# Killer Apps and Killer Robots

## An Ethical Framework for Answering Questions of Automation

Peter McKay, Nisansa de Silva, Elizabeth Fuller, Bradley Green, Gautam Sondur
Information and Computer science Department
University of Oregon
CIS 590 - Winter 2016

*Abstract*— **In the following paper, we carry out an examination of two related ethical quandaries concerning the increasing degree [1] of automation in the modern world. First, we consider the ethics of existing systems, exploring the case in which automated systems make choices apparently on their own, and what kind of moral reasoning we face in such situations. We consider the idea that systems may act as moral agents, or as proxy agents [2] for the humans involved in their design and deployment. Secondly, we examine a somewhat more fundamental issue: is it right to automate these tasks in the first place? Rather than trying to achieve answers both general and correct, we will instead lay out a framework which may be put to use in the pursuit of answerss for specific questions.**

## I. Introduction

## II. Background

Later on in our paper, we will put forth an ethical framework for use in answering questions pertaining to automation and related tecchnologies. As an important part of that framework, we aim to analyze who may be affected by changes in the state of the art of technology, and how those effects may ripple out to change society at large. While these analyses would be much simpler were it not the case, the fact of the matter is that technological innovation is part and parcel of human society. Advances in technological innovaton cannot be extricated from the social conditions which gave rise to such innovation, nor can they be separated from the the social consequences which result from the adoption of that technology. If we wish to base our analyses in historical fact, it seems prudent to first review some basic history with an eye towards the factors we intend to consider. If we wish to see a preview of the impacts that automation will have on the workforce, and consequently the society at large, it seems prudent to look at the ways automation shaped society as its effects on the economy were first made manifest.

Let us consider, for a moment, the steam hammer. The steam hammer, a powered hammer for forging and stamping of metal parts, was invented in the mid-1800s as increasingly large and complex machines necessitated the fabrication of larger and heavier components. It made routine the construction of many works that, only years before, would have been considered marvels of engineering and concerted human effort. It contributed to the construction of great ships, the production of delicate clockwork, and a massive increase in the industrial productivity of Western Europe that saw an increase in the availability of cheaply manufactured goods formerly restricted to only the richest members of society. It also contributed [3] to a sharp decrease in the leverage that skilled engineers held over their employers, resulting in drastic consequences for working environments, and sharp increases in child-labor. James Nasmyth, one of those credited with the invention of the steam hammer, made the following statement regarding the power of automation in the modern (as of 1851) factory:

> The characteristic feature of our modern mechanical improvements is the introduction of self-acting machinery. What every mechanical workman has now to do, and what every boy can do, is not to work himself, but to superintend the beautiful labour of the machine. The whole class of workmen that depend exclusively upon their skill is now done away with. Formerly, I employed four boys to every machine. Thanks to these new mechanical combinations, I have reduced the number of grown-up men from 1,500 to 759. The result was a considerable increase in my profits. [4]

From the specific, we proceed to the general. The example of the steam hammer (in particular, the contrast between its positive and negative social effects) raises one of the first difficulties we will be forced to surmount in the course of developing our ethical framework. In the case of the industrialization of the workforce of Western Europe, those who wished to deploy greater degrees of automation in their factories were directly at odds with those who work in those factories. This is reflective of the effects of automation, not just on the market for the product they create, but also on the labor market. Labor, like any other commodity, has a price, which fluctuates based on supply and demand. When the owners of the factory can produce just as much value from less work, the demand for that labor goes down, and its price drops. Those workers have less leverage, and, as observed by Gaskell, their working conditions suffered.

When balancing the gains in technology and standards of living that have resulted from this march of progress, we see that there are two sides of this coin we must consider. The society as a whole may benefit, but what of the workers that toil in the factory? Does the gain outway the harm? Fortunately, in this particular case, we are not forced to make such a decision: social and cultural pressures intervened. The labor movement, in an internationally coordinated effort, slowly but surely pushed many governments into enacting legislation that protected workers from unfair practices on the part of their employers.

And yet, we are faced with a somewhat tougher questions, when we step back and consider the broader context around the industrialization of Western Europe. Industrialization is a somewhat titanic effort, it requires the investment of massive amount of capital to begin to see the advantages resulting from leveraging the economies of scale that make automation profitable. Different societies have achieved such a concentration of wealth in different ways. China, in the course of The Great Leap Forward, began industrialization on a massive scale by nationalizing all production, taking central command of labor and distributing it according to a grand plan for industrialization of the workforce as a whole. The Soviet Union took a similar approach, nationalizing industry and placing it under central command, but then decentralized the industry geographically, exporting raw materials to nearby states as it absorbed them, thus creating a demand for the growth of industry in those formerly independent states that drove their growth further outward until they started bumping up against NATO countries. Western European countries accumulated wealth in privately held trading companies by establishing an early foothold in mercantile capitalism through the Triangle Trade. That is, they establishing trade routes between Europe, Africa, and the Americas, alternating in the shipment of slaves, raw materials (such as sugar cane grown by slaves), and manufactured goods (such as rum produced from last year's sugar cane).

In each of these situations, there exists a contradiction between the forces acquiring enough wealth to assemble an industrial base, and the people that had to give up that wealth. In China and the Soviet Union, the costs were paid by the citizens of those countries, and they reaped the benefits slowly as their standards of living increased fairly evenly. In Europe, the standards of living and availability of luxury goods skyrocketed much more rapidly: after all, the cost of automaton was offloaded to somewhere else: Africa and the American colonies. In the Soviet Union and China, the analysis looks simple: did the country gain more than it lost? China suffered a number of famines as the Great Leap Forward encountered organizational difficulties, but over a period of 30 years, the average lifespan of a Chinese citizen nearly doubled.

When, however, we look at a situation like British industrialization, we now find that we must take looks at a much more diverse selection of groups in analyzing the impact of technology on the society. As globalization has blurred the lines between nations, so too has information technology led to a vastly more interconnected world than we have ever seen. A change in technology today can lead to nearly instantaneous changes through enormous swaths of society. In the immediately following sections, we will further examine the current

state of the world as it has evolved, the approaches that some groups take in analyzing that world, and the steps they deem necessary due to such an analysis.

## III. The Current Scenario

## IV. A Review of Present Solutions

### A. The Hope of the Singularity

### B. Bill Joy

### C. Jason Lanier

### D. The Amish

### E. Deep Green Resistance

Deep Green Resistance is a group primarily identified by their beliefs and behavior with regard to specifically environmental matters, rather than more general opinions about technology. Specifically, DGR subscribes to the notion that industrial civilization poses too great of a threat to life on the planet earth, both human and otherwise, to be allowed to survive. The ultimate goal of DGR is the destruction of industrial civilization and a return to an earlier stage of societal and economic development.

DGR's philosophy draws from the Deep Ecology movement, which holds that an anthropocentric analysis of ecology, defining other forms of life in terms of their utility to humankind, does not do an adequate job of describing the complexity of ecological systems. DGR takes this a step further, in declaring all life to be equal to human life. Starting at the problem of inequality among life, they turn to ideals similar to the political stance of anarcho-primitivism: the idea that human society, industrial capitalism in particular, cannot be reformed into a more beneficial form.

Such a philosophy stems from an agreement with Friedrech Engal's seminal anthropological work on early civilizations [5], specifically the assertion that early gatherer-hunter societies were unable to produce structural inequality without the ability to acquire surplus wealth. Unlike Engals, modern Marxists, and even their fellow anarchists, DGR and anarcho-primitivists do not believe that a society can ever reach a level of egalitarianism and freedom from oppression that they are comfortable with. Rather, they believes that the only truly equal societies that have ever existed were only able to function without the trappings of civilization. Therefore, they are willing to advocate any means to right the wrongs they see with the world. To them, technology is not evil per se. Rather, technology is another product of a system with which they cannot make peace, and as such, it must fall by the wayside.

Later in our paper, we advance an ethical framework for the analysis of decisions regarding technology and automation. In so doing, we emphasize that we must look to the past when examining possible consequences of an action, basing, wherever we can, our reasoning on historical fact. In this case, we would be remiss if we did not qualify our description of the views of DGR with the caveat that their views ultimately reduce to something chillingly familiar. A return to pre-industrial levels of technology would necessitate either the death of billions, or the restriction of new births to a fraction of their current levels. Without modern medical technology, it becomes a disturbing fact that disabled people will not be able to survive in such a world: this means DGR endorses either genocide or eugenics.

## V. A Coherent Ethical Framework

### A. Identifying Stakeholders

### B. General Classes of Stakeholders

### C. Defining a Coherent Value System

### D. Synthesis

## VI. Self-Driving Cars: an Application of our Framework

### A. Background

Long a fascination for inventors, the seed for autonomous cars was sewn in the 1960s with the Stanford Cart [6]. The cart navigated small spaces by taking photos of its surroundings and using a simple computer program to analyze the photos and choose a path of travel. Two decades later, a vision-guided Mercedes-Benz van designed by Ernst Dickmanns navigated empty streets at speeds

of up to 39mph. By 2009, Google began privately testing anonymous vehicles on closed streets, and by 2012, changes in various states' laws allowed for testing on city streets. As of 2016, Google claims over 1.3 million autonomous miles have been driven, although Google's cars are limited to 25mph and must always have a human driver present. [7] As driverless car technology improves and becomes more widely implemented, the true societal impact of this technology will be felt.

### *B. Application*

We're in an interesting time in regard to autonomous vehicle technology. It's clear that the technology will become widely used, but it has yet to actually happen. This presents an opportunity to consider how, as a society, we should approach the ethical ramifications of this potentially society-altering technology. To do so, we'll apply the ethical framework laid forth earlier in this paper to autonomous cars.

#### three main stakeholders: public, government, corporations

To begin, it is important to identify the three main stakeholders in driverless car technology: the public, the government, and corporations. The stakeholders are universal, but each stakeholder's value systems and the weight given to each stakeholder can vary based on cultural norms and/or economic status. In the interest of concision, this application will focus on a generalized idea of American culture, values, and economic status.

Because the United States is such a demographically, economically, and culturally diverse country, it is difficult to make generalizations of the country as a whole. That being said, it is possible to extrapolate mainstream value systems of its public, government, and corporations. The public values individualism, corporations value growth and the acquisition of capital/power, and the government acts as a mediator. Each stakeholder has the power to act on the other and force change. With the stakeholders, value systems, and power structure identified, it is possible to apply our ethical framework to autonomous vehicle technology in the United States.

#### Do self-driving cars satisfy a utilitarian?

Autonomous vehicles have a lot to offer a utilitarian. A recent Virginia Tech study [8] shows that the crash rate of Google's self-driving cars is lower than the national crash rate of conventional cars. Google's cars experience 3.2 crashes per million miles while conventional cars experience 4.2 crashes per million miles. With autonomous vehicle technology continually improving, the technology will become even safer in the years to come. But is a full embrace of the technology the best course of action? For a utilitarian, this may be an easy answer, and therein lies a problem: autonomous vehicles face a variety of moral dilemmas, and a one-size-fits all ethical framework doesn't cover all of these possibilities.

#### Example 1:

Consider the following example: during the November 2015 terror attacks in Paris, ride-sharing app Uber's algorithmic design dramatically raised prices in the areas near the attacks due to an increase in ride requests from people trying to flee the chaos. In this situation, the public and the corporation were in direct opposition to each other. Uber saw the situation as a moneymaking opportunity, while the public saw Uber as price gouging and profiting from a disaster. Eventually, Uber gave in to public demand and lowered prices. Situations such as this will become the norm in the age of autonomous cars. In fact, Uber is aggressively seeking driverless car technology for use in their taxi company.

#### How would our framework handle this scenario?

The moral framework presented in this paper allows for exceptions that strict utilitarianism does not. In the case of Uber raising the prices on people fleeing a disaster, utilitarianism supports profit maximization as a means to the greatest societal good. Conversely, our framework allows an exception on the basis of protecting an individual's welfare and well-being.

#### Example 2:

Consider a hypothetical example: you're riding down the highway in your autonomous car and a large object falls off of a truck in front of you. Your

car cannot stop itself in time and must decide on the best course of action. If it continues straight and runs into the object, it places your life in serious jeopardy. There are two motorcyclists on either side of your car; one is wearing a helmet and the other is not. Swerving into one of the motorcyclists will likely save your life, but jeopardize the life of one of the motorcyclists.

With strict utilitarianism, your car would likely swerve into the motorcyclist wearing the helmet because it creates the best chance of survival amongst all of the possible scenarios, but is it fair to punish a motorcyclist for being responsible and wearing a helmet? What are the legal ramifications of such a programmed decision? In such a scenario, our ethical framework would fall back onto the autonomy of the individual to decide how their car is programmed to react in cases such as these.

Conclude that using our framework makes self-driving cars more morally acceptable.

While far from foolproof, our ethical framework allows for important exceptions to utilitarianism in extreme cases. The implementation of driverless cars will undoubtedly lead to many such extreme cases, therefore the existence of âĂIJfallbackâĂİ rules will be critical. In the previous examples, the âĂIJfallback rulesâĂİ of our ethical framework placed health and well-being above financial profit as well as the autonomy of a human over the autonomy of a machine. These are only two examples of the types of moral dilemmas that will become ever-present in our society, and it is critical that our society agree on an even more extensive ethical framework to confront these issues.

## VII. Conclusion

## References


[1] Carl Benedikt Frey and Michael A Osborne, "The future of employment: how susceptible are jobs to computerisation," September 2013.

[2] Deborah G Johnson and Thomas M Powers, "Computers as surrogate agents," *Information technology and moral philosophy*, pp. 251–269, 2008.

[3] Peter Gaskell, *The manufacturing population of England, its moral, social and physical conditions*, Baldwin & Cradock, 1833.

[4] Karl Marx, "The strife between workman and machine," in *Capital*, vol. 1, chapter 15. 1867.

[5] Friedrich Engels and Tristram Hunt, *The Origin of the Family, Private Property and the State*, Penguin UK, 2010.

[6] Hans P Moravec, *The Stanford cart and the CMU rover*, Springer, 1990.

[7] Alex Davies, "GoogleâĂŹs self-driving car caused its first crash," *Wired*, 2016.

[8] Myra Blanco, Jon Atwood, Sheldon Russell, Tammy Trimble, Julie McClafferty, and Miguel Perez, "Automated vehicle crash rate comparison using naturalistic data," Tech. Rep., Virginia Tech Transportation Institute, 2016.