Machine Ethics

At its core, a computer executes instructions, the code we build into it. Therefore, one can argue that computers do not have their own discretion beyond what humans incorporate into such systems. Intentional actions of any code are essentially limited only to the extent its writer chooses. In this mode, computers can be described as tools, or enablers, of what the users want to do. The entire accountability for ethical conduct rests with its creators.

This is not to say that humans do not use these enablers in an immoral way. During the 1970s, the leadership of Equity Funding, a US life insurance company, decided to create fake records of life policies, reinsure them and obtain cash from reinsurers by declaring some of these fake insured entities as dead. They hid these data from their auditors in what they called “File 99.” Manipulating the code or data to commit a criminal or immoral act is possible even when computers are no more than tools and the perpetrators come from the line of users or creators.

But, an era of advanced computing in which computers are increasingly taking over far more sophisticated roles has arrived and continues to expand. Robotics and nanotechnology are just two examples of developing disciplines that will push the role of computers and computing well past the era of enablers. According to Ray Kurzweil, by the year 2045, “human intelligence will enhance a billion-fold thanks to high-tech brain extensions.” He refers to this phenomenon as the “singularity,” a point at which humans and computers will merge. This sort of “one in two” will create serious challenges in the allocation of moral accountability between the two.

To develop insights into ethical dilemmas of the new world of advanced technologies and their applications, a whole new field, called moral machines or machine ethics, is emerging. Broadly, machine ethics is a discipline that attempts to address the ethics of artificial intelligence (AI). While AI has slowly been transitioning from fiction and movies (e.g., Transformers) to the real world over the past several decades, attempts to articulate its moral dimensions are relatively recent. Moral Machines: Teaching Robots Right From Wrong and Machine Ethics are two significant publications offering a discussion of morality in the context of smart machines. This leap from computer ethics to machine ethics is necessary due to the elevated status of computers from mere enablers to intelligent collaborators with humans.

James Moor considers computing machines that are basically enablers of tasks as normative agents, but not necessarily ethical agents, because they merely perform the tasks as specified and their performance can be objectively assessed. Any development of machines beyond this state requires consideration of the ethical dimension that the embedded intelligence should reflect in its design. For this, Moor suggests three ways to classify issues of moral values in machines: ethical impact agents, implicit ethical agents and explicit ethical agents. Each category progressively assigns a greater moral role to machines. Here, we discuss machine ethics using Moor’s classification.

**ETHICAL IMPACT AGENTS**

In addition to executing instructions, machines sometimes subtly influence the moral dimension of the situation, thus operating as an agent that delivers ethical impact. For example, now that the use of mobile phones is pervasive, we are concerned about whether to take a call on the cell phone while driving. There is enough evidence that the use of cellphones while driving entails a greater chance of harm to oneself and others on the road. In places where the law does not require action, one needs to consciously determine whether to take the call or let the answering system handle it while driving a vehicle (or any moving equipment you are actively controlling; exception: driverless cars). Of course, the latter option alleviates the risk while the former may actualize it. The dilemma is caused entirely by the technological innovation in the form of cellphones and wireless telecommunication. Improvement in communication—anytime, anywhere—results in
a side effect in terms of a new ethical dilemma. Thus, mobile computing can be seen as an agent with ethical impact; it does what it is programmed to do, but it also has an impact on moral dimensions of human behavior.

Jim Lewis describes an interesting scenario of ethical impact. It has to do with camel jockey robots, about two feet high, with a right hand to bear the whip and a left hand to pull the reins. Hardware is encased in aluminum and plastic (35 lbs.) and software: a 400-MHz processor running Linux and communicating at 2.4 GHz; GPS-enabled, monitoring the camel’s heart rate. Camel jockeys were used in a desert camel race outside Doha, Qatar. Until the robots displaced human jockeys, the natives of Doha used enslaved Sudanese boys and starved them to keep them lean. Every robot camel jockey bopping along on its improbable mount means one Sudanese boy freed from slavery. The robots built in Switzerland had an ethical impact outside of just running the camel race.

Interestingly in the case of mobile phones, the ethical impact resides outside of the design of the cellphone; thus, one might argue that the phone itself has not been delegated any moral responsibilities. Nevertheless, cellphone technology has the potential to transform mobile phones into ethical impact agents—they can be programmed to make the decision given the owner’s state, whether the owner is driving or is not driving a vehicle at the time of a call. This may not become widely popular among the users and salespeople. Thus, it is clear that evolving technologies offer new ways of addressing ethical dilemmas as they improve ways to satisfy human needs and wants. For example, the global presence of the Internet, supposedly neutral in its design, has created a new debate on privacy and equal access. In 2012, Facebook conducted a psychological experiment on 689,003 users to gauge their emotional state. The purpose was to determine if Facebook could alter the emotional state of its users and prompt them to post either more positive or negative content. We did not invest as much time on these issues in the pre-Internet era.

Let us return for a moment to the computers that perform normative tasks; there is nothing inherent in their design that would touch our conscience. Looking to the future, as we become surrounded by intelligent machines, responses to our moral dilemmas take new shapes and often become even more complex. In the past, we never thought we would have any ethical issues in answering a phone call, but with wireless technology, we now do.

**IMPLIED ETHICAL AGENTS**

Ethical impact agents create ethical contexts that are new or different for humans to solve. Since AI programs are sophisticated in the level of intelligence, why not have them sort out the ethical dilemma they pose? Thus, we adjust the situation from one of ethical impact to ethical decision making by the computer, albeit from a portfolio of choices that programmers embed into the software. Instead of opting to sort out ethical questions emerging from ethical impact agents, an option is to build into the AI products and services anticipated optimal responses to ethical questions. If moral choices are programmed with proper reflection on what is appropriate under specified conditions, one would expect the anticipated behavior to occur consistently. Unlike humans, machines are not capable of breaking their resolve to behave ethically.

**EXPLICIT ETHICAL AGENTS**

The next stage recognizes the agent as not only able to execute moral decisions, but also to make ethical decisions based on a repertoire of decision rules. Thus, one need not assign the right thing to do, but rather how to determine what is the right thing to do in a specific context. Take, for example, drones (unmanned combat air vehicles) used in combat.

One would want the drone to mitigate the risk by destroying the threat, while at the same time noticing innocent civilians nearby whose lives should be spared. If a drone is able to make such judgments based on a specific scenario that develops on the spot, it would be an explicit ethical agent.

Explicit ethical agents are equal, or nearly equal, to humans. Some even argue that they could be much more reliable and consistent in their moral behavior.

While these may seem rather premature ideas, the truth is this: AI applications are here to stay. Newer applications of AI are more sophisticated and some even self-learning. The complexity of the new world driven by technology is bewildering, although ethical precepts of the past still apply. The challenge of modern times lies in reliable comprehension of the new contexts in light of the eternality of ethics.
we do not have a comparable context from the past, the application of ethical precepts becomes confusing, much like when lawmakers try to craft a bill for something like net neutrality, which is new and nearly incomparable to existing law. Universal open access to Internet without regard to the media, devices, platforms or nature of use is hard to pull from history. No matter how the debate on net neutrality is resolved, it will create more ethical issues to address as we move forward. And as the gap between humans and machines closes over time, we will have to give serious consideration to new ways of implementing ethics.

ENDNOTES
2 Wallach, Wendell; Colin Allen; Moral Machines: Teaching Robots Right From Wrong, Oxford University Press, UK, 2008
3 Anderson, Michael; Susan Leigh Anderson (Eds.); Machine Ethics, Cambridge University Press, UK, 2011
4 Ibid., p. 13-20
5 Lewis, Jim; “Robots of Arabia,” Wired Magazine, November 2005
6 Albergotti, Reed; “Furor Erupts Over Facebook Study,” The Wall Street Journal, 30 June 2014