Book Reviews

New laws of robotics: defending human expertise in the age of AI, by Frank Pasquale,

Cambridge, Harvard University Press, 2020, 330 pp., ISBN: 9780674975224

Ethical and legal challenges brought about by advances by artificial intelligence (AI), broadly construed, are a hot topic from policy circles to academia. Perhaps because of how wide-ranging and troubling these challenges are, much of the academic and popular discussion about them is carried on by people that do not always want to or know how to stay within the lanes of their expertise. Ethicists swerve into computer science, computer scientists swerve into epistemology, legal scholars swerve into economics, and so on, causing what may appear as a general blurring of the issues at the heart of rapid advances in AI technologies. The loss of clarity can be offset by a coherent “big picture” narrative, which shows us how ethics, computer science, law, economics, etc., individually tell a part of the story of a world undergoing changes because of advances and implementations of AI. This “big picture” narrative is not often given convincingly enough to persuade people that stay strictly within the boundaries of their disciplines that this sort of work is valuable. Frank Pasquale’s ‘New Laws of Robotics’ does give us a big picture narrative and in a way that should be clearly of value to anyone working in philosophy, law, or computer science. Its biggest selling point is a set of concrete and easily understandable proposals for legal instruments that can mitigate the undermining of human expertise. An additional bonus is an extensive interdisciplinary bibliography, impressively detailed and informative footnotes, and a cautiously optimistic tone. What is perhaps most valuable in this book, is the legal perspective and scholarly know-how that Pasquale brings to issues that are typically addressed by people in without legal expertise. Given this, ‘New Laws of Robotics’ will inevitably be useful to a wide range of experts that are looking for the legal aspects of what they are researching. One can only hope that its insights will influence policy. The book can also serve as a comprehensive introduction for people relatively new to these debates, but who aim to develop a scholarly project. Given its approachable style, it will also find a readership among non-academics looking for a big picture narrative about the impact of artificial intelligence and machine learning technology on society and its institutions.

Altogether, Pasquale’s book gives us eight chapters, six of which take on a different sphere of human expertise that is being replaced by AI/ML technologies. For each of these six chapters we get a unique legal dimension of the ongoing process of replacement of expertise and Pasquale’s take on ways to mitigate it using his four “new laws”:

1. Robotic systems and AI should complement professionals, not replace them.
2. Robotic systems and AI should not counterfeit humanity.
3. Robotic systems and AI should not intensify zero-sum arms races.
4. Robotic systems and AI must always indicate the identity of their creator(s), controller(s), and owner(s).

These appear largely depending on what Pasquale takes to be the most valuable element of human expertise worth protecting in the domain under discussion. For example, the third law, which is concerned with AI arms races, is present in context of a discussion of autonomous weapons, but also in context of mitigating the spread of misinformation.

The first domain Pasquale focuses on is health care. The discussion anticipates many of the problematic developments of AI-aided medicine, but without losing sight of AI’s potential for addressing health-related problems facing a globalized world, including the timely issue of pandemics. We get a well-informed, but admittedly non-technical, discussion of opacity of machine learning and the troubling legal consequences that it will bring to medical professionals and patients. Some technical specialists may find this discussion to be glib. AI in the medical context is not a magic eight ball that turns out to be right ninety percent of the time, but a set of relatively well-understood predictive modelling techniques. That said, even they will likely find Pasquale’s legal insight about machine learning in medicine valuable. Automated diagnostics, where the expertise of the medical professional is gradually replaced with algorithms need legislative solutions that can protect the patient and the doctor, perhaps even changing the way that liability and malpractice are assessed in court altogether. AI-aided medicine leaves the legal risk with the medical professionals, while effectively taking away their autonomy to collaboratively decide on courses of care with their patients. Pasquale’s solution, and one that we will see later in the book in various forms, is to put the human decision-making process in a central position with respect to AI-aided medicine, thus preserving the values we most care about. Institutions should embrace one or more of Pasquale’s New Laws and implement them in specific legislation or expect the negative consequences to catch up with them. For example, it is difficult to disagree that handing over massive amounts of patient data to for-profit companies is a sound way to preserve accountability in health-care systems, especially in a context where the position of the patient is ever-more like that of a consumer and the position of the doctor more like that of a salesperson.

The second domain of concern in the book is education. This section is especially pertinent today, given all the ways in which technologically mediated forms of education have become the norm during the Covid-19 pandemic. Pasquale identifies the overall trend in AI-aided education as an gradual increase in control, parametrization, and dehumanization and then carefully shows how these trends flush out of education the values that rendered it available to the general public in the first place. The goal of general public education in democracies was, at least in part, to create and maintain a well-informed population that is able to critically assess and decide who should represent them in government. The AI-aided trends undermine this goal, by turning education into a numbers game that has features of a conformist rat race for employment. Pasquale’s analysis here banks on the value of educating citizens, as opposed to merely skilled workers. Whether indeed metrics of success average out all the uniqueness out of any single student and ultimately form the basis for ever more dehumanizing educational technologies, is to some extent an empirical question. That said, Pasquale’s analysis could not have forecast what is indeed starting to emerge as the new status quo in higher education, namely, the video-conference classroom and remote testing via the internet. We should assume that at the time *New Laws of Robotics* was being written, educators were not yet fully aware of just how dehumanizing and exhausting teaching over the internet can be, nor just how much privacy will have to be sacrificed to ensure that online exams administered to students stuck at home can be proctored for exams. We also did not have a full picture of the cost in mental health that mass online education can have on students and staff. These unanticipated developments may ultimately make Pasquale’s arguments for valuing expertise of the teacher that much stronger. We should wait for the data in this case and prepare for the worst.

The third domain is media, which, again, is a particularly timely topic. Misinformation, conspiracy theories, and AI-aided amplification of narratives are real and present dangers to democracies around the world. Pasquale again brings his legal insight to bear on the issue and illuminates the way in which the attention economy, mixed with machine learning and absentee ownership by Big Tech created a legal limbo. It is not clear to what extent content delivery platforms can be held accountable for the nature of the content. It is not clear to what extent it is their legal obligation to worry that their content is radicalizing teenagers or facilitating crime. Pasquale’s analysis here shows that the dramatic consequences of the decline of newspapers in the wake of internet “content providers” taking over the flow of information on jobs and the quality of public discourse was just the opening act of a process of social epistemic degradation. In of a world in which trusted sources are replaced by filter bubbles the most pernicious actors get to exploit human foibles and manipulate public opinion at will. So, the decline of the epistemic expert is not just a sad state of affairs, it poses a very real and imminent threat to societies everywhere, and especially democracies. The current media landscape funnels whatever individual decision-making powers citizens may have had to the highest bidder for the latest media platform. This chapter is, arguably, the most depressing in the book, since its implications may make Pasquale’s recommendations and suggestions look somewhat Pollyannaish. If the media landscape continues in this direction, the legal instruments Pasquale sees as possible solutions will simply never come about. They will be smothered by misinformation or organized campaigns to discredit them. Happily, Pasquale spares the reader a discussion of the consequences that are likely to transpire if experts are not put back in the role of gatekeeps in the media landscape.

The media chapter naturally leads to the next theme in the book, which at least in part is the legal domain itself. AI and predictive policing are forecast to do the same to expertise as in the media, and play along with the forces that shape it. Predictive analysis of human behavior, including law enforcement, finances, etc., is literally everywhere, and there is a number of important works on the subject that Pasquale judiciously mentions throughout this chapter. Algorithmic bias, social credit systems, and so on, are all here. For people already aware of these topics and that read the relevant scholarly and popular literature on these topics, this chapter may offer the least compared to the others. It is worth quoting Pasquale at length to demonstrate how his legal insight adds an important legal perspective to the existing voices:

As Kiel Brennan-Marquez has explained, a jurisprudence of well-founded suspicion (largely arising out of Fourth Amendment law in the US) demands that authorities give a *plausible*, and not merely probabilistic, statistical, or artificially intelligent, account of the reasons why they investigate suspects. We need to understand not simply *that* we are being watched, but also *why*. This is a limit on the power of the state, which may be all too tempted to use advanced surveillance technology to achieve complete control over citizens. (p 129)

In other words, the legal instruments to protect us from the state may be the same ones that will protect us from AI.

In case you have not been frightened yet, you should be after finishing Chapter 6, which concerns AI in war. The most apocalyptic science fiction scenarios of a robotic arms race are indeed around the corner and Pasquale delivers a sober, crisp, and matter-of-fact assessment, full of insight into the legal and political ramifications of taking human beings out of the process of killing in war. This is, however, the least worked-out chapter in the book, and one which arguably deserved much more serious legal analysis. For one, the situation of an arms race in this domain has been going on for some time. It is not an accident that the Department of Defense and various military agencies have been major sponsors of AI research in the US and abroad for decades. Autonomous sentry robots have been deployed on the Korean border last century. Most major militaries in the world have are actively using autonomous underwater drones, autonomous tanks, or simply a chassis with a machine gun. So, here Pasquale’s “new laws” are fighting windmills at best and whitewashing reality at worst. What we need is a set of arguments that will make it clear to decision-makers even in the most militaristic and oppressive regimes in the world, that developing autonomous weapons is against their own interests. The only arguments to that end will focus on the immediate, short-term consequences, not the long-term ones, which may or may not even transpire. Militaries are preparing for the next war, not next world. We know from at least the time of the Cold War, that the logic of war can comfortably fit with total world annihilation or not next world at all.