

ESSAY

Humor, Ethics, and Dignity: Being Human in the Age of Artificial Intelligence

Sean Kanuck

Information and communications technologies (ICTs) are inherently new means to facilitate traditional ends of human endeavor. The quintessential function of ICTs is to help humans create, store, aggregate, analyze, process, transmit, and utilize information more effectively over greater distances and within shorter periods of time. But artificial intelligence (AI)—or, more properly, machine learning algorithms and expert systems—takes this a step further, and is quickly replacing human beings for a wide range of activities from pure mathematical computation to medical diagnostics and even sports writing.¹ These changes inevitably raise questions about any remaining comparative advantages of humans in the future as well as notions of preserving “humanity” and the conduct of “humane” behavior.

Given the imminent ubiquity of AI applications in our lives, reflection on such normative principles in the context of this technology’s implementation should be approached as a moral question. The progressive adoption of AI will affect how we define ourselves and prioritize our values. The ethics of AI should not, however, be considered a *sui generis* field of its own, because it is just one more instrumentality of human action. How human beings design, deploy, utilize, and rely upon AI will be an expression of our will and intentionality, even if mediated by a highly capable and seemingly independent computer processor. We are the masters of, and therefore responsible for, our own creations. The same principles that provide the foundation for both deontological and consequentialist reasoning about human activity must also be applied to AI. Accordingly, the normative priorities assigned to ICTs (including AI software and the entirety of cyberspace) must

Ethics & International Affairs, 33, no. 1 (2019), pp. 3–12.

© 2019 Carnegie Council for Ethics in International Affairs

doi:10.1017/So892679418000928

benefit from interdisciplinary scholarship in law, political economy, and social philosophy.

The very first sentence of the Universal Declaration of Human Rights (UDHR) recognizes the inherent dignity and equality of all human beings as the “foundation of freedom, justice and peace in the world,” and for over seventy years these ideals have guided much of the literature and practice on human rights around the world.² The principles laid out in the UDHR are viewed as fundamental and universal truths. Today, however, we stand on the cusp of a new age of AI in which the very concept of who and what is counted as human may once again be up for debate. In 2014, for example, the Hong Kong venture capital firm Deep Knowledge Ventures appointed a computer algorithm named “Vital” to its board of directors.³ Three years later, Saudi Arabia made history by granting citizenship to a humanoid robot called Sophia.⁴ Though one might dismiss these moves as publicity stunts, they beg for a moment (or more) of reflection. If we are to protect universal human rights, who and what are we protecting? What makes us distinctly human in an age of AI? This essay aims to take such a moment of reflection, examining the human condition in order to understand the far-reaching role of AI for ethical considerations.

ON BEING UNIQUELY HUMAN

Humor

Humor and love are often considered to be uniquely human behavioral characteristics. But while canines and many other animals have been shown to exhibit qualities of fidelity or affection, only primates have been assessed to appreciate humor. The reason is that humor reflects higher cognitive functions that are able to assess information and juxtapose unrelated schema. The dominant psychological theory of humor since the eighteenth century, the so-called “incongruity theory,” asserts that laughter stems from the perception of something incongruous, that is, “something that violates our mental patterns and expectations.”⁵ Modern-day comedians refer to this shift as a “punch line,” a final statement that often diverges radically from what preceded it in the joke’s story line.

Immanuel Kant, Arthur Schopenhauer, Søren Kierkegaard, and many other philosophers and psychologists have subscribed to this theory. In fact, Aristotle’s *Rhetoric* even advised public speakers to get a laugh by creating an expectation in the audience and then violating it.⁶ Schopenhauer provides an

example of this in his writings, quoting an old Austrian joke that proceeds as follows: “You like walking alone; so do I; therefore we can go together.”⁷ In that case, what is a seemingly logical connection between two similarities actually yields an untenable result. Or for purposes of the discussion here, a higher level of cognitive analysis of the proposition is required to detect the incongruity and undesirability of the outcome, and hence to understand the humor.

Other modes of humorous expression also require the juxtaposition of unrelated concepts through language. Puns leverage either multiple meanings of the same word or syntactical or aural similarities between different words. In order to appreciate a pun, one’s mind must instantaneously compare the given words with the entirety of one’s linguistic knowledge set. A music teacher’s puns such as “Handel with care” or “Haydn go seek” would be unamusing, or even incomprehensible, to someone without contextual knowledge of these famous composers.

Irony and sarcasm are even more complicated, as they are achieved through expressions that have literal meanings that are the opposite of the sentiment that is actually being articulated. A significant level of contextual understanding about the individuals involved or previous experience is required to determine that a particular statement is meant to provide ironic humor. A listener, or computer program, without any such background knowledge would simply take the comment at face value and proceed with exactly the wrong intent.

Just as most animals cannot appreciate humor, it has also proven a repeated stumbling block for AI programs. Machine learning algorithms and expert systems are designed to perform specified functions on data sets, but not to consider seemingly unrelated information. It is that critical ability to perform lateral thinking—to identify similarities between seemingly unrelated entities, or to identify distinctions between seemingly related entities—that sets human beings apart.

That cognitive capability is such a part of our essence as human beings that we in fact demand extraneous considerations to be part of our personal, social, and political interactions. The principle of human dignity demands nonlinear reasoning, which in turn is dependent on the acquisition and transmission of vast troves of information on a broad range of disparate topics.

Ethics and Law

Ethical standards convey normative priorities within any social system, and legal rules seek to enforce value judgments upon human interactions. Notions of “justice” and “fairness” are incredibly complex objectives that mandate higher-order

cognition. And in almost every case, we will find that incongruities, extraneous information, and unrelated events can play a critical role in shaping our ethical judgment.

The British social philosopher and ethicist Philippa Foot can be credited with best illustrating how circumstantial knowledge can thwart formulaic reasoning, whether it be deontological or consequentialist. In her famed “trolley problem,” she juxtaposed different personages (and combinations thereof) on a forked railway track and demanded that her reader select down which path the trolley should proceed, thereby killing those assembled on that branch.⁸ The victims could range from innocent children to Adolf Hitler, and vary in number, age, and health through alternate permutations. Possible distinctions between action and inaction can even be offered to challenge deontologists. Foot’s brilliant exposition of the nuanced information that humans hold relevant even for utilitarian calculi underscores why algorithms are so difficult to substitute for human judgment in ethical considerations.

Social choice theory has encountered similar challenges in addressing extraneous information in voting contests while trying to ensure the fairness of democratic processes. Common intuition would suggest that the ranking of preferences for two candidates (or other options) should not be affected by the introduction of additional candidates. In other words, a third entry may be preferred more, less, or somewhere between the original two options, but it should not change the pairwise relationship between the original two. However, Nobel laureate Kenneth Arrow demonstrated in his 1951 book *Social Choice and Individual Values* that no rank-order voting system could satisfy a set of mutually desirable (yet mathematically incompatible) fairness criteria that included such an intuitive requirement.⁹ Arrow’s Impossibility Theorem, as it is now known, shows the cognitive importance of what are termed “irrelevant alternatives” to any decision-making process.

Law is no more immune to the impact of contextual information than moral philosophy or political economy. For example, during the fifteenth and sixteenth centuries a Court of Chancery developed in England to provide equitable remedies that were complementary to the more strict and rigid rules of the common law courts.¹⁰ In the interest of fairness, principles of “equity” were established to adapt to the increasingly complex set of issues faced by the court system.¹¹ The Supreme Court of Judicature Act of 1873 formally merged law and equity in Great Britain, and today common law jurisprudence goes to great lengths to compare factual scenarios with—or distinguish them from—precedential cases so that nuanced details can be taken into account.

Similarly, Islamic law has draconian rules, but also vital exceptions to those rules. Stealing is punishable by amputation of the hand under Shari'a law, although hunger can operate "as a form of necessity relieving the thief of legal responsibility . . . if he was not going to steal absent his hunger."¹² Therefore, simply knowing the nature of the act that took place is not enough; the judge must also have access to contextual information about the parties, their intentions, and possibly external circumstances. One can observe analogous leniency based on equitable considerations during the sentencing phase of criminal proceedings in common and civil law court systems as well.

International humanitarian law, as partially embodied in the Geneva Conventions, further elaborates on the notion of human dignity by adopting a principle of humanity that "forbids the infliction of all suffering, injury or destruction not necessary for achieving the legitimate purpose of a conflict."¹³ Recognition of and respect for the value of each person is always supposed to be taken into account, even in national security decision-making. In order to make sound judgments regarding military necessity and proportionality in a time of war, decision-makers need to be able to process contextual data, such as whether bystanders are civilians or combatants.

THE CASE FOR CONTEXT

The common factor in all of the preceding discussions about humor, moral dilemmas, voting mechanisms, and judicial systems is that human cognition relies upon attention to collateral information that exists beyond any specific situation that may be at hand at any given time. As AI technology now affords a level of data processing that is unprecedented in human history, global society is poised to revisit all the moral and legal quandaries mentioned above. Now, however, software code is being written and machine learning algorithms are being trained that will make those decisions *for* human beings. Nonetheless, the very trait that such programs cannot yet exhibit is the ability to leverage vast amounts of unrelated data from inductive experience that can alter the desired (or expected) outcome of a situation. In short, AI does not yet understand the context within which it is operating.¹⁴ There are clear reasons why no one should want to be judged by a jury of computers or sit through an AI stand-up comedy routine.

What follows are two case studies that will be used to illustrate the importance of maintaining human cognition and values to provide context for ethical

decision-making. We must understand the tensions inherent in the application of this new technology lest humans resign themselves to the equivalent of law without equity or to utilitarianism without the ability to measure and calculate costs.

Case Study 1: Driverless Cars

The transportation sector is being revolutionized by AI, and driverless cars will soon be the rule and not the exception. These vehicles will constantly sense and communicate with each other as well as with roads and other infrastructure. Lawrence Lessig's prescient discussion of "code is law" will come to fruition as the options for automobiles become more and more circumscribed so that they do not exceed the speed limit, violate traffic rules, or otherwise cause accidents.¹⁵ Concerns over safety and systemic efficiency will delineate the range of possible actions.

This leaves the interesting question, among others, of how a driverless car would validate and respond to a medical emergency affecting one of its passengers. Will all driverless cars have manual override switches so that a heart attack victim or a pregnant woman entering labor can get to the hospital in time? That is, how will the driverless cars of the future collect, process, and account for highly relevant contextual information that is unrelated to their programmed tasks?

Next, what will happen when there is a mechanical failure or natural hazard that cannot be safely predicted and avoided? Who is programming the driverless car for how to react in the modern analogues of Philippa Foot's trolley problem? If a young boy runs into the street after an errant ball, will the car choose to strike him, swerve off a cliff, or drive head-on into traffic in the adjoining lane. For argument's sake, we can presume that the boy, the car's own passenger(s), or the passenger(s) in the oncoming vehicle would be killed in each respective case. That raises a plethora of disturbing ethical considerations and dystopian scenarios that are not just futuristic hypotheticals in the classroom; in 2016, Mercedes-Benz announced that its driverless vehicles will prioritize the safety of their occupants over pedestrians.¹⁶

How should the software programmer, who likely comes from a different cultural background than the boy and passenger(s) of either car, value the lives of different individuals she will never know? Would one want to process the exact number and identity of the persons involved before reaching such an important moral decision? Would one therefore want driverless cars to broadcast the identity of their passenger(s) to nearby cars in case one of them needed to make such a critical determination? Biometric security devices on mobile telephones or in

the cars themselves would certainly make that a possibility. What if individuals living in countries with “social credit” scores (an increasing reality) had their lives prioritized based on how well they have conformed to communal political views?

It could become even more disturbing if there were no standardized regulations for the decisions that driverless cars were supposed to make under specific circumstances. How many individuals would prefer to purchase or ride in a brand of vehicle that was known to prioritize the lives of other people over the safety of its own passengers? And even if regulations were enacted, what would prevent a car manufacturer from misrepresenting its software code, similar to how Volkswagen willfully used computer chips that cheated under emission testing conditions?¹⁷ Many of these specific moral dilemmas may remain purely academic, however, for the future of road transport will likely evolve to resemble rail transport, where the network has the right-of-way over trespassers. After all, no one expects a train to jump the rails to avoid a person in its path. Nevertheless, the questions raised here will be broadly applicable to AI-human interactions.

Case Study 2: Lethal Autonomous Weapons

International humanitarian law (IHL) is formulated by national governments and interpreted by esteemed scholars, but it is often implemented (or violated, as the case may be) by military personnel with little or no specialized legal training. They rely on instructions regarding the appropriate rules of engagement, but are also expected to exercise their own judgment to reject orders from superiors that would contravene international law. Respect for the value of human life and the preservation of human dignity mandate that every soldier considers plausible exceptions to every rule or instruction he or she receives.

Posted signs or audible warnings may threaten the use of lethal force against persons approaching too close to military installations in a combat zone, but one would certainly hope that individuals appearing to be blind or deaf would not be strictly held to that standard. Likewise, young children may not comprehend the apparent danger, and should therefore be spared as well. These simplistic examples illustrate the challenges associated with codifying rules of behavior into automated systems. Just as with driverless cars, the trolley problem, or common law, adherence to predetermined algorithms for decision-making may not properly account for the particular nuances of any given situation.

The technological progress toward lethal autonomous weapon systems (LAWS) is bringing new focus to this issue. LAWS are now the subject of regular meetings of the United Nations Group of Governmental Experts, and much consideration is being given to keeping “humans in the loop” for any life-or-death decisions.¹⁸ The perceived problem with LAWS is that, in the interest of expediency, certain functions are being delegated to systems that cannot think laterally or accommodate the kind of irrelevant alternatives that lead to Arrow’s Impossibility Theorem. Perhaps the best historical example of the need to acquire and use contextual information came on September 26, 1983, when Stanislav Petrov adjudged the automated reports of a nuclear missile launch from the United States to be erroneous. Petrov is often credited with averting a Soviet retaliatory strike, earning the moniker “the man who saved the world” for using his own intuition during the potential crisis.

But what information did Petrov rely upon to make his judgment? How did he come to believe that his missile warning system was producing a false positive? In interviews after the fact, Petrov recalled that he had repeatedly been told in trainings that any first strike by the United States would be massive, designed to knock out all Soviet capabilities with one blow. The system was only showing five missiles, which seemed to him an illogical move for the Americans. There was also no corroboration from the Soviet ground radar. Thus, as it turned out, his rationale was predominantly based on contextual information that only a human would consider, along with a lack of any positive feedback from other systems.¹⁹

In order to comply with the IHL principles of necessity, distinction, and proportionality, military leaders need to know about much more than just a potential target. And any LAWS or strategic AI proxy for those human decision-makers would need the same access and ability to process contextual information. In the civilian context, Internet of Things devices will have ever-increasing demands for data to make the most efficient allocation of resources. The more cyberspace becomes an automated environment and the more we allow AI into our lives, the more pressing these ethical questions will become.

CONCLUSION

The higher human faculties that distinguish us from other life forms and man-made computational machines all derive from, and are dependent on, the ability to process incongruous information. These competencies enable humor and

ethically preserve human dignity. They also engender the visceral aversion to “fail-safe” cyber weapons and driverless cars that will immutably perform preordained tasks with higher fidelity than any human could. We want Petrov to be able to violate his assigned protocols because it may end up being our own human dignity that he protects. Even William Blackstone’s famous deontic adage that “it is better that ten guilty persons escape, than that one innocent suffer” favors false negatives over false positives.²⁰ The principles of dignity and humanity demand as much room for context and nuance as possible before making substantial decisions, and especially irreversible ones.

Humanity must not forego its most essential and unique qualities in order to conform to the limitations of its own inventions. There is a choice to be made, then, about whether we impose limitations on these new technologies in favor of maintaining human control or whether we seek to replicate ethical reasoning and lateral thinking in the machines we create. The answer will have profound effects not only on how we interact with AI but also on how we interact with one another.

NOTES

- ¹ See, for example, “AI, Radiology and the Future of Work,” *Economist*, June 7, 2018, www.economist.com/leaders/2018/06/07/ai-radiology-and-the-future-of-work; and Stacey Liberatore, “Your Days Could Be Numbered If You’re a Sports Writer: The Associated Press Is Using AI to Write Minor League Baseball Articles,” *Daily Mail*, June 30, 2016, www.dailymail.co.uk/sciencetech/article-3668837/Your-days-numbered-sports-writer-Associated-Press-using-AI-write-Minor-League-Baseball-articles.html.
- ² UN General Assembly, “Universal Declaration of Human Rights,” Res. 217 (III) A, Preamble, December 10, 1948, www.un.org/en/universal-declaration-human-rights/.
- ³ See “Algorithm Appointed Board Director,” *BBC News*, May 16, 2014, www.bbc.com/news/technology-27426942; and Sophie Brown, “Could Computers Take Over the Boardroom?,” *CNN Business*, October 1, 2014, www.cnn.com/2014/09/30/business/computers-ceo-boardroom-robot-boss/index.html.
- ⁴ See Zara Stone, “Everything You Need to Know about Sophia, the World’s First Robot Citizen,” *Forbes*, November 7, 2017, www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/.
- ⁵ John Morreall, “Philosophy of Humor,” in *Stanford Encyclopedia of Philosophy*, revised September 28, 2016, Section 4, plato.stanford.edu/entries/humor/.
- ⁶ *Ibid.*
- ⁷ *Ibid.*
- ⁸ See generally Judith Jarvis Thomson, “The Trolley Problem,” *Yale Law Journal* 94, no. 6 (1985), p. 1395.
- ⁹ See generally Kenneth J. Arrow, *Social Choice and Individual Values*, 3rd edition (New Haven: Yale University Press, 2012).
- ¹⁰ See “Chancery Division,” *Encyclopedia Britannica*, revised October 19, 2018, www.britannica.com/topic/Court-of-Chancery.
- ¹¹ *Ibid.*
- ¹² David F. Forte, “Islamic Law and the Crime of Theft: An Introduction,” *Cleveland State Law Review* 34 (1985–1986), p. 47, engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1000&context=fac_articles.
- ¹³ International Committee of the Red Cross, “What is IHL?” September 18, 2015, www.icrc.org/en/document/what-ihl.

- ¹⁴ For example, IBM's "Watson" supercomputer does not know whether it is playing a televised game show like *Jeopardy* or solving corporate problems. It lacks the situational awareness of a human being.
- ¹⁵ See generally Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999).
- ¹⁶ See Michael Taylor, "Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians," *Car and Driver*, October 7, 2016, www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/.
- ¹⁷ See, for example, United States Department of Justice, "Former CEO of Volkswagen AG Charged with Conspiracy and Wire Fraud in Diesel Emissions Scandal," May 3, 2018, www.justice.gov/opa/pr/former-ceo-volkswagen-ag-charged-conspiracy-and-wire-fraud-diesel-emissions-scandal.
- ¹⁸ See UN Web TV, "Second 2018 Meeting of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems," August 27, 2018, webtv.un.org/meetings-events/treaty-bodies/watch/second-2018-meeting-of-the-group-of-governmental-experts-on-emerging-technologies-in-the-area-of-lethal-autonomous-weapons-systems-/5827311154001/?term=&lan=original.
- ¹⁹ David Hoffman, "I Had A Funny Feeling in My Gut," *Washington Post*, February 10, 1999, www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm.
- ²⁰ See Harvard Law School Library, "Words of Justice: Roof Garden Wall – Right Panel," library.law.harvard.edu/justicequotes/explore-the-room/south-4/.

Abstract: The growing adoption of artificial intelligence (AI) raises questions about what comparative advantage, if any, human beings will have over machines in the future. This essay explores what it means to be human and how those unique characteristics relate to the digital age. Humor and ethics both rely upon higher-level cognition that accounts for unstructured and unrelated data. That capability is also vital to decision-making processes—such as jurisprudence and voting systems. Since machine learning algorithms lack the ability to understand context or nuance, reliance on them could lead to undesired results for society. By way of example, two case studies are used to illustrate the legal and moral considerations regarding the software algorithms used by driverless cars and lethal autonomous weapons systems. Social values must be encoded or introduced into training data sets if AI applications are to be expected to produce results similar to a "human in the loop." There is a choice to be made, then, about whether we impose limitations on these new technologies in favor of maintaining human control, or whether we seek to replicate ethical reasoning and lateral thinking in the systems we create. The answer will have profound effects not only on how we interact with AI but also on how we interact with one another and perceive ourselves.

Keywords: humor, ethics, dignity, human rights, artificial intelligence, Philippa Foot, law, philosophy, driverless cars, lethal autonomous weapons

Copyright © Carnegie Council for Ethics in International Affairs
2019