launch-all-seeing-domain-awareness-system-real-time-cctv-license-plate-monito [https://perma.cc/XK4U-WTEZ].

44. *See id.*

45. *See generally* United States v. Jones, 132 S. Ct. 945, 963 (2012) (Sotomayor, J., concurring) ("The availability and use of these and other new [monitoring] devices will continue to shape the average person's expectations about the privacy of his or her daily movements.").

46. *See generally* ETZIONI, *supra* note 40.

writes, "Humans, who are limited by slow biological evolution, couldn't compete and would be superseded."[47] Similarly, an op-ed written by scholars including Stephen Hawking states: "One can imagine [AI] outsmarting financial markets, out-inventing human researchers, out-manipulating human leaders, and developing weapons we cannot even understand. Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all."[48] *The Washington Post* reports that

> Nick Bostrom's favorite apocalyptic hypothetical involves a machine that has been programmed to make paper clips (although any mundane product will do). This machine keeps getting smarter and more powerful, but never develops human values. It achieves 'superintelligence.' It begins to convert all kinds of ordinary materials into paper clips. Eventually it decides to turn everything on Earth—including the human race (!!!)—into paper clips.[49]

AI doomsayers cite science fiction movies such as *The Terminator*, *The Matrix*, *2001: A Space Odyssey*, and *Transcendence*,[50] as indications that such an AI-driven Armageddon can be imagined.

One can readily see that if smart instruments malfunction, they could cause untold harm; a nuclear plant's malfunctioning AI program, for example, could wreak a great deal of havoc. However, it is far from evident that these instruments could develop a "will" of their own to dominate their makers, let alone humanity. Granted, one cannot rule out such a rebellion of smart instruments at some point in the remote future, but what follows from such a statement? Should the government outlaw the development of smart instruments and forfeit the many and rapidly growing benefits, such as making instruments that serve us more efficiently, cost less, and are available to more people? Moreover, could such bans be enforced on a global level?

Historically, new technologies upon which we now rely attracted doomsayers who turned out to be false prophets.[51] Nevertheless, all

---

47. Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC (Dec. 2, 2014), http://www.bbc.com/news/technology-30290540 [https://perma.cc/3GLM-TYCV].

48. Stephen Hawking, Max Tegmark, Frank Wilczek & Stuart Russell, *Transcending Complacency on Superintelligent Machines*, HUFFINGTON POST (June 19, 2014), http://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html [https://perma.cc/3WN5-AVNW].

49. Joel Achenbach, *The A.I. Anxiety*, WASH. POST (Dec. 27, 2015), http://www.washingtonpost.com/sf/national/2015/12/27/aianxiety/ [https://perma.cc/Q9JW-G4CU].

50. Hawking, Tegmark, Wilczek & Russell, *supra* note 48.

51. The most famous case is of the Luddites who smashed mechanical looms during the Industrial Revolution. *See* Adrian J. Randall, *The Philosophy of Luddism: The Case of the West of England Woolen Workers, ca. 1790-1809*, 27 TECH. & CULTURE 1, 1–17 (1986) (discussing the Luddites smashing mechanical looms during the Industrial Revolution).

operational AI programs should be subject to continual oversight to ensure that their conduct does not stray from the boundaries set by human agents. This, to reiterate, can be accomplished only by oversight provided by other AI programs, which we dubbed AI Guardians. These AI Guardians will need to become smarter just as operational AI programs are improving. Because growth in human intelligence is unlikely to keep pace with growth in artificial intelligence, humans may have little choice but to draw on AI to check AI—and to seek to increase oversight of artificial intelligence as the intelligence of the programs they oversee grows.

VI. CONCLUSION

Thoughtful people have asked for centuries, "Who will guard the guardians?"[52] We have no new answer to this question, which has never been answered well. For now, the best we can hope for is that all smart instruments will be outfitted with a readily locatable off-switch to grant ultimate control to human agents over both operational and oversight AI programs.[53]

52. The question "*Quis custodiet ipsos custodes*" was first posed by the Roman author Juvenal. *See* JUVENAL, SATIRE VI, at 65 (Lindsay Watson & Patricia Watson eds., 2014).

53. This last line may seem to contradict an earlier statement that oversight programs should be protected from human override. This previous statement refers to the use of smart instruments, but not to avoiding their use. Thus, as long as one drives a car, one will be subject to its monitoring program. But both its operational and oversight program can be avoided if one idles the car or if one stops using it. The same should hold for all instruments.