

"The point is to change it"

On the new demise of ethics

Dr. Elke Schwarz, Queen Mary University of London

On the new demise of ethics

Dr. Elke Schwarz, Queen Mary University of London

This essay is about the disappearance of ethics in an era which is characterised by the capacity to inflict large-scale violence at machine-speed. In what follows, I argue that in a technologically saturated military environment, where artificial intelligence (AI) is fast becoming a foundational infrastructure, ethics has taken a leave of absence. This should worry us, because the erosion of ethics in the domain of war and conflict might also be a harbinger for the demise of ethical thinking more broadly in public life.

At the time of writing, the world is witness to two aggressive international wars that have been raging on with breathtaking levels of human and material destruction. These two wars, in Ukraine and Gaza, have accelerated the development and rollout of new AI technologies that strive for greater and faster delivery of force, and reports of AI-enabled autonomous weapons systems being fielded are beginning to proliferate. The systems of greatest concern are those in which the human decision-maker plays an increasingly marginal role. In such times, which stand under the tragic sign of expanding violence, escalating conflict, and rising geopolitical tensions, the evaporation of what it means to give ethical consideration to other human beings should give us pause for thought.

At a 2024 AI Expo conference in Washington DC, a panel of experts with military, government, and defence industry backgrounds contemplated the wars of the future. The panel's core message was that future wars will be both devastating and incredibly lethal, but they will also be precise and humane because of new AI-enabled weapons. With this, the experts seemed to suggest that responsible powers must acquire AI weapons to deter irresponsible powers from using such weapons, and if responsible powers must use such weapons, they will do so, of course, responsibly. The ethos of this peculiar justification is perhaps best summed up by the remarks of one of the panellists — the CEO of an AI company that specialises in offensive targeting — who proclaimed with a remarkable

degree of enthusiasm: 'The peace activists are actually the pro-war activists, and we're the peace activists!' — proposing the argument that, if you don't want war, somehow the responsible action is to gear up for one. You would be forgiven for not following this logic. It makes a somewhat muddled allusion to the mechanisms of deterrence theory, a Cold War game theory-inspired way to justify stockpiling that most absurd of all weapons, the nuclear bomb. It is not a coherent discourse. But it is one in which the erosion of the status of ethics is patently demonstrated. This erosion has taken place gradually, concurrent with the ascendancy of computational rationality across various domains of human affairs.

In his 1954 book, *The Human Use of Human Beings* (Wiener, 1954), the grandfather of cybernetics Norbert Wiener makes some incisive observations about the co-constitution of humans and autonomous machines. He saw the perils of cybernetic technology for society as twofold. First, its inner logic might be wielded by one group of humans as a form of power and control over others. Second, this logic might also lead humans to believe it would be better to outsource complex and difficult moral decisions to machine authority. As an eminent scientist of technology and mathematical thinker, Wiener was ill at ease with the way in which technological thinking was fast becoming a dominant mode of philosophical and social reasoning. An insight we appear not to be heeding today.

Indeed, cybernetics, especially in the form of computational processes, has captured the collective imagination of the west since the 1950s and 60s, prompting many to understand the world and everything in it through feedback logics and probabilistic reasoning. Cybernetics is, first and foremost, a theory of processes and mechanisms. It is a theory of functionality in which inputs are translated into outputs. In cybernetics, the world is conceived as a collection of information that is to be transmitted. Information is understood as calculable data and even humans are conceived as objects of information, rendering them functional elements in a system determined by process. Cybernetics is unique in that it enables material and mental phenomena to be conceived in the same computational register. This makes its technological logic especially alluring.

It is unsurprising, then, that certain, prominent strands of moral philosophy were seduced into this schema. Derek Parfit's approach to applied ethics, for example, was thoroughly informed by the probabilistic logics of expected utility and game theory, and various modes of analytic moral reasoning about justifications for harm — from the Trolley Problem onwards — are characterised, in form, by approaching ethics as an algorithmic

puzzle. All this fits hand-in-glove with modern approaches to warfare, which are themselves more and more dependent on computational reason.

Every epoch has its own paradigmatic ways of conceptualising wars and of justifying the harms these paradigms produce. Ours stands under the sway of workflow efficiency. Current military doctrine and practice increasingly mirrors the logic of computational optimisation, prioritising cost-benefit calculations in the interest of speed and scale. All ethically salient considerations are swept up in this logic. The notion of permissible civilian harm, for example, is rerouted through the process of calculating 'collateral damage' — the percentage of civilians (men, women, children) that might be harmed if a situation necessitates it. A notoriously controversial calculus that clearly echoes the parameters of the Trolley Problem.

The Trolley Problem and its many variations are early iterations of algorithmic rationality. It is also fundamentally a problem about justifying harm. The basic scenario is fixed around the same dilemma: there is a runaway trolley, five people are tied down on one track, one person — either on another track or a footbridge or in some other perilous situation — will die. A decision must be made — will one person be sacrificed or shall five die? Each variation sets out to test to what degree different variables in the scenario matter to the moral outcome. New variables are introduced to change the moral calculus — a lever to divert the trolley; a tractor to rescue the five but which will then crush a bystander; the option to push someone off a bridge to stop the trolley. Technological instruments are always posited to make a difference to the justification of the lethal action. The role technologies play in shaping moral choices is almost never acknowledged. But crucially, this mode of reasoning implicitly constitutes an objectification of all that is human. Every subject becomes object in a world of computable things.

To be fair, Philippa Foot's original version of the problem was not intended to be used in such a formulaic manner, nor did Foot seek to spawn an entire 'Trolley-ology' industry. One might argue that utility-focused approaches to ethical reasoning originated precisely from a moral concern with social reform, refracted through a then radically new mode of accounting for egalitarian and universal values. (Although, it must be said, even in Bentham's era, the idea of smoothing the plurality of human life into a single calculus was controversial.) But the point at issue is this: ethical reasoning does not happen in a vacuum. It develops and changes alongside ways of envisioning and practising human life. In our time, these visions and practices stand under the sign of computation. Consequently, our forms of ethical reasoning are increasingly inflected by Bayesian statistics and game theory.

Nowhere is this more clearly felt than in the context of warfare. As military weapons become more computationally sophisticated, ethics is pressed further into a schema of technicity, to the point where ethics itself becomes technical, an exercise in which systematic process matters more than substance. Analytic approaches to ethics facilitate this. We can see this within abstracted conceptions of 'just war' reasoning, where the language of calculable utility dominates and the human is objectified, quite literally, in one of two ways: he or she is either the recipient of a moral decision and becomes a set of data points within a wider configuration of variables; or he or she is tasked with administering the moral calculation about who may die and in this way becomes a functional element in the wider data infrastructure. In their most abstract forms, analytic modes of just war reasoning become indistinguishable from technical discourse. They abound with mathematical syntax, taking on the pure form of algorithmic rationality itself: if X condition applies, then, *ceteris paribus*, Y must logically follow.

Algorithmic rationality seems tailor-made for thinking through ethical problems associated with autonomous, potentially lethal machines. The MIT Moral Machine project, for example, wants to identify the most agreeable ethical principles with which to programme self-driving cars, so that these cars can make the 'right' kill decision autonomously. Other research efforts by well-known technology companies are underway to design 'AI morality', so that humans can receive 'ethical guidance' on issues such as who should receive a kidney. When humans are rendered as objects of information, it is only logical that the most complex ethical calculations should be left to machines.

Abstraction marginalises the ethos of ethics altogether, especially in and for warfare. The aforementioned deterrence theory, popularised during the Cold War, is one such form of abstract reasoning about lethal technologies. It assumes that all actors are purely rational; that all actions and consequences of these actions are, to some degree, predictable; and that, based on this, the best outcome can be ascertained by drawing on expected utility and decision theory. This is ethics as economic science, as a technique for solving optimisation problems. In the context of nuclear deterrence, it overlooks entirely that the rational grounding of this approach rests on a threat of civilisational annihilation. In the contemporary context, with AI-enabled weapons, moral reasoning fits effortlessly into a workflow process, and it equally becomes ever more remote from recognising the ethos of the ethical problem itself, which is almost always an intractable social problem, resistant to optimisation efforts.

The abstraction is not accidental, but precisely the point. Abstraction decouples empathy from action. It decouples the consequences of actions from the moral weight of taking responsibility. It objectifies all those embroiled in the morally relevant calculus as data. It removes the most fundamental element of ethics — that of understanding one’s responsibility in relation to a decision which affects others. It also works to render the ‘rational’ so fantastically technological that surely, only savvy experts can be trusted with engineering the difficult choices. This directs our gaze onto ‘solvable’ ethical problems and away from those that do not sit so comfortably within a techno-economic logic.

However, our morality, and thus our moral decision-making, is anchored in a shared history of human plurality and human social relations. It is this condition that makes us not just functional actors, but moral actors. Moral agency means the capacity not simply to have, but also to take moral responsibility in relation to others. And that means having the capacity and space to reflect on one’s position and that of others in any given configuration of moral relevance, particularly in warfare. The ever-present potential for open-endedness in moral decision-making and responsibility-taking does not lend itself to the parameters of systems that function best with and within closures, and at speed and scale, as is the case with AI.

Analytic approaches to ethics are, of course, only one way of thinking about ethics as it relates to war and technologies of war. Instead of an algorithmic programme, ethics might be understood as a relational practice. Ethics as a practice is open-ended, deliberative, and ultimately concerned with broader questions concerning human relations, not technical processes. It rests on a foundation of human relations, anchored in both vulnerability and trust. As pivotal parameters for social and political life change, ethical guidelines, their associated practices, and our human relationship with ethical values and beliefs become subject to new inquiry. Ethical thought should inform the laws, guidelines, and rules that govern social and political life. But it must also serve to prompt critical reflection about the ways in which morality might become fossilised in these laws and rules, to avoid the risk that the form supersedes the content. I am reminded of Hannah Arendt’s observation about morality and mores: ‘What people get used to is less the content of the rules [than] the possession of the rules under which to subsume particulars.’

Having a set of codified ethical principles in place does not guarantee ethics in practice. This is tragically confirmed in our present moment when the law appears to be inadequate to curb the worst excesses of the ongoing conflicts, let alone stymie their expansion. Neither just

war principles nor international humanitarian law were intended as a checklist that could be satisfied, or optimised, in purely procedural ways. Rather, the intent with early just war thinking was to grapple with and understand that actions and circumstances need to be considered carefully in morally charged environments. That there are always broader ethical questions to consider in any war, questions about how we wish to live with one another, about the possibility for peace once fighting ends.

There is no easy answer to the question of how many innocent people may be killed to achieve a goal in warfare, nor is it sufficient to think of the ethics of war in mere numerical terms, counting only the dead. Those that make a decision to kill must wrestle with the weight of responsibility attached to such actions. For this, we must remain humans, not become objects ourselves, and we must tolerate the discomfort of an irresolvable ethical challenge, despite being embedded within a network of machine functionality. Herein resides moral responsibility.

The ethos of international humanitarian law or the just war tradition is still intuitive to most of us: those blameless should not have to suffer, violence should not be a first resort, and there should be a mandate for restraint to violence because otherwise peace cannot be restored. Where ethics becomes automated, this ethos is exchanged for variables that can be adjusted based on preference, interest and aim. In other words, it becomes subject to power, which ultimately is a feature of nihilism. But that is where we are today. The accelerated rollout of AI-enabled technologies of warfare, their imperative to make war faster and more lethal, and their logics of process and efficiency leave neither space nor reason for ethics to thrive, not even as a calculation. No wonder, then, that ethics is on its way out as an uncomfortable necessity, with the many intractable, irresolvable, difficult challenges that arise in war. Ethics is about meaning; AI is about function. The two are not easy bedfellows.

Two points are worth stressing in closing. First, AI's emphasis on speed is utterly incompatible with ethical deliberation. I have written elsewhere on how the iterative logic of failing and learning through accelerated processing power is both the foundation of AI as well as the technology industry itself. The motto 'move fast and break things' is not just a PR slogan; it is an ideology. Increasingly too, it seems, for the defence domain. With this shift, ethics is necessarily sidelined. The second point is that computational logics always objectify their environments, and that means objectifying humans. When an AI system identifies a human as a target-object, that human is immediately objectified, and thus

dehumanised. And from the history of mass violence, we know that dehumanisation plays a significant role in facilitating an erosion of moral restraint.

And algorithmic ethics cannot address either of these issues. Quite the contrary. Taking the ethos of ethics seriously would mean, first and foremost, not succumbing to the speed and scale mandate of killing as a systematic process. It would also mean that we ask more of those involved in AI-enabled war, not less; that we should consider maximally restrictive ideas about the use of violence rather than follow the logic of machine functionality. This means that those involved in the business of war should become habituated into thinking about humans as humans, not objects, depicted on an interface.

The ethical stakes are patently high, in warfare and otherwise. By functionalising the notion of agency and by marginalising the human as a decision maker, we diminish our ability to understand what moral agency, and responsibility, means in practice. To be sure, moral agency can be theorised and defined in different ways, and I accept that perhaps our idea of what constitutes agency should be revised in an AI-saturated environment. But agency and *moral* agency are not the same. Moral relations do not arise *ex nihilo*. As humans, we understand and are able to judge the contextual nuances of human relations in a way that technological artefacts are simply not able to. Technological artefacts can comprehend neither the complexities and incongruities of human life nor its meaning.

AI-enabled weapon systems are such technological artefacts. And they are also technologies of violence. Instruments of violence are designed to inflict damage on human bodies and human property. Too often we discuss these systems as though they were somehow separate from this violence against human life; as though war is primarily an engineering challenge and as though they somehow carry within themselves their own justification for the violence they enable. Algorithmic ethics facilitates this configuration.

Rethinking ethics not as systematic process, but as practice rooted in human relations, is an urgent task. If we let ethics disappear, and with that the intrinsic, mutual value of human life, we are readying ourselves for a future of violence, not peace.

References:

- Wiener, N. (1954). *The human use of human beings: Cybernetics and society*. New York: Da Capo Press.