Robotics, AI and Humanity

Russell A. Newman, Emerson College and Harvard University's Berkman Klein Center for Internet and Society

Book Review: Frank Pasquale (2020) *New Laws of Robotics: Defending Human Expertise in the Age of AI.* Cambridge, MP: Belknap Press (344 pages).

In early March 2021, the National Security Commission on Artificial Intelligence, co-chaired by Eric Schmidt, published a report arguing that the U.S. position on AI lagged behind that of China and Russia. A number of initiatives were proposed to remedy the situation. This development followed controversies surrounding Google's letting go of two prominent and respected ethicists. Meanwhile, robot police dogs were trialed in New York City's Bronx borough, and the global pandemic forced higher education to reckon with its own future amidst an already-dire student loan crisis. Into this environment, Frank Pasquale offers *New Laws of Robotics,* following his important *The Black Box Society.* In the earlier work, Pasquale made a compelling and chilling clarion call against the opaque processes of data collection and interpretive judgment that have encroached on social life. The present book builds upon the previous effort to advance a cohesive optimistic program, although it begs deeper theorization of the object being addressed.

If the 'robot question' in the 1960s focused on the automation of manufacturing jobs, now "the computerization of services is top of mind" (197). Economists dominate this debate, rendering it largely in terms of cost-benefit analysis that "prioritizes capital accumulation over the cultivation and development of democratically governed communities of expertise and practice" (197). Indeed, "Conversations about robots usually tend toward the utopian ('machines will do all the dirty, dangerous, or difficult work') or the dystopian ('and all the rest besides, creating mass unemployment'). But the future of automation in the workplace—and well beyond—will hinge on millions of small decisions about how to develop AI" (14).

To this end, he proposes four 'laws' for artificial intelligence: 'complementarity,' 'authenticity,' 'cooperation,' and 'attribution.' "A humane agenda for automation," he argues, "would prioritize innovations that complement workers in jobs that are, or ought to be, fulfilling vocations" (4). Pointing to the emergence of chatbots, appointment-assistants, and more, he suggests that "robotic systems and AI should not counterfeit humanity" (7). With an eye in particular toward military conflicts and automated, 'smart' policing, "Robotic systems and AI should not intensify zero-sum arms races" (9).

Finally, "Robotic systems and AI must always indicate the identity of the creator(s), controller(s), and owner(s)" (11). These 'laws' are "not meant merely to rein in mad scientists or put some guard rails around rambunctious corporate titans. Rather, the purpose is to preserve certain human values in health, education, journalism, policing, and many other fields" (171).

Pasquale is not just pushing back against neoliberalism's core drives of austerity and efficiency wrought through the plunging of populations into ever more deeply embedded market structures; he is also resisting calls for a kind of left 'accelerationism' that might lead to the 'end of work.' Similarly, he questions literary narratives that evoke hope in automation in itself as offering social equity. All such viewpoints strike him as equally pernicious in 'eliminating the human' in important ways. He seeks to preempt 'scientific freedom' or even 'free speech' as justifications for AI development. These justifications are seen to "unleash […] commercial pressures and robotic mimicry to colonize, replace, and rule every form of human labor," in the process "radically reorganiz[ing] society" (24). Pasquale leads us on a sweeping tour of health, education, what he terms 'automated media,' finance, defense, policing, and even the discursive realms of the arts and literature.

Pasquale roots his concerns in what has been termed 'cost disease.' According to dominant strains of economic theorizing, certain sectors seem impervious to efforts to render them more efficient or productive, such as services and the arts. Thus, in relative terms, these sectors become costlier in comparison to everything else. The "cure" for cost disease is "standardizing, automating, even robotizing service-intensive sectors like health care, finance, and education. From massive open online courses to robotic companions for the elderly, these innovations aim to displace workers (or at least work) with mechanical imitations of human labor" (196). The "cost disease" narrative has underwritten virtually every technological and communications policy debate for the better part of decades now, even as, in Pasquale's terms, it represents "an ensemble of bad metaphors and dubious assumptions" (187). This problem is heightened as a series of neoliberal administrations have taken the dicta to heart: "[N]eoliberal progressives were consistently pushing policies to disrupt health care and education, reducing the flow of funds into these sectors, accelerating degree programs, and tightening the budgets of hospitals and other caregiving institutions. Some even explicitly connected this agenda to the alleged need for more military spending. Centrists around the world have adopted these policies, as have more conservative parties" (189).

The book truly shines when it revisits the former territory Pasquale treaded in *Black Box Society*. The most striking revelations are found when he is thinking through "machines judging humans" via facial recognition, in regard to finance and in policing. Here, his sweep draws together converging developments in China and America. But this is not to leave aside the rewards of his other journeys. The neoliberalization of higher education, he worries, has boiled the enterprise down to 'pragmatism and behaviorism' driven by "an ideology of austerity" (85) rather than any deeper purpose. Technologies introduced to the classroom have contributed to building what amounts to a move beyond grades to a "total behavioral record" that could potentially measure students "along axes of friendliness, attractiveness, and more" from which it is difficult to escape (75). As Ruha Benjamin, Safiya Noble, Nick Couldry and Ulises Mejias have noted, inclusion in the present environment is quite complex, and Pasquale provides a very effective takedown of its predatory, creepy, and subordinating dimensions. "Without proper guardrails, there will be a race to the bottom in both data sharing and behavior shaping as more individuals compete for better deals" (133) across multiple sectors.

We are offered numerous policy suggestions across these areas. The through-line is a call to rejuvenate the idea of professions as valuable in themselves, emphasizing the principle of 'subsidiarity' that "commends a devolution of responsibility to the most local entity capable of handling [an issue] well" (175-176). It is similarly a call to recognize and appreciate complexity so as to not fall for illusions of clear distinctions that can be readily drawn by AI systems. Yet this is no call to eliminate AI; it is to recognize a balance between those systems that make sense to automate and those systems which are better suited to augmenting professional intelligence. These are social questions, not technological ones. The solutions to the conundrums he describes lie less in some 'more effective' use of these technologies than in reorganizing policy around the normative roles of humans and humanity going forward: "[W]e can redesign social systems to change these incentives and pressures, rather than trying to change people to fit into increasingly dysfunctional systems" (p. 55). In fact, technical solutions to current problems, he notes, actually supply new points of vulnerability and thus targets for attack. Relatedly, the regimentation sought by China's citizenship scoring and methods of censorship (alongside their developing equivalents outside China) ultimately disrupt the very feedback loops necessary to continue systemic improvements. Such improvements require human input. Labor looms large, with collective bargaining agreements providing opportunities for worker participation in how automation systems are implemented: "AI should not entrench deep disparities in the power of workers, managers, and capital owners. Rather, it can help unions and worker associations bring more prerogatives of self-governance to the workplace" (p. 176). He also encourages the resuscitation of union schools, which drove home lessons not just in skill training but in regard to the structural forces that shaped shifting labor conditions (173). Any rethink of higher education would be wise to go beyond 'skill training' in this way itself. Given the increasing employer surveillance of employees, he calls for the creation of regulatory bodies to push back against "abusive uses of technology" and to enforce clear rules (131). Additional such bodies, such as Pasquale's proposed "Education Technology Agencies" modeled upon drug regulatory agencies, might also be considered. In media, he encourages governments to require communications platforms like Facebook to "label, monitor, and explain hate-driven search results and other deeply offensive content" (108). Logics of engagement as a commercial end in itself need to be dealt with—especially given that the tech giants have themselves become de facto regulators.

Useful as these proposals are, of perhaps even greater value are Pasquale's observations that certain debates require expansion. Conversations about ethics in artificial intelligence desperately need to be supplemented (or even supplanted) by broad revisioning. He argues that analysis of AI needs to move from merely "reforming" mechanisms of scoring to more basic questions of "turn[ing] to limits on its scope, judging AI itself as falling short in crucial ways" (141). The epistemological dimensions here become crucial, as "we need to start grappling with the thought of public intellectuals who question the entire reformist discourse of imparting ethics to military robots. …[W]e need a different path—to cooperation and peace, however fragile and difficult its achievement may be" (168-169). This is particularly true as AI is inserted into military-industrial complexes. Even as the National Security Commission on Artificial Intelligence chose to use a lens of competing national systems in its analysis, Pasquale recognizes instead that a more appropriate view on the "advance of AI guard labor" is "less a rivalry among great powers than a globe-encompassing project by corporate and government elites to maintain hegemony over restive populations. The powerful often have far more in common with one another than with the average citizens of their home countries" (167). Nationalist lenses offer a "convenient veneer of conflict

over a common interest in promoting the stable extraction of value from labor unable to demand a larger share of the economy's productive gains" (167). Debates surrounding emergent media need similar expansion; the now common tropes of filter bubbles and echo chambers have been rendered quite quaint. Thinking through algorithmic bias and production as in fact an "alien intelligence" requiring new forms of interrogation, we should go as far as to "echo the critical theory penned in reaction against the authoritarianism of the 1930s"—particularly as algorithmic threats to the contemporary public sphere are already exploited by authoritarian populists worldwide (117).

On the strength of such calls, it is perhaps surprising that the critical theory proposed is quite constrained. In Pasquale's telling, less emphasis is placed on the trajectory of present systems than the current logics at play. These are two different things, and the distinction is crucial. This issue is made plain by his animating questions: "How far should machines be entrusted to take over tasks once performed by humans? What is gained and lost when they do so? What is the optimal mix of robotic and human interaction? And how do various rules—from codes of professional ethics to insurance policies to statutes—influence the scope and pace of robotization in our daily life?" (14).

These questions are certainly pressing; however, the effort to answer them—while thorough and certainly contributing to important conversations of our time—is frustratingly undertheorized. Pasquale is well aware that something more than "vague commitments to corporate social responsibility and regulation" (131) will be necessary; he presses the need for a "more sustainable economy, with automation ordered primarily toward creating ample goods and services for all, while reducing that production's negative impact on the environment" (171), something that will require a "revitalization of tax, competition, labor and education policy" (172). Further, he notes that "the history of US decline over the past half-century could be written as a story of the systematic transfer of once-public funding for infrastructure, education, and health care back to private individuals and corporations" (185). All the same, one worries that we still lack a clear rubric here for theorizing our times beyond the expansion of corporate and elite power, even as its logics, incentives, and practices prove noxious and pervasive. This is to ask, what is the cost when the *critique* of emergent artificial intelligence incompletely captures the neoliberal (or even possibly *post*-neoliberal) tendencies of the teleological ends being prepared via its adoption?

There are real consequences to approaching this realm exclusively as a snapshot of current logics. Pasquale's incomplete theorization of the trajectory within which this reorganization is taking place is revealed in the assumption that underwrites much of this project—that it is seemingly always cost-effective to replace the labor of humans with that of machines. This is to leave out the findings of the likes of Ekbia and Nardi (2017). In their description of 'heteromation,' the insertion of humans into technical processes occurs *precisely because of the cost savings of using human labor.* The punishing warehouse labor by not-yet-robotized third party logistics operators who fulfill online orders is one such example; Amazon's 'Mechanical Turk' labor force provides another. Similarly, when Google's pervasive Captchas are used to confirm that a user is human, this user is providing what the autonomists have termed 'free labor' for the tech giant. That is, the challenges posed are as likely to clear up a puzzle Google's AI cannot solve itself (such as identifying crosswalks, railroads, and buses in its stored imagery, or resolving unclear text in a scanned book) as they are to confirm what it already 'knows.'

Pasquale's critique, stemming more from the snapshot approach as opposed to an *historical* one, foreshortens the horizon toward which this process steers. Surely even platform (or even surveillance) capitalism, fast becoming the most common and perhaps deceptively attractive trope describing our time, is hardly the end but part of a process, an early envoy of a structure to come

which impinges on our social being and how we conceive of freedom within it. Pasquale is a student of the law; and his chosen instruments to which he turns reflects this. An emphasis on Modern Monetary Theory and a strong faith that "careful application of legal and ethical principles can lead to nuanced, contextual judgments"—here, describing positively the structures supporting the EU's controversial "right to be forgotten"—perhaps does not take into account enough the ways that such positive initiatives are themselves being transformed, or corrupted.

This might seem like scathing critique, but it is not intended to detract from the value of the work: rather, it is intended to acknowledge that Pasquale cracks open the door to enable the consideration of such big questions. This is his overarching contribution, and it is no small thing. The final chapter, an analysis of the world of art and literature and the ways in which AI has been grasped there, acknowledges the cultural importance of envisioning our collective horizon. Perhaps most importantly, the book joins a broader effort to reclaim ground from those economists who have dominated the making of innovation policy. Pasquale notes that "Other social science and humanities scholars have vital insights…A renewed political economy of automation must be built upon these plural foundations, unafraid of exacting the substantive value judgments encoded in the new laws of robotics" (196). Pasquale has set himself an incredibly ambitious task and he is to be lauded for it.

Author bio

Russell A. Newman is Associate Professor at Emerson College and Faculty Associate at Harvard University's Berkman Klein Center for Internet and Society. His work explores the intersections of the political economy of communication, neoliberalism, the epistemological foundations of media policymaking, commercial and governmental surveillance, and activism surrounding communications policy. He is the author of *The Paradoxes of Network Neutralities* (MIT Press, 2019); with Robert McChesney and Ben Scott, he co-edited *The Future of Media: Resistance and Reform in the 21st Century* (Seven Stories Press, 2005).

References

Ekbia HR and Nardi BA (2017) *Heteromation and Other Stories of Computing and Capitalism*. Cambridge, MA: MIT Press.