When Robots Make Legal Mistakes

Susan C. Morse

WHEN ROBOTS MAKE LEGAL MISTAKES

SUSAN C. MORSE[*]

Abstract

The questions presented by robots' legal mistakes are examples of the legal process inquiry that asks when the law will accept decisions as final, even if they are mistaken. Legal decision-making robots include market robots and government robots. In either category, they can make mistakes of undercompliance or overcompliance. A market robot's overcompliance mistake or a government robot's undercompliance mistake is unlikely to be challenged. On the other hand, government enforcement can challenge a market robot's undercompliance mistake, and an aggrieved regulated party can object to a government robot's overcompliance mistake. Robots will have an incentive to make decisions that will avoid the prospect of challenge, especially if they cannot defend their legal decisions due to a lack of explainability. This incentive could encourage counterintuitive results. For instance, it could encourage market robots to overcomply and government robots to undercomply with the law.

Table of Contents

213

Introduction

This Essay is concerned with what happens when robots make legal mistakes. When an automated intelligent system makes a legal mistake, will it stand as a valid and final interpretation of the law?

A robot, like any legal decision maker, has to defend decisions only when those decisions are challenged. And not all decisions are equally open to challenge. As it stands now, some robot legal decisions might be challenged as unlawful. Others will not be. I argue here that, all else equal, robot legal decision makers will prefer decisions that are less vulnerable to challenge. This may cause the development of the law to drift in new and different directions.

As Henry Hart and Albert Sacks observed and taught,[1] legal process helps determine when to respect an institution's production of law. Some of these procedures for challenging decisions could change in response to artificial intelligence, since the rules we have do not necessarily line up with the robot capabilities we predict. For example, consider prosecutors and defense lawyers. As one paper in this issue observes, artificial intelligence systems, or robots, might eventually act as prosecutors and defense lawyers in criminal cases.[2] The job of a prosecutor might require more moral judgment when deciding whom to prosecute, for instance, compared to the single-minded zealous advocacy of a defense lawyer. Other contributions to this issue might take this example and suggest that robots' discretionary or moral prosecutorial decisions should be more open to challenge than the similar decisions of human prosecutors.[3] Perhaps, in the future, law will establish an avenue to challenge prosecutorial discretion decisions made by robots. But this would be a change from existing law.

---

1. HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 3-9 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994) (discussing the idea of institutional settlement and the interaction between private and government decision-making).

2. *See generally* Stephen E. Henderson, *Should Robots Prosecute and Defend?*, 72 OKLA. L. REV. 1 (2019).

3. *See generally* Joshua P. Davis, *Artificial Wisdom? A Potential Limit on AI in Law (and Elsewhere)*, 72 OKLA. L. REV. 51 (2019); Chris Chambers Goodman, *AI/Esq: Impacts of Artificial Intelligence in Lawyer-Client Relationships*, 72 OKLA. L. REV. 147 (2019); W. Bradley Wendel, *The Promise and Limitations of Artificial Intelligence in the Practice of Law*, 72 OKLA. L. REV. 21 (2019).

My goal here is not to propose changes that would make it easier to challenge robot legal decisions. Instead, I use existing law and existing process to illustrate the interaction between available legal process and the content of robots' legal decisions, and to argue that, all else equal, robots will prefer to make legal decisions that are less likely to be challenged. The perspective here is ex post and focused on review of robot decisions.[4]

The analysis can be illustrated by considering both a government robot and a market robot.[5] An example of a government robot that makes legal decisions is an automated system, developed or purchased by the government, that determines an applicant's eligibility for welfare or disability benefits. An example of a private-market robot is an automated system, such as TurboTax, that generates tax returns.

In the case of the government robot that determines welfare benefits, a decision might overcomply with the law, for instance by demanding unnecessary information before granting benefits. A welfare applicant illegally denied benefits might challenge this overcompliance decision.[6] However, decisions that undercomply with the law, for instance by granting benefits without all of the legally required information, will not be challenged. The recipient of the benefit has no reason to challenge the government's undercompliance, and no one else has standing to do so.

---

4. In contrast, I would characterize the ideas of certifying robots proposed in one of this volume's papers as an ex ante idea. Susan Saab Fortney, *Online Legal Document Providers and the Public Interest: Using a Certification Approach to Balance Access to Justice and Public Protection*, 72 OKLA. L. REV. 91 (2019). Another paper proposes using robots to review human decisions, while my take is the other way around and focuses on human review of robot decisions. Anita Bernstein, *Minding the Gaps in Lawyers' Rules of Professional Conduct*, 72 OKLA. L. REV. 123 (2019).

5. Other robot systems that are more securely located in private law, such as the systems for will preparation, are considered in another paper in this volume. Emily S. Taylor Poppe, *The Future is ~~Bright~~ Complicated: AI, Apps, and Access to Justice*, 72 OKLA. L. REV. 183 (2019). These might also have analogous incentives to reach decisions that would minimize the risk of challenge. But I focus here on robots whose legal decisions arise in the area of regulatory compliance.

6. *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256 (2008) (describing the Colorado Benefits Management System, which incorrectly denied benefits to eligible welfare applicants); Marc Cohan & Mary R. Mannix, *National Center for Law and Economic Justice's SNAP Application Delay Litigation Project*, 46 CLEARINGHOUSE REV. J. POVERTY L. & POL'Y 208, 211 (2012).

For a market robot, such as TurboTax, the situation is reversed. A decision that undercomplies with the law, perhaps by allowing certain borderline expenses as trade or business tax deductions, could be challenged by the government. A decision that overcomplies with the law, for instance by disallowing borderline expenses as trade or business deductions, will not be challenged by the government.

When private-market robots or government robots make the kind of legal mistakes that I have described, sometimes we want to know how they make their decisions because a valid legal process can persuade us to respect the decisions themselves as final. For instance, reason-giving can show the reasonableness or non-negligence of a private market decision. Or it can show that a government decision followed constitutional due process requirements or the Administrative Procedure Act.

Robots face a challenge in responding to requests for reason-giving because robot decisions may not be transparent or explainable. When a robot's legal decision is challenged, the "transparency" or "explainability" of the decision can provide a path to respecting the decision of a robot as final and valid. If decisions are not considered explainable, robotic systems may not be able to invoke the validity of their own internal process as a way to defend their decisions. This may give robots, more than humans, the incentive to avoid any request to explain why they did what they did.

Because of the difficulty robots may face in reason-giving, they may have an incentive to generate decisions that run contrary to usual expectations. Government robots may begin to interpret the law to include fewer requirements and provide a more generous interpretation of the law. Private-market robots may begin to interpret the law to include more requirements and provide a less generous interpretation of the law. This is because it is more difficult to challenge overcompliant private-market decisions and undercompliant government decisions. Instead, overcompliant private-market decisions (such as declining to claim a borderline tax deduction) and undercompliant government decisions (such as granting benefits without collecting all required information) tend to stand as valid and final, simply because the paths available to challenge them are limited or nonexistent.

Compliance robots face other incentives in addition to the incentive to avoid legal challenge. For instance, market robots generally have a profit incentive. If this profit incentive encourages a market robot to undercomply with the law, then the incentive to avoid legal challenge may provide a helpful offset, and reduce undercompliance rather than causing overcompliance. On the other hand, if a market robot's profit incentive

encourages overcompliance with the law, for instance because users prefer safe legal positions, then the incentive to avoid legal challenge may increase the market robot's tendency to overcomply. In the case of TurboTax, for example, profit incentives may encourage the software to break taxpayer confidentiality laws but comply with substantive tax law.[7] I do not mean to suggest that the incentive to avoid legal challenge is always an overriding consideration, only that it is one consideration that helps predict the direction of robot-made law.

This paper proceeds as follows. Part I discusses the existing concerns in the law literature about the transparency and explainability of automated intelligent systems, or robots. Part II categorizes the legal mistakes of automated intelligent systems and shows that mistakes of overcompliance or undercompliance could be made by either a private-market robot or a government robot. Part III uses the examples of negligence liability and procedural due process for administrative decisions to illustrate how explainability could enable the law to decide that a robot's decision should be respected as legal, even if it is mistaken. A conclusion follows.

I. Transparency and Explainability

Transparency and explainability are questions that arise when robots make legal decisions. There are narrower and broader definitions of "robot." Some definitions say a robot is an "object."[8] Others define robots to include intelligent automatic systems that arrive at results and take actions without human intervention.[9] I will use the broader definition here. An automated system that independently makes legal decisions qualifies as a robot under this broader definition. It is useful to categorize TurboTax, for instance, as a robot.

Legal scholars and policy makers worry about the explainability of robot legal decisions. A law in the European Union attempts to provide a "right to explanation" of automatic legal decisions.[10] Legal scholars struggle to

---

7. *See* Susan C. Morse, *When Will a Tax Compliance Robot Follow the Law?*, 1 OHIO ST. TECH. L.J. (forthcoming 2019) (suggesting that tax software preparation programs may violate taxpayer confidentiality law and comply with substantive tax law).

8. A. Michael Froomkin, *Introduction* to ROBOT LAW x, xi (Ryan Calo, A. Michael Froomkin & Ian Kerr eds., 2016).

9. Mark A. Lemley & Brian Casey, Remedies for Robots (forthcoming 2019) (advocating broader definition).

10. *See* Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 82 (2017) (arguing that the law does not succeed

reconcile existing legal tools for validating decisions with the black-box nature of automatic decisions. How can we determine the negligence of an algorithm too complex for human processing or evaluate whether due process is met when machine learning manipulates data to maximize the correlation of different variables through a process very difficult or impossible to reverse engineer? Legal scholars recommend access to the inner workings and data inputs of automated systems as a solution to the problem of explainability.[11] Some suggest that the right approach is for law to evolve and develop the expertise necessary to evaluate the quality of the data input and machine-learning models used by automatic systems, just as law has earlier developed the capacity to evaluate the quality of statistical analysis.[12] Others suggest that automated systems may prove unable to fulfill law's requirements for explainability.[13]

Some objects or goals of automated system explainability have been identified in the scholarship that explores it. The goals involve reason-giving and providing tools that will help a rightful plaintiff develop her case challenging the automated system's legal result. But the function of explainability—or not—in the development of the law created by robots still deserves further study.

The question of explainability has to do with institutional competence, or institutional settlement, which is to say the legal process issue of when we will treat a decision made by a particular legal institution as a final and lawful decision. Robots act as instruments of other existing institutions, in particular the "fourth branch," also known as administrative agencies;[14] and private ordering, which is sometimes called the "fifth branch," especially when it engages in explicit self-regulation.[15]

---

in part because it does not engage the question of whether the right relates to an "ex ante explanation" of system logic or functionality or an "ex post" explanation of how a particular decision was reached).

11. *See, e.g.*, Citron, *supra* note 6, at 1305-13 (recommending "audit trails," the release of source code, testing suites and public participation in the building of automated systems).

12. *See* Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 14-16 (2019) [hereinafter Coglianese & Lehr, *Transparency*] (citing Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017)).

13. *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

14. *See, e.g.*, Peter P. Swire, *Incorporation of Independent Agencies into the Executive Branch*, 94 YALE L. J. 1766, 1766 (1985) (citing use of the term "fourth branch" in Humphrey's Ex'r v. United States, 295 U.S. 602 (1935)).

15. *See, e.g.*, Harold I. Abramson, *A Fifth Branch of Government: The Private Regulators and Their Constitutionality*, 16 HASTINGS CONST. L. Q. 165 (1989).

Existing law helps us see when robot decisions—whether by fourth-branch, or government, robots or fifth-branch, or private-market, robots—will be open to challenge and when we will respect them as final without giving any avenue for challenge. We care about explainability when decisions are open to challenge, because explainability determines the ability of a robot to give reasons that may persuade us to accept its decision, even if a different institution, like a court, might make a different decision. Generalizing the question in this way allows us to observe the vulnerability of robot decisions that are (i) not explainable and (ii) open to challenge.

II. A Typology of Robot Legal Mistakes

A. Market and Government Robots, Overcompliance and Undercompliance Mistakes

This Part seeks to show that robots' legal mistakes can go in different directions. In other words, mistakes can be overinclusive or underinclusive. I use the example of compliance with an environmental emissions regulation to argue that either overcompliance or undercompliance mistakes could be made by either a government robot or a private-market robot. That is, a fourth-branch administrative agency or government robot could either overcomply, by restricting emissions too tightly, or undercomply, by allowing too much pollution. Likewise, a fifth-branch private ordering or market robot could allow either too little or too much pollution.

The below two-by-two matrix summarizes the mistakes a robot could make in interpreting a regulation that allows a certain emissions level for a certain pollutant. The robot's task in this example involves legal judgment calls at one or more steps in its compliance process. For instance, the robot might measure source data about the pollutant, including making decisions about the validity of the data. Or, the robot might compare the data to the permitted emissions level to see if the emissions were "compliant," perhaps translating continuous data about emissions to a standard written using a different measurement unit. Or, the robot might face a pollutant with a somewhat different chemical makeup than that mentioned in guidance. Other judgment calls might present when the robot prepares and submits a report to the government verifying compliance or admitting noncompliance.

*Table 1: Undercompliance/Overcompliance
in Market/Government Robots*

	Market robot	Government robot
Undercompliance	(1) Allows more pollution than the law permits. Remedy: Government enforcement. More likely to be challenged.	(3) Allows more pollution than the law permits. Remedy: Requires standing, varies with subject area. Less likely to be challenged.
Overcompliance	(2) Restricts pollution more than the law requires. Remedy: Private law. Less likely to be challenged.	(4) Restricts pollution more than the law requires. Remedy: Regulated party challenge to government decision. More likely to be challenged.

   The matrix reveals four types of mistakes: (1) market undercompliance;
(2) market overcompliance; (3) government undercompliance; and (4)
government overcompliance.
   Category (1), market undercompliance, is perhaps the easiest to
understand. It makes sense that a robot designed by and for the private
market would pursue the goal of undercompliance. Classic economic theory
suggests that a person subject to the law will undercomply to save
compliance costs so long as the probability of getting caught is less than
one hundred percent.[16]

---

   16. *See* Gary Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON.
169, 198-99 (1968) (considering situation where penalties not increased to offset imperfect
enforcement).

Category (4), government overcompliance, is also straightforward. When government administrators are charged with implementing the law, they may play it safe by requiring more than is strictly necessary under the law. For instance, government may use safe-harbor guidance to describe behavior that will definitely and clearly comply.[17]

Category (3), government undercompliance, is predicted by the theory of public choice and the idea of capture. An industry group, such as a group of firms that produce a pollutant, might lobby a government agency to adopt a sympathetic interpretation of the law when the agency programs a legal compliance robot or chooses the data set on which the robot will base decisions.[18] Capture and undercompliance could result.

Category (2) describes market overcompliance. It suggests, in other words, the existence of a phenomenon that we can call reverse capture, which causes market robots to make decisions that favor the government. A market robot might prefer to make conservative decisions about the content of the law for several reasons. One possibility is that transaction costs are lower because the robot can directly import government guidance, which may have conservative or safe-harbor bent.[19] Or, the market robot's users may prefer less aggressive positions.[20] Or, the market robot's users may not know that the robot is taking less aggressive positions.

---

17. *See* Susan C. Morse, *Safe Harbors, Sure Shipwrecks*, 49 U.C. DAVIS L. REV. 1383, 1392-94 (2016) (giving examples of safe-harbor guidance).

18. *See, e.g.*, James Q. Wilson, *The Politics of Regulation*, *in* SOCIAL RESPONSIBILITY AND THE BUSINESS PREDICAMENT 135, 142 (James W. McKie ed., 1974) (describing a "small, relatively homogeneous beneficiary group" that "impos[es] unobtrusive costs on large numbers of others").

19. *See* Joshua D. Blank & Leigh Osofsky, *Simplexity: Plain Language and the Tax Law*, 66 EMORY L.J. 189, 229-31 (2017) (giving examples of TurboTax repeating government guidance verbatim); Shu-Yi Oei & Leigh Z. Osofsky, *Constituencies and Control in Statutory Drafting: Interviews with Government Tax Counsel*, 104 IOWA L. REV. 1291, 1317-18 (2019) (reporting that statutes and guidance are often drafted with a tax software audience in mind).

20. There could be a division of opinion within an institution like a firm about the right level of under- or overcompliance. For instance, less risk-averse shareholders might prefer undercompliance, while more risk-averse employees in the compliance department of a firm might prefer overcompliance. *Cf.* GEOFFREY PARSONS MILLER, THE LAW OF GOVERNANCE, RISK MANAGEMENT AND COMPLIANCE 168-69 (2014) (describing the developing multidisciplinary "compliance industry" and the emergence of a new compliance profession).

B. Challenging Robots' Legal Mistakes

The type of mistake made by a robot influences the avenue that can be used to challenge the mistake. Market overcompliance mistakes and government undercompliance mistakes are less likely to be challenged. Market undercompliance mistakes and government overcompliance mistakes are more likely to be challenged.

Consider category (2) mistakes, where the market robot overcomplies with the law, and, for example, emits too little pollution. There is no public-law mechanism to correct this problem of overcompliance. However, a private-law mechanism could reverse this decision for the future. For instance, a contract between a robot and a user could require the robot to take more aggressive positions, or different legal-compliance robots could market different levels of compliance.

With respect to category (3), government undercompliance, it is sometimes the case that no challenge is possible. Even if a tax provision illegally favors a particular group of taxpayers, other taxpayers (whose taxes presumably will increase as a result) generally lack standing to challenge the provision directed at the favored group.[21] In contrast, in environmental law, avenues exist for the general public to claim standing and challenge such examples of government undercompliance, [22] even though these avenues may be narrow.[23]

There is generally an established procedure for challenging category (1), market undercompliance. This avenue is government enforcement. Partly because of limited resources, government underenforces the law. But the process is there and might be made more efficient if the government directly targeted market legal-compliance robots.[24]

---

21. *See, e.g.*, Frothingham v. Mellon, 262 U.S. 447, 487 (1923) (denying standing to a taxpayer objecting to provision of tax law). *See generally* Linda Sugin, *Invisible Taxpayers*, 69 TAX L. REV. 617 (2016). *But see generally* Nancy Staudt, *Taxpayers in Court: A Systematic Study of a (Misunderstood) Standing Doctrine*, 52 EMORY L. J. 771 (2003) (reporting cases of taxpayer standing in municipal tax cases).

22. *See, e.g.*, Massachusetts v. EPA, 549 U.S. 497, 504-06 (2007) (failure of EPA to regulate carbon dioxide emissions); Friends of the Earth, Inc. v. Laidlaw Envtl. Servs., 528 U.S. 167, 173-74 (2000) (failure of government regulator to challenge corporation's violation of Clean Water Act).

23. *See, e.g.*, Lujan v. Defs. of Wildlife, 504 U.S. 555, 560 (1992) (requiring a "concrete and particularized" "injury in fact," a "causal connection," and the capacity of a court decision to redress the harm).

24. *See generally* Susan Morse, *Government-to-Robot Enforcement*, 2019 U. ILL. L. REV. (forthcoming 2019).

Another mistake likely to be challenged is type (4), government overcompliance—when the government robot requires more of a regulated party than the law does. Examples of government robots improperly denying welfare benefits[25] or allegedly terminating a public-school teacher[26] fall into this category. If the harm is big or salient enough, the individual harmed by the government decision has an incentive to challenge it.

III. When Are Robot Decisions Final, Even if They Are Mistaken?

A. Explainability and Finality of Robot Decisions

When an institution—like the federal or a state government, the executive, a court, a regulatory agency, or private ordering—makes a decision or takes an action, the law faces a choice. One option is to leave the decision or action alone, thus implicitly respecting it as legal. Another option is to provide an avenue to challenge the decision.[27]

We can observe this pattern in the categories of decisions described above. For market robots, undercompliance mistakes might be challenged by the government. But overcompliance mistakes are unlikely to be challenged.

For government robots, the pattern is reversed. Overcompliance mistakes might be challenged by aggrieved regulated parties. But undercompliance mistakes are less likely to be challenged, despite the existence of narrow theories of standing in areas such as environmental law.

Transparency and explainability help us decide whether the action or decision is lawful and final. They have a role when a robot's action or decision is challenged. For instance, they are relevant when a market robot makes an undercompliance mistake and faces government enforcement. Transparency and explainability are also relevant when a government robot makes an overcompliance mistake and faces a challenge from a regulated party. Especially without explainability, market robots will be poorly

---

25. Citron, *supra* note 6, at 1256 (describing a state government automated law system that incorrectly denied benefits to eligible welfare recipient).

26. *See, e.g.*, Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1171 (S.D. Tex. 2017).

27. This choice is one object of Hart and Sacks' classic investigation of the sources of law and their idea of institutional settlement. *See* HART, JR. & SACKS, *supra* note 1, at 158 ("What is each of [the various U.S. lawmaking institutions] good for? How can it be made to do its job best? How does, and how should, its working dovetail with the working of the others?").

equipped to defend the validity of undercompliance mistakes, and government robots will be less able to defend the validity of overcompliance mistakes. Thus, there is an incentive for robot legal decision makers to favor the government if they are private market robots and to favor the private market, or regulated parties, if they are government robots.

The claim is not that this incentive to avoid legal challenge will control robot legal decisions. For instance, it may only partly counteract a market robot's incentive to profit by violating the law. The goal here is simply to observe the incentive to avoid legal challenge and to state its counterintuitive bent. Robots will face this incentive especially if they cannot explain their actions, because a lack of explainability makes it more difficult to defend an action under familiar legal standards.

B. Applying a Negligence Standard to Market Undercompliance

Private ordering, or the market, is sometimes labeled the "fifth branch" of government, particularly when it takes the form of an institution that seeks to self-regulate. This label appears to place private ordering after the legislature (the first branch), the executive (second), the judiciary (third), and government bureaucracy (fourth). Nevertheless, private ordering typically comes first in time when new legal issues are presented. That is, private ordering initially determines who shall bear liabilities and make payments and enjoy benefits, and it is the role of the other branches to limit or change those private decisions.

The Hart and Sacks treatment of the capacity of legal process to validate institutions' legal decisions begins with private ordering.[28] The authors' setup is to investigate the government's role in limiting the market decisions that will be respected as lawful, or in helping with gap-filling when private agreements are incomplete. They include one private-ordering case involving a railroad that required a cattle drover to waive all rights to a negligence claim in exchange for the drover's "free" transport aboard the train that also carried the stock in his charge.[29] Citing the railroad's common-carrier status and the greater power of the railroad relative to the customer, the Supreme Court refused to allow the railroad to waive negligence liability.

---

28. *Id.* at 7 ("[T]he individual himself, acting alone or in concert with other private individuals, is in the front line of decision, and this fact is of far-reaching importance in the whole theory of social ordering.").

29. *See* N.Y. Cent. R.R. Co. v. Lockwood, 84 U.S. (17 Wall.) 357, 384 (1873).

Similarly, commentators who study automated systems have begun to suggest that platforms should bear responsibility for some unreasonable decisions. Danielle Citron and Benjamin Wittes, for instance, make this suggestion with respect to platform liability for terrorist messages or exploitative sexual material on the internet.[30] The issue is covered by § 230 of the Communications Decency Act, which as enacted provides that platforms like websites or search engines will not be considered "the publisher or speaker of any information provided."[31] Citron and Wittes describe the judicial interpretation of this provision as a First Amendment-based permission for sites advertising illegal commercial sex or spreading defamatory rumors.[32] They recommend a statutory amendment that would require "reasonable steps to prevent . . . unlawful use of . . . services" as a prerequisite to the conclusion that a platform did not act as a publisher or speaker.[33]

If this recommendation became law, it would become necessary to decide what "reasonable steps" means. The authors help by giving a good example and a bad example. The good example is Twitter, which takes seriously "complaints that accounts are being run by designated foreign terrorist groups.[34]" The number of pro-terrorist accounts removed by Twitter approaches 400,000. The bad example is Omegle, which apparently only warns users that when it matches users with "new friends," those new friends might be sexual predators.[35] If Omegle responded to complaints of predation by investigating and removing accounts, the authors suggest, perhaps it too would be eligible for immunity from liability under § 230.[36]

If the idea of "reasonable steps" followed the established path of other similar inquiries in law, it would require some understanding of how the automated system processes the complaints. What criteria does a platform use to determine that an account should be removed? Does it automatically remove an account in response to a complaint, or does it investigate the particulars of the charge? Understanding how the platform compares the evidence in a complaint to the content of the law prohibiting certain

---

30. *See generally* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401 (2017).

31. 47 U.S.C. § 230(c)(1) (2018).

32. Citron & Wittes, *supra* note 30, at 413-14 (giving examples of providers granted immunity under § 230).

33. *Id.* at 419.

34. *Id.* at 418.

35. *Id.* at 401.

36. *Id.* at 417-18.

behavior on the internet presumably forms the content of a "reasonable steps" analysis. It is like comparing a particular doctor's actions to accepted practice to determine whether the doctor demonstrated a reasonable degree of care, skill, and learning in connection with a malpractice case.

Without such an analysis of the reasonableness of a platform's procedures, which requires some degree of explainability of its automatic decisions, what would happen? In the absence of such information about its process, perhaps the platform would not be able to defend a decision to leave an account open after receiving a complaint. It might look instead for a way to avoid legal challenge altogether. If the platform could avoid legal challenge by showing that it always takes down content in response to complaints, perhaps it might adopt an automatic takedown policy instead.[37] If no one sues the platform to complain about removing content, then a policy of overcompliance via automatic takedown is an attractive option. Such a policy would cause the law to drift in the opposite direction from its initial bias. The platform's content might become too censored, rather than not censored enough.

C. Applying Procedural Due Process Requirements to Government Overcompliance

Administrative agency bureaucrats participate in the "fourth branch" of government. They present the question of when to respect agency legal decisions as final determinations of law. Administrative procedure plays an important role in this issue. For example, procedural due process precedent requires not only fair procedures but also procedures that minimize error.[38]

Cary Coglianese and David Lehr provide an optimistic view of how administrative procedure law and automatic regulation might together evolve to allow explainable robot decision-making that satisfies procedural due process requirements.[39] They review the three-prong *Mathews v. Goldberg* balancing test, which considers the private interest at stake; the

---

37. *See* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1185-86 (2008) (noting the problem of possible negative externalities resulting from the filtering and organizing of search results by algorithms).

38. Reason-giving is an element of defending an administrative action against Administrative Procedure Act claims, such as claims of arbitrary and capricious action, *see* 5 U.S.C. § 706(2)(A) (2018), as well as against constitutional claims such as procedural due process claims.

39. *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision-Making in the Machine Learning Era*, 105 GEO. L.J. 1147, 1184-91 (2017) [hereinafter Coglianese & Lehr, *Regulating by Robot*].

risk of error as a result of the administrative procedure and the likely error-correction benefit of a proposed change; and the Government's interest, including the resources required to administer the law using different procedures.[40] Coglianese and Lehr point out that the first point (the private interest) is "exogenous" to whether a government adjudication is made by an algorithm or a human. The third, they say, will weigh in favor of using efficient automatic or robotic adjudication.

The second *Mathews* point, relating to the risk of error and the possibility that a different process would reduce errors still further, is the hard one. In Coglianese and Lehr's view, a government robot can and will evolve to allow the validation and "cross-examination" of its automatic decision-making process. If it does so, the government robot will have achieved sufficient transparency and explainability to defend its procedures and provide enough reasons to defend its decisions against regulated parties' charge of mistakes of overcompliance.[41]

The same authors[42] refer to a recent Texas case about the firing of Houston schoolteachers based on the results of a secret algorithm that considered, among other factors, the test scores of a student in a teacher's classroom. The court denied the school district summary judgment on the teacher union's procedural due process claim. The results given by the algorithm for a certain teacher could not be replicated to check for errors, in part because the results for one teacher were related to the results for other teachers, so that changing a data input for one teacher would affect everyone else's score.[43]

One question facing the Houston school district on this issue is whether to engage or avoid the procedural due process question. It could perhaps give plaintiffs access to the internal information and programming to allow the adjudication of such a procedural due process issue. But it could also change its decision-making process to reduce the chance of such a claim, for instance by forgoing the interrelated ranking feature which now appears to be part of its review process.

If the Houston school district chooses to defend its decision by sharing information, it will be an example of government fulfilling the promise of algorithmic governance by adjusting to the demands of existing law—that is, by making its procedures explainable. A court could ease this process by

---

40. Mathews v. Eldridge, 424 U.S. 319, 335 (1976).

41. *See* Coglianese & Lehr, *Regulating by Robot*, *supra* note 39, at 1184-91.

42. *See* Coglianese & Lehr, *Transparency*, *supra* note 12, at 37-38.

43. Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1172 (S.D. Tex. 2017).

adopting a flexible and modern concept of reason-giving, as Coglianese and Lehr recommend.

But the Houston school district could also choose the latter route, for instance by abandoning its commitment to rank teachers relative to one another, in order to make its decisions easier to explain and defend. This latter, defensive route would cause a change in the development of the law. For instance, it might increase teachers' job security, because that change would reduce the chance of challenge to government decisions about teacher retention.

D. What If Robot Processes Are Not Explainable?

If robot processes are sufficiently transparent and explainable, then both market robots and government robots may defend the legality and finality of their decisions by showing the validity of their process. For instance, market robots might persuade a court that their process of decision-making is reasonable, or nonnegligent. Government robots might persuade a court that their process of adjudicating a claim adequately guards against error.

If robot processes are not sufficiently transparent and explainable, then it will be more difficult for market and government robots to defend their decisions as legal and final. Automated systems will have an incentive to make decisions that will not face legal challenge. Market robots would be more likely to overcomply with the law, and government robots would be more likely to undercomply with the law. This result is contrary to the usual prediction that private ordering tends to undercomply with the law while government guidance may overcomply.

Whether this counterintuitive prediction comes true depends on what other incentives robots face. For instance, a private-market robot's profit motive is another important factor.[44] The problem of explainability and the incentive to avoid legal challenge may reduce, but not erase, a market robot's incentive to violate the law in order to increase profits. It might offset, but not eliminate, tax preparation software companies' incentive to violate taxpayer confidentiality law. On the other hand, the incentive to avoid having to explain results might exacerbate a market robot's tendency

---

44. In addition, sometimes the robot's customer is not the regulated party. In consumer credit reports, the consumer is arguably the regulated party, since applicable rules determine how her credit report should be generated. But the profit motive of a robot-generated credit report runs to the banks that purchase the reports. *See* Angela Littwin, *Escaping Battered Credit: A Proposal for Repairing Credit Reports Damaged by Domestic Violence*, 161 U. PENN. L. REV. 363, 381-83 (2013) (noting that credit reporting agencies' "main revenue source is the [lenders] who purchase credit reports").

to overcomply, especially if overcompliance will not materially reduce user revenue. It might increase tax preparation software companies' tendency to take safe positions under substantive tax law.[45]

In circumstances where a market robot overcomplies or a government robot undercomplies because of the incentives described here, the question may arise as to whether this new drift in the law is normatively good. Most likely, the answer to this question will depend on context. For instance, take the pollution example used to illustrate the different categories of robot legal mistakes above in Part II. Context helps determine whether to favor a move toward market overcompliance or government undercompliance.

Assume, for example, that a person believes that environmental law is not strict enough, that is, that it allows too much pollution. Assume further that private-market systems dominate the area of pollution-emissions compliance. The person who prefers less pollution would welcome a result that would cause a market robot to overcomply with the law and limit or restrict pollution even more than required. This person may prefer nonexplainability so that the market system cannot defend its emissions decisions by proving that they are reasonable.

Now take the same person who prefers less pollution, but change the hypothetical to assume that government robot systems, rather than market robots, dominate the area of pollution-emissions compliance. Now, this same person might prefer that government robots are sufficiently transparent and explainable. Then, government can defend its decisions against the challenge of a regulated party that asserts that the government decisions are invalid because they cannot be explained. If the government lacks this defense, it may decide to avoid challenges from regulated parties by undercomplying rather than overcomplying with the law.

Another way to see the value of explainability is to observe that it would protect existing pathways for different government institutions to provide a check on each other. Government enforcement helps check market decisions. Judicial review of regulated party challenges helps check government decisions. If we think that the legal controversies resulting from these checks benefit the development of the law, then we should prefer to sustain them as methods of confirming institutional competence. This in turn is a reason to encourage robots to develop ways of explaining how they make decisions, and to encourage the law to develop ways of

---

45. *See* Morse, *supra* note 7 (explaining different tax software compliance incentives for taxpayer confidentiality law compared with substantive tax law).

understanding robots' legal process. Otherwise, robot legal decisions may develop in a way that avoids legal challenge altogether.

Conclusion

The questions presented by robots' legal mistakes are familiar to legal scholarship. They are part of the general inquiry about when we will accept legal decisions, like those made through private ordering or administrative action. Decision-making robots might be market robots or government robots. In either category, they can make mistakes of undercompliance or overcompliance.

Some robot mistakes are more reviewable than others. A market robot's overcompliance mistake or a government robot's undercompliance mistake is relatively unlikely to be challenged. On the other hand, government enforcement can challenge a market robot's undercompliance mistake, and an aggrieved regulated party can object to a government robot's overcompliance mistake.

Explainability is relevant to the question of whether to accept a robot's decision as a final statement of the law. This is because explainability can justify the process of decision as legitimate, for instance because it is reasonable or nonnegligent. Explainability provides a pathway for validating robot action.

If robot systems are not explainable, or if courts do not recognize their way of explaining as legitimate under the law, then robots will be left without this means of defending their actions. They may be more likely to make counterintuitive decisions that will avoid the prospect of legal challenge. For instance, subject to other relevant factors including the profit motive of market robots, a lack of explainability is likely to encourage market robots to overcomply and government robots to undercomply.