Deadly algorithms

Can legal codes hold software accountable for code that kills?

Susan Schuppli

Algorithms have long adjudicated over vital processes that help to ensure our well-being and survival, from pacemakers that maintain the natural rhythms of the heart, and genetic algorithms that optimise emergency response times by cross-referencing ambulance locations with demographic data, to early warning systems that track approaching storms, detect seismic activity, and even seek to prevent genocide by monitoring ethnic conflict with orbiting satellites.[1] However, algorithms are also increasingly being tasked with instructions to kill: executing coding sequences that quite literally execute.

Guided by the Obama presidency's conviction that the War on Terror can be won by 'out-computing' its enemies and pre-empting terrorists' threats using predictive software, a new generation of deadly algorithms is being designed that will both control and manage the 'kill-list,' and along with it decisions to strike.[2] Indeed, the recently terminated practice of 'signature strikes', in which data analytics was used to determine emblematic 'terrorist' behaviour and match these patterns to potential targets on the ground, already points to a future in which intelligence-gathering, assessment and military action, including the calculation of who can legally be killed, will largely be performed by machines based upon an ever-expanding database of aggregated information. As such, this transition to execution by algorithm is not simply a continuation of killing at ever greater distances inaugurated by the invention of the bow and arrow that separated warrior and foe, as many have suggested.[3] It is also a consequence of the ongoing automation of warfare, which can be traced back to the cybernetic coupling of Claude Shannon's mathematical theory of information with Norbert Wiener's wartime research into feedback loops and communication control systems.[4] As this new era of intelligent weapons systems progresses, operational control and decision-making are increasingly being outsourced to machines.

Computing terror

In 2011 the US Department of Defense (DOD) released its 'roadmap' forecasting the expanded use of unmanned technologies, of which unmanned aircraft systems – drones – are but one aspect of an overall strategy directed towards the implementation of fully autonomous Intelligent Agents. It projects its future as follows:

> The Department of Defense's vision for unmanned systems is the seamless integration of diverse unmanned capabilities that provide flexible options for Joint Warfighters while exploiting the inherent advantages of unmanned technologies, including persistence, size, speed, maneuverability, and reduced risk to human life. DOD envisions unmanned systems seamlessly operating with manned systems while gradually reducing the degree of human control and decision making required for the unmanned portion of the force structure.[5]

The document is a strange mix of Cold War caricature and Fordism set against the backdrop of contemporary geopolitical anxieties, which sketches out two imaginary vignettes to provide 'visionary' examples of the ways in which autonomy can improve efficiencies through inter-operability across military domains, aimed at enhancing capacities and flexibility between manned and unmanned sectors of the US Army, Air Force and Navy. In these future scenarios, the scripting and casting are strikingly familiar, pitting the security of hydrocarbon energy supplies against rogue actors equipped with Russian technology. One concerns an ageing Russian nuclear submarine deployed by a radicalized Islamic nation-state that is beset by an earthquake in the Pacific, thus contaminating the coastal waters of Alaska and threatening its oil energy reserves. The other involves the sabotaging of an underwater oil pipeline in the Gulf of Guinea off the coast of Africa, complicated by the approach of a hostile surface vessel capable of launching a Russian short-range air-to-surface missile.[6]

These Hollywood-style action film vignettes – fully elaborated across five pages of the report – provide an odd counterpoint to the claims being made throughout the document as to the sober science, political prudence and economic rationalizations that guide the move towards fully unmanned systems. On what grounds are we to be convinced by these visions and strategies? On the basis of a collective cultural imaginary that finds its politics within the CGI labs of the infotainment industry? Or via an evidence-based approach to solving the complex problems posed by changing global contexts? Not surprisingly, the level of detail (and techno-fetishism) used to describe unmanned responses to these risk scenarios is far more exhaustive than that devoted to the three primary challenges which the report identifies as specific to the growing reliance upon and deployment of automated and autonomous systems:

1. Investment in science and technology (S&T) to enable more capable autonomous operations.
2. Development of policies and guidelines on what decisions can be safely and ethically delegated and under what conditions.
3. Development of new Verification and Validation (V&V) and T&E techniques to enable verifiable 'trust' in autonomy.[7]

As the second of these 'challenges' indicates, the delegation of decision-making to computational regimes is particularly crucial here, in so far as it provokes a number of significant ethical dilemmas but also urgent questions regarding whether existing legal frameworks are *capable* of attending to the emergence of these new algorithmic actors. This is especially concerning when the logic of precedent that organizes much legal decision-making (within common law systems) has followed the same logic that organized the drone programme in the first place: namely, the justification of an action based upon a pattern of behaviour that was established by prior events.

The legal aporia intersects with a parallel discourse around moral responsibility; a much broader debate that has tended to structure arguments around the deployment of armed drones as an antagonism between humans and machines. As the authors of the entry on 'Computing and Moral Responsibility' in the *Stanford Encyclopedia of Philosophy* put it:

> Traditionally philosophical discussions on moral responsibility have focused on the human components in moral action. Accounts of how to ascribe moral responsibility usually describe human agents performing actions that have well-defined, direct consequences. In today's increasingly technological society, however, human activity cannot be properly understood without making reference to technological artifacts, which complicates the ascription of moral responsibility.[8]

When one poses the question, under what conditions is it morally acceptable to deliberately kill a human being, one is not, in this case, asking whether the law

permits such an act for reasons of imminent threat, self-defence or even empathy for someone who is in extreme pain or in a non-responsive vegetative state. The moral register around the decision to kill operates according to a different ethical framework: one that doesn't necessarily bind the individual to a contract enacted between the citizen and the state. Moral positions can be specific to individual values and beliefs whereas legal frameworks permit actions in our collective name as citizens contracted to a democratically elected body that acts on our behalf but with which we might be in political disagreement. While it is, then, much easier to take a moral stance towards events that we might oppose – US drone strikes in Pakistan – than to justify a claim as to their specific illegality given the anti-terror legislation that has been put in place since 9/11, assigning moral responsibility, proving criminal negligence or demonstrating legal liability for the outcomes of deadly events becomes even more challenging when humans and machines interact to make decisions together, a complication that will only intensify as unmanned systems become more sophisticated and act as increasingly independent legal agents. Moreover, the outsourcing of decision-making to the judiciary as regards the validity of scientific evidence, which followed the 1993 *Daubert* ruling – in the context of a case brought against Merrell Dow Pharmaceuticals – has, in addition, made it difficult for the law to take an activist stance when confronted with the limitations of its own scientific understandings of technical innovation. At present it would obviously be unreasonable to take an algorithm to court when things go awry, let alone when they are executed perfectly, as in the case of a lethal drone strike.

By focusing upon the legal dimension of algorithmic liability as opposed to more wide-ranging moral questions I do not want to suggest that morality and law should be consigned to separate spheres. However, it is worth making a preliminary effort to think about the ways in which algorithms are not simply reordering the fundamental principles that govern our lives, but might also be asked to provide alternate ethical arrangements derived out of mathematical axioms.

Algorithmic accountability

Law, which has already expanded the category of 'legal personhood' to include non-human actors such as corporations, also offers ways, then, to think about questions of algorithmic accountability.[9] Of course many would argue that legal methods are not the best frameworks for resolving moral dilemmas. But then again nor are the objectives of counter-terrorism necessarily best serviced by algorithmic oversight. Shifting the emphasis towards a juridical account of algorithmic reasoning might, at any rate, prove useful when confronted with the real possibility that the kill list and other emergent matrices for managing the war on terror will be algorithmically derived as part of a techno-social assemblage in which it becomes impossible to isolate human from non-human agents. It does, however, raise the 'bar' for what we would now need to ask the law to do. The degree to which legal codes can maintain their momentum alongside rapid technological change and submit 'complicated algorithmic systems to the usual process of checks-and-balances that is generally imposed on powerful items that affect society on a large scale' is of considerable concern.[10] Nonetheless, the stage has already been set for the arrival of a new cast of juridical actors endowed not so much with free will in the classical sense (that would provide the conditions for criminal liability), but intelligent systems which are wilfully free in the sense that they have been programmed to make decisions based upon their own algorithmic logic.[11] While armed combat drones are the most publicly visible of the automated military systems that the DOD is rolling out, they are only one of the many remote-controlled assets that will gather, manage, analyse and act on the data that they acquire and process.

Proponents of algorithmic decision-making laud the near instantaneous response time that allows Intelligent Agents – what some have called 'moral predators' – to make micro-second adjustments to avert a lethal drone strike should, for example, children suddenly emerge out of a house that is being targeted as a militant hideout.[12] Indeed robotic systems have long been argued to decrease the error margin of civilian casualties that are often the consequence of actions made by tired soldiers in the field. Nor are machines overly concerned with their own self-preservation, which might likewise cloud judgement under conditions of duress. Yet, as Sabine Gless and Herbert Zech ask, if these 'Intelligent Agents are often used in areas where the risk of failure and error can be reduced by relying on machines rather than humans … the question arises: Who is liable if things go wrong?'[13]

Typically when injury and death occur to humans, the legal debate focuses upon the degree to which such an outcome was foreseeable and thus adjudicates on the basis of whether all reasonable efforts and pre-emptive protocols had been built into the system to mitigate against such an occurrence. However, programmers cannot of course run all the variables that combine to produce machinic decisions, especially when the degree of uncertainty as to conditions and knowledge of events on the ground is as variable as the shifting contexts of conflict and counter-terrorism. Werner Dahm, chief scientist at the United States Air Force, typically stresses the difficulty of designing error-free

systems: 'You have to be able to show that the system is not going to go awry – you have to disprove a negative.'[14] Given that highly automated decision-making processes involve complex and rapidly changing contexts mediated by multiple technologies, can we then reasonably expect to build a form of ethical decision-making into these unmanned systems? And would an algorithmic approach to managing the ethical dimensions of drone warfare – for example, whether to strike 16-year-old Abdulrahman al-Awlaki in Yemen because his father was a radicalized cleric; a role that he might inherit – entail the same logics that characterized signature strikes, namely that of proximity to militant-like behaviour or activity?[15] The euphemistically rebranded kill list known as the 'disposition matrix' suggests that such determinations can indeed be arrived at computationally. As Greg Miller notes: 'The matrix contains the names of terrorism suspects arrayed against an accounting of the resources being marshaled to track them down, including sealed indictments and clandestine operations.'[16]

Intelligent systems are arguably legal agents but not as of yet legal persons, although precedents pointing to this possibility have already been set in motion. The idea that an actual human being or 'legal person' stands behind the invention of every machine who might ultimately be found responsible when things go wrong, or even when they go right, is no longer tenable and obfuscates the fact that complex systems are rarely, if ever, the product of single authorship; nor do humans and machines operate in autonomous realms. Indeed, both are so thoroughly entangled with each other that the notion of a sovereign human agent functioning *outside* the realm of machinic mediation seems wholly improbable. Consider for a moment only one aspect of conducting drone warfare in Pakistan – that of US flight logistics – in which we find that upwards of 165 people are required just to keep a Predator drone in the air for twenty-four hours, the half-life of an average mission. These personnel requirements

are themselves embedded in multiple techno-social systems composed of military contractors, intelligence officers, data analysts, lawyers, engineers, programmers, as well as hardware, software, satellite communication, and operation centres (CAOC), and so on. This does not take into account the R&D infrastructure that engineered the unmanned system, designed its operating procedures and beta-tested it. Nor does it acknowledge the administrative apparatus that brought all of these actors together to create the event we call a drone strike.[17]

In the case of a fully automated system, decision-making is reliant upon feedback loops that continually pump new information into the system in order to recalibrate it. But perhaps more significantly in terms of legal liability, decision-making is also governed by the system's innate ability to self-educate: the capacity of algorithms to learn and modify their coding sequences independent of human oversight. Isolating the singular agent who is directly responsible – legally – for the production of a deadly harm (as currently required by criminal law) suggests, then, that no one entity beyond the Executive Office of the President might ultimately be held accountable for the aggregate conditions that conspire to produce a drone strike and with it the possibility of civilian casualties. Given that the USA doesn't accept the jurisdiction of the International Criminal Court and Article 25 of the Rome Statute governing individual criminal responsibility, what new legal formulations could, then, be created that would be able to account for indirect and aggregate causality born out of a complex chain of events including so called digital perpetrators? American tort law, which adjudicates over civil wrongs, might be one such place to look for instructive models. In particular, legal claims regarding the use of environmental toxins, which are highly distributed events whose lethal effects often take decades to appear, and involve an equally complex array of human and non-human agents, have been making their way into court, although not typically with successful outcomes for the plaintiffs. The most notable of these litigations have been the mass toxic tort regarding the use of Agent Orange as a defoliant in Vietnam and the Bhopal disaster in India.[18] Ultimately, however, the efficacy of such an approach has to be considered in light of the intended outcome of assigning liability, which in the cases mentioned was not so much deterrence or punishment, but, rather, compensation for damages.

Recoding the law

While machines can be designed with a high degree of intentional behaviour and will out-perform humans in many instances, the development of unmanned systems will need to take into account a far greater range of variables, including shifting geopolitical contexts and murky legal frameworks, when making the calculation that conditions have been met to execute someone. Building in fail-safe procedures that abort when human subjects of a specific size (children) or age and gender (males under the age of 18) appear, sets the stage for a proto-moral decision-making regime. But is the design of ethical constraints really where we wish to push back politically when it comes to the potential for execution by algorithm? Or can we work to complicate the impunity that certain techno-social assemblages currently enjoy? As a 2009 report by the Royal Academy of Engineering on autonomous systems argues,

> Legal and regulatory models based on systems with human operators may not transfer well to the governance of autonomous systems. In addition, the law currently distinguishes between human operators and technical systems and requires a human agent to be responsible for an automated or autonomous system. However, technologies which are used to extend human capabilities or compensate for cognitive or motor impairment may give rise to hybrid agents … Without a legal framework for autonomous technologies, there is a risk that such essentially human agents could not be held legally responsible for their actions – so who should be responsible?[19]

Implicating a larger set of agents including algorithmic ones that aid and abet such an act might well be a more effective legal strategy, even if expanding the limits of criminal liability proves unwieldy. As the 2009 ECCHR *Study on Criminal Accountability in Sri Lanka* put it: 'Individuals, who exercise the power to organise the pattern of crimes that were later committed, can be held criminally liable as perpetrators. These perpetrators can usually be found in civil ministries such as the ministry of defense or the office of the president.'[20] Moving down the chain of command and focusing upon those who participate in the production of violence by carrying out orders has been effective in some cases (Sri Lanka), but also problematic in others (Abu Ghraib) where the indictment of low-level officers severed the chain of causal relations that could implicate more powerful actors. Of course prosecuting an algorithm alone for executing lethal orders that the system is in fact designed to make is fairly nonsensical if the objective is punishment. The move must, then, be part of an overall strategy aimed at expanding the field of causality and thus broadening the reach of legal responsibility.

My own work as a researcher on the Forensic Architecture project, alongside Eyal Weizman and several others, in developing new methods of spatial and visual investigation for the UN inquiry into the use of armed drones, provides one specific vantage point for considering how machinic capacities are reordering the field of political action and thus calling forth new legal strategies.[21] In taking seriously the agency of things, we must also take seriously the agency of things whose productive capacities are enlisted in the specific decision to kill. Computational regimes, in operating largely beyond the thresholds of human perception, have produced informatic conjunctions that have redistributed and transformed the spaces in which action occurs, as well as the nature of such consequential actions themselves. When algorithms are being enlisted to out-compute terrorism and calculate who can and should be killed, do we not need to produce a politics appropriate to these radical modes of calculation and a legal framework that is sufficiently agile to deliberate over such events?

Decision-making by automated systems will produce new relations of power for which we have as yet inadequate legal frameworks or modes of political resistance – and, perhaps even more importantly, insufficient collective understanding as to how such decisions will actually be made and upon what grounds. Scientific knowledge about technical processes does not belong to the domain of science alone, as the *Daubert* ruling implies. However, demands for public accountability and oversight will require much greater participation in the epistemological frameworks that organize and manage these new techno-social systems, and that may be a formidable challenge for all of us. What sort of public assembly will be able to prevent the premature closure of a certain 'epistemology of facts', as Bruno Latour would say, that are at present cloaked under a veil of secrecy called 'national security interests' – the same order of facts that scripts the current DOD roadmap for unmanned systems?

In a recent ABC Radio interview, Sarah Knuckey, director of the Project on Extrajudicial Executions at New York University Law School, emphasized the degree to which drone warfare has strained the limits of international legal conventions and with it the protection of civilians.[22] The 'rules of warfare' are 'already hopelessly outdated', she says, and will require 'new rules of engagement to be drawn up': 'There is an enormous amount of concern about the practices the US is conducting right now and the policies that underlie those practices. But from a much longer-term perspective and certainly from lawyers outside the US there is real concern about not just what's happening now but what it might mean 10, 15, 20 years down the track.'[23] Could these new rules of engagement – new legal codes – assume a similarly preemptive character to the software codes and technologies that are being evolved – what I would characterize as a *projective* sense of the law? Might they take their lead from

the spirit of the Geneva Conventions protecting the rights of noncombatants, rather than from those protocols (the Hague Conventions of 1899, 1907) that govern the use of weapons of war, and are thus reactive in their formulation and event-based? If so, this would have to be a set of legal frameworks that is not so much determined by precedent – by what has happened in the past – but, instead, by what may take place in the future.

Notes

1. See, for example, the satellite monitoring and atrocity evidence programmes: 'Eyes on Darfur' (www.eyesondarfur.org) and 'The Sentinel Project for Genocide Prevention' (http://thesentinelproject.org).
2. Cori Crider, 'Killing in the Name of Algorithms: How Big Data Enables the Obama Administration's Drone War', *Al Jazeera America*, 2014, http://america.aljazeera.com/opinions/2014/3/drones-big-data-waronterror obama.html; accessed 18 May 2014. See also the flow chart in Daniel Byman and Benjamin Wittes, 'How Obama Decides Your Fate if He Thinks You're a Terrorist,' *The Atlantic*, 3 January 2013, www.theatlantic.com/international/archive/2013/01/how-obama-decides-your-fate-if-he-thinks-youre-a-terrorist/266419.
3. For a recent account of the multiple and compound geographies through which drone operations are executed, see Derek Gregory, 'Drone Geographies', *Radical Philosophy* 183 (January/February 2014), pp. 7–19.
4. Contemporary information theorists would argue that the second-order cybernetic model of feedback and control, in which external data is used to adjust the system, doesn't take into account the unpredictability of evolutive data internal to the system resulting from crunching ever-larger datasets. See Luciana Parisi's Introduction to *Contagious Architecture: Computation, Aesthetics, and Space*, MIT Press, Cambridge MA, 2013. For a discussion of Weiner's cybernetics in this context, see Reinhold Martin, 'The Organizational Complex: Cybernetics, Space, Discourse', *Assemblage* 37, 1998, p. 110.
5. DOD, *Unmanned Systems Integrated Roadmap Fy2011–2036*, Office of the Undersecretary of Defense for Acquisition, Technology, & Logistics, Washington, DC, 2011, p. 3, www.defense.gov/pubs/DOD-USRM-2013.pdf.
6. Ibid., pp. 1–10.
7. Ibid., p. 27.
8. Merel Noorman and Edward N. Zalta, 'Computing and Moral Responsibility,' *The Stanford Encyclopedia of Philosophy* (2014), http://plato.stanford.edu/archives/sum2014/entries/computing-responsibility.
9. See John Dewey, 'The Historic Background of Corporate Legal Personality', *Yale Law Journal*, vol. 35, no. 6, 1926, pp. 656, 669.
10. Data & Society Research Institute, 'Workshop Primer: Algorithmic Accountability', *The Social, Cultural & Ethical Dimensions of 'Big Data'* workshop, 2014, p. 3.
11. See Gunther Teubner, 'Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law,' *Journal of Law & Society*, vol. 33, no.4, 2006, pp. 497–521.
12. See Bradley Jay Strawser, 'Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles,' *Journal of Military Ethics*, vol. 9, no. 4, 2010, pp. 342–68.
13. Sabine Gless and Herbert Zech, 'Intelligent Agents: International Perspectives on New Challenges for Traditional Concepts of Criminal, Civil Law and Data Protection', text for 'Intelligent Agents' workshop, 7–8 February 2014, University of Basel, Faculty of Law, www.snis.ch/sites/default/files/workshop_intelligent_agents.pdf.
14. Agence-France Presse, 'The Next Wave in U.S. Robotic War: Drones on Their Own', *Defense News*, 28 September 2012, p. 2, www.defensenews.com/article/20120928/DEFREG02/309280004/The-Next-Wave-U-S-Robotic-War-Drones-Their-Own.
15. When questioned about the drone strike that killed 16-year old American-born Abdulrahman al-Awlaki, teenage son of radicalized cleric Anwar Al-Awlaki, in Yemen in 2011, Robert Gibbs, former White House press secretary and senior adviser to President Obama's re-election campaign, replied that the boy should have had 'a more responsible father'.
16. Greg Miller, 'Plan for Hunting Terrorists Signals U.S. Intends to Keep Adding Names to Kill Lists', *Washington Post*, 23 October 2012, www.washingtonpost.com/world/national-security/plan-for-hunting-terrorists-signals-us-intends-to-keep-adding-names-to-kill-lists/2012/10/23/4789b2ae-18b3–11e2–a55c-39408fbe6a4b_story.html.
17. 'While it might seem counterintuitive, it takes significantly more people to operate unmanned aircraft than it does to fly traditional warplanes. According to the Air Force, it takes a jaw-dropping 168 people to keep just one Predator aloft for twenty-four hours! For the larger Global Hawk surveillance drone, that number jumps to 300 people. In contrast, an F-16 fighter aircraft needs fewer than one hundred people per mission.' Medea Benjamin, *Drone Warfare: Killing by Remote Control*, Verso, London and New York, 2013, p. 21.
18. See Peter H. Schuck, *Agent Orange on Trial: Mass Toxic Disasters in the Courts*, Belknap Press of Harvard University Press, Cambridge MA, 1987. See also: www.bhopal.com/bhopal-litigation.
19. Royal Academy of Engineering, *Autonomous Systems: Social, Legal and Ethical Issues*, RAE, London, 2009, p. 3, www.raeng.org.uk/societygov/engineeringethics/pdf/Autonomous_Systems_Report_09.pdf.
20. European Center for Constitutional and Human Rights, *Study on Criminal Accountability in Sri Lanka as of January 2009*, ECCHR, Berlin, 2010, p. 88.
21. Other members of the Forensic Architecture drone investigative team included Jacob Burns, Steffen Kraemer, Francesco Sebregondi and SITU Research. See www.forensic-architecture.org/case/drone-strikes.
22. Bureau of Investigative Journalism, 'Get the Data: Drone Wars', www.thebureauinvestigates.com/category/projects/drones/drones-graphs.
23. Annabelle Quince, 'Future of Drone Strikes Could See Execution by Algorithm', *Rear Vision*, ABC Radio, edited transcript, pp. 2–3.