

Boston University School of Law

Scholarly Commons at Boston University School of Law

Faculty Scholarship

2017

On Questioning Automation

Woodrow Hartzog

Boston University School of Law

Follow this and additional works at: https://scholarship.law.bu.edu/faculty_scholarship



Part of the [Law Commons](#)

Recommended Citation

Woodrow Hartzog, *On Questioning Automation*, in 48 *Cumberland Law Review* 1 (2017).
Available at: https://scholarship.law.bu.edu/faculty_scholarship/3101

This Article is brought to you for free and open access by Scholarly Commons at Boston University School of Law. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarly Commons at Boston University School of Law. For more information, please contact lawlessa@bu.edu.



ON QUESTIONING AUTOMATION

WOODROW HARTZOG*

INTRODUCTION

Given the rapid pace of innovation and adoption, it can be hard to make sense of automated technologies.¹ New products that leverage algorithms and artificial intelligence seem to be both promising and frightening. Law and policymakers, as well as the general public, are grappling with when to be excited and when to be concerned. If you're confused, you're not alone.

People making decisions related to technology law, policy, and ethics have not faced such uncertainty since the advent of the Internet.² Virtual assistants can interact with us as though they were human. But what type of relationship is appropriate to form with them?³ Technologies that leverage data and algorithms can fine tune what vendors recommend to us and help save us time and money.⁴ At the same time, it is unclear what the costs are of using systems that can predict behavior and infer things about us. Then, there are automated cars. They can save millions of lives and make commuting a dream. But what will we lose when we stop driving ourselves and relegate important decisions about crashes to companies and the will of the al-

*Professor of Law and Computer Science at Northeastern University School of Law and College of Computer and Information Science. I wish to thank Ryan Calo and Evan Selinger for their comments and Lindsey Catlett, Matt Hambrick, and the staff of the Cumberland Law Review for all their hard work in putting this issue together.

¹ See Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 402 (2017) (Providing a good introduction on artificial intelligence law and policy) (“If the terminology, constituent techniques, and hopes and fears around artificial intelligence are not new, what exactly is? FalseFirst, as is widely remarked, a vast increase in computational power and access to training data has led to practical breakthroughs in machine learning, a singularly important branch of AI FalseSecond, policymakers are finally paying close attention.”).

² See generally Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. L. REV. 513 (2015).

³ See Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MD. L. REV. 785 (2015); Kate Darling, *Extending Legal Protection to Social Robots*, IEEE SPECTRUM (Sept 10, 2012), <https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots>; KATE DARLING, EXTENDING LEGAL PROTECTIONS TO SOCIAL ROBOTS: THE EFFECTS OF ANTHROPOMORPHISM, EMPATHY, AND VIOLENT BEHAVIOR TOWARDS ROBOTIC OBJECTS (M. Froomkin, R. Calo, I. Kerr, Edward Elgar, 2016).

⁴ See generally KENNETH CUKIER & VIKTOR MAYER-SCHONBERGER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* (2014).

gorithms that they promote?⁵

It is also hard to know how to weigh the costs of automation. Law enforcement officers are attempting to use robots, big data, and algorithms to snuff out crime before it even happens (with questionable efficacy).⁶ But can we embrace predictive crime paradigms without sacrificing privacy, equal justice, and due process?⁷ Digital currencies can leverage public blockchains to enable micropayments and anonymous online transactions. And yet, are they worth it if they also can generate black markets and drain precious energy resources?⁸ There are many other comparable examples and they all prompt the same questions. How can people keep their eye on what's important and how can lawmakers embrace all the digital age has to offer while creating rules that promote safety and stand a chance at being sustainable?

This issue of the Cumberland Law Review is dedicated to helping resolve some of the confusion around automation. The authors have contributed pieces to chart the path forward and provide a means of identifying problems and solutions to complex issues such as predictive analytics and big data, automated decision-making, and blockchain technology.

For my part, I would like to introduce these articles with a short and simple proposal to help make sense of automated technologies

⁵ See generally Bryant Walker Smith, *How Governments Can Promote Automated Driving*, 47 N.M. L. REV. 99 (2017); Bryant Walker Smith, *Slow Down that Runaway Ethical Trolley*, STAN. L. SCH. CTR. FOR INTERNET & SOC'Y: BLOG (Jan. 12, 2015, 3:42 PM), <https://cyberlaw.stanford.edu/blog/2015/01/slow-down-runaway-ethical-trolley>; HELEN NISSENBAUM, DEREGULATING COLLECTION: MUST PRIVACY GIVE WAY TO USE REGULATION? (May 1, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282; Patrick Lin, *The Ethics of Autonomous Cars*, THE ATLANTIC (Oct. 8, 2013), <https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>; Patrick Lin, *Why Ethics Matters for Autonomous Cars*, AUTONOMOUS CARS 69–85 (Markus Maurer et. al. eds., 2016), https://link.springer.com/chapter/10.1007/978-3-662-48847-8_4.

⁶ See generally ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING (2017).

⁷ See generally *id.*; Pro Publica piece; Julia Angwin, Jeff Larson, Surya Mattu, & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

⁸ Patrick Sawyer, *How Mining Bitcoin is 'Killing the Planet'*, THE TELEGRAPH (Dec. 16, 2017), <http://www.telegraph.co.uk/news/2017/12/16/virtual-bitcoin-production-killing-planet/>; Ezra Marcus, *Here's How People are Actually Using Bitcoin*, VICE (Aug. 31, 2016), https://www.vice.com/en_us/article/5gq3ga/bitcoin-testimonials-black-market-dispatches; Andy Greenberg, *The Silk Road's Dark-Web Dream is Dead*, WIRED (Jan. 14, 2016), <https://www.wired.com/2016/01/the-silk-roads-dark-web-dream-is-dead/>.

and the rules that should govern them. I believe that the best way for us to understand the role that automated technologies should play in our lives is also the best approach for understanding any confusing and complex concept: we must ask critical and probing questions.⁹ These questions should help us identify the values implicated by automated technology, the specific nature of human and machine conduct to be encouraged or discouraged, the acceptable and unacceptable costs of automation, and the degree to which people's agency is affected. Not all of these questions will be relevant in every context, but they can act as brief signposts and reference points to help assess the role of automation in specific situations.

What Values Are Implicated?

Of course, it is one thing to recognize the importance of asking questions and quite another to determine which questions should be prioritized. A good starting point is to ask why we want automated technologies in the first place and why we might not. In other words, we should assess the values implicated by any particular technology. Identifying relevant values will help us move beyond that vague, nagging feeling that automated technologies create problems and cause us to focus on the specific conduct that brings about troubling behavior.¹⁰

So, what kinds of values should we focus on? In their work on "value sensitive design," Batya Friedman, David Hendry, and Peter Kahn identify several human values with ethical importance that are relevant to technologies, including property, privacy, freedom from bias, informed consent, universal usability, trust, autonomy, identity, calmness, and environmental sustainability.¹¹ I have argued in work

⁹ See generally GARY T. MARX, *WINDOWS INTO THE SOUL: SURVEILLANCE AND SOCIETY IN AN AGE OF HIGH TECHNOLOGY* (1st ed. 2016) (foregoing prescriptions about surveillance in favor of a series of questions which can help justify surveillance practices).

¹⁰ A related phenomenon is the tendency to label disturbing new technologies as "creepy." See Evan Selinger, *Why Do We Love to Call New Technologies "Creepy"?*, SLATE (Aug. 22, 2012), http://www.slate.com/articles/technology/future_tense/2012/08/facial_recognition_software_targeted_advertising_we_love_to_call_new_technologies_creepy_.html.

¹¹ See VALUE SENSITIVE DESIGN, <http://www.vsdesign.org/> (last visited Jan. 23, 2018). Pioneered by Professors Batya Friedman, Peter Kahn, David Hendry, and others, value sensitive design focuses on the method of creating information and computer systems to account for human values in a principled and comprehensive manner throughout the design process. Philosopher Helen Nissenbaum developed the values in design theory and method as "a way of considering human life that explores how the values we think of as societal may be expressed in technological designs, and how these designs in turn shape our social values." *What is Values in Design?*, VALUES IN DESIGN, <http://www.nyu.edu/projects/nissenbaum/vid/about.html>. See also IVAR HOLM, IDEAS AND

with Neil Richards, Evan Selinger, and Fred Stutzman that the value of privacy could be better conceptualized by focusing on the values of trust, obscurity, and autonomy.¹² European data protection regimes in many ways value dignity, and consumer protection regimes in the US value the ability to participate in markets and information symmetries.¹³

Perhaps the best place to start (or finish) a conversation about automated technologies is with agency. People need to be able to meaningfully exert power and control over automated systems.¹⁴ Loss of agency and autonomy is probably the most deep-seated fear of users and the most pressing issue for policymakers. For example, the decision as to whether automated cars have a steering wheels or pedals is important and intense.¹⁵ The steering wheel itself is a visceral and salient embodiment of user control over autonomous systems. Fully autonomous cars are debatably safer because there is great potential for error in the “handoff” between user and AI.¹⁶ But cars without steer-

BELIEFS IN ARCHITECTURE AND INDUSTRIAL DESIGN: HOW ATTITUDES, ORIENTATIONS, AND UNDERLYING ASSUMPTIONS SHAPE THE BUILT ENVIRONMENT (2006).

¹² WOODROW HARTZOG, *PRIVACY’S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGY* (2018); Woodrow Hartzog & Evan Selinger, *Surveillance as Loss of Obscurity*, 72 WASH. & LEE L. REV. 1343, 1355 (2015); Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CAL. L. REV. 1, 4 (2013); Neil Richards & Woodrow Hartzog, *Privacy’s Trust Gap: A Review*, 126 YALE L. J. 1180 (2017); Neil Richards and Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016).

¹³ See James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L. J. 1151 (2004); see also <http://jolt.law.harvard.edu/articles/pdf/v18/18HarvJLTech229.pdf>.

¹⁴ See generally BRETT FRISCHMANN & EVAN SELINGER, *RE-ENGINEERING HUMANITY* (2018).

¹⁵ See Timothy B. Lee, *The Most Important Question About a Self-Driving Car is Whether it has a Steering Wheel*, VOX (April 27, 2016), <https://www.vox.com/2016/4/27/11518396/autonomous-vehicle-steering-wheel>; Larry Greenemeier, *Driverless Cars Must Have Steering Wheels, Brake Pedals, Feds Say*, SCIENTIFIC AMERICAN (Mar. 16, 2016), <https://www.scientificamerican.com/article/driverless-cars-must-have-steering-wheels-brake-pedals-feds-say/>; Ben Guarino, *The Case for the Steering Wheel (and Against Truly Driverless Cars)*, INVERSE INNOVATION (Oct. 27, 2015), <https://www.inverse.com/article/7465-the-case-for-the-steering-wheel-and-against-truly-driverless-cars>; Cadie Thompson, *Ford is building cars without steering wheels, gas or brake pedals*, BUSINESS INSIDER (Aug. 22, 2016), <http://www.businessinsider.com/fords-self-driving-cars-wont-have-steering-wheels-2016-8>; Joann Muller, *Look, Ma, No Steering Wheel Or Pedals In GM’s Robo-Taxi, Coming in 2019*, FORBES (Jan. 12, 2018), <https://www.forbes.com/sites/joannmuller/2018/01/12/look-ma-no-steering-wheel-or-pedals-in-gms-robo-taxi-coming-in-2019/#4475ea9a3d82>.

¹⁶ John Markoff, *Robot Cars Can’t Count on Us in an Emergency*, THE NEW YORK TIMES (June 7, 2017), <https://www.nytimes.com/2017/06/07/technology/google-self->

ing wheels and pedals mean people must be more submissive to the will of the machine. Once loss of agency is normalized, it's hard to reverse. Thus, we should identify when agency is lost and how it might be preserved elsewhere within autonomous systems.¹⁷ Human agency is a value that should not be lightly relinquished.

The articles in this issue ask about values as well. Mark MacCarthy asks hard questions about the value of “fairness.”¹⁸ Should law and policymakers emphasize *group* fairness in order to help ensure statistical parity and equal group error rates that aim to reduce the subordination of disadvantaged groups? Or should the rules value *individual* fairness to avoid the arbitrary misclassification of individuals, with little concern about how groups are affected?

Hideyuki Matsumi examines the value of “accuracy” and how predictions of traits and future behavior threaten it.¹⁹ Is there a meaningful distinction between forecasting (predictions about future actions) and verifiable guesses about personal traits? Should these kinds of “predictive” data be treated as legally exceptional? Richard Warner and Robert Sloan advocate for focusing on the value of human control in profile-driven decision-making that harmonizes with contextually-sensitive information norms.²⁰

Katie Szilagyi critically examines the value of property with respect to the cryptocurrency Bitcoin. This technology attempts to mimic rivalry—one of the defining characteristics of tangible property, in a web of digital 1s and 0s as a public ledger called a blockchain. Szilagyi conceives of the property value of bitcoin as patrimonial, or “an envelope of all of the property rights associated with a particular person.”²¹ These sorts of inquiries will help lawmakers focus on key challenges that automated technologies pose and how to meet them head on.

What Kind of Conduct is Culpable?

The only way to produce effective rules for the creation, sale,

driving-cars-handoff-problem.html.

¹⁷ See generally Woodrow Hartzog, et. al., *Inefficiently Automated Law Enforcement*, 2015 MICH. ST. L. REV. 1763 (2015).

¹⁸ Mark MacCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 CUMB. L. REV. 67 (2018).

¹⁹ Hideyuki Matsumi, *Predictions and Privacy: Should There be Rules About Using Personal Data to Forecast the Future?*, 48 CUMB. L. REV. 149 (2018).

²⁰ Richard Warner & Robert H. Sloan, *The Ethics of the Algorithm: Autonomous Systems and the Wrapper of Human Control*, 48 CUMB. L. REV. 37 (2018).

²¹ Katie Szilagyi, *A Bundle of Blockchains? Digitally Disrupting Copyright Law*, 48 CUMB. L. REV. 9, 30 (2018).

and use of automated technologies is to have a clear sense of what conduct should be encouraged and discouraged. When I teach torts, I ask my students to articulate the specific culpable conduct in every case. Asking this question not only helps us identify who should be held responsible in the creation, sale, and use of automated technologies, but it also forces lawmakers, industries, and the public to consider which actions create adverse outcomes and whether those outcomes were predictable.

It is useful to distinguish between different types of conduct to ensure that a rule and its enforcement can be targeted and proportional to its effects and the degree of culpability. For example, there is a clear and compelling case for rules that prevent companies from programming robots to deceive us for financial gain.²² Scams only benefit the scammers. But what about negligent programming that creates entirely foreseeable adverse consequences for people? Do we want to make sure that companies follow the right steps in creating automated technologies? Or is the only thing that matters the risk presented or actual harm caused by the finished product?²³

Should those who build technologies bear full responsibility for their products? What about entities that collect and supply the data that powers the automated technologies?²⁴ What about others in the system, such as procurers and those who implement systems for others? Should those that order malicious AI also be held responsible? Finally, what should we expect of people who use AI? Those who use automated technologies to influence the fates of others should be bound to act reasonably and refrain from harming them. But do people have a duty to protect *themselves* when using robots, AI and other automated technologies? If so, when and on what basis?

Articles in this issue also articulate conduct to be encouraged or discouraged. For example, Matsumi proposes a legal regime that would constrain data controllers as they create and act upon predictive information—essentially speculation about people's future behavior and personal characteristics. MacCarthy highlights rules that would require companies to test algorithms at the development stage and in use for potential bias—a process known as a disparate impact assess-

²² See generally Hartzog, *supra* note 3.

²³ See generally KRISTEN THOMASEN, DRIVING LESSONS: LEARNING FROM THE LEGAL HISTORY OF AUTOMOBILE SAFETY TO INFORM DRONE PRIVACY REGULATION (2015) (paper presented at the We Robot Conference, April 10, 2015).

²⁴ See generally, WOODROW HARTZOG, WILLIAM SMART, & CINDY GRIMM, AN EDUCATION THEORY OF FAULT FOR AUTONOMOUS SYSTEMS (2017) (paper presented at the We Robot: Conference on Legal and Policy Issues Related to Robotics, Yale Law School, New Haven, CT April 2017).

ment.²⁵ Warner and Sloan highlight actions by those that design algorithmic systems that violate informational norms in opaque ways and encourage legal rules that guide the informational norm-creating process to desirable outcomes.

What Are the Costs We Can Live With?

We cannot create automated technologies that have no externalities. The use of algorithms, artificial intelligence, and robots will always have some costs. So, while it is admirable to pursue things like “perfect” and “guaranteed” privacy, too much focus distracts from reality and the difficult tradeoffs that must be made in building and implementing automated technologies.²⁶ We’ve got to pay something if we want to benefit from automated technologies and we must be particularly careful that we do not pay too much. Policymakers should take care to accurately gauge society’s cost thresholds in advance so that they can recognize them, accept what society (including the most vulnerable among us) is willing to pay, and know when to shut it down.

For example, automated cars are going to crash. The salient question is whether they will crash less often (or less harmfully) than human-driven cars.²⁷ Virtual assistants, bots, and androids are going to try to manipulate us with their charm. But it will be important for us to know just how much manipulation we should tolerate and be able to identify when the bots have crossed the line. Automated decision-making systems will affect discrete populations in different ways. But we should not accept a significantly disparate impact on vulnerable, minority, and underprivileged populations.

By having a frank conversation about tradeoffs, we can get a more realistic picture of whether building or using technologies in a given context is a good idea. Society will likely need to come to terms with acceptable amounts of risk and inevitable damage as well as identify the things that are worth preserving at all costs. Moreover, the argument that regulation dooms innovation should be subject to greater scrutiny.²⁸ If we want to get automated technologies right,

²⁵ See MacCarthy, *supra* note 18.

²⁶ Woodrow Hartzog & Evan Selinger, *It’s time to give up on the ideal of perfect privacy online*, AEON (Nov. 24, 2015), <https://aeon.co/ideas/if-online-privacy-is-dead-will-obscurity-do-instead>.

²⁷ See generally Bryant Walker Smith, *How Governments Can Promote Automated Driving*, 47 N.M. L. REV. 99 (2017); Bryant Walker Smith, *Slow Down that Runaway Ethical Trolley*, STAN. L. SCH. CTR. FOR INTERNET & SOC’Y: BLOG (Jan. 12, 2015, 3:42 PM), <https://cyberlaw.stanford.edu/blog/2015/01/slow-down-runaway-ethical-trolley>.

²⁸ NISSENBAUM, *supra* note 5; see generally Chris Hoofnagle, Federal Trade Commis-

they must be subject to limits. If the costs of safe and sustainable AI are too high for companies to stomach, then we are better off without it.

In this issue MacCarthy notes that even though pursuing group fairness might require a small sacrifice of equal accuracy if some vulnerable groups are to be protected.²⁹ Matsumi notes that treating predictive information differently than other kinds of personal information might impose costs on the ability of companies to process data and engage in business models that require targeting and profile-based decision making. Warner and Sloan propose an approach to assessing the right balance of costs and benefits centered around informational norms. And Szilagyi argues that the costs (and benefits) of cryptocurrency are better conceptualized in terms of an “indivisible system of property rights tied to a purpose” rather than a set of individual rights.³⁰

Automated technologies are the next great movement for technology law and policy. They are complex and unpredictable. We must have a clear vision of what we want them to do and what we want to preserve. At the time this issue was published, the rules for their governance were quite unclear. In fact, law and policy makers and the general public are still not quite sure what to make of the coming revolution of algorithms and artificial intelligence.

But fear not. The blueprint to make sense of these problems has existed as long as critical thought. So the next time you find yourself flummoxed by some new automation technology, try stepping back and identifying some fundamental touchstones to help find your way out. Values. Conduct. Cost. While your inquiries will not have one single correct answer, with a lot of rigor and a little bit of luck, the path for a safe and sustainable future will become clearer.

sion Privacy Law and Policy (2016).

²⁹ MacCarthy, *supra* note 18 at 68.

³⁰ Szilagyi, *supra* note 21 at 34.