Automation Rights: How to Rationally Design Humans-Out-of-the-Loop Law

*Orly Lobel**

* * *

Introduction

This essay begins with the following puzzle: in sharp contrast to significant evidence demonstrating the effectiveness of AI-based automation in high-stakes spheres—health care, transportation, national security, finance, workplace safety, public administration, and more—the contemporary impulse is to legally require a human in the loop. This impulse is heightened the higher the stakes of the activity or decision. Indeed, the legislation emerging in both the European Union (EU) and the United States ironically showcases the assumption that when it comes to AI, high stakes equal high risk of tackling the stakes through the most advanced technology. Moreover, while there are hundreds of bills, reports, and executive orders that seek to prohibit or restrain certain uses or applications of AI, there are virtually no equivalent frameworks, or even language, that would mandate automation when such a shift has been empirically shown to be the safest or most consistent in achieving agreed upon goals or courses of action.

This Essay argues for the development of more robust—and balanced—law that focuses not only on the risks, but also the potential, that AI brings. In turn, it argues that there is a need to develop a framework for laws and policies that incentivize and, at times, mandate transitions to AI-based automation. Automation rights—the right to demand and the duty to deploy AI-based technology when it outperforms human-based action—should become part of the legal landscape. A rational analysis of the costs and benefits of AI deployment would suggest that certain high-stakes circumstances compel automation because of the high costs and risks of not adopting

* Orly Lobel is the Warren Distinguished Professor of Law and Director of the Center for Employment and Labor Policy (CELP) at the University of San Diego. She graduated from Tel-Aviv University and Harvard Law School. Named as one of the most cited legal scholars in the United States, and specifically the most cited scholar in employment law and one of the most cited in law and technology, she is influential in her field. Professor Lobel has served on President Obama's policy team on innovation and labor market competition, has advised the Federal Trade Commission (FTC), and has published multiple books to critical acclaim. Her latest book, The Equality Machine, is an Economist Best Book of the Year.

the best available technologies. Inevitably, the rapid advancements in machine learning will mean that law soon must embrace AI; accelerate deployment; and, under certain circumstances, prohibit human intervention as a matter of fairness, welfare, and justice.

The Essay suggests that the thinness of legal thinking on automation rights in both policy and scholarship can be connected to several flaws in AI debates. First and foremost is the dearth of comparative analysis of automation versus a human decision-maker, a comparison that must be made along at least six related yet distinct axes:

1. *Machine v. current human performance in relation to the desired outcomes (accuracy, consistency, safety, speed).*

2. *Machine v. human scalability and access.*

3. *Machine v. human black-box opacity and explainability.*

4. *Machine v. human traceability and detection of failures.*

5. *Machine v. human learning and improvement prospects.*

6. *Machine v. human liability schemes.*

Instead of developing such a comparative matrix, debates about AI tend to focus on the risks and failures of the technology in absolute terms, resulting in a double standard and a strong bias toward human action. Further propelling the failure to adopt a comparative-advantage analysis when it comes to AI are broadly documented behavioral biases. These biases include the *status quo bias*—the human tendency to favor what is currently in place and to fear change; the related *loss aversion effect* (psychologists Daniel Kahneman and Amos Tversky's famous dictum that losses loom larger than gains, in which people impute greater value to a given endowment when they give it up); the human tendency to *distrust what is perceived as artificial* as opposed to human or natural; and *the holier than thou bias*—the tendency to overestimate one's own performance, despite statistical evidence of high human error. This Essay argues that the responsible way for policymakers to adapt the law to the era of AI is to consider the costs and harms of staying static, just as we consider the risks of an AI shift itself. Moreover, policymakers should assume the role of aiding the public to adopt a more rational relationship with AI applications, adopting laws designed to mitigate both algorithmic adoration and algorithmic aversion.

The lack of robust AI-human comparative analysis is exacerbated by two additional flaws in contemporary law and policy. First is the privileging of privacy, including deontological or theoretical privacy rights as trumping other individual rights and social goals.

The outsized and often misleading fear of the loss of privacy has contributed to [a dearth of law pertaining](#) to mandates on fuller data collection. Second is a conflation between technological readiness and the effects of technological deployment. For example, questions about the safety of autonomous trucks have been muddled with the separate, albeit important, questions about the inevitable job losses that will result from their legal deployment on the roads.

I. Giving Up the Wheel

A striking number of recent laws have sought to ban or slow down the adoption of welfare-enhancing, and even lifesaving, AI-based technology in transportation, medicine, law, criminal justice administration, employment, finance, and education. With the rapid advancements in machine learning, the legal landscape is witnessing [calls for bans, moratoriums, and limits](#) on the deployment of the technology. Legal scholars call for an overarching [precautionary rule](#) when it comes to AI and have even [proposed](#) "a system of 'unlawfulness by default' for AI systems."

In a forthcoming Article, *[The AI Regulatory Toolbox,](#)* I show that prohibitory AI laws have been skewed to the top of the regulatory pyramid, focused on bans and command-and-control prohibitions with little attention to other forms and methods of regulation and governance, including standardization, public investment, assessment, and incentives to learn and develop best-practices. In 2023, the United States saw countless proposals to ban AI technologies, ranging from biometrics and monitoring technologies to weapon systems, autonomous vehicles, and the use of AI in decision-making processes like criminal justice administration, hiring, or loan approvals. [The federal Algorithmic Accountability Act](#) would require an assessment of the need for "guard rail[s] for or limitation on certain uses or applications of the automated decision system."

The newly enacted [EU Artificial Intelligence Act](#) (EU AI Act) includes bans on certain "high-risk" AI practices. The European Parliament [summarized these practices](#) as including: "biometric categorisation systems that use . . . political, religious, philosophical beliefs, sexual orientation, race"; "untargeted scraping of facial images"; "emotion recognition"; "social scoring based on social behaviour or personal characteristics"; "AI systems that manipulate human behaviour to circumvent their free will"; and "AI used to exploit the vulnerabilities of people (due to their age, disability, social or economic situation)." There is no upshot path to automate when, for example, AI is used to protect against the exploitation of the vulnerabilities and fallibilities of humans.

The EU AI Act differentiates between high- and low-risk AI systems, providing that "[h]uman oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse." The regulation requires high-risk AI systems to be "designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons." It also bans certain uses of AI that create "unacceptable risks," though it does not specify what those risks are. It further divides the world of AI risk into high- and low-risk, subjecting high-risk AI systems to elaborate risk regulation and funneling lower-risk AI systems into the abstract requirement of an acceptable level of risk. There are no principles in the Act on what acceptable levels of risks are or what requirements are necessary to compare to risks emanating from the status quo of human processing.

One important example of attempts to ban lifesaving AI comes from the debates over the deployment of AI in transportation. Tesla, Waymo, and Uber are at the forefront of developing autonomous vehicles (AVs) with AI-enabled navigation, obstacle detection, and decision-making in real-time traffic scenarios. The primary advantage of AVs is their potential to significantly reduce accidents caused by human error. According to the National Highway Traffic Safety Administration, [94% of serious crashes are due to human error](). AVs, through their precise and consistent operation, could lower this statistic dramatically. Automation on the roads is the antidote to driver fatigue and risky maneuvers, potentially saving lives while boosting efficiency and lowering shipping costs. The trajectory towards mandating AI applications in such an area where they are set to outperform humans is both inevitable and responsible. Yet even the legality of their deployment is currently in question.

In 2023, California [Assembly Bill 316]() aimed to put the brakes on self-driving trucks by mandating the presence of human safety operators in these trucks. Proponents of the bill championed it as both a shield against job losses for the state's sixty thousand truckers and a safety measure. Governor Newsom [vetoed the bill](), rightfully in my opinion, pointing to the [evidence]() that self-driving trucks actually improve safety and that a human trucker override could reduce, rather than augment, road safety. Newsom [vetoed]() a nearly identical bill in September 2024. He issued a message about the veto to the Members of the California State Assembly:

> California leads the nation with some of the strongest worker protection laws. Our state also is renowned globally

as a leader in technological innovation. We reject that one aim must yield to the other, and our success disproves this false binary. But advancing both priorities requires creativity, collaboration, and a willingness to work together to identify pragmatic solutions. Toward that end, my office offered multiple rounds of suggested amendments, which were unfortunately not accepted.

Deployment of other AI-based transportation advancements has also been slowed down by laws with questionable rationality. The [Federal Aviation Administration rules](#) requiring drone operation to occur within a human operator's view have slowed down important advancements in delivery and other services. The California law prohibiting opt-in AI-based driving safety monitoring technologies has [prevented lifesaving technology](#) even as a voluntary, private, insurance-based measure.

The job loss question should not muddle the evidence about safety. Much of the fear and resistance to AI has to do with our fear of being replaced by machines. These are valid fears. We are facing undeniable waves of seismic changes in the labor market. We should anticipate and address these changes that will occur in every industry. For example, like transportation, healthcare is facing seismic change. Already, medical AI outperforms the work of healthcare professionals in a wide variety of diagnostics and patient care, including [AI-assisted surgery](#), [AI–nursing assistants](#), and [telemedicine](#). Parallel changes are happening in the legal field, in art and entertainment, and in science and technology.

It is critical that we keep the questions distinct: Is AI safe and ready? Does it outperform human decision-making? Is it less prone to accidents and inconsistencies? And separately: What must the new social contract look like in the face of rapid job loss and changes in employment patterns? For everyone to enjoy the benefits of AI, the labor market effects need to be tackled with investment in reskilling, education, taxation, social welfare, public options, public procurement, and access to these revolutionary technological capabilities.

The late psychologist Daniel Kahneman [predicted in an interview](#) that "[b]eing a lot safer than people is not going to be enough. The factor by which [AVs] have to be more safe than humans is really very high." That is an alarming prediction. Yet, in another high-stakes transportation context, the international community has already agreed that automation is much safer in high-states circumstances: air travel. The entirety of the international aviation industry operates with the gold standard of autopilot when weather conditions are harsh. In the riskiest conditions faced by commercial

aviation, the international community has long established humans-out-of-the-loop rules. Reduced Vertical Separation Minimums [allow flights](#) to have a small vertical separation (one thousand feet) if the instruments meet certain accuracy requirements and the airplanes are operated using autopilot. Pilots undergo required training but give up controls in riskier proximities.

If regulators, pilots, and passengers are comfortable with this standard, there is no reason to believe that we cannot learn to love "a lot safer" autonomous cars—as well as fully autonomous commercial planes. An acceptance of an "a lot safer" standard instead of a "really very high" factor of better performance requires law and policy, research, design, and education. Although the willingness of people to give up control of the wheel may be different across generations, once our individual rights and public goals can be clearly better realized with technology, a right and duty to automate become not only possible but morally correct.

II. A Right to AI Decision-Makers

Where AI is proving to be more accurate and effective than human decision-making, law requires new directions for analyzing the necessity of mandating AI applications in certain fields. For example, in the medical field, particularly in diagnostic procedures, recent advancements in AI have enabled systems to diagnose certain conditions, such as skin and breast cancer or retinal diseases, more accurately than human doctors. In the financial sector, algorithms are now able to analyze market data and consumer behavior with a precision that far surpasses human capabilities. For example, JPMorgan Chase's AI program, COIN, can interpret commercial loan agreements in seconds, a task that takes legal professionals 360,000 hours annually.

By automating such tasks, AI not only increases efficiency but also minimizes human errors and bias that could have significant consequences. Failing to develop laws prohibiting human decision-making under certain circumstances is a normative failure with serious costs. Yet again, with regard to recommendation algorithms and sorting applications, we witness a slew of laws that are aimed to prohibit such technologies, while no equivalent laws mandating automation currently exist.

In 2023, for example, [a New York state bill](#) was introduced to prohibit the use of algorithmic decision-making in hiring and job screening decisions without the involvement of a final human decision-maker. Article 22 of the EU [General Data Protection Regulation](#) (GDPR) states that the "data subject shall have the right not to be

subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." The exceptions are narrow and require a showing of necessity, special authorization, or "the data subject's explicit consent."

Again, no equivalent duties to adopt AI have been contemplated in our recent laws. In particular, when it comes to civil rights, access to justice, and public administration, AI regulation should include mandates to automate. In these spaces, AI systems have been proven safer, fairer, or more accurate than human decision-making and existing systems:

- **Good Governance**: Cary Coglianese and I argue in a forthcoming article *Algorithmic Administration as Constitutional Governance* that government agencies, as part of their constitutional duty to ensure good government, should shift to digitization and automation when those systems are more efficient and better at achieving public goals.

- **Clean Slate**: Colleen Chien has offered a compelling argument that to make laws that increase the eligibility of people with criminal backgrounds to clear their records and regain the right to vote a reality, we need to shift "administrative burdens from the defendant [ ] onto the state and algorithms through automation, standardization, and ruthless iteration."

- **Criminal Justice**: In criminal justice, AI systems like Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) have been used for sentencing and assessing the likelihood of a defendant reoffending. Their application has sparked debates around the biases inherent in AI algorithms and their impact on fairness and justice, but the empirical evidence demonstrates that shifting to such algorithmic application reduces overall racial bias and the harms of arbitrary policing.

- **Child Welfare**: In a recent empirical study of algorithmic tools that assist caseworkers in investigating child abuse or neglect, Amit Haim found that such tools improved the abilities of caseworkers and lowered the risks of invasive investigations.

- **Pay Equity**: In recent research on gender and racial pay gaps, I have shown that to tackle longstanding, stagnating pay inequities, AI-based software should not only be voluntary, but also become an industry gold standard and be required by law.

- **Patenting**: Using AI can automate patent-application drafting, error checking, distinguishing filed patent claims from what came before, flagging mismatches between the specification and the claimed invention, providing first-draft specification language, and placing a patent application in a better condition to be granted. Expanded and equitable access to technology can increase patent access, quality, efficiency, and equity.

- **Freedom of Information:** If agencies adopt clear rules requiring automation in scanning, digitizing, and making government documents searchable and findable, the mandate of freedom of information will be more of a reality. In a recent ACUS report, my co-authors and I have called for the accelerated development of digital tools to realize the mandate of affirmative disclosure of agency materials.

- **Public Safety:** California has enacted laws that temporarily prohibit the use of facial recognition technology in body cameras worn by law enforcement officers. Similarly, cities like San Francisco, Oakland, and Boston have passed ordinances banning the use of facial recognition technology by city departments, including police departments. Portland, Oregon, passed one of the most stringent bans, prohibiting private entities from using facial recognition technology in public places. The similar Facial Recognition and Biometric Technology Moratorium Act, proposed in Congress, seeks to impose a moratorium on federal use of the technology until certain conditions are met. Ironically, avoiding profiling, arbitrariness, and exclusion more often than not requires more complete data collection and publicly available datasets. The specific resistance to biometrics and facial recognition and the privileging of broad notions of individual privacy come at significant costs that have not yet received enough attention in law and legal scholarship.

III. Operationalizing Automation Rights

The adoption of AI advancements is not just a technological inevitability, but a societal responsibility. In private law—from torts to occupational safety regulations, anti-discrimination to environmental protection—doctrines about duty of care and standards about state-of-the-art safety and compliance are already embedded in existing laws. In public law, core values of constitutional law include equal treatment, consistency, and effective governance.

Automation rights would also include a focus on how to foster rational trust in AI. Beyond the assessment of the effectiveness and accuracy of the AI models themselves, law and policy should pay more

attention to how individuals—users, passengers, patients, citizens—
assess AI. The nascent field of behavioral human-machine interactions
indicates the push and pull of the resistance to AI: driven by both the
perception that AI is a mysterious "black box" and the illusion
(perhaps delusion) that human decision-making is easy to understand.

Both in lay and expert settings, people routinely overestimate
their ability to perform and decide accurately without bias. In
experimental research, patients are willing to forego better health care
to have a human, rather than an AI, decision-maker. Our current laws
have likely contributed to such irrationalities. As I argued in *The Law
of AI for Good*, "[r]equiring humans to be the final decision-makers in
high stakes processes is not only a flawed solution in contexts where AI
has clearly reached comparative advantages, but it also risks
perpetuating irrational fears about AI instead of helping debias
citizens about the comparative risks of technology."

In a new article, *Do We Need to Know about Artificiality:
Unpacking Disclosure and Generating Trust in an Era of Algorithmic
Action*, I argued that there should not be a default overarching right to
know that a decision or content was generated by AI. The right to
know that you are interacting with a bot, or that you are subject to
automated decision-making, is a centerpiece of EU/U.S. legislative
proposals. Both GDPR and the California Consumer Privacy Act
(CCPA) already include the right to know if AI is making decisions
about an individual and to request explanations for those decisions.

Under the new EU AI Act, consumers will have a right to know
if they are chatting with or seeing images produced by AI. Title IV,
Article 52 states: "Providers shall ensure that AI systems intended to
interact with natural persons are designed and developed in such a
way that **the AI system, the provider itself or the user informs** the
natural **person exposed to an AI system** that they are interacting
with an AI system … unless this is obvious from the circumstances and
the context of use." Quebec has also passed a law that requires
individuals to be informed when automated decision-making tools are
being used. Other private and public declarations about ethical AI
similarly emphasize such disclosures about artificiality as a keystone
of AI governance.

For example, a recent FTC post warns businesses not to make
humanizing claims about AI applications: "Your therapy bots aren't
licensed psychologists, your AI girlfriends are neither girls nor friends,
your griefbots have no soul, and your AI copilots are not gods.
We've warned companies about making false or unsubstantiated
claims about AI or algorithms." While certainly consumer protection
laws must be applied in full force in the age of AI, we need to be clear

about what are the most effective ways to apply these laws to new applications and when disclosures about artificiality may have counter-productive effects of enhancing irrationality or undue distrust of a system.

In recent surveys, most people in the United States want to know when they are interacting with AI. And yet, I argue the reasons for disclosing artificiality are complex, and often in tension with other goals of generating rational decision-making, trust, and safety. AI law and policy should be based on analysis, not epithets. Automation rights are inevitable, but we have not even begun to conceptualize, develop terminology, and consider the complex design of the approaching legal landscape.

* * *

Orly Lobel is the Warren Distinguished Professor of Law and Director of the Center for Employment and Labor Policy (CELP) at the University of San Diego. She is the author of The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future.