Socio-Economic Considerations in the Rush to Embrace AI

An Ethics-Based Perspective

Blake White
Founder, Strategic Technology Institute

&

Gerald Harris
President and Managing Director, Quantum Planning Group, Inc.

Preliminary

Contents

Introduction

This paper calls for a significant re-evaluation of the ethics of distribution and the professional responsibility of engineers and scientists in the context of the risks and rewards presented by innovative applications of Artificial Intelligence (AI) technologies. The technology industry is selling AI as a solution to business problems to an increasing number companies who each intends to exercise the maximalist uses of AI technology. Recent literature looks at complementary uses of technology that augment human decisions vs those that replace human activity on tasks once considered uniquely human. Some explore the hidden costs of granting software control over our work and leisure, and examine the limits of AI delegated to human decision-making in critical situations, like autopilots in aircraft, driverless cars, and chemical plants. What none of these adequately address are the moral issues and ethical obligations of safety implications and the socio-economic considerations in a society that ties money to a job that may be increasingly threatened by AI.

We identify ethical standards that are already well known, and promoted by professional societies, and applies them to the emerging world of AI. Using the ethics principles of John Rawls and established product liability concepts, we seek to answer the big question as to whether there is a better solution for businesses and society as a whole other than the standard economic innovation argument. As Rawls would have it, designers and developers of AI technology must proactively seek to maximize the most benefits for the largest number of people, while delivering the most benefit to those most negatively impacted, or likely to be negatively impacted, by the unintentional consequences of complex technology. This renewed ethical imperative would lead to technological research and product designs for the most positive consequences, rather than settling on the current approach of minimizing the maximum regret.

Finally, to the degree that AI contributes to socio-economic dislocation and/or increased safety risks, its societal costs and remedies need to be reassessed. Redress, prevention, and prohibition might leverage:

- Civil and/or Criminal Penalties
- Product Liability Law
    - Design Defect
    - Failure to Warn
    - Manufacturing Defects
    - Design Defect
- Targeted Taxation
- Increased Regulation

Stakeholder Perspectives

There is a push (based on trying to make money) that is selling AI as a solution to business problems (improving productivity, increasing efficiency, cutting cost, adding new features for competitive advantage). Paraphrasing PwC's definition of AI, let us agree that AI is a collective term for computer systems that can sense their environment, think, learn, and take action toward their objectives in response to what they are sensing. PwC found that 22% of executives surveyed believe that AI will be the most disruptive technology to their business model, and 30% believe it will be the most disruptive to their industry, in the next 5 years.

The AI market is growing quickly, attracting more than $3 billion in venture funding in 2016, according to a PwC/CB Insights MoneyTree™ Report. PwC's 2017 Global Digital IQ Survey notes that the number of companies that are investing or have plans to invest in AI is second only to the number of companies that are investing in the IoT. Today, 54% of the companies surveyed are investing in AI, while 63% say they plan to do so in three years. "North American and Western European companies lead the way in pioneering and adopting AI technologies, as do insurers, entertainment and media companies, and healthcare payers. In three years, information and communications firms, asset managers, capital markets, hospitality and leisure, and professional services firms expect to throw their hats into the AI ring," according to PwC.

In the 2017 book, *When Machines Do Everything: How to Get Ahead in a World of AI, Algorithms, Bots and Big Data,* Cognizant's Malcolm Frank, Paul Roehrig, and Ben Pring, observe that Artificial intelligence has left the laboratory (and the movie lot) and is in your building. It's in your home. It's in your office. From Alexa to Nest to Siri to Uber to Waze, we are surrounded by smart machines running on incredibly powerful and self-learning software platforms. Google's autonomous cars have logged thousands of miles on American highways. IBM's Watson soundly beat the best human Jeopardy! players. Frank, Roehrig, and Pring address the inevitability of AI-enabled automation, and its elimination of certain jobs. But, they also stress that AI can be the next level of productivity for business and they argue for embracing the innovative uses of AI and big data for business advantage and competitiveness. "AI should stand for augmented—not artificial—intelligence," notes David Kenny, IBM's senior vice president for Watson and the company's cloud platform.

If it is true that every society determines reality, truth, beauty, and values in accordance with its own worldview and its unique historic path, then the common view among casual observers of Silicon Valley is that its extreme technology emersion can lead to a one-dimensional perspective of the world and its problems. According to San Jose State anthropologist, Jan English-Lueck, in Silicon Valley, people transfer engineering and entrepreneurial approaches to their understanding of the social world, such that efficiency, utility, instrumentality, and economic rationality become the philosophical underpinnings of the Silicon Valley worldview (English-Lueck 74-77). She notes that, "In Silicon Valley, people view the daily conflicts of life as 'social engineering problems' that can be 'solved' if given thoughtful and systematic appraisal" (English-Lueck 76). As such, critics of Silicon Valley note that, while it is clear that technology has the power to enhance lives, it is not always as clear to the developers and consumers of high technology products that the same beneficial technology might also lead to an oversimplified public discourse of social problems.

Recently, broadcast and print news, technology journals, and social media are filled with warnings about Big Data and AI. Julia Bossmann, President of the Foresight Institute, writing for The World Economic Forum, lists "*Top 9 Ethical Issues in Artificial Intelligence"* in a 2016 publication. Several 2016-2017 stories in CNN, ExtremeTech, WIRED, and the Daily Mail raised concerns about consumer tracking and data privacy of Google Play, Google Home, Amazon's Alexa, and Bose headphones. CBS ran a story about government concerns over "spying billboards" using phone data to track shoppers. TechCrunch reported that AI data-monopoly risks are to be probed by UK parliamentarians. In July 2017, The Wall Street Journal examined the monopoly power of the tech giants to go beyond disruption of the creative economy to target the service economy next. Dustin McKissen, writing for LinkedIn, comments on the disruption of the workforce in *"My Father-In-Law Won't Become a Coder, No Matter What Economists Say"* and The Register reports that Europe mulls treating robots legally as people. Even Scientific American asked in a February 2017 article, "Will Democracy Survive Big Data and Artificial Intelligence?"

Figure 1 - Industry, consumer, and government concerns about Big Data and AI.
(*For illustrative purposes only. All copyrights and trademarks retained by their respective owners.)*

The literature also has other cautionary perspectives. The 2014 book, *The Glass Cage: Automation and Us*, by Nicholas Carr, addresses the hidden costs of granting software control over our work and leisure, and explores the limits of AI delegated to human decision-making in critical situations, like autopilots in aircraft, driverless cars, and chemical plants. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, by Erik Brynjolfsson & Andrew McAfee in 2014, looks at complementary uses of technology that augment human decisions vs those that replace human activity on tasks once considered uniquely human. This gets into more areas that traditionally require higher levels of training like medicine where AI has the potential to diagnose diseases more accurately than doctors can.

Does Silicon Valley's worldview take into consideration the untidy emotional factors inherent in the social 'ends' that justify the technological 'means?' Ian Barbour sees the danger, not in technology as such, but in uncritical preoccupation with technological goals and methods (Barbour 65). Some of the less enlightened engineers have fostered a gee-whiz attitude of applying technology either for technology's sake or for the short-term profits of employers. Robert Pool observes that engineers do not think of what they do in social terms. However, as technologies become more complex, engineers will find it increasingly necessary to take human performance and, eventually, organizational factors into account in their designs (Pool 287).

We live in a society that rapidly diffuses technology, each with intended and aggregative unintended consequences on the well-being of society to an increasing number of rights claimants who each exercises the maximalist uses of technology. What none of the above points of view adequately address are the safety implications of over reliance on AI in critical operations and the socio-economic implications in a society that ties money to a job that may be increasingly threatened by AI. Also, modern technology has taught us that if anyone is to be negatively affected by an experiment, new process, or withholding of benefits, it will be the poor, the powerless, and those of

color. The anti-globalization/Brexit/Trump election is calling all of this into question as the losers do actually lose and they count.

Left unchecked, this can ultimately lead to the "destruction of the commons" and degradation of the overall social fabric. In the case of Artificial Intelligence or Augmented Intelligence (AI), uncritical application of the technology risks harming the public interest through increased risks to:

- Human safety (such as unproven driverless cars or over reliance on autopilots)
- Public infrastructure (such as AI-controlled power grids, communication systems, and financial records)
- Mass dislocation of workers due to AI-based automation
- Misinforming policy decision-makers through bad data inputs and improper analytic algorithms.

Business is being left off the hook in terms of the social impact of the products and services they are producing. Tesla and SpaceX CEO Elon Musk has been called to task about the impact of self-driving vehicles on jobs and has provided a typical response about how it will improve society in other ways that will balance those costs. To his credit, Musk also sounded the alarm in his July 2017 comments to the US Governors Association, noting that, in the case of AI, the risks are too high to allow AI to develop unfettered. AI calls for precautionary, proactive government intervention.

So maybe the big question is whether there is a better solution for businesses and society as a whole other than the standard argument. Can AI be applied, monitored, regulated — with real enforcement? Applying the basic tenets of product liability to the introduction of use of AI and slanting it toward social responsibility might also be a good structure.

- Manufacturing defects are those that occur in the manufacturing process and usually involve poor-quality materials or shoddy workmanship.
- Design defects occur where the product design is inherently dangerous or useless (and hence defective) no matter how carefully manufactured; this may be demonstrated either by showing that the product fails to satisfy ordinary consumer expectations as to what constitutes a safe product, or that the risks of the product outweigh its benefits.
- "Garbage in, garbage out." AI's weakness is the pool of data it draws from to create its algorithms. If that pool of data is socially suspect (in general poorly structured and maintained) then the results can lead to social harm.
- Even when AI works as intended, there may be unexpected side-affects that have social and moral issues.
- Failure-to-warn defects arise in products that carry inherent nonobvious dangers which could be mitigated through adequate warnings to the user, and these dangers are present regardless of how well the product is manufactured and designed for its intended purpose.
- Should there be actual limits on where AI is applied (parallel to those preventing certain bio-technology experiments on humans)?
- Can the benefits be shared with those who are harmed (and if so how, via taxes, special fees)?

This paper identifies some ethical standards that are already well known and understood and applies them to the emerging world of AI. The moral case is grounded in harmful and damaging social impacts and a basic sense of unfairness (disproportionate spread of downside costs on a particular community or group). A starting list of these impacts includes, but is not limited to:

- loss of opportunity to participate equally in a benefit
- imposition of higher costs to get benefits
- the emergence of disruptive or costly externalities with no chance of redress

Historical Context

As one traces the development of the social underpinnings of technology through agricultural, industrial, post-industrial, and information ages, one is forced to re-evaluate the validity of the dominant philosophical paradigm as one moves from one age to the next. In fact, new ages and new technologies stand on the foundations of previous technologies, philosophies, and ages. As such, acquisitive capitalism, as the primary driver of technological development and the criteria for distribution, could become anachronistic in the worldview that may be enabled by modern technology.

Just as scarcity of raw materials, labor, and land drove the economics of the industrial revolution that reaped the benefits of the earlier agricultural age's plenty, modern western culture has shaped a technologically-saturated information-based society that is not constrained by scarce information, nor scarcity of basic goods, nor scarcity of basic medical care, nor scarcity of food. The modern technologically advanced, wealthy, and information-rich society of Western Europe and North America needs to develop a new worldview that is not constrained by the overwhelming influence of raw capitalist win-lose economic philosophies that developed during the industrial age.

Britain was the first major country to base itself on an industrial and commercial economy. As the population migrated from the countryside into the cities in the 18th and 19th centuries, social and political institutions formed to deal with this new situation (Budge 7). (One might also note that political institutions also sought to meet the interests of the commercial class and their sustained need for workers.) The Labour Party was formed to address many of the needs of the working class and it adopted a socialist constitution in 1918 that was committed to common ownership of the means of production, distribution, and exchange. However, Labour opposed Soviet-type communism. Their ethical principles were the basis of their socialism, rather than Marxism (Childs 9).

One of the most significant arguments put forth by the trade union movement was that workers' wages were not just a commodity price to be set by the market. Competing workers driving the costs down also meant increased human misery in terms of poverty, health, family stability, crime, and problems that would affect the whole society (Budge 48). This was more than lobbying by the unions; it was a powerful ethical argument of the intrinsic worth of the individual that set the foundation for an accepted policy of a social safety net. In addition, according to Budge, "The mass unemployment of the 1930s had demonstrated how inadequate social protection was in the absence of comprehensive state aid" (Budge 625).

An economics of cooperative long-term social value and an ethics of just, but not necessarily equal, distribution of benefits and costs is called for if modern society is to reduce the threats to all of humanity by economic dislocation that grow out of unbounded application of AI. Former Apple executive and current chairman and chief executive of Sinovation Ventures, Kai-Fu Lee, suggests that it is "unavoidable that large chunks of the money created by AI will have to be transferred to those whose jobs have been displaced. This seems feasible only through Keynesian policies of increased government spending, presumably raised through taxation on wealthy companies."

In this regard, the pursuit of happiness must be insured such that the ability to support oneself and one's family is available for anyone that is willing to work. That will require a redefinition of the right to work in a global economy.

Ethical Underpinnings

To start, let us agree that technology is not neutral, and therefore can be held to moral and ethical standards. Historically, technology, or technique, has been, and continues to be, driven by the underlying cultural values of society. Those values have been derived from the worldview of a society, which includes the dominant philosophical paradigms of what is known (science), what is believed (religion), and what is desired (self-interest). Neither science, religion, or self-interest are unbiased and they certainly actualize in the real geo-political economy as non-neutral and often unfair. Suffice it to say for our purposes that technology is science plus purpose. While science is the study of the nature around us and subsequent development of scientific laws, technology is the practical application of those laws, in sometimes non-rigorous ways, toward the achievement of some purpose -- usually material (Dorf, 1).

In this context, is AI really just "augmented human intelligence" or are there more aspects to it? Expanding on PwC's definition, we see that AI works in at least four ways:

- ***Automated intelligence:*** *Automation of manual/cognitive and routine/non-routine tasks.*
- ***Assisted intelligence:*** *Helping people to perform tasks faster and better.*
- ***Augmented intelligence:*** *Helping people to make better decisions.*
- ***Autonomous intelligence:*** *Automating decision-making processes without human intervention.*

Regardless of the type of application, how and where to use this entire process remains a human decision. Thus, the human component must be subject to the normal checks and balances established by society.

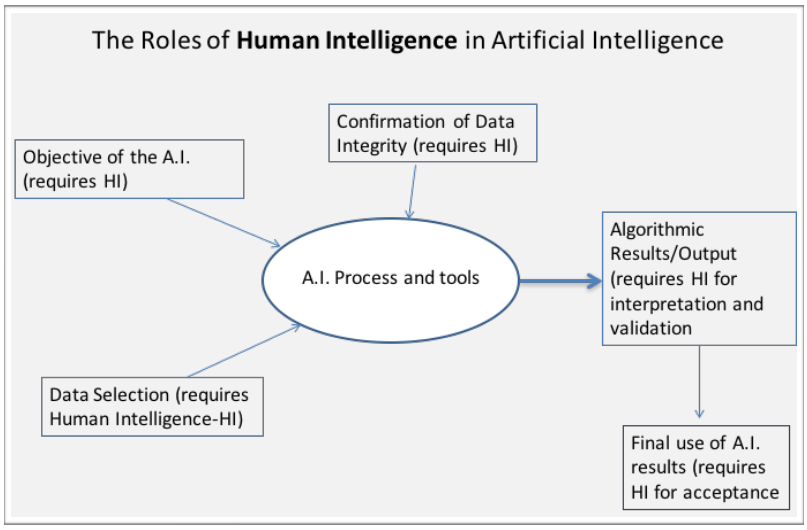

Figure 2 – Human Intelligence roles in AI

The day is long gone when technologists could arguably claim to have no responsibility for how their innovations are used. It is hollow solace for scientists and engineers to assume that they have left their belief systems, prejudices, fears, needs for security, egos, friendships, and enemies at the door of the laboratory. Given that technology, and the engineering profession that practices it, is a function of these non-neutral drivers, the potential impact of technology on the well-being of global society is, likewise, determined by our knowledge, beliefs, and desires.

In a complex modern technological society, one whose interconnected systems threaten to spin out of control, we must collectively ask technologists, "… are you living up to the proper engineering codes of ethics or have you delegated your responsibility to business interests and government ideologues?[1] Stanford professor Robert McGinn described several ethical problems facing modern 21st Century engineering practitioners. These problems include execution problems, such as unfair distribution of benefits and costs, the fear of whistle blowing, and lack of consideration of long-term effects. He also described communication problems, such as fraud and misrepresentation (McGinn, Ethics 18-26). Scientists and engineers have also erred by having misplaced loyalties. They have become servants to organizations rather than to the public. The basic canons of professional ethics have been subverted to gain employment and to preserve national power structures.

There are existing codes of ethics held and promoted by engineering professional societies. As an example of the types of traditional codes of ethics, occasionally (and sometimes routinely) ignored by technologists, consider the following:

- The **National Society of Professional Engineers** declares itself "to hold paramount the safety, health and welfare of the public" in the performance of their professional duties.[2]

- The **Institute of Electrical and Electronics Engineers** (IEEE) declares that its members must "accept responsibility in making decisions consistent with the safety, health, and welfare of the public, and to disclose promptly factors that might endanger the public or the environment."[3]

- In recognition that computers have a central and growing role in commerce, industry, government, medicine, education, entertainment and society at large, the **Association for Computing Machinery** (ACM) acknowledges that software engineers have significant opportunities to do good or cause harm, to enable others to do good or cause harm, or to influence others to do good or cause harm. As such, and in accordance with their commitment to the health, safety and welfare of the public, software engineers shall:
    - act consistently with the public interest,
    - act in a manner that is in the best interests of their client and employer consistent with the public interest,
    - advance the integrity and reputation of the profession consistent with the public interest,
    - maintain integrity and independence in their professional judgment,

---

[1] **Assumptions**

- Technology is not neutral, and therefore can be held to moral and ethical standards
- Codes of ethics are held and promoted by engineering professional societies
- The rights of stakeholders must be bounded by the constraints of the modern technological society and, in certain special cases, be restricted
- John Rawls' ethics principles ground the moral case if harmful and damaging social impacts and a basic sense of unfairness (disproportionate spread of downside costs on a particular community or group) occurs. A starting list of these impacts includes, but is not limited to:
    - loss of opportunity to participate equally in a benefit
    - imposition of higher costs to get benefits
    - the emergence of disruptive or costly externalities with no chance of redress

These moral responsibilities provide a paradigm shift away from merely cost reduction or harm reduction to a combination of maximization of benefits within the context of minimizing harm. This renewed ethical imperative would lead to scientific research and product designs for the most positive consequences, rather than settling on the current approach of minimizing the maximum regret.

[2] https://www.nspe.org/resources/ethics/code-ethics

[3] http://www.ieee.org/about/corporate/governance/p7-8.html

  - subscribe to and promote an ethical approach to the management of software development and maintenance, and
  - meet the highest professional standards possible.[4]

The argument of the supposed neutrality of scientists and engineers is no longer an acceptable shield behind which technologists can hide. Given that technologists must get directly involved in technology policy issues, it is timely and proper that a renewal of professional ethics is also in order.

- Engineers as a group and as individuals have special responsibilities as citizens, which go beyond those of non-engineer citizens. “All citizens have an obligation to devote some of their time and energies to public policy matters. Minimal requirements for everyone are to stay informed about issues that can be voted on, while stronger obligations arise for those who by professional background are well grounded in specific issues as well as for those who have the time to train themselves as public advocates," as put forth by Philosopher Mike Martin and Engineer Roland Schizinger (Martin 29l).

- Technologists must do so as an act of allegiance to their professions’ commitments to social justice as the primary goal, and hold other allegiances to employers, trade associations, profit motives, and self-advancement secondary. Failure to do so will continue to place the profession in a reactive mode to ever-increasing negative aggregative consequences, competing claims of “rights holders,” mistrust by the public, degradation of the profession, and ultimately governmental regulation.

In the case of AI, what are the industry issues and concerns of digital transformation without a moral framework that considers customer centricity, brand alignment, engineering ethics, liability risks, and political/regulatory assessment?

- **Digital Disruption & Dislocation** -- Economic value moving from owners of content (NBCU, Time Warner) to dominant platforms (Comcast, AT&T/DIRECTV, iTunes, Netflix). Will dominant platform businesses overturn the service sector next? Is impulse to share beating impulse for privacy?

- **Because we can build it … we will** -- AI has the potential to diagnose diseases equal to or more accurately than doctors can. Uncritical application of the technology risks harming the public interest through increased risks to public infrastructure, such as AI-controlled power grids, communication systems, and financial records.

- **Tyranny of Data** -- Will insurance companies move from discounts for health-monitor bracelets, like Fitbit or Apple Watch, to requiring you to do so?

- **Herd Mentality** -- Rushing into early-stage AI product development and marketing hype without proper vetting and plans.

What are the implications of digital transformation without a moral framework that considers customer centricity, brand alignment, engineering ethics, liability risks, and political/regulatory assessment?

- **Mass dislocation of workers due to AI-based automation**
  - The World Economic Forum estimates that AI, robotics, and automation could replace 5 million jobs around the world by 2020.

[4] http://www.acm.org/about/se-code

  - PwC estimates that 38% of jobs in the U.S. are at "high risk" of being replaced by robots and artificial intelligence over the next 15 years
  - EY expects graduate recruitment at auditors and accountants could fall by as much as 50% by 2020 due to the impact of artificial intelligence.
  - 2.9 million truckers and delivery drivers, 674,000 bus drivers, 181,000 cab drivers and chauffeurs could be impacted by driverless cars and trucks.

- **Misinforming policy decision-makers through bad data inputs and improper analytic algorithms.** Garbage in = garbage out. Be wary of paradigm overshooting as regards the use of analytical methods in human decision-making contexts

- **Human safety risks**, such as unproven driverless cars, or over reliance on autopilots. A 2010 FAA study of 10 years of airline crash data concluded that "Pilot errors had been involved in over 2/3 of all crashes; and automation has made such crashes more likely" by degrading situational awareness and weakening hand-flying skills.

To the degree that AI contributes to socio-economic dislocation and/or increased safety risks, its societal costs, redress, prevention, and prohibition will need to be reassessed by governments … and your customers.

- **Civil and/or Criminal Penalties** — acting in a way that one knows (or should have known) will harm (or pose an unreasonable risk of harming) the public interest through undue risks

- **Product Liability Law**
  - Design Defect — product design is inherently dangerous or useless (and hence defective) no matter how carefully manufactured
  - Failure to Warn — inherent non-obvious dangers which could be mitigated through adequate warnings to the user
  - Manufacturing Defects — poor-quality materials or shoddy workmanship
  - Even when the product works as intended, there may be unexpected side affects that have social and moral issues

- **Targeted Taxation** — Can the benefits be shared with those who are harmed (and if so how, via taxes, special fees)?

- **Increased Regulation** — Can AI be applied, monitored, regulated — with real enforcement? Should there be actual limits on where AI is applied (parallel to those preventing certain bio-technology experiments on humans)?

The ethics of John Rawls may be applied here.

- In a Rawlsian society, differences in wealth and social position can be tolerated only when they can be shown to benefit everyone and to benefit those who have the fewest advantages, in particular. A just society, according to Rawls, is not one where everyone is equal, but one in which inequalities must be demonstrated to be legitimate.

- In addition, in a Rawlsian society, everyone must be given a genuine opportunity to acquire membership in a group that enjoys special benefits (Munson 22-23).

- The implications of Rawls' principles are that: (1) everyone is entitled to the same public goods and services, (2) inequalities in the socio-economic system can be justified only if those in most need can benefit from them, and (3) to the degree that the previous two

conditions are not met, a wholesale reform is called for that would provide public goods and services to those who are unable to take advantage of the benefit (Munson 24-25).

Traditional professional society codes of ethics cite a series of actions and practices that a professional engineer or scientist should not engage in. It is a "thou shalt not" approach to ethics. Most codes are so general that they rarely give the practitioner any tangible guidance as to how research and development should be performed and the deontological admonitions give the practitioner a mistaken belief that, perhaps, one can perform any task that is not explicitly prohibited. Since most codes are non-binding and only the most glaring of offenses become publicly known, very little guidance is offered to the engineer who wants to work in the spirit of best practices. To this end, Stanford's Robert McGinn has identified a series of **Fundamental Moral Responsibilities (FMRE)** that provide a much more concrete and proactive approach to engineering ethics (McGinn, Moral Responsibilities 6-19). Those FMREs include:

- FMRE1 – Not act in any way that one knows (or should have known) will harm (or pose an unreasonable risk of harming) the public interest.

- FMRE2 – To try to prevent (or prevent the repetition of) preventable harm (or the creation of an unreasonable risk of harm) from being done to the public interest.

- FMRE3 – Assure that all parties likely to bear non-trivial risks from one's engineering work are adequately informed about them upstream and given a realistic chance to give or withhold their consent to their subsequent imposition.

- FMRE4 – Work to the best of the engineer's ability to serve the legitimate business interests and objectives of the employer or client.

From these FMREs, there are certain **Derived Moral Responsibilities (DMR)** advocated by McGinn that include:

- Disclose to the employer or client any unrecognized options,
- Help the employer or client reach a clarified definition of problems originally presented to the engineer in distorted form,
- Insure that all prerequisite conditions for the safe operation of a technology transferred from a more to a less developed society are satisfied,
- Be wary of paradigm overshooting as regards the use of analytical methods in innovative engineering contexts,
- Establish a precautionary organizational culture as regards the formal approval of integrity-related product changes,
- Assure in engineering work akin to social experimentation, that human subjects likely to be put at risk of harm are informed about those risks and given a meaningful opportunity to give or withhold consent to their imposition.

These moral responsibilities provide a paradigm shift away from merely cost reduction or harm reduction to a combination of maximization of benefits within the context of minimizing harm. From a quantitative analysis perspective, McGinn is proposing the optimization of two simultaneous equations (Anderson 350-352, 372-373):

- **Maximax** – Select the decision that maximizes the maximum payoff (do the most good for the most people)
- **Minimax Regret** – Minimize the maximum regret, or opportunity loss, associated with a decision (do no harm)

This is an improvement over traditional approaches that minimize harm (regret) or maximize profit (payoff), but rarely attempt to do both. Designers and developers of technology must proactively seek to maximize the most benefits for the largest number of people, while delivering the most benefit to those most negatively impacted, or likely to be negatively impacted, by the unintentional consequences of complex technology, in this case AI. This renewed ethical imperative would lead to scientific research and product designs for the most positive consequences, rather than settling on the current approach of minimizing the maximum regret.

In addition, the rights of stakeholders must, at a minimum, be bounded by the constraints of the modern technological society and, in certain special cases, be restricted. McGinn builds a convincing case for restricting those rights when they harm society (McGinn, Technology, 14-15). Among the conditions for restriction are:

- If the very existence of society is called into question
- If continued social functioning is threatened
- If some natural resource vital to society is threatened
- If a seriously debilitating financial cost is imposed on society
- If a significant aesthetic, cultural, historical, or spiritual value to a people is jeopardized, or
- If some highly valued social amenity would be seriously damaged

The Case of AI

In the case of AI, uncritical application of the technology risks harming the public interest through increased risks to:

- **Human safety**
  - Carr summarizes a 2010 FAA study of 10 years of airline crash data – "Pilot errors had been involved in over 2/3 of all crashes; and automation has made such crashes more likely" by degrading situational awareness and weakening hand-flying skills.
  - Carr's warnings echo earlier admonitions by Robert Pool — both suggesting that, when it comes to complex systems, the emphasis needs to be on making operators of technology more effective, instead of just making machines more effective. The industry should consider systems that inform humans, in great and varied detail, rather than blindly automate and delegate important and risky operations to machines.

- **Public infrastructure**
  - According to Manimaran Govindarasu, Professor of Electrical and Computer Engineering at Iowa State University, and Adam Hahn, Assistant Professor of Electrical Engineering and Computer Science at Washington State University, the **U.S. electricity grid** is the "largest interconnected machine," a complex digital and physical system crucial to life and commerce in this country. It is composed of more than 7,000 power plants, 55,000 substations, 160,000 miles of high-voltage transmission lines and millions of miles of low-voltage distribution lines. This web of generators, substations and power lines is organized into three major interconnections, operated by 66 balancing authorities and 3,000 different utilities. As the grid has become more dependent on computers and data-sharing, it has become more responsive to changes in power demand and better at integrating new sources of energy. But its computerized control could be abused by attackers (humans and "bots") who get into the systems. "Until 2015, the threat was hypothetical. But now we know cyberattacks can penetrate electricity grid control networks, shutting down

power to large numbers of people. It happened in Ukraine in 2015 and again in 2016, and it could happen here in the U.S., too."

- The **Internet of Things (IoT)** is in full flower. According to a Pew report, while 49% of the world's population is connected online, an estimated 8.4 billion connected things are in use worldwide. These include: IoT in cars, voice-activated assistants, appliances and other home systems, health-monitoring devices, road sensors, public-safety and security devices, smart meters and personal fitness and health trackers, thermostats, pressure gauges, pollution detectors, cameras, and microphones. Dirk Helbing and team estimate that in 10 years' time there will be 150 billion networked measuring sensors, 20 times more than people on Earth. Then, the amount of data will double every 12 hours. Pew notes that the very connectedness of the IoT leaves it open to security and safety vulnerabilities, noting, "In September 2016 at DEF CON, one of the world's largest security conferences, 47 vulnerabilities affecting 23 IoT-enabled items (door locks, wheelchairs, thermostats and more) from 21 manufacturers were disclosed."

- Today 70% of all **financial transactions** are performed by algorithms. News content is, in part, automatically generated (Helbing et al). Alston Ghafourifar writing in Venturebeat, notes that some 1,360 hedge funds rely on computer models to trade stocks and other investments. These funds represent $197 billion dollars of investor money being directed by lines of computer code. Most of these funds represent traditional 'quant' (quantitative) funds that use computer models to predict share price movements and determine trades. But they also remind us about potential glitches -- "In August 2012, a trading program at one fund 'ran amok,' creating losses of $10 million a minute. It took nearly an hour for the human team to identify and solve the problem, and the firm lost $440 million in the process. Two years earlier, an algorithmic trade caused a "flash crash," in which U.S. share and future indices dropped 10 percent within minutes. Some say those incidents are telling preludes to disaster. A rogue algorithm at one of the country's major banks, or a cascading failure in which multiple big banks are derailed by faulty programs, could lead to a catastrophic crash."

- **Mass dislocation of workers** due to AI-based automation might be the hallmark of the "Fourth Industrial Revolution." Just as machines replaced many manual labor jobs in the industrial revolution, advances in computing mean many service sector jobs risk being replaced by computers. The World Economic Forum estimates that AI, robotics, and automation could replace 5 million jobs around the world by 2020. Citi has also estimated that 35% of jobs in the UK are at risk of being replaced by automation, 47% of US jobs are at risk, and across the OECD as a whole an average of 57% of jobs are at risk. In China, the risk of automation is as high as 77%. 38% of jobs in the U.S. are at "high risk" of being replaced by robots and artificial intelligence over the next 15 years, a separate estimate by consulting and accounting firm PwC found. Rice University computer scientist Moshe Vardia observes -- "Society needs to confront this question before it is upon us: if machines are capable of doing almost any work humans can do, what will humans do?" (Business Insider)

  - Companies from Ford to Tesla to Uber are investing in **automated car technology**. Automated cars pose an existential threat to the many Americans who drive for a living: 2.9 million truckers and delivery drivers, 674,000 bus drivers, 181,000 cab drivers and chauffeurs (Portland Press Herald).Vardi suspects that truck drivers will be the first victims of automation applied to driving. Automated trucks,

such as those being developed by Uber's Otto startup, can be programmed to go from one warehouse to another, plying express lanes reserved for trucks that let them avoid interacting with human drivers. Vulnerable truckers have much in common with the factory workers who've been ousted by machines over the past several decades — They tend to be white men, middle-aged or older, with high school-only educations. "These things do not happen without political consequences," Vardi said. (Portland Press Herald)

- **Accountants** — HR Block is using IBM's Watson. What is the impact on accountants during tax season? Smacc offers small and medium-sized enterprises a platform to digitize and automate accounting and financial processes. Customers submit their receipts to Smacc, which turns them into a machine-readable format, encrypts them, then allocates them to an account. The system checks against some 64 data points, verifies the invoice, checking, for example, that the math adds up, and even if the VAT-ID and its issuer are correct. The platform gradually also self-learns, tracking invoices, sales and costs, as well as their liquidity. Graduate recruitment at auditors and accountants could fall by as much as 50% by 2020 due to the impact of artificial intelligence, according to Steve Varley, chairman and managing partner for the UK and Ireland at "Big Four" accounting firm EY. (Business Insider)

- **Doctors** — As Brynjolfsson & McAfee's examples in medicine show, AI has the potential to diagnose diseases equal to or more accurately than doctors can. In the diagnosis of Diabetic Retinopathy, the fastest growing cause of blindness with nearly 415 million diabetic patients at risk worldwide, if caught early, the disease can be treated; if not, it can lead to irreversible blindness. Google's algorithm's performance, using Machine Learning and Computer Vision, was shown to be on-par with that of ophthalmologists (Google Research Blog). HealthTap's Dr. A.I. platform compiles a list of the most and least likely causes for the symptoms and ranks them by order of seriousness. It digitizes health care triage, which is the process of assessing the level of medical risk facing a patient and the first step in the treatment pathway. The software relies on individual patients' medical profiles and knowledge gleaned from 105,000 physicians who are triage experts over more than half a decade (Fortune).

- **Misinforming** through bad data inputs and improper analytic algorithms. AI draws on a large data base and creates some useful and predictive software on the incidents of crime.
  - Jonathan Zittrain, speaking at an AI event at Harvard's Berkman Klein Center for Internet & Society, summarized major questions posed by using automated processes for decision-making. He cited an example of current concerns over bias in the AI systems already in use by some U.S. judges to guide parole or sentencing decisions. AI systems can and do reflect human biases.
  - Because of the way the data is collected, poor and African-American areas are designated high crime and police are assigned there. This leads to over-policing and higher levels of police brutality. As we know drug use in White and Black communities is the same, but drug arrests are higher in African-American communities. AI will just lock all of this in with more data.
  - AI may be like the impact of red-lining and its long-term impacts on Black wealth accumulation because Blacks were forced to live in housing areas with lower chances of appreciation and limited access to better schools. AI systems might extend those kinds of impacts into many other markets.

Responsibilities of the Industry

Since the technology industry benefits from publicly funded research, government-granted patents and tax breaks, and since it makes products vitally important to public commerce, it should be accountable to society at large, rather than just to its shareholders. In the technology-intense globalized world of the early 21st Century, the basic skills needed to support one's family need to be reassessed; and to the degree that AI contributes to socio-economic dislocation and/or increased safety risks, its societal costs and remedies need to be reassessed. Redress, prevention, and prohibition might include the following actions tied to Fundamental Moral Responsibilities (FMRE):

- **Civil and/or Criminal Penalties** — Not act in any way that one knows (or should have known) will harm (or pose an unreasonable risk of harming) the public interest through undue risks to human safety, risks to public infrastructure, and mass dislocation of workers due to AI-based automation without proper redress or retraining.

- **Design Defect aspects of Product Liability Law** — To try to prevent (or prevent the repetition of) preventable harm (or the creation of an unreasonable risk of harm) from being done to the public interest, as listed above.

- **Targeted Taxation —** If economic dislocation cannot be adequately prevented, under Rawls, as well as the Harm Principle and the Welfare Principle, the AI companies would not be permitted to exploit displaced workers without redress. Perhaps a better solution in the real world of capitalism is not to forbid AI, but to tax certain applications heavily.

    - Microsoft Founder and philanthropist Bill Gates makes an argument that robots who replace human workers should incur taxes equivalent to that worker's income taxes. "Right now, the human worker who does, say, $50,000 worth of work in a factory, that income is taxed . . . If a robot comes in to do the same thing, you'd think that we'd tax the robot at a similar level." Gates argues that these taxes, paid by a robot's owners or makers, would be used to help fund labor force retraining in health services, education, or other fields where human workers will remain vital.

    - California already imposes a luxury tax on automobiles over a certain dollar price tag and charges vehicle license fees on a sliding scale based on the value of the car, in a recognition of the burdens cars place on the environment. We do not prohibit Hummers; we tax them. We also tax the lifestyle demons, such as liquor and tobacco, based on their long-term medical burdens on society. Why not tax accounting software, medical databases, driverless cars and trucks, automated financial trading, and drone delivery companies? Tax them heavily and provide benefits to the millions of Americans out of work due to these systems or contribute the proceeds to fund retaining at local colleges and trade schools (since in a Rawlsian society, everyone must be given a genuine opportunity to acquire membership in a group that enjoys special benefits).

- **Failure to Warn aspects of Product Liability Law** — Assure that all parties likely to bear non-trivial safety or dislocation risks from one's engineering work are adequately informed about them upstream and given a realistic chance to give or withhold their consent to their subsequent imposition.

- Likewise, the technology industry has Derived Moral Responsibilities (DMR) to:
  - **Manufacturing Defect aspects of Product Liability Law —** Insure that all prerequisite conditions for the safe operation of AI technology are satisfied,
  - **Failure to Warn aspects of Product Liability Law** — Assure in engineering work akin to social experimentation, that human subjects or groups likely to be put at risk of harm or dislocation are informed about those risks and given a meaningful opportunity to give or withhold consent to their imposition,
  - **Design Defect aspects of Product Liability Law —** Be wary of paradigm overshooting as regards the use of analytical methods in human decision-making contexts,
  - **Design Defect aspects of Product Liability Law —** Establish a precautionary organizational culture as regards the formal approval of AI-related product changes.

Assessment Approach

What is STI's approach to assessing and advising on digital transformation? On behalf of our clients, we seek cooperative long-term customer value and a brand ethics of just distribution of benefits and costs that will reduce AI's threats of economic dislocation that grow out of unbounded application of technology. Figure 3 below summarizes the steps in our process.

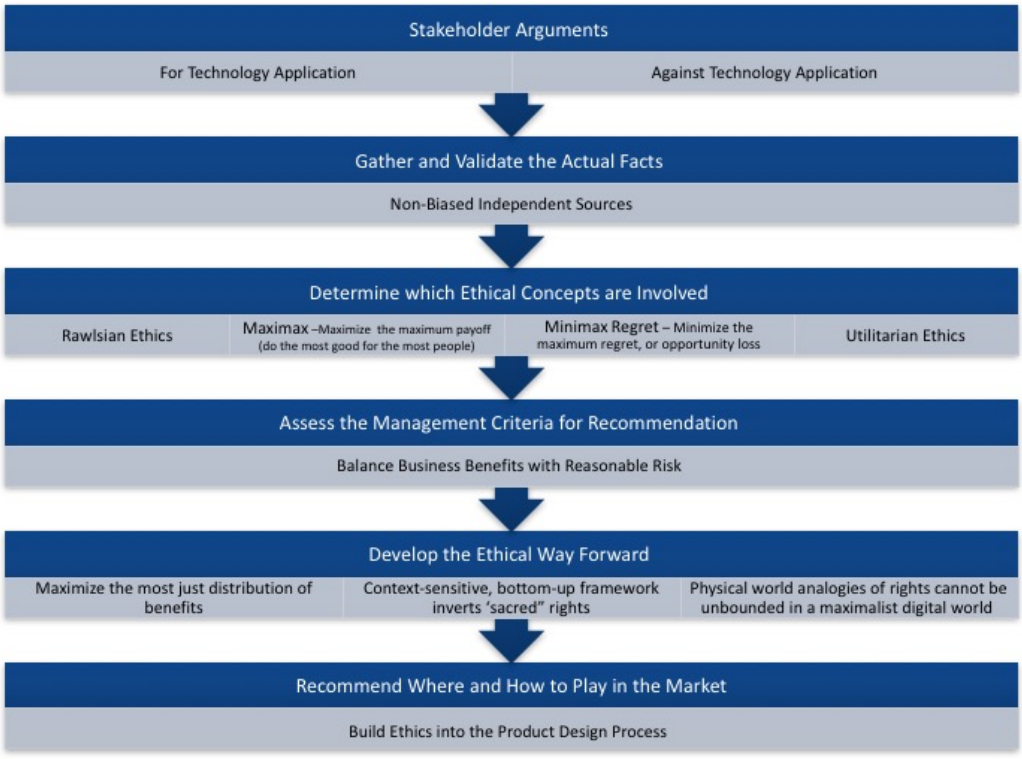

Figure 3 – STI Ethics-Based Technology Assessment

We also use scenario planning to help clients decide where their AI investments need to be in the matrix of market vs. regulatory protections.

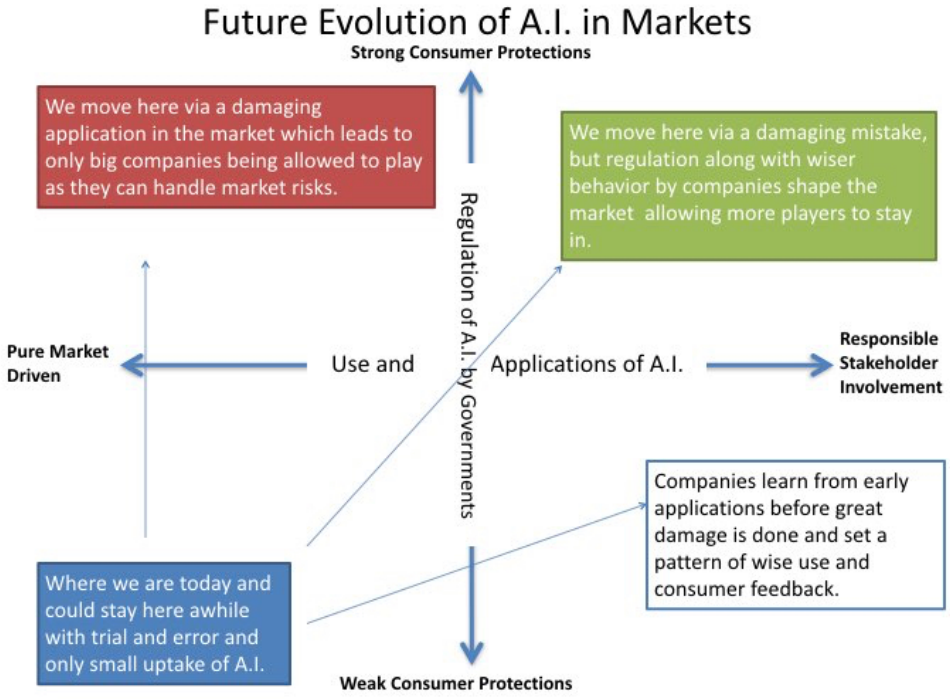

Figure 4 – Possible market positions for AI applications and companies

The assessment approach needs to be combined with a long-term learning loop to sustain and continuously improve ethical consumer engagement.

Figure 5 – AI Learning Loop

Carrot or Stick?

Is AI ethics driven by proactive risk avoidance or mandatory regulations? Will technology companies really value (and pay for) pre-emptive brand protection and risk/penalty avoidance? Or, as we typically see in the valley's short-term thinking, they will take the risk, leave it up to PR and Legal, and assume it is up to the next CEO to fix it. We have examples that provide signals as to where the wind might be blowing. The accounting profession has a growing business because there are government regulations and Wall Street requirements that force companies to use CPAs. Whereas, cybersecurity was a fledgling practice until the Sony hack; then it took off. Will proactive ethics around AI will have the same problem?

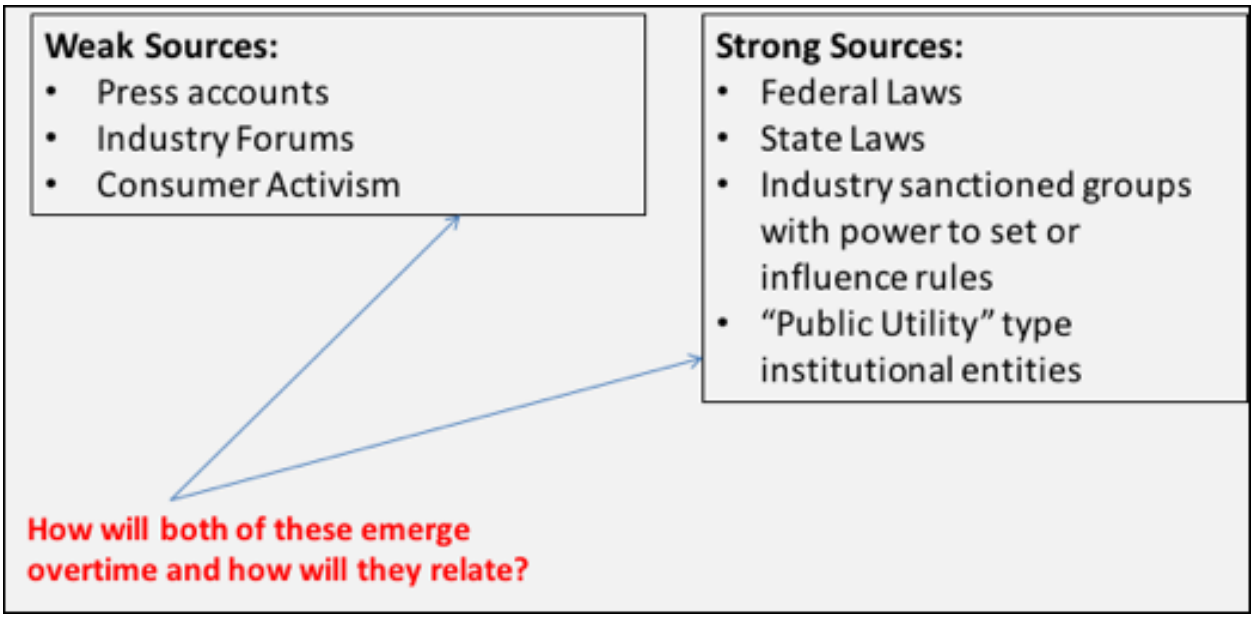

Figure 6 – Strong vs. weak regulatory forces

In the current US political environment, we are more likely to see reduced or weakened regulation, rather than a "stick" approach. So, engineering ethics might require a "quasi-regulatory" advocacy group to drive the demand. Examples of such groups include:

- **The Electronic Frontier Foundation** -- The U.S. 501(c)(3) nonprofit Electronic Frontier Foundation (EFF) is the leading organization defending civil liberties in the digital world. For decades EFF has worked in the courts and with activists to protect the freedom to hack, tinker, remix, and create for the long term good of society. EFF defends free speech on the Internet, fights illegal surveillance, supports freedom-enhancing technologies, and much more. The EFF team consists of a unique blend of activists, technologists, and attorneys that work to ensure that the rights and freedoms we enjoy are enhanced, rather than eroded, as our use of technology grows. https://www.linkedin.com/company-beta/20885/

- **Electronic Privacy Information Center** - EPIC is an independent non-profit research center in Washington, DC. EPIC works to protect privacy, freedom of expression, democratic values, and to promote the Public Voice in decisions concerning the future of the Internet.

EPIC pursues a wide range of program activities including public education, litigation, and advocacy. EPIC routinely files amicus briefs in federal courts, pursues open government cases, defends consumer privacy, organizes conferences for NGOs, and speaks before Congress and judicial organizations about emerging privacy and civil liberties issues. EPIC works closely with a distinguished advisory board, with expertise in law, technology and public policy. https://www.linkedin.com/company-beta/8202888/

- **Internet of Things Privacy Forum** -- A non-profit serving as a crossroads for industry, regulators, academics, government and privacy advocates to discuss the privacy challenges of the Internet of Things. The term "Internet of Things" can be vague, so, broadly speaking, the Forum focuses on the billions of connected devices that speak autonomously and semi-autonomously to each other and to host systems, and the privacy issues that result from their ever increasing collection of personal data. This means connected cars, wearable medical devices, smart homes, smart meters, smart cities, quantified self devices, TVs that listen to your conversations, autonomous drones and everything in between. The Forum's mission is to produce guidance, analysis and best practices to enable industry and government to reduce risk and innovate responsibly. https://www.iotprivacyforum.org

There is also an innate value in transparency and trust. Cognizant observes, "To succeed in today's digital age, companies must think beyond dollars, cents and convenience, and focus on data ethics. As malfeasance, blunders and mishandling of consumers' personal information reaches epic proportions, trust is the new battlefield for companies to seize the digital high ground, our latest research reveals." So, maybe a better "carrot" for the industry is trust among its customer base.

The industry challenge is to engage with customers at an individual, personalized level... in a privacy-assured manner. In fact, unlike the all too often tendency for media technology platform operators to surreptitiously "spy" on their customers, earlier research by PwC found that a trust-based relationship encourages consumers to voluntarily share more data:
• 76% of respondents are willing to share personal information when they were offered free benefits
• 80% of respondents said they were willing to share personal information if the company lets them know upfront how they are going to use it
• A Consumer Privacy Bill of Rights might actually increase consumers' willingness to share information
• 87% of survey respondents want to be able to manage what and how personal information is used

Recent Cognizant studies further validate these findings. While 91% of consumers worry about who sees their personal information, and 57% won't do business with a company that misuses their data, 45% will share personal data with trusted companies that openly explain how the data will be used.

Figure 7 – Cognizant consumer sentiment around sharing personal data

Benefits of Proactive Steps

What are the business and brand benefits of digital transformation in the context of a moral framework that considers customer centricity, brand alignment, engineering ethics, liability risks, and political/regulatory assessment?

- Reinforcement of Sustainable Brand Value to Customers, Partners, Investors, and Public
- Focuses Investments in Sustainable Products and Markets
- Reduction of Risks of Civil/Criminal Penalties
- Reduction of Need for "Crisis Management" Costs and PR Damage
- Minimize Need for Government Intervention and Regulation

Conclusion

This paper calls for a significant re-evaluation of the ethics of distribution and the professional responsibility of engineers and scientists, in the context of the risks and rewards presented by innovative applications of Artificial Intelligence technologies. An interdisciplinary approach, combining economics, social justice, Rawlsian ethics, and regulatory, tort, and criminal law can be used to analyze the options for further development of distributive justice.

The development of a new economic paradigm, especially one based on a techno-centric social structure, will require a firm understanding of the cultural values underlying technologies that lay the foundation for the evolving economic system. We suggest a framework for a more advanced, globally fair, and sustainable economic system that is appropriate to the realities of a technologically-intensified society, that otherwise threatens to spiral out of human control.

Sources Cited

Anderson, David R., Sweeney, Dennis J., Williams, Thomas A. **An Introduction to Management Science: Quantitative Approaches to Decision Making.** St. Paul: West Publishing Co. 1976.

Barbour, Ian G. **Science & Secularity, The Ethics of Technology**. New York: Harper & Row, 1970.

Bossman, Julia. Top 9 ethical issues in artificial intelligence, World Economic Forum. October 2016. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Budge, Ian; Crew, Ivor; McKay, David; and Newton, Ken. **The New British Politics**, Third Edition. Essex: Pearson Education Limited, 2004.

Butcher, Mike. *Goodbye accountants! Startup builds AI to automate all your accounting.* TechCrunch.com. June 28, 2016. https://techcrunch.com/2016/06/28/goodbye-accountants-startup-builds-ai-to-automate-all-your-accounting/

Childs, David. **Britain Since 1939: Progress and Decline**, Second Edition. Basingstoke: PALGRAVE, 2002.

Domonoske, Camila. NPR. *Elon Musk Warns Governors: Artificial Intelligence Poses 'Existential Risk*', July 17, 2017. http://www.npr.org/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk

Dorf, Richard C. **Technology, Society and Man**. San Francisco: Boyd & Fraser , 1974.

English-Lueck, Jan. A. **Cultures @ Silicon Valley.** Stanford, CA: Stanford University Press, 2002.

Ghafourifar, Alston; Ghafourifar, Mehdi; and Walker, Brian. *AI-powered trading raises new questions.* **Venturebeat**, May 6, 2017. https://venturebeat.com/2017/05/06/ai-powered-trading-raises-new-questions/

Govindarasu, Manimaran and Hahn, Adam. *Cybersecurity of the power grid: A growing challenge*, **The Conversation.com**. February 23, 2017. http://theconversation.com/cybersecurity-of-the-power-grid-a-growing-challenge-73102

Grut, Oscar-Wiiliams. **Business Insider.** *AI could destroy hiring in one of the biggest industries for graduates*. May 10, 2016. http://www.businessinsider.com/ai-could-reduce-graduate-hiring-at-big-four-accountants-by-50-2016-5

Helbing, Dirk; Frey, Bruno S.; Gigerenzer, Gerd; Hafen, Ernst; Hagner, Michael; Hofstetter, Yvonne; van den Hoven, Jeroen; Zicari, Roberto V.; and Zwitter, Andrej. *Will Democracy Survive Big Data and Artificial Intelligence?* **Scientific American**, February 25, 2017. https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

Lee, Kai-Fu. T*he Real Threat of Artificial Intelligence*, **The New York Times**, June 24, 2017. https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html

Martin, M. W. and Schinzinger, R. **Ethics in Engineering**. New York: McGraw-Hill, 1983.

McGinn, Robert. *Ethics, Science, and Technology.* 1990.

McGinn, Robert. *Moral Responsibilities of Professional Engineers: Empirical and Theoretical Approaches.* Presentation given at the Engineering Ethics Forum, University of Nagoya, Japan. December 8, 2002.

McGinn, Robert. *Technology, Demography, and the Anachronism of Traditional Rights.* Journal of Applied Philosophy, Vol. 11, No. 1, Spring, 1994, pp. 57-70.

Morris, David Z. *Bill Gates Says Robots Should Be Taxed Like Workers*. Fortune, Feb 18, 2017. http://fortune.com/2017/02/18/bill-gates-robot-taxes-automation/

Munson, Ronald. **Intervention and Reflection: Basic Issues in Medical Ethics**, Sixth Edition. Belmont, California: Wadsworth/Thomson Leaning, 2000.

Mukherjee, Sy. **Fortune.com.** *You Can Now Download an Artificial Intelligence Doctor.* Jan 10, 2017. http://fortune.com/2017/01/10/healthtap-dr-ai-launch/

Peng, Lily and Gulshan, Varun. **Google Research Blog**. *Deep Learning for Detection of Diabetic Eye Disease.* Nov 29, 2016. https://research.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html

Pool, Robert. **Beyond Engineering: How Society Shapes Technology**. Oxford: Oxford University Press, 1997.

Portland Press Herald. *Automated cars threaten pro drivers' jobs*. November 3, 2016. http://www.pressherald.com/2016/11/03/automated-cars-threaten-pro-drivers-jobs/

PricewaterhouseCoopers, *Briefing: Artificial intelligence*. August 1, 2017. http://usblogs.pwc.com/emerging-technology/briefing-ai/

Rainie, Lee and Anderson, Janna. *The Internet of Things Connectivity Binge: What Are the Implications?* **Pew Research Center: Internet & Technology**. June 6, 2017. http://www.pewinternet.org/2017/06/06/the-internet-of-things-connectivity-binge-what-are-the-implications/

Talbot, David. AI Advance: *A Community Convening at Harvard Law School to advance the Ethics and Governance of Artificial Intelligence Initiative*. **Berkman Klein Center**, June 1, 2017. https://medium.com/berkman-klein-center/ai-advance-may-15-2017-2c36ee9d8dc8

Taplin, Jonathan. *Can the Tech Giants Be Stopped?* **The Wall Street Journal**, July 14, 2017. https://www.wsj.com/articles/can-the-tech-giants-be-stopped-1500057243

World Economic Forum. *The Future of Jobs.* Jan. 20, 2016. https://www.weforum.org/reports/the-future-of-jobs

About the Authors

Blake White, Founder, Strategic Technology Institute
www.strategic-tech.org

Blake is an experienced Silicon Valley and Hollywood industry executive, who has also held leadership positions in some of the world's most respected technology companies (HP, Apple, SGI), consulting firms (PwC, Ascent Media, Cognizant), and innovative start-ups. He is the author of *The Technology Assessment Process: A Strategic Framework for Managing Innovation,* and several industry publications. He is a graduate of NC State, Xavier, and Stanford.

Gerald Harris, President and Managing Director, Quantum Planning Group, Inc.
http://www.artofquantumplanning.com/

The Quantum Planning Group is a boutique consulting group specializing in helping organizations plan for the future using scenario planning and other strategic planning tools. Gerald has experience across a wide range of industries and planning challenges, gain through experiences with PG&E, the Global Business Network, the Monitor Group, and as an advisor to start-ups; plus he is the author of *The Art of Quantum Planning: Lessons from Quantum Physics for Breakthrough Strategy, Innovation and Leadership*. He is a graduate of Morehouse College and the University of Chicago.

About STI

Founded in 1985, the Strategic Technology Institute (STI) is a network of independent consultants that provides executive-level management advisory services focused on corporate strategies, often facilitated by technology innovation. STI is also a virtual 'think tank' that investigates the business and public policy issues raised by science and engineering. Aligning technology roadmaps to corporate strategy, STI has the following lines of business:

- Advisory Services
- Public Policy
- IP Development & Licensing
- Program Management