Cutter IT Journal

The Journal of Information Technology Management

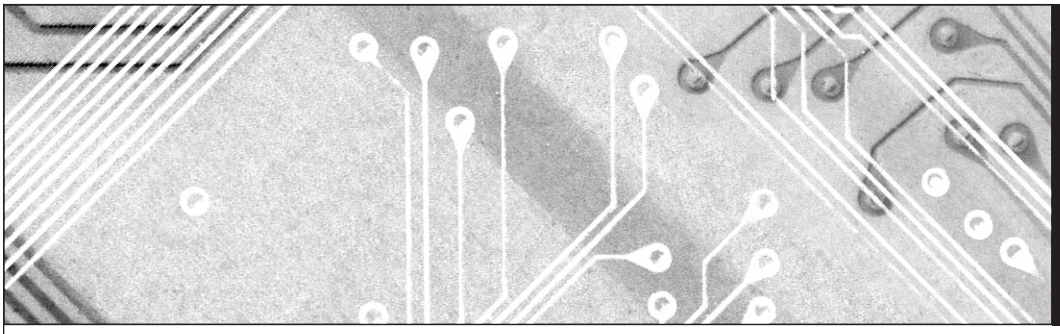
Vol. 29, No. 5
May 2016

"It should be readily apparent that information systems and technology pose a wide range of thorny ethical questions. Software, given its invisibility, provides tempting opportunities for unethical behavior."

— Robert N. Charette, Guest Editor

The Role of Ethics in Algorithm Design

CUTTER CONSORTIUM



Cutter IT Journal

About Cutter IT Journal

Part of Cutter Consortium's mission is to foster debate and dialogue on the business technology issues challenging enterprises today, helping organizations leverage IT for competitive advantage and business success. Cutter's philosophy is that most of the issues that managers face are complex enough to merit examination that goes beyond simple pronouncements. Founded in 1987 as *American Programmer* by Ed Yourdon, *Cutter IT Journal* is one of Cutter's key venues for debate.

The monthly *Cutter IT Journal* and its companion *Cutter IT Advisor* offer a variety of perspectives on the issues you're dealing with today. Armed with opinion, data, and advice, you'll be able to make the best decisions, employ the best practices, and choose the right strategies for your organization.

Unlike academic journals, *Cutter IT Journal* doesn't water down or delay its coverage of timely issues with lengthy peer reviews. Each month, our expert Guest Editor delivers articles by internationally known IT practitioners that include case studies, research findings, and experience-based opinion on the IT topics enterprises face today — not issues you were dealing with six months ago, or those that are so esoteric you might not ever need to learn from others' experiences. No other journal brings together so many cutting-edge thinkers or lets them speak so bluntly.

Cutter IT Journal subscribers consider the *Journal* a "consultancy in print" and liken each month's issue to the impassioned debates they participate in at the end of a day at a conference.

Every facet of IT — application integration, security, portfolio management, and testing, to name a few — plays a role in the success or failure of your organization's IT efforts. Only *Cutter IT Journal* and *Cutter IT Advisor* deliver a comprehensive treatment of these critical issues and help you make informed decisions about the strategies that can improve IT's performance.

Cutter IT Journal is unique in that it is written by IT professionals — people like you who face the same challenges and are under the same pressures to get the job done. *Cutter IT Journal* brings you frank, honest accounts of what works, what doesn't, and why.

Put your IT concerns in a business context. Discover the best ways to pitch new ideas to executive management. Ensure the success of your IT organization in an economy that encourages outsourcing and intense international competition. Avoid the common pitfalls and work smarter while under tighter constraints. You'll learn how to do all this and more when you subscribe to *Cutter IT Journal*.

Cutter IT Journal®

Cutter Business Technology Council: Rob Austin, Ron Blitstein, Tom DeMarco, Lynne Ellyn, Vince Kellen, Tim Lister, Lou Mazzucchelli, Ken Orr, and Robert D. Scott

Editor Emeritus: Ed Yourdon
Publisher: Karen Fine Coburn
Group Publisher: Chris Generali
Managing Editor: Karen Pasley
Production Editor: Linda Dias
Client Services: service@cutter.com

Cutter IT Journal® is published 12 times a year by Cutter Information LLC, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA (Tel: +1 781 648 8700; Fax: +1 781 648 8707; Email: citjeditorial@cutter.com; Website: www.cutter.com; Twitter: @cuttertweets; Facebook: Cutter Consortium). Print ISSN: 1522-7383; online/electronic ISSN: 1554-5946.

Subscription rates are US $485 a year in North America, US $585 elsewhere, payable to Cutter Information LLC. Reprints, bulk purchases, past issues, and multiple subscription and site license rates are available on request.

Opening Statement

by Robert N. Charette, Guest Editor

> We are rushing headlong into the robotics revolution without consideration for the many unforeseen problems lying around the corner. It is time now to step back and think hard about the future of the technology before it sneaks up and bites us when we are least expecting it.
>
> — Noel Sharkey, Foundation for Responsible Robotics

If we consider recent events, it's clear that we had better heed Sharkey's plea to step back and think hard about the robotics revolution very soon. For in April, China announced the details of its plan to triple its robotic production in the country over the next five years.[1] A *Financial Times* story on the implications of China's robot initiative stated that China has already bought more industrial robots each year than any other country since 2013, including a quarter of the world's total supply last year alone.[2] By the end of this year, the *FT* notes, China will be the largest operator of industrial robots in the world, surpassing Japan.

There are several driving factors behind China's wholehearted embrace of robots. One is that China's working-age population is predicted to fall over the next three decades (due in part to its former one-child policy), and industrial automation is aimed at filling the expected labor shortfall. Another factor is that other manufacturing countries are investing heavily in industrial robots as a way to undercut China's current competitive edge in manufacturing. For instance, last year Japan's government announced a major initiative to create a "robotic revolution" that would "spread the use of robotics from large-scale factories to every corner of our economy and society."[3] South Korea's government, in response to Japan's action, immediately announced a US $2.69 billion investment in its local robotics industry in order to keep competitive.[4] A third factor is the development of AI knowledge: China doesn't want to be left behind in the emerging "AI arms race" that the US and Russia, as well as 80 other countries, seem prepared to embark on and that has Sharkey and other leading technologists and scientists concerned.[5, 6]

There may be good reason for their concern. In a recently published report titled "Autonomous Weapons and Operational Risk," Paul Scharre, a former US Department of Defense official who helped establish US policy on autonomous weapons, warns that such weapons could lead not only to significant civilian casualties and fratricide, but to "unintended escalation" during a precarious international political confrontation if their software algorithms malfunction or their security is compromised by intruders.[7] An open ethical question is whether such autonomous weapon systems can — and will — be programmed to follow the international rules governing modern warfare. For example, what would an autonomous fighting robot in a war zone do if a tall child approached carrying a bucket? Would the robot comprehend that it was a child and that the bucket was full of mud and not explosives? Or what if that same child approached but attempted to cover the robot's sensors and cameras with the mud from the bucket? The worry is that autonomous weapon systems will be deployed before these risks are, to echo Sharkey's words, thought long and hard about.

As a freshly published Obama administration study makes abundantly clear, the "potential of encoding discrimination" in algorithmically driven decisions is a real and growing risk.

The question of ethical algorithms doesn't just affect autonomous robotic operations, either. As more devices are being increasingly connected into an "Internet of Things," how, when, and why should the gathered information be used and to whom should it be made available? For example, as home appliances, heating and electrical systems, and home security systems interconnect to form a "smart home," will insurance companies intent on reducing their risk profile begin to penalize homeowners who don't have these smart devices in place, or decide because of the types of devices owned, the homeowner can afford to pay more for their insurance?[8] As a freshly published Obama administration study titled "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights" makes abundantly clear, the "potential of encoding

discrimination" in algorithmically driven decisions is a real and growing risk.[9]

Then there are the more "common" IT system ethical situations, as highlighted by Volkswagen's use of software to cheat emission control tests[10] or by the host of banks that conspired to manipulate global bank interest rates.[11] Software, given its invisibility, provides tempting opportunities for unethical behavior.

Computing systems and related technologies create choices — and resulting ethical questions or dilemmas — that did not arise before computers.

WHAT IS COMPUTER ETHICS?

It should be readily apparent that information systems and technology pose a wide range of thorny ethical questions. And yet the generally agreed definition of computer ethics is still the one developed in 1985 by philosophy professor James Moor, which states that it is:

> ... the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology.[12]

One notices quickly that this definition focuses on the "ethical use" of IT, as Moor makes the strong case that what is or is not "ethical" cannot be universally prescribed in advance. Indeed, what is ethical is highly context-sensitive.

What Moor's definition really is trying to highlight is that computing systems and related technologies create choices — and resulting ethical questions or dilemmas — that did not arise before computers. For instance, it is entirely legal for US banks to reorder their depositors' withdrawals and deposits such that the withdrawals are paid before the deposits are credited, instead of by chronological order.[13] The banks claim they are doing their depositors a favor by ensuring the timely payment of important bills like mortgages and car payments. However, the rearrangement also increases the likelihood that their depositors will incur multiple overdraft fees, which can make the bank millions of dollars as a result. Few (other than the banks themselves) defend the practice as being entirely ethical. Furthermore, without the aid of their computer systems, banks wouldn't be able to efficiently process deposits and withdrawals in any order other than chronologically. Nor would it be possible for a criminal to be able to physically steal 80,000,000 paper-based health records files all at once, which is relatively easy when those health records are digitized.[14]

EVER-INCREASING ETHICAL CHOICES

More and more over the past 70 years, computing technology and systems have changed what decisions humans are able to make and how they make them. As important, they have changed the perceived significance and value of the decisions and activities we want to undertake. We are now entering a period where computing systems are increasingly going to take decision

making away from humans because, frankly, those systems will be better than humans at:

- Uncovering more decision alternatives
- Quickly choosing the one(s) that have the highest likelihood of success utilizing the fewest resources, and
- Implementing them

However, as has often been said, with great power comes great responsibility to ethically manage the risks that power can create if misused deliberately, accidentally, or by not understanding its unintended consequences. The growing power of computing algorithms and the technology that supports them may have massive impacts on the employability of hundreds of millions of people worldwide over the coming decades. One study has predicted that half of current US jobs — or some 60 to 70 million — could theoretically be automated by 2035, for example.[15] The same prospects of job destruction face every country that is investing heavily in automation. Even the Chinese government acknowledges that its investment in industrial robotics in manufacturing might start eliminating the traditional jobs path out of poverty used by tens of millions of its citizens living in the countryside.[16] It is an ethical imperative for governments to help those likely to be affected by automation, which for the most part will be individuals already at the bottom of the economic ladder.

Similarly, as government services become ever more automated, it is critically important to ensure that the automation works properly, because when it doesn't, those least able to afford it again end up suffering the most. For example, when the state of North Carolina decided in 2013 to go live with its new $484 million benefits system NC FAST before it was fully tested, tens of thousands of food-assistance recipients were not able to receive their benefits for weeks, and some for months, until the system's flaws were adequately fixed.[17] While government officials apologized for the "inconvenience" caused, they never perceived their behavior as possibly being unethical.

IN THIS ISSUE

In this issue of *Cutter IT Journal*, we have assembled four articles that address different choices created by information systems, along with the many ethical questions raised by the algorithms that underpin them. Our first article is by Cutter Senior Consultant Paul Clermont, who explores "the boundary between machine capabilities and what once seemed uniquely human." Clermont provides clarity on the areas where computers and algorithms seemingly have the edge over humans and those where humans are likely to be needed for a long time yet. For example, humans are able to apply common sense when an unexpected situation arises and can be held accountable for their unethical activities.

Next, Darren Dalcher focuses on the need to think deeply about what ethics, trust, and responsibility mean in an age of smart machines. For instance, what does it mean to trust a robot? Should the same level of ethical behavior be expected from a robot as a human? What are the responsibilities of a robot's designers in ensuring that the robot acts in a safe (i.e., reliable, responsible, and ethical) manner? How should policy makers react to smart machines? Do they need to define what is illegal or unethical for smart machines as we do currently for people? Dalcher examines these various questions and more and discusses the societal risk-reward tradeoffs that arise.

Our third article is by Hal Berghel, who takes a look at what he calls "bad faith technology." Technology is generally viewed as being ethically "neutral" — the way it is used defines the ethics involved. For example, a kitchen knife is neither good nor bad in itself, but it can be used for either good or bad purposes. Berghel then asks a provocative question: "is it possible to design a technology with unethical use in mind from the start?" With this as a starting point, he looks at what would characterize bad faith technology and how we can recognize and prevent it.

We conclude the issue with an article by Jesse Feiler, who discusses making ethical considerations a required part of software system development. He describes multiple opportunities for interjecting ethical thinking into system development, such as when stakeholders are initially defining the system, or just as importantly, in the maintenance of existing software systems. Feiler also offers insight into the types of ethical issues that should be considered and practical ways to address them.

I hope you will enjoy the articles in this issue. I think you'll find them especially thought-provoking.

ENDNOTES

[1]Ning, Victor. "China Issues Guidelines on Robotics Development." China Radio International, 27 April 2016 (http://english.cri.cn/12394/2016/04/27/4202s925750.htm).

[2]Bland, Ben. "China's Robot Revolution." *The Financial Times*, 28 April 2016 (www.ft.com/intl/cms/s/2/1dbd8c60-0cc6-11e6-ad80-67655613c2d6.html).

[3]Bremner, Brian. "Japan Unleashes a Robot Revolution." *Bloomberg BusinessWeek*, 28 May 2015 (www.bloomberg.com/news/articles/2015-05-28/japan-unleashes-a-robot-revolution).

[4]Sadauska, Andrew. "South Korean Government Responds to Japan's 'Robot Revolution' with $2.69 Billion Investment in Robotics." *Smart Company*, 28 July 2015 (www.smartcompany.com.au/technology/42997-south-korean-government-responds-to-japan-s-robot-revolution-with-2-69-billion-investment-in-robotics).

[5]Tucker, Patrick. "Every Country Will Have Armed Drones Within 10 Years." *Defense One*, 6 May 2014 (www.defenseone.com/technology/2014/05/every-country-will-have-armed-drones-within-ten-years/83878/?oref=d-skybox).

[6]Rundle, Michael. "Musk, Hawking Warn of 'Inevitable' Killer Robot Arms Race." *Wired*, 27 July 2015 (www.wired.co.uk/news/archive/2015-07/27/musk-hawking-ai-arms-race).

[7]Scharre, Paul. "Autonomous Weapons and Operational Risk." Center for a New American Security, 29 February 2016 (www.cnas.org/autonomous-weapons-and-operational-risk).

[8]Ambasna-Jones, Marc. "The Smart Home and a Data Underclass." *The Guardian*, 3 August 2015 (www.theguardian.com/media-network/2015/aug/03/smart-home-data-underclass-internet-of-things).

[9]"Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights." Executive Office of the President, May 2016 (www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf).

[10]Gates, Guilbert, et al. "Explaining Volkswagen's Emissions Scandal." *The New York Times*, 28 April 2016 (www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html?_r=0).

[11]McBride, James, Christopher Alessi, and Mohammed Aly Sergie. "Understanding the Libor Scandal." *CFR Backgrounders*, 21 May 2015 (www.cfr.org/united-kingdom/understanding-libor-scandal/p28729).

[12]Moor, James H. "Reason, Relativity, and Responsibility in Computer Ethics." *ACM Computers and Society*, Vol. 28, No. 1, March 1998 (http://dl.acm.org/citation.cfm?id=277355).

[13]Migoya, David. "Customers Challenge the Way Banks Reorder Debits in Order to Rack Up Overdraft Fees." *The Denver Post*, 20 August 2010 (www.denverpost.com/ci_15847179).

[14]Hilzik, Michael. "Anthem Is Warning Consumers About Its Huge Data Breach. Here's a Translation." *Los Angeles Times*, 6 March 2015 (www.latimes.com/business/la-fi-mh-anthem-is-warning-consumers-20150306-column.html).

[15]Frey, Carl Benedikt, and Michael A. Osborne. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" Oxford Martin Programme on the Impacts of Future Technology, 17 September 2013 (www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf).

[16]Bland (see 2).

[17]Charette, Robert N. "North Carolina Continues to Wrestle with Wayward IT Projects." *IEEE Spectrum*, 20 December 2013 (http://spectrum.ieee.org/riskfactor/computing/it/north-carolina-continues-to-wrestle-with-wayward-it-projects).

Robert N. Charette is a Fellow with Cutter Consortium's Business Technology & Digital Transformation Strategies practice. He is also President of ITABHI Corporation, a business and technology risk management consultancy. With 40 years' experience in a wide variety of international technology and management positions, Dr. Charette is recognized as an authority regarding IS, IT, and telecom risk management. His experience in both government and business provides a unique perspective on addressing the risk management issues confronting today's business and public sectors. He can be reached at rcharette@cutter.com.

Robots, Algorithms, Ethics, and the Human Edge

by Paul Clermont

Robots have fascinated people ever since the Czech author Karel Capek introduced the word in his 1920 play *R.U.R.*, so worrying about robots and how they'll interact with (or replace) us is nothing new. Early conceptions of robots were eerily anthropomorphic, but the first practical robots were anything but. Introduced to factories in the 1980s to do specific repetitive tasks like welding or painting, they replaced muscle power and resembled disembodied arms with multiple elbows and wrists. Their "intelligence" was provided by algorithms their "masters" — programmers — coded. (Household robots can now vacuum floors and mow grass.)

An early concern about robots was their ability to replace human labor, leaving the question of what unemployed welders and painters would do, a question as old as the first time a beast of burden carried a load that would have required several people. That remains important from a societal perspective (more on this later), but it has recently been eclipsed by bigger questions as the related field of artificial intelligence (AI) has, in essence, begun to robotize tasks requiring brain power applied in prescribed and structured ways, such as lending decisions, legal research, and medical diagnoses. (I have taken the liberty of extending the concept of "robotizing" to the use of algorithms to execute nonphysical tasks once the exclusive province of people.)

Computer scientists in labs and universities have pursued AI for more than half a century, making advances apace with the ever-growing power of computers. They made headlines in 1997 when IBM's Deep Blue defeated the world's greatest chess player. Language translation has evolved from the butt of humor to something serviceable like Google Translate. The latest headline accomplishment was the AlphaGo computer beating the world's best *Go*[1] player about 10 years before experts believed that would be possible. AlphaGo succeeded by incorporating heuristics — not even today's supercomputers have the horsepower to evaluate exhaustively the myriad combinations by brute force — and these heuristics had the ability to improve themselves through trial and error and pattern recognition without human intervention. One somewhat disquieting aspect of AlphaGo's victory is that it made some very unusual, almost unorthodox, moves that would not occur to most human *Go* masters but which proved highly effective. AlphaGo's designer was himself surprised.[2]

Much sensationalist (as well as thoughtful) press has appeared about computers exhibiting such seemingly superhuman capabilities in human pursuits.[3] Impressive as AlphaGo and its ilk are, however, we need not yet fear robots and computers taking over the world. *Go*, like chess, is well suited to AI. It involves complex and rigorous "thought" processes that are natural to computers but make most people's heads hurt. One hundred percent of the relevant information is available. There is no possibility of cheating. The goal is unambiguous and morally and ethically neutral. In other words, it's not like much of real life.

We should be scared — very scared — if and when a computer starts beating Las Vegas poker champions or even our buddies from a friendly game. Poker is only peripherally about cards and odds; it's really a psychological contest in which players try to read their opponents' "tells" while concealing or falsifying their own. In other words, it's like real life distilled down to its most stressful intensity. Such a computer would pass the Turing test with an A+!

THE EMERGENCE OF ROBOETHICS

The purpose of this article is to explore the boundary between machine capabilities and what once seemed uniquely human. That boundary has certainly moved over the years, justifying concerns that the relatively new field of *roboethics* addresses. Roboethics goes beyond job losses and looks at the impact of robotization on society as a whole; that is the major topic here. (I will address job losses at the end.)

An algorithm can be unethical in both obvious and subtle ways. It could be illegal, as may have been the case with Volkswagen's engine management algorithms for its "clean" diesel engines. It could be unethical in the sense that it violates a sense of fair play, as I discuss in the "Overdraft Handling" sidebar.

More subtly, an algorithm could take on decision-making roles that a human is better equipped to play, thereby yielding unethical results. While algorithms are better at minimizing stereotyping and personal prejudice in decision making, and they guarantee thorough and complete data collection and analysis, people still offer critical strengths. I call these the "human edge":

- Nuanced judgment based on circumstances and context, differentiating between situations that are the same only technically or on the surface
- Emotional intelligence and empathy
- Plain old common sense applied when the algorithm produces absurd or unjustifiable results
- Intuition, imagination, and creativity
- A sense of fairness, decency, and the golden rule — the essence of ethics — and the ability to apply it when an algorithm would violate that sense based on data that a human could recognize as stray, incorrect, or irrelevant
- Being accountable for results without the defense that "the algorithm made me do it"

I use the term "edge" in two senses, as both a boundary and an advantage, and I suggest that the boundary will prove robust for a very long time.

A corollary of this is that algorithmic *approaches* don't necessarily involve computers and AI. Consider, for example, mandatory sentencing rules that take over part of the traditional role of judges.

Another conclusion is that algorithms are only part of this emerging discussion because most algorithms depend on data about the situation at hand, plus knowledge developed from large volumes of statistics related to that situation. Data don't just appear; they have to be collected, primarily from us, often without our knowledge. The sheer power of IT to collect, store, transmit, analyze, and distribute exabytes (a billion billion bytes) of data — all of these capabilities growing exponentially — has raised possibilities for abuse and misuse only now imaginable and well outside the scope of laws and regulations developed to address yesterday's issues. Today's data collection can provide real and important benefits to individuals and society as a whole, but we must not ignore the potential for data misuse and abuse (a subject that merits an article of its own).

OVERDRAFT HANDLING

Say you unintentionally write a number of checks that hit your bank on the same day, when the available funds are insufficient to clear them all. Your account might be with any of three generically named banks. Tenth National's algorithm processes the checks in a sequence that minimizes the number of overdrafts, each incurring a typical $35 fee, usually accompanied by the payee's $25 fee for a bounced check. Fifth State processes them in no particular order. Third City processes them in the order that maximizes the number of overdrafts and thus the fees it collects.

A strong argument could be made that Tenth National's algorithm is ethical; it extracts an appropriate penalty but reflects a level of decency. Third City's algorithm, by that standard, is unethical by design; it takes advantage of an opportunity to increase its profits by piling fees onto what are probably its least well-off depositors. Fifth State's algorithm is not by design either ethical or unethical, but when ethics *could* be designed in, the bank's failure to do so borders on unethical.

ALGORITHMS USED AND MISUSED

The following examples show how robotizing activities that call for nuanced thinking, fine judgment, common sense, and/or plain decency is not just unethical, but often absurd and abusive. Some of the examples have nothing to do with computers but still represent robotizing.

Home Mortgages: Then and Now

- In 1969 and 1978, my wife and I were granted mortgages based not just on verifiable data, but by coming across to the banker who interviewed us as serious and responsible young people. The officer had to be careful, since the loan would stay in the bank's portfolio for multiple years. It's possible our results would have been different — even if everything else were the same — had our skin been darker or had my wife been instead a male "housemate."
- In 1993, despite a large down payment and my wife's great job, my self-employment status raised an algorithmic red flag (a false negative in this case). As a result, we had to jump through numerous hoops, some of them ridiculous, and engage a mortgage broker to advocate for us. We never met a single employee of the bank that ultimately granted the

mortgage, but then why would they have bothered, given that they sold it off in a few months?

- The collapse of the housing bubble was largely the fault of false positives — granting huge mortgages to people who had little chance of paying them off. This was driven in part by the extraordinary profitability of subprime loans *if* they are repaid, a very big "if" as it turned out. The incentive structure virtually guaranteed the horrendous days of reckoning we experienced, which almost collapsed the world's financial system.

The $400 Hammer

A number of years ago, a US defense contractor was publicly pilloried by a Senate committee for charging the Air Force $400 for a hammer. By chance, I was consulting with that contractor at the time, and their CFO showed me in detail how that figure was developed using the Defense Department's own algorithms for calculating the price. Strictly speaking, nothing unethical happened, but when an algorithm can create the appearance of unethical behavior, it makes sense for a human to override it, as the contractor thereafter did.

Criminal Justice

During the 1970s and 1980s, when US crime rates were soaring, concerns were raised that some judges were "coddling criminals" by turning them loose with nominal or suspended sentences. In response, both US state and federal legislatures passed minimum sentencing laws, imposing algorithms on judges.

Laws are blunt instruments that address offenses, not individual offenders. The traditional role of judges has been to exercise — yes — judgment, recognizing that two people convicted of the same offense can pose drastically different risks to society if not locked up. The result of these laws is a far higher rate of incarceration in the US than ever in the past or than in any other developed country, largely for nonviolent offenses. The toll of lives ruined by excessive algorithmic sentences fairly screams "unethical."

Sex Offender Registries

Few crimes are more heinous than the torture, rape, or murder of children. Sometimes perpetrators, after finishing even long sentences, do it again. The emotional outcry when this happens is understandable: "How can we protect our children when we don't know these monsters might be lurking right next to a school?" Laws get passed, often named for the victim, requiring that sex offenders no longer in prison register their whereabouts and live a safe distance from possible prey. Unfortunately, in the implementation, some jurisdictions have used very expansive definitions of sex offenses to include consensual sex between minors and even children playing "doctor." A recent *New Yorker* article[4] cited a number of cases where behavior of minors that's at worst undesirable or unwise gets classified as a sex crime, casting a permanent pall over the lives of these minors when their names show up on publicly accessible registries. Even when appeals to common sense have expunged the names from the official registry, there is no requirement for private websites that have copied public registries to update their copies. Even some of the most vocal proponents of registry laws have decried the gross injustices — aka unethical algorithms — built into their implementation.

When an algorithm can create the appearance of unethical behavior, it makes sense for a human to override it, as the contractor thereafter did.

Speed Traps

Soon after automobiles were introduced, it was clear that they could go a lot faster than was safe in the circumstances, hence speed limits. After not too many years, it also became clear that catching people speeding could be a good source of non-tax revenue for local jurisdictions. The advent of radar and lasers — plus quotas — turned speed traps into an industry. Perversely, they were usually placed to catch people in locations where modest speeding would not be dangerous — a technical violation unrelated to the intent of the law. You're caught and fined by algorithm because it's easy, unlike detecting and dealing with driving that's actually dangerous. While the practice is not technically unethical, the bigger picture is more questionable, as scarce police resources are diverted from public safety to collecting government revenue. The police practices uncovered in Ferguson, Missouri, crossed the ethical line unambiguously when African-Americans became a special target of enforcement.

Airport Security

As a frequent flyer with a metal hip, I get to spend a fair bit of time being examined at airports. Not only am I taken aside and 100% wanded to locate the metal and

assure its harmlessness, which is perfectly reasonable, I am also patted down, a "service" only rarely offered to people who don't set off the metal detector. Two algorithmic approaches crash head-on here. One is avoiding even the appearance of ethnic profiling, devoting the same level of attention to all, regardless of any clues about the likelihood of being an actual terrorist (such as age and, yes, ethnicity). The other algorithm would favor efficient use of security resources, which would argue that clues be taken into account when determining the level of attention. Recent test findings showing that contraband is still getting onto flights with an alarming frequency suggests the balance between the competing algorithms may be a bit off.[5]

The threat comes from robotic and algorithmic approaches to situations where the human edge is critical to ensuring results that are fair and beneficial to individuals and society at large.

"Zero Tolerance"

Zero tolerance policies, often applied in schools to students based on behavior or possessions, bespeak a no-nonsense approach: no exceptions, applying equally to all, and so on. Unfortunately, this rejection of human discretion can lead to absurdities such as a five-year-old being accused of making a "terroristic threat" for talking in the bus line about her Hello Kitty bubble gun.[6] Clearly such policies should sound an alarm to anyone who seriously ponders the ethics of algorithms. They are a flat-out denial that the human edge can add value, and they're coming under critical scrutiny.[7]

WHAT TO DO

Roboethics owes its existence as a new discipline to robots and algorithms, but these are not themselves the real ethical threat. Rather, the threat comes from robotic and algorithmic *approaches* to situations where the human edge is critical to ensuring results that are fair and beneficial to individuals and society at large. Computers may or may not be involved; it's the approach that matters. Addressing the threats needs to happen at multiple levels.

Public Policy

- Only legislation or judicial decisions can deal with existing laws such as mandatory minimum sentences or the overinclusive definition of sexual offenders. This means recognizing that justice is not the same as law enforcement. No matter how necessary or well intentioned, a statute cannot make the fine distinctions that justice calls for if lives are not to be unnecessarily blighted.
- Governments need to embrace the notion that fines should be levied as punishment for infractions with the goal of minimizing occurrence of those infractions — not as a source of predictable revenue. (Good luck with this!)
- Unethical algorithms need to be exposed and dealt with by, for example, consumer protection agencies.
- New laws should better protect whistle-blowers who call out ethical issues with algorithms.
- New laws should mandate that third-party repositories of official data keep their copies of that data up to date when the official source changes, with penalties for failure to do so.

Media, Watchdog, and Advocacy Groups

Such organizations can play a constructive role by highlighting laws that result in unethical outcomes so as to generate popular support for change. They can also play a part in naming and shaming businesses that deploy unethical algorithms, such as Third City's handling of overdrafts, with the goal of banning them. By building awareness, such publicity makes it worthwhile for better-behaved companies like Tenth National to incorporate their "code of ethics" into their marketing. (Refer back to sidebar.)

Businesses and Governments

Businesses and governments need to remove robotic algorithms from jobs where the human edge matters. Algorithms can be tremendously helpful in decision making up to making recommendations, but not actually deciding in cases where the human edge plays an important role in ensuring fairness and applying common sense. Explicit liability for bad robotic decisions is needed.

These entities also need to recognize that as algorithms become more sophisticated, they may generate unpredictable results à la AlphaGo. This suggests a need for the equivalent of the nuclear industry's containment vessels[8] to avoid algorithms going out of control, as may have contributed to the home mortgage meltdown in 2007-8.

Military Policy

To the extent that autonomous weapons replace physically present soldiers who have clear visibility into the scene (the "fog of war" notwithstanding) such that they can exercise judgment, common sense, and decency, robots as soldiers would be another example of unethical use of algorithms. Robots fighting robots sounds like a lot of games and films.

IT Practitioners Need a Code of Ethics

Under an IT code of ethics, practitioners would:

- Refuse to participate in illegal IT (e.g., VW's emission test–cheating software)
- Call out attempted misuse based on robotizing activity where the human edge is critical
- Call out algorithms that offend standards of human decency in pursuit of profit (e.g., Third City's overdraft handling)
- Establish "containment vessel" processes for recognizing unpredictable and possibly erroneous algorithmic outcomes in time to enable human intervention
- Avoid premature public release of applications when the likelihood of problems adversely affecting users is more than very low (unless accompanied by explicit warnings and waivers)
- Ensure the security of sensitive personal data

Is this idealistic? Of course. Companies are not democracies. IT professionals have mortgages to pay and children to educate, making pressure to build something of dubious ethics extremely difficult to resist. When whistle-blowers reveal that they were asked — or, more accurately, told — to do something illegal, they may have the satisfaction of knowing they did the right thing, but too often at great cost to their careers and their families.

Just Because We Can

… should we build it? Concerns over new technologies can be overblown by the media and politicians, but they should not be reflexively dismissed. Yes, such concerns can slow down innovation, but that is not necessarily a bad thing. DDT and thalidomide did their intended jobs beautifully — but then we saw their devastating side effects. The pressure to move fast is particularly intense in IT, where speed to market is critical and tech executives with a libertarian bent want governments and public interest groups to stay out of the way. That doesn't mean every idea should proceed at full throttle, though, assuming nasty flaws will be kind enough not to materialize. When members of the public could be adversely affected by things going wrong, prudence and caution are in order.

WHAT ABOUT THE JOBS?

Every technology in human history has destroyed some jobs. They have also created new jobs, requiring more skill, for people who improve, manage, and operate the technology. With respect to jobs — not social impact — robots and AI are simply the latest manifestation of this phenomenon. In this article, I argue that we are far, perhaps a bridge too far, from computers that truly mimic the capabilities I've lumped together as the human edge. What we can expect is that jobs requiring *no element* of the human edge (e.g., welding and painting on an assembly line) will continue to disappear, along with even more skilled work (e.g., a lot of legal research and basic accounting) as clever technologists figure out how to do it by machine. But the jobs won't disappear entirely. Some may simply never be economical to robotize, like a lot of housework. (Who knows, we may see a resurgence in the use of live-in domestic servants, in one stroke filling up the McMansions and relieving the housing shortage in affluent areas!)

When members of the public could be adversely affected by things going wrong, prudence and caution are in order.

Jobs requiring the human edge will not disappear and in fact could increase as people take back work that has been overly delegated to algorithms. Naturally, the more that jobs require multiple elements of the human edge, the more interesting and emotionally (and financially) rewarding they become, as the element of scut work is minimized by the use of algorithms.[9] For example, in the financial industry, the labor-saving benefit of algorithms that analyze large volumes of data will be offset by time spent doing more analyses and improving the algorithms.

Concerns about job displacement are legitimate, and the labor market left to its own devices may not be a match for the dislocation. Society as a whole has to ensure fair and responsible solutions, but that's above the pay grade of technologists.

IN CONCLUSION

Where technologists need to play a real role, both in terms of actions and advocacy, is in dealing with the ethical issues we face as IT ever more thoroughly infiltrates our whole lives, not just our jobs. Future technologies — perhaps prodded by laws and regulations and social pressure — may help address some of the ethical concerns we have about today's technologies, but they will likely add new concerns. Responding to them requires clear thinking about the encroachment of computers and algorithms into the realm of the human edge. In this article, I have tried to contribute to that clarity.

ENDNOTES

[1]*Go* is an ancient game from Asia played on a 19x19 matrix. Players have either black or white stones that they place to mark territory and capture their opponent's stones. It seems simple until you try to play it!

[2]Millar, Jason. "The Momentous Advance in Artificial Intelligence Demands a New Set of Ethics." *The Guardian*, 12 March 2016 (www.theguardian.com/commentisfree/2016/mar/13/artificial-intelligence-robots-ethics-human-control).

[3]Lanchester, John. "The Robots Are Coming." Review of *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies* by Erik Brynjolfsson and *Average Is Over: Powering America Beyond the Age of the Great Stagnation* by Tyler Cowen. *London Review of Books*, Vol. 37, No. 5, 5 March 2015 (www.lrb.co.uk/v37/n05/john-lanchester/the-robots-are-coming).

[4]Stillman, Sarah. "The List." *The New Yorker*, 14 March 2016 (www.newyorker.com/magazine/2016/03/14/when-kids-are-accused-of-sex-crimes).

[5]Jenkins, Brian Michael. "TSA Flunked Its Security Test Big Time — Now What?" *The Hill*, 15 June 2015 (http://thehill.com/blogs/pundits-blog/homeland-security/244957-tsa-flunked-its-security-test-big-time-now-what).

[6]Edwards, Meridith. "Pennsylvania Girl, 5, Suspended for Talk of 'Shooting' a Hello Kitty 'Bubble Gun.'" CNN, 22 January 2013 (www.cnn.com/2013/01/21/us/pennsylvania-girl-suspended).

[7]Berwick, Carly. "Zeroing Out Zero Tolerance." *The Atlantic*, 17 May 2015 (www.theatlantic.com/education/archive/2015/03/zeroing-out-zero-tolerance/388003).

[8]I owe this analogy to my former Nolan Norton colleague Bruce Rogow.

[9]Just as I got senior enough to delegate spreadsheet work, VisiCalc came on the market!

Paul Clermont is a Senior Consultant with Cutter Consortium's Business Technology & Digital Transformation Strategies practice. He has been a consultant in IT strategy, governance, and management for 30 years. Before going into individual practice, he was a Principal with Nolan, Norton & Co., a boutique consultancy that became part of KPMG. His clients have been primarily in the financial and manufacturing industries, as well as the US government. Mr. Clermont has spoken and written and taught executive MBA courses on the challenges of getting significant and predictable value from IT investments. Before joining Nolan, Norton & Co., he directed IT strategy at a major Boston bank and launched its IT executive steering committee. His undergraduate and graduate education at MIT's Sloan School of Management was heavily oriented toward operations research. Mr. Clermont is based in San Francisco and can be reached at pclermont@cutter.com.

Rise of the Robots: Rethinking Ethics, Trust, and Responsibility in the Age of Autonomous Machines

by Darren Dalcher

Hosting the Olympic Games is viewed as an opportunity to celebrate achievements and showcase new technologies. In October 1964, prior to the Tokyo Olympic Games, the revolutionary Maglev bullet train was unveiled as the fastest train in the world. Throughout their nearly 52 years of operation, the trains have often traveled through typhoons and earthquakes, using sophisticated detection and alarm systems to transport over 10 billion passengers without suffering a single incident involving loss of life.

To celebrate the return of the Olympic Games to Tokyo in 2020, visitors will be introduced to a robotic experience. Over 1 million visitors to the Odaiba district will be hosted in a futuristic village where robots will hail taxis, fetch luggage, administer check-in desks, operate hotels, offer instantaneous translation services, and ferry visitors to their destinations. According to Japanese Prime Minister Shinzo Abe, Japan is even planning to stage a Robot Olympics alongside the summer games. Whilst the games may once again offer a glimpse into the future of a new technology, it is a controversial future that demands a greater trust in autonomous robots and their ability to make safe and ethical decisions.

This article explores and repositions ethics, trust, and responsibility in the age of autonomous machines. It begins by posing new technology-induced ethical dilemmas and suggesting that there are different moral resolution systems. Frankenstein's monster and Asimov's Laws of Robotics represent early attempts to come to terms with the ethics of autonomous robots, yet the role of humans and their ability to intervene invite ethical considerations about the impact of autonomous machines. The new uncertainties that come with the use of innovative technologies require paying more attention to the responsibility of designers for their creations. The article concludes by identifying different levels of responsibility and proposing a finer balance between trust, safety, and reliability as well as considering the interaction between users, designers, and their creations.

WE'VE GOT A TICKET TO RIDE

Modern technological advances continue at an unprecedented pace, proudly displaying greater autonomy and decision-making skills embedded into an expanding range of technologies, including artificially intelligent robots, self-driving cars, drone delivery, ubiquitous mobile supercomputing, implantable technologies, and smart cities. As the new technologies are deployed to undertake tasks as diverse as eliminating enemies, delivering parcels, and educating, entertaining, driving, guiding, satisfying, and informing us, they also take over a growing number of repetitive and dangerous duties and chores that humans previously handled. However, when such technology is given full autonomy for making decisions, it can also introduce a new kind of computer-assisted error, where a system designed to make us safer is directly responsible for causing an accident. In abrogating responsibility for mundane decisions to new technologies, we are increasingly relying on their ability to deal with risk, uncertainty, ambiguity, and the greater unknown.

Japan has long viewed robots as a major pillar of its economic growth strategy and an important aid for a rapidly aging society. In preparation for the 2020 Tokyo Olympics, Robot Taxi is field-testing its new driverless taxi service. Starting in March 2016, 50 residents of Fujisawa, known as Japan's first sustainable smart city, are regularly being driven between their homes and the city's supermarkets, some two miles away. The autonomous cars combine GPS, radar, stereovision cameras, and image analysis systems to navigate around town. Successful trials are expected to lead the way to the use of thousands of robot taxis to ferry spectators around the venues of the 2020 games.

Capable robots with 360-degree vision, full awareness of the environment, and perfect driving skills may yet force a redefinition of humans in cars as cargo. But until such precision instruments replace all drivers, driving will require interaction with other road users, which

may often mean flashing lights, gesticulating, or making eye contact. However, as future human cargo, we should all be interested in the choices made by autonomous driverless cars. As the car you are transported in careers toward a junction at the bottom of a hill, whilst gathering speed as the brakes fail, would you expect your autonomous taxi to come to a halt after running down the elderly lady standing at the bus stop, or would you prefer that it head straight for the five young men in the open area by the town square? Would the decision change if the woman at the bus stop is your beloved grandmother? If one of the young men is your youngest son? At any rate, should you the passenger care as long as you are kept safe?

Future designers may well be expected to categorize the overriding value system underlying the logic of their creation, so that resident philosophers can interpret its behavior and potential intentions.

WHOSE MORAL REASONING?

Moral philosophers have utilized the "trolley problem" thought experiment to debate choices and uncover the circumstances under which it is acceptable to harm others. This dilemma questions whether it is equally acceptable:

- To divert a trolley about to kill five people on a train track toward a single individual on another track in order to save the five
- To push a single "fat" man onto the track to save the five

Different formulations can be used to highlight the modern quandaries created by robotics and new technology advances.

The human responses to such trolley dilemmas typically highlight the difference between consequentialist and categorical moral reasoning. Consequentialists focus on the results of an action to determine whether it is right or wrong, implying a need to maximize good — or "less bad" — results. Categorical moralists take issue with each act, investigating its appropriateness, arguing that some actions are categorically wrong. The former position locates morality in future consequences, while the latter takes issue with the act itself, thereby locating morality in certain duties and rights and being guided by rules and absolutes. In contrast, emphasizing harm avoidance, harm minimization, or utilitarian maximization may lead to very different outcomes.

In June 2010, the US military lost control of a helicopter drone for over 30 minutes and 23 miles as it swerved toward Washington, DC, potentially threatening the White House and other civil and military assets in direct contravention of established airspace restrictions. Relevant agencies would certainly have benefited from knowing whether it was armed with missiles or parcels, as well as what moral system it might be deploying. In programmed artifacts, the system of preference needs to be coded and acknowledged, as the different moral systems each invoke a particular method of reasoning about safety, responses, and consequences. Future designers may well be expected to categorize the overriding value system underlying the logic of their creation, so that resident philosophers can interpret its behavior and potential intentions.

FROM FRANKENSTEIN TO ASIMOV

Consideration of the role of ethics in emerging technologies is not new. Fear of dealing with robotic creations and their unpredictable behaviors has repeatedly been featured in literature. Mary Shelly's *Frankenstein*, often used as an allegory for the folly of scientific experimentation, actually tells the story of an "assembled" powerful creature — capable of extreme and destructive violence but who also learns to speak, secretly cares for a poor family, reads literature, and yearns for a soul mate — and his struggle to reconcile power, autonomy, and feelings.[1] It can also be read as a commentary on an irresponsible creator who fails to recognize and embrace his responsibilities to his creation and society at large.

In the 1950s, Isaac Asimov formulated the Three Laws of Robotics:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.[2]

Asimov's writing is concerned with the safe behavior of autonomous robotic machines and their greater impact on individuals and society. His fictional stories explore the dilemmas of unexpected events, counterintuitive

behavior, unexplored boundary conditions, and the unintended consequences of applying such laws.

Ethics is often slow to catch up with technological developments. A key question that emerges from the writing of both Shelley and Asimov is whether machines can act as moral agents. If robots are to take on added autonomous roles, they must be programmed with moral decision-making responsibility — but whose morality do they take on?

Shelly's dystopic tale of demonized technology emphasizes the result of execution gone awry, in a consequentialist tradition. Meanwhile, Asimov's use of the three laws as a literary device exemplifies the difficulty in enforcing a categorical value system through the use of absolute rules and prohibitions.

HUMANS, FEEDBACK, AND STAYING IN CONTROL

In many human endeavors, intelligent automation is replacing some of the tasks and roles traditionally performed by human agents. The key reasons for employing intelligent technology are superior computational capability, elimination of human error, reduction in work overload, and increased dependability. However, replacing humans may increase system vulnerability, especially with regard to unanticipated perturbations, which cannot be foretold or specified.

Reliance on autonomous machines requires *total* trust in the ability of the system to make safe or rational decisions (unless they encompass human override capabilities and/or human overseers). Replacing human decision makers with "responsible" mechanical alternatives would therefore presuppose that all possible failure modes had been foreseen and that specific actions were included to mitigate their effects. Given that many contexts cannot be fully specified in advance, solutions often need to evolve through experience by dynamically consulting domain-specific experience. Artificial intelligence (AI) proponents contend that autonomous cars are able to learn from incidents, and the resulting corrections and changes then apply to the entire class, not just to a single autonomous agent. In other words, all cars become smarter following an accident. Yet overreliance on technology often results in ignoring the human element.

It is expected that visitors to the 2020 Olympic Games will be staying in robot-staffed hotels. The first such hotel, Henn-na, said to be the first hotel totally run by robots, opened in Nagasaki in July 2015. The hotel is operated by robot receptionists (with a choice between English-speaking dinosaurs or Japanese-speaking female androids), robot porters, and other electronic creatures, coupled with facial recognition technology and a multitude of sensor panels. Check-in is available from 3 pm; visitors who arrive early and attempt to engage the robots will encounter a human who comes out of his small room to announce that the machines will become operational at 3 pm. Even a complete system may thus require occasional corrective intervention.

Ultimately, if a function is to be automated, then the system must be supplied with enough variety and control to cope with any situations that might arise. In order to generate suitable control responses to address unexpected conditions, every controller must be provided with:

1. Sufficient control responses, *and*
2. Decision rules for generating all the control responses, *or*
3. Authority to become a self-organizing system in order to respond to unexpected events, *or*
4. A resident human ready to address lapses

If robots are to take on added autonomous roles, they must be programmed with moral decision-making responsibility — but whose morality do they take on?

SHOULD DESIGNERS ANTICIPATE SURPRISES?

Inherent uncertainties and unexpected external conditions can lead to surprises, often necessitating urgent interventions. Yet, as we will see, adding interventions may itself lead to complications.

On 1 June 2009, an Airbus A330 equipped with the latest "glass cockpit" controls entered an aerodynamic stall from which it could not recover and crashed into the Atlantic Ocean, killing 228 passengers and crew members. The wreckage of this Air France flight from Rio de Janeiro to Paris was discovered five days later near Saint Peter and Saint Paul Archipelago in the central equatorial Atlantic Ocean. The accident report concluded that the crash occurred after temporary inconsistencies between the airspeed measurements caused the automatic pilot to disconnect. Crew members struggling to regain control in a sudden emergency misread the situation and reacted incorrectly, ultimately causing the aircraft to enter the fatal stall.

Imposition of new technology can also change the balance in the environment itself and introduce potential breakdowns in communication. Indeed, David Woods of Ohio State University, an expert on human interaction with technology, suggests that automation can potentially make flying more difficult for pilots.[3] British aviation expert David Beaty also documents typical automation errors that have precipitated accidents,[4] while other aviation experts[5, 6] observe that the modern cockpit has solved a great many problems but created some new ones as well, a direct parallel to driverless cars. Beaty believes that pilots are increasingly being pushed out of the control loop.[7] In the Airbus 330 accident, bringing pilots back into the system in an emergency may actually have escalated the failure scenario.

Robots are only as reliable as the systems and people that build them.

Difficult risk issues are rarely about risks alone, as events pertaining to hazards interact with psychological, social, institutional, and cultural processes in ways that heighten or attenuate perceptions of risk and shape behavior. News media reports, personal biases, and societal conventions may demand (or impede) protective actions and lead to responses ranging from ignoring a risk to mass hysteria, especially when we interact with unknown/uncertain systems or innovative technology. Charles Perrow of Yale University has asserted that given the interactive complexity and tight coupling characteristics of certain systems, "normal accidents" (failures caused by multiple and unexpected interactions between humans and these complex, tightly interconnected technologies) are inevitable.[8]

In normal accidents:

- Operators are confronted by unexpected and mysterious interactions among failures (so anticipation is of limited use with complex and interactive technologies).
- Great events have small beginnings.
- Organizations and management play a major part in causing (and preventing) accidents and failures.
- Fixes, as well as safety devices, add to the inherent complexity and hence to the likelihood of accidents.

Components can thus affect each other unexpectedly and are also capable of spreading problems. Adding safety components may increase the range and scope of potential interactions and therefore the number of potential ways for something to go wrong. In other words, safety interventions can redistribute the burden of risk rather than reduce it. This redistribution may be unpredictable and uncontrollable, suggesting that shifting risks may be more dangerous than tolerating them.

A QUESTION OF RESPONSIBILITY?

The deployment of new technologies has always invoked concerns regarding their potential harm and their impacts on humans, civil society, and the wider environment. However, the rapid growth of automation, AI, and machine learning is raising important new questions about the moral responsibilities associated with using such technologies.

Indeed, if a machine is no longer a tool or instrument used by a human agent, whose morals and conventions does it follow? Moreover, given the autonomy invested in such systems, how are the ethics and rules programmed? Considering the profound uncertainty, complexity, and interconnectedness of such technologies, the inability to predict all potential future outcomes, and the disruptive potential of inconsistencies (as exhibited in the Airbus crash), who bears ultimate responsibility for the impacts of such interaction between technology and society?

Just because we can design all types of programs does not necessarily mean that we *should*. Given the safety implications of the artifacts being delivered, it is important to ensure that the impacts of change are analyzed and understood prior to deployment. Driverless cars, drones, and other autonomous creations are likely to transform and revolutionize roads, delivery systems, and most other entities they interact with. They demand extensive thinking up front when the business case is proposed, significant testing and adjustment periods during development, and education and preparation of stakeholders throughout.

It is therefore important to promote responsibility for the autonomous nature of new technology. Robots are only as reliable as the systems and people that build them. Developers are integral to shaping the new technological revolution, and they need to be held to account for their role in delivering it. Ultimately, when it comes to execution, systems will follow the

programmed instructions that comply with the system specification. The tricky part is to make the system comprehensive enough to cover all eventualities and to exhaustively test it to ensure it is safe. Consumers need to know that it is reliable and trustworthy, or that additional safeguards have been included.

It is difficult for policy makers and consumers to keep up with rapid developments in autonomous technology and robotics, yet it is crucial to make certain that consumers are protected. Whether we recognize it or not, assumptions related to risk and uncertainty are embedded into all the artifacts that we develop. If we send our children to school in a driverless taxi, we would like assurance that it will minimize risks on the journey. Does that mean that before entering a junction or joining a roundabout, the driverless car will wait forever for absolutely risk-free entry? Alternatively, do we want to specify an acceptable level of risk, or engage a philosopher to identify the ideal system for engaging with risk?

Asimov explicitly commanded robots to do no harm, but that act requires recognition of harm and consequences. Indeed, why assume that robots will seek to inflict damage in the first place? The trolley dilemmas show that when we reason about potential hazards, we need to understand the worldview of developers and determine if they are looking at situations as consequentialists or categorical moralists. In the same way that assumptions encapsulate certain decisions that flow and impact actions downstream, moral positions can also color perceptions, limit options, and lead to particular courses of action. Asimov's command indicates recognition of the importance of the thinking system embedded in designed systems. Yet responsibility for the actions of designed systems ultimately lies with the designers and promoters of new technology.

While responsibility entails owning up to acts, effects, and consequences, one can discern different types of responsibility:

- **Causal responsibility** — bringing something about either directly or indirectly (e.g., by ordering someone else to do it).
- **Legal responsibility** — fulfilling the requirements for accountability under the law.
- **Moral responsibility** — having a moral obligation or fulfilling the criteria for deserving blame or praise for a morally significant act, or omission, and the resulting consequences.
- **Role responsibility** — performing duties that are attached to particular professional, societal, or even biological roles. Failure to fulfill such duties can expose the role holder to censure, which can be moral, legal, or constitutional.

Moral responsibility normally assumes some degree of causal responsibility. Therefore, a professional can be held morally responsible for failing to act. When we take control from human experts, such as pilots, and offer it to machines, we redesign the responsibility equation. Thus, developers could potentially bear causal, legal, and moral responsibility for events. They may also be held accountable under the obligations of role responsibility. Indeed, as we engage with new technologies, apportioning responsibilities may become a key activity.

DESIGNING FOR TRUST

Can we trust new technologies? Trust is often established on the basis of the reliability of a system. This is a problem for new systems with no known track record. It is also a problem when safety features are added and are expected to operate in concert with existing components. Moreover, as noted above, the addition of new safety features can also decrease the reliability of a system by introducing new modes of failure. With safety viewed as an emergent property of an entire system, the relationship between the different properties can be depicted as in Figure 1. Trust in the safety of a system would require fundamentally different tests from those needed to establish trust in the reliability of the individual components. Safety is established at the system level; tests would be required to address the interaction between the system (including the different components), its users, and the wider environment. Such tests can also encompass consideration of completeness, emergent behavior, and the ability to intervene and change course.

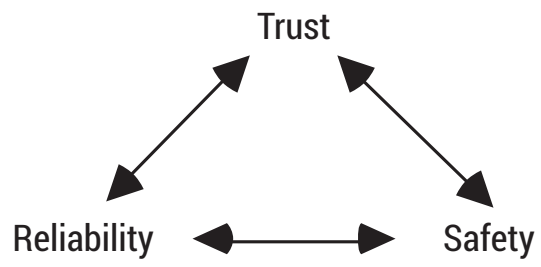


Figure 1 — Reliability, safety, and trust: relationship between different properties of a system.

Some of the new autonomous technologies seem to require a maxim of "trust until proven otherwise," but this is a dangerous position. System safety depends on the interaction between components (rather than on past history) as well as considering the place of humans within the system. Trust in a system needs to be built on the basis of taking *all* aspects and their relationships into account, prior to the release of the technology.

Trust is not simply a function between the client and the product system. It is a complex mechanism that involves the developers and should balance the different concerns and account for different ways (according to the different moral reasoning systems) of considering consequences, impacts, and the range of permitted and forbidden operations and interactions. A new model depicting the dynamic relationships is shown in Figure 2.

The model suggests that trust applies to the relationships between the user and the developer, the user and the product, and the developer and the product (depicted as the outer arrows). Developers have a direct influence over the safety level of the products and systems they devise. Users are expected to place their trust in systems and products, but in order to do so, they must build trust in the innovators, designers, and developers. Track record and reputation shape the interaction and the resulting adoption rates of new technologies and are therefore also likely to influence the way developers and inventors are perceived. Developers discharge their responsibility to their client (hopefully fostering additional trust) by developing trusted products.

Risk management is central to balancing and trading off among acceptability, responsibility, and safety levels, thereby enabling direct relationships between risk and acceptability, risk and responsibility, and risk and safety. For example, rather than insist on absolute safety, the user can view safety as a measure of the acceptability of some degree of risk. Decisions about acceptable risks associated with new technology may recognize that while the technology is novel, and the *safety* track record associated with it remains unclear, the developers have built a reputation for *responsibility*, which has been generally *accepted* by user groups. In this way, users can trade off some of the unknowns about the technology with the knowns about those responsible for delivering it. In other words, trust can leverage other aspects of responsibility, safety, and acceptability, as depicted in Figure 2, to facilitate tradeoffs. Knowledge of the moral reasoning applied (i.e., categorical or consequentialist assumptions and the likely implications) may feed into the risk management calculations and can be reflected in the balance among the attributes of safety, responsibility, and acceptability associated with a new technology.

Trust and acceptability are also coupled with the ability to control systems, hazards, and risk levels. Systems delivered to users should combine the elements of trust and acceptability with an agreed level of control. Wresting control from operators and users should imply attaining their full trust in the system (or providing sufficient and acceptable modes of recovery). The model thus offers a new way of reasoning about the

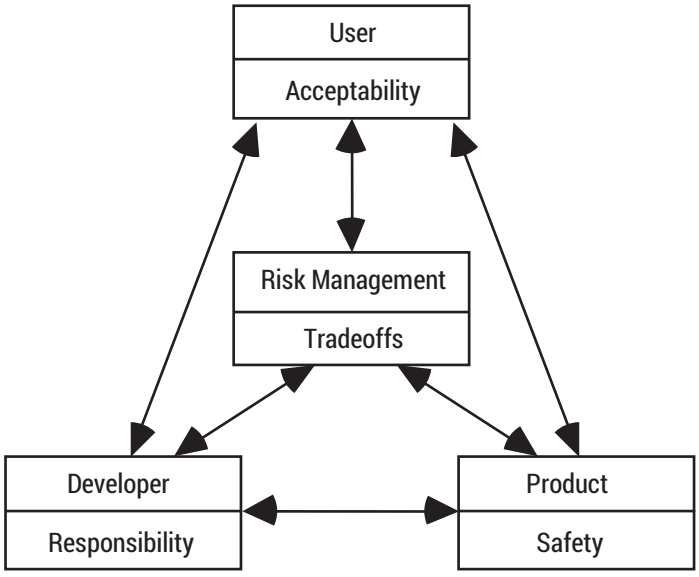


Figure 2 — The dynamic balance of trust.

adequacy of designed systems. The colocation of the Robot Olympics alongside the main games is a clever attempt to enhance the acceptability of new technology by spectators who are already staying in automated hotels and traveling in Robot Taxis. Having experienced the safety of the transport and accommodation system, they can begin building up trust in the responsibility of the designers and the impact of the technology itself. Indeed, the 2020 Tokyo Olympics represent a major trust-building exercise in the future of robot technology.

TRUST REPRISED

Trust is fragile: it is created slowly and can be destroyed in an instant. Trust builds up over time as a result of complex, multidimensional interactions.

To establish trust, there is a need for designers to take their role into account. Ultimately, the responsibility for a developed system lies with the developer. Some 3,570 years ago, Hammurabi, King of Babylon, recognized the perils of design and enacted a building code that clarified the responsibilities of designers:

> If a builder has built a house for a man and his work is not strong, and if the house he has built falls and kills the householder, that builder shall be slain.

While the sentiment may seem harsh, it is useful to apply a personal test to new technology. Would you place your child in the hands of the new technology you are about to design, sell, or commission?

The final word on the topic is reserved for US astronaut Alan B. Shepard, who observed, "It's a very sobering feeling to be up in space and realize that one's safety factor was determined by the lowest bidder on a government contract." As we embark on our own journey into the realms of uncharted technology that will transform our future, we should humbly join Shepard in reflecting on the potential impact of a scary new technology, our limited knowledge of its working, and the trust that we must engender in its ability to do good (or at least to do the least harm).

ENDNOTES

[1]Shelley, Mary Wollstonecraft. *Frankenstein; Or the Modern Prometheus*. Lackington, Hughes, Harding, Mavor, & Jones, 1818.

[2]Asimov, Isaac. *I, Robot*. Gnome Press, 1950.

[3]Woods, David D. "The Cognitive Engineering of Problem Representations." In *Human-Computer Interaction and Complex Systems*, edited by George R.S. Weir and James L. Alty. Academic Press, 1991.

[4]Beaty, David. *The Naked Pilot: The Human Factor in Aircraft Accidents*. Crowood Press, 1995.

[5]Wiener, Earl L., and David C. Nagel (eds.). *Human Factors in Aviation*. Academic Press, 1988.

[6]Owen, David. *Air Accident Investigation*. Patrick Stephens, 2001.

[7]Beaty (see 4).

[8]Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*. Basic Books, 1984.

Darren Dalcher is Professor of Project Management at the University of Hertfordshire and is the Founding Director of the National Centre for Project Management, an interdisciplinary center of excellence operating in collaboration with industry, government, charities, NGOs, and the learned societies. He has been named one of the top 10 "movers and shapers" in project management by the Association for Project Management and has also been voted Project *magazine's Academic of the Year for his contribution in "integrating and weaving academic work with practice." Dr. Dalcher has built a reputation as a leader and innovator in the area of practice-based education and reflection in project management and has worked with many major industrial, commercial, and charitable organizations and government bodies. He is active in numerous international committees, steering groups, and editorial boards. Dr. Dalcher is heavily involved in organizing international conferences and has delivered many keynote addresses and tutorials. He has written over 200 refereed papers and book chapters on project management and software engineering. He is Editor-in-Chief of* Journal of Software: Evolution and Process *and Editor of the* Advances in Project Management and Fundamentals of Project Management *book series published by Routledge. Dr. Dalcher is a Chartered Fellow of the British Computer Society and a Fellow of the Chartered Management Institute and the Royal Society of Arts. In 2011, he was awarded a prestigious Honorary Fellowship from the Association for Project Management for his outstanding contribution to project management. He can be reached at d.dalcher2@herts.ac.uk.*

Bad Faith Technology

by Hal Berghel

When discussing the morality of technology and its use, people tend to fall into two camps. On one side of the debate we have "anthropomorphists," who hold that technology can easily take on a moral character. They cite the atom bomb and Nazi gas chambers as examples of inherently immoral technology. On the other side are "amoralists," who claim that technology is ethically neutral and that the only ethical considerations derive from its actual use. This position has been likened to a soft form of social determinism whereby one would look to the social and cultural contexts in which a technology arises for ultimate judgment on its value and effect. In this account, it is the *use* of the technology in context that takes on moral character: it makes no more sense to attribute morality to technology than it does to rocks and fallen timber. However, there is an important intermediate case that both sides overlook: *bad actor technology* or technology offered in *bad faith*. In this article, I will try to expand upon the last clause of Kranzberg's First Law of Technology: "Technology is neither good nor bad; nor is it neutral."[1]

Those who claim technology is ethically neutral tend to focus attention on the underlying processes, to the exclusion of the original motivation. While I lean in this direction,[2] I recognize that bad actor/bad faith technology provides an important exception.

To amoralists, judging technology in ethical terms is a category mistake — it is as if we adjudge shoes by the paths traveled. However, in order to deal adequately with bad faith technology, we need to inject some measure of substantive assessment and recognize that as with due process of law, we must not ignore the circumstances that give rise to technology and the intention of the developers. As human artifacts, technologies must be evaluated in the social and political contexts in which they are embedded. An example or two will make the concepts clearer.

BAD FAITH TECHNOLOGY

Let's return to the original position described above — that some technology is unethical. Zyklon B, Nazi death chambers, the atomic bomb, land mines, torture devices of sundry stripe and form, and chemical and biological weapons are frequently used as examples of immoral technology. This turf is, of course, both slippery and slopey. Many lethal substances only become so when removed from natural settings (e.g., extracting ricin from the castor bean). This is usually not of practical concern, as intentions are normally clear. For example, there aren't too many biological weapons that I know of that were originally intended to be food flavorings or medicines. On the other hand, we can't blame the bean for the ricin.

So what would be our criterion for judging a technology as unethical? It would appear that we have to go beyond mere use to intention. If a technology itself is blameworthy, it derives its immoral status from the inception of the idea and whatever intention the designer had in mind. It is only in this way that we may separate technology with harmful effects from those that are legitimate candidates for immorality. If we don't take this step, it would appear that nuclear physicists, biological chemists, and weapons manufacturers would all have to share some responsibility for the ultimate effects. This is not to deny that there are people who *do* hold such positions; people who argue that ammunition manufacturers are partly responsible for homicides come to mind. However, in this article, we're going to focus on specific technologies and not deal with claims of inherent immoralities by category.

Let's frame our question thusly: is it possible to design a technology with unethical use in mind from the start? Phrased in this way, the inclination is to assent, yet examples that might qualify for the label of unethical may be difficult to find. What might qualify? I'll suggest that they will be technologies that (a) are inherently capable of being used in ways that society would adjudge unethical, immoral, or illegal, and (b) that the full intent of the designer(s) was not disclosed to the stakeholders (users, customers, stockholders, regulators, etc.) when the technology was developed. In this sense, we may say that the technologies were developed covertly. With these two conditions, we can still maintain the ethical neutrality of

the controversial technologies mentioned above, as all of the stakeholders were fully aware of their intended use.

WINNER'S THESIS

So what are purported bad faith technologies? As it turns out, political scientist Langdon Winner has given considerable thought to this topic.[3] He observes that "Unfortunately, a great many of the technical devices and systems that surround us are designed, built, and deployed in flagrant disregard of humane principles.... includ[ing] the waste of material resources; the destruction of living species and ecosystems; pollution of the air, land, and water; surveillance as a means of social control; and militarism as first response to disagreement and conflict."[4]

Winner is clearly correct in criticizing technology boosters who "have insisted that the 'biggest and best' that science and industry made available were the best guarantees of democracy, freedom, and social justice."[5] Today these boosters seem to champion the Internet as the liberating technology flavor of our time. Some have even given a name to the enthusiastic belief in technology's power to liberate: the Google Doctrine. In what is arguably the single most important piece of scholarship on this topic, Evgeny Morozov easily disposes of this naive doctrine as another case of foolish technopomorphism that seeks to imbue technology with such qualities as intention, resolve, purpose, and single-mindedness.[6] Morozov justifiably challenges the bogus claims of social media's liberating power in Iran's 2009 "Twitter Revolution." With the passage of time, we now see that Twitter and all of its Internet siblings had little enduring effect on the power elite — in the end tyranny, not Twitter, won the day.

Technologists should see through the hyperbole and recognize that the Internet (qua technology) is no more likely to set people free than rubbing a lamp will produce a genie. But the public gets caught up in the spin and becomes lulled into supporting foolish beliefs and counterproductive policies. Were similar claims made of fiberoptic technology or integrated circuits, they would be immediately dismissed as folly. But because of the ubiquity of the Internet and its importance in our daily lives (online shopping, video chatting, etc.), the claim attracts serious attention that it doesn't deserve.

Winner understands the absurdity of this technopomorphism and goes one step further by claiming that technology may on occasion take on an unethical quality that may go virtually undetected. If we adopt his broad sense of the "inhumane" — meaning not taking into account human concerns when a technology is designed or operated — then the Internet would be a prime example of this phenomenon. The Internet was built for technologists by technologists, whose primary concern in the early days was getting something to work, not anticipating that it would morph into what it is now.

In "Technologies as Forms of Life," Winner writes:

> the important question becomes, as we "make things work," what kind of world are we making? This suggests that we pay attention not only to the making of physical instruments and processes … but also to the production of psychological, social, and political conditions as part of any significant technical change.[7]

Once the Internet started to take off, these issues were pretty much ignored, so now we have an Internet that, according to Winner's definition, is inhumane: insecure, exploitative, and providing surveillance over hundreds of millions of individuals by private enterprise and the state. People are reportedly changing how they use the Internet for these three reasons; they also explain why there is a movement by many leading computer scientists to rebuild the Internet from scratch.

It's All About Intention

So what would constitute technological bad faith? Winner directs us to search for socially unacceptable ulterior motives behind the design and implementation of a technology, rather than study end use. Consequently, we can dismiss most of the world's great man-made disasters like the Tacoma Narrows Bridge collapse and the Fukushima Daiichi nuclear meltdown, as both still qualify as good faith technology efforts. These disasters may have involved human error, a lack of understanding, poor craftsmanship, or outright criminal neglect, but we can still reasonably characterize the results as unforeseen or unintended consequences.

Winner gets at the intention behind unethical technologies in two ways: "First are instances in which the invention, design, or arrangement of a specific technical device or system becomes a way of settling an issue in a particular community.... Second are cases of what can be called inherently political technologies, man-made systems that appear to require, or to be strongly compatible with, particular kinds of political relationships."[8] For example, Baron Haussmann built the broad Parisian thoroughfares, so admired today, at the direction of Louis Napoléon as part of the renovation of Paris. Their width was at least partly dictated by the desire to prevent the reoccurrence of street riots, such as those that occurred before and during the 1848 February Revolution that brought Louis Napoléon to power.

To give another example, Winner claims that New York urban architect Robert Moses attached a social meaning to the curiously low-hanging overpasses he built on the parkways of Long Island. Moses, acting out of social class bias and racial prejudice, fully intended his parkways to be a barrier to public transportation access to the borough by constructing the overpasses too low for buses to pass under them. The goal seems to have been to render public assets, such as popular Jones Beach, useless to the tired, poor, and huddled masses. Simply put, this is de facto segregation by class: the upper classes don't use mass transportation. Winner makes a good case that the low-hanging overpasses took on an unethical character as tokens of bad faith technology. I would note in passing that such examples tend to support a hard technological determinism: in this case the overpass technology directly impacted social and cultural norms. In a phrase, the overpasses exhibited bad faith design.

Langdon Winner rightly shows that we need also be concerned about the ethical intent of technology.

Winner and others suggest that designing immorality into a project is not that unusual. To the extent that this is true, my default position that technology is inherently neutral must be considered incomplete. My emphasis was on the ethical use of technology. Winner rightly shows that we need also be concerned about the ethical *intent* of technology. Consider two other examples Winner cites: the introduction of pneumatic molding machines to the McCormick reaper manufacturing plant in the mid-1880s and the introduction of the mechanical tomato harvester in California in the 1960s. In both cases, he argues, the technologies were specifically introduced to undermine the effectiveness of union organization by replacing the skilled workers who were union members. In the case of McCormick, the introduction of the new pneumatic molding technology actually created a loss for the company for three years, but by then union organization was broken. At that point, McCormick ripped out the technology.

In both of these instances, Winner argues, undisclosed political advantage was also in attendance. It is of considerable importance for society to address the extent of this phenomenon, in which technology is claimed to be employed for the user's benefit (e.g., to "improve customer convenience"), but is actually used to their detriment (e.g., secretly harvesting customers' personal information for future exploitation).

DIESELGATE

Careful observation will support Winner's thesis. By now we have all heard of the Volkswagen "dieselgate" scandal, which revealed that Volkswagen intentionally altered the control code in its turbocharged direct injection diesel engines to circumvent accurate testing of nitrogen oxide (NOx) exhaust emissions on as many as 11 million vehicles manufactured from 2009 to 2015. The cheating involved sensor-induced control code routines that resulted during the emissions testing procedure. This was detected by a university research team that was testing emissions during actual road trips rather than in stationary emission testing settings.[9] Volkswagen US President and CEO Michael Horn admitted that the company used a "software program that served to defeat the regular emissions testing regime,"[10] so there's no doubt at this point that the engine control program contained code that was specifically included to circumvent emissions compliance tests in violation of air quality laws. While VW has at times tried to diminish the scale of the culpability, it has not denied wrongdoing since the exposure.[11]

VW's infamous engine control system was apparently a descendant of some code changes developed by Audi engineers in 1999 as a means to quiet diesel engines.[12] Audi engineers found that their "acoustic mode" of operation not only silenced the engine, but also increased NOx emissions, so they shelved the software modifications. Parties as yet unidentified at VW apparently resurrected Audi's concept of multi-modal operation of the control system, but this time in reverse. The normal operation would be something like the "acoustic mode," but during emissions testing, the control system would sacrifice performance for compliance and switch to deceit mode. German auto parts maker Bosch GmbH warned VW in 2007 not to use software modifications to its engine management system to defeat emissions testing[13] but was ignored. It has since been discovered that VW was fined for using defeat devices to disable pollution control systems in 1973.[14] Further, an internal VW PowerPoint presentation has recently surfaced that discusses the deception regimen,[15] so emissions cheating takes on legacy status at VW. The Volkswagen diesel scandal is a recent confirmation of Winner's thesis that some technology is just unethical from the start.

We should note that this scandal satisfies both of our conditions for unethical technology in that it was inherently capable of being used in ways that society would

adjudge unethical, immoral, or illegal and that the full intent of the designer(s) was not disclosed to the stakeholders (users, customers, stockholders, regulators, etc.) when the technology was developed. In the VW case, there were no irregularities to be found in the supply chain. Nor did blatant consumer frauds take place. From a business process perspective, everything was in order after the design. What was not in order was the willful and intentional attempt to subvert environmental protection laws. At bottom, VW's innovative code might be meaningfully said to be unethical.

CONCLUSION

And so Winner's thesis is demonstrable. There are technologies that are created in bad faith by bad actors. Two questions arise. First, where would such technologies likely be found, and, second, what should be done about them?

In the IT world, likely places to look would be hardware and software associated with high stakes enterprises with little oversight and regulation in name only. My current candidates would include direct-recording electronic voting systems, flash trading systems, encryption and security products, government and contracted surveillance systems (e.g., Xkeyscore) and databases (e.g., no-fly lists, the NSA's PRISM program), commercial software with back doors, government-issued malware, ransomware (drive-by infectious websites), and so on. In general, the perpetrators of such threats are likely to be found in the nexus of big government and big money. The only reason that more VW-level scandals haven't been exposed is that tight control is maintained over proprietary information by big government and big business. In both cases, the default is to conceal anything that might prove embarrassing or encourage litigation. I predict that many more fruitful confirmations of Winner's thesis are only one whistle-blower away from public view.

So what should be done about such technologies? By the time some university researchers discovered the VW deception, scores if not hundreds of VW, Audi, and Bosch employees were aware of the problem, yet no one blew the whistle. Fear of reprisal is as great a deterrent to public disclosure of bad faith technology as it is to disclosure of government malfeasance. That's not likely to change unless whistle-blowing becomes far more lucrative.

A more subtle, but perhaps in the long run more effective, tactic might be to head off bad faith technology through the certification process. We can see that VW's behavior in its engine maintenance system code is in direct violation of portions of the ACM's Code of Ethics and Professional Conduct, to wit:

> Section 1.1 When designing or implementing systems, computing professionals must attempt to ensure that the products of their efforts will be used in socially responsible ways, will meet social needs, and will avoid harmful effects to health and welfare. In addition to a safe social environment, human well-being includes a safe natural environment. Therefore, computing professionals who design and develop systems must be alert to, and make others aware of, any potential damage to the local or global environment.
>
> Section 1.2 To minimize the possibility of indirectly harming others, computing professionals must minimize malfunctions by following generally accepted standards for system design and testing. Furthermore, it is often necessary to assess the social consequences of systems to project the likelihood of any serious harm to others.... In the work environment *the computing professional has the additional obligation to report any signs of system dangers that might result in serious personal or social damage. If one's superiors do not act to curtail or mitigate such dangers, it may be necessary to "blow the whistle" to help correct the problem or reduce the risk*.
>
> Section 2.3 ACM members must obey existing local, state, province, national, and international laws unless there is a compelling ethical basis not to do so.... If one decides to violate a law or rule because it is viewed as unethical, or for any other reason, one must fully accept responsibility for one's actions and for the consequences.
>
> Section 3.1 Because organizations of all kinds have impacts on the public, they must accept responsibilities to society. Organizational procedures and attitudes oriented toward quality and the welfare of society will reduce harm to members of the public, thereby serving public interest and fulfilling social responsibility. Therefore, organizational leaders must encourage full participation in meeting social responsibilities as well as quality performance.
>
> Section 4.2 Adherence of professionals to a code of ethics is largely a voluntary matter. However, if a member does not follow this code by engaging in gross misconduct, membership in ACM may be terminated.[16]

Without belaboring the point, a casual review of these fragments of the ACM code shows that the software developers involved in the VW scandal were poster children for distorted ethics and misplaced loyalties. Perhaps a solution to the bad faith technology challenge would be through increased awareness of ethical principles and closer scrutiny of applicable standards. This is best addressed at the university and even the high school level. The new high school Advanced Placement (AP) courses in computer science, for example, teach ethical principles as part of their curriculum.

Kranzberg's First Law, while true, is not really helpful. Perhaps a more useful guide would be to reevaluate the ethical standards we set for ourselves in terms of such things as ethical codes, duties, procurement policies, and the like. Perhaps if Volkswagen had been slapped with a 10% import tax after the 1973 disclosure, the 2015 disclosure wouldn't have happened. It is axiomatic that when ethical violations lead to no unpleasant consequences, we can expect a good deal more of them. Let the word go forth that bad faith technology is both real and unworthy of us.

ACKNOWLEDGMENT

My thanks to Howard Rheingold for encouraging me to revisit my naive default position that technology is inherently neutral.

ENDNOTES

[1]Kranzberg, Melvin. "Technology and History: 'Kranzberg's Laws.'" *Technology and Culture*, Vol. 27, No. 3, July 1986 (www.jstor.org/stable/3105385?seq=1#page_scan_tab_contents).

[2]Berghel, Hal. "Net Neutrality vs. Net Neutering." *IEEE Computer*, March 2016 (www.computer.org/csdl/mags/co/2016/03/mco2016030073.pdf).

[3]Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus*, Vol. 109, No. 1, Winter 1980 (http://innovate.ucsb.edu/wp-content/uploads/2010/02/Winner-Do-Artifacts-Have-Politics-1980.pdf).

[4]Langdon Winner's Home Page (http://homepages.rpi.edu/~winner/).

[5]Winner (see 3).

[6]Morozov, Evgeny. *The Net Delusion: The Dark Side of Internet Freedom*. Public Affairs, 2011.

[7]Winner, Langdon. "Technologies as Forms of Life." *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. University of Chicago Press, 1989.

[8]Winner (see 3).

[9]Bigelow, Pete. "West Virginia Researcher Describes How Volkswagen Got Caught." autoblog, 23 September 2015 (www.autoblog.com/2015/09/23/researcher-how-vw-got-caught/).

[10]Chappell, Bill, "'It Was Installed for This Purpose,' VW's US CEO Tells Congress About Defeat Device." NPR, 8 October 2015 (www.npr.org/sections/thetwo-way/2015/10/08/446861855/volkswagen-u-s-ceo-faces-questions-on-capitol-hill).

[11]Cremer, Andreas. "VW Says CO_2 Emissions Scandal Not as Bad as Feared." Reuters, 9 December 2015 (www.reuters.com/article/us-volkswagen-emissions-carbon-idUSKBN0TS12I20151209).

[12]Ramsey, Jonathon. "Audi Invented the VW's Dieselgate System in 1999." The Drive, 21 April 2016 (www.thedrive.com/news/3104/audi-invented-the-vws-dieselgate-system-in-1999).

[13]Arvinth, Karthick. "VW Scandal: Carmaker Was Warned by Bosch About Test-Rigging Software in 2007." *International Business Times*, 8 September 2015 (www.ibtimes.co.uk/vw-scandal-carmaker-was-warned-about-test-rigging-software-2007-1521442).

[14]Gardella, Rich, and Mike Brunker. "VW Had Previous Run-In over 'Defeat Devices.'" CNBC, 23 September 2015 (www.cnbc.com/2015/09/23/vw-had-previous-run-in-over-defeat-devices.html).

[15]Ewing, Jack. "VW Presentation in '06 Showed How to Foil Emissions Tests." *The New York Times*, 26 April 2016 (www.nytimes.com/2016/04/27/business/international/vw-presentation-in-06-showed-how-to-foil-emissions-tests.html?ref=world&_r=1).

[16]"ACM Code of Ethics and Professional Conduct." ACM Council, 16 October 1992 (www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct).

Hal Berghel is currently Professor of Computer Science at the University of Nevada, Las Vegas, where he has previously served as Director of both the Schools of Computer Science and Informatics, and as Associate Dean of the College of Engineering. His research interests are wide-ranging within the binary and digital ecosystem, ranging from logic programming and expert systems to relational database design, algorithms for non-resolution-based inferencing, approximate string matching, digital watermarking and steganography, and digital security and privacy. Since the mid-1990s, Dr. Berghel has applied his work in digital security to law enforcement and intelligence gathering, particularly with respect to digital crime, digital money laundering, information warfare, and trusted identities. His research has been supported by both industry and government for over 30 years. In addition to his academic positions, Dr. Berghel is also a popular columnist, author, frequent talk show guest, inventor, and keynote speaker. For nearly 15 years, he wrote the popular "Digital Village" column for the Communications of the ACM, *and he has written the "Out-of-Band" column for* IEEE Computer *since 2011. Dr. Berghel is a Fellow of both the IEEE and the ACM and serves both societies as a Distinguished Visitor and Distinguished Lecturer, respectively. He can be reached at hlb@berghel.net.*

Making Ethics Considerations a Required Part of System Development

by Jesse Feiler

Ethics is "hot" now. You can buy ethically sourced coffee at Starbucks, Product Red merchandise to help eliminate HIV/AIDS in Africa, and dolphin-safe tuna, grass-fed beef, and locally grown food in many places. You can review codes of conduct and audits that some companies (Walmart, Apple, Ikea, etc.) publish regarding their ethics and standards for workers, contractors, and suppliers. These trends are sometimes grouped together as "consumer ethics" or "ethical consumerism."

Sustainability and environmental ethics expand the traditional focus of ethics on humans and their behaviors to include the non-human world. When you start to look at specific cases and actions, there often appears to be no hard and fast division between these fields. In this article, "ethics" is used in its broadest sense, to include consumer ethics, sustainability, and environmental ethics as well as bioethics and political, military, business, and public sector ethics.

Now a conversation is taking shape about the role of ethics in software projects, all the way from the algorithms that implement those projects to the projects themselves. Interest is especially focused on autonomous vehicles, the cyber espionage and cyber warfare fields, and many consumer products such as the Web and video authoring tools that have played essential roles in new media and publishing projects — as well as terrorism and child pornography.

Note that "software projects" and "system development" are used in a very broad way to encompass apps on mobile devices, complex systems running across an elaborate network of hosts, and websites and Web apps of all types and sizes.

ORGANIZING ETHICS IN SYSTEM DEVELOPMENT

We can look at system development from the viewpoint of ethics and identify several important areas to explore.

Ethical Work

Most organizations have an explicit or implicit code of ethics to cover basic behaviors such as lying and cheating. System development projects fall under these guidelines. In beginning a discussion about ethics in system development, it can be helpful to remind yourself and others that ethics are already involved in this sense, but beyond that, there are specific ethical issues in software projects and system development.

Ethical Sourcing

Ethical sourcing refers to sources of raw materials and labor. Consumers are often sensitive to ethical sourcing and, in some cases, are willing to pay a premium price for products that are produced organically (in the case of food) or by workers and contractors who are paid living wages. Ethical sourcing applies to system development particularly when it comes to outsourcing production and development to countries where prevailing wages are substantially lower than at home. There have been a number of media stories (and accompanying outrage) after exposés of working conditions for workers subcontracting to high-tech companies. Protests against the use of offshore developers have been voiced by domestic developers. Many people are starting to think more and more about how their products are developed, who is developing them, what wages they are paid, and what working conditions they experience.

The idea that a product you use is produced by people who are forced to work 60 or 70 hours a week and who are paid salaries that force them to live so far away from their workplace that an extra hour or more is added to each already long day is very troubling. (This describes the living and working conditions in some parts of Silicon Valley. Were you envisioning another place?)

Ethical Projects and Products

This is one of the big ones, because it includes deciding whether to participate in a system development project in the cyber warfare or cyber terrorism sectors. Perhaps it is an indicator of the ethical issues involved that some people will not talk about cyber warfare or cyber terrorism, and they may even deny that such sectors exist. They certainly exist in Hollywood movies and TV shows, and they are topics of interest in many publications ranging from mainstream to marginal. Such basic questions as whether the US has ever been attacked or has launched its own attack using cyber tools are very hard to answer because the usual sources for such answers will not or cannot talk.

Software developers and technologists often wind up on the front lines of ethical quandaries in their roles as go-to people for data access.

Another recent example of the confusing issues at play is the US Department of Justice's demand that Apple break its own encryption in order to access an iPhone used by a terrorist in the San Bernardino attack. The demand was withdrawn after the FBI reportedly paid (somewhere in the vicinity of $1 million, according to rumor) to have the iPhone hacked open. Here are some of the unresolved ethical issues in this case, which will probably recur in other cases:

- Was Apple behaving ethically in refusing to break its own encryption?
- Was the FBI ethical in paying someone to break Apple's encryption? If that payment was to a US citizen, does that matter? If that payment was to a foreign company, does that matter?
- Would matters be different if it was another country that paid a hacker to bypass the encryption?
- Was it ethical for the hacker to perform the act whether or not it was paid for it?

Software developers and technologists often wind up on the front lines of ethical quandaries in their roles as go-to people for data access. A common example is the sharing of data, especially mailing lists. How were those addresses collected? What promises were made or implied to those whose data was collected? For what uses can they be repurposed? Although management should be answering those questions, it is not uncommon for the questions to go unasked. And if they are asked, it's the technical support people who have to ask them.

Sustainability

Sustainability and environmental ethics are gaining more and more attention. In the high-tech world, they have already influenced the packaging of products. In terms of system development, sustainability plays into three areas:

1. **Hardware upgrades.** Are hardware upgrades defensible from a sustainability point of view? What happens to the old hardware? Manufacturers and retailers such as Apple, Verizon, Best Buy, and others have started aggressively implementing and promoting recycling programs.
2. **Software upgrades.** While software upgrades are not examined so often, they also pose ethical issues. The most obvious question is whether it is fair to require people to pay for software upgrades, particularly if they are required for continued operation of the product. When it comes to priced upgrades, do all users need to purchase them even if the new functionality is something they won't use? Or something the customers don't care about but the developer thinks they should have anyway? (Unfortunately, security features are frequently in this category of must-haves from the developer's point of view and unnecessary from the user's point of view.)
3. **System maintenance.** Sustainability also poses some interesting questions about system maintenance. Does a company providing a product or service (even a Web-based service) have the resources to continue to provide it in the future? This is also a marketing consideration that has been very important in the world of cloud computing. Questions are being asked about who owns purchased media and user-provided content after the demise of the user or the vendor, but the conversations are not over.

ADDING ETHICS TO EXISTING SYSTEMS AND OPERATIONS

Many people and organizations still think of systems in the old mainframe model. People enter data, a program runs, and some output is produced. Today's systems are dynamic: there may be little distinction between data entry and reporting — it all happens together in multiple threads of processing. Thinking in terms of the clear divisions between data entry, processing, and reporting just doesn't match today's reality.

Unfortunately, all too often existing procedures and even laws still envision the mainframe pattern. Users simply ignore those outdated laws and policies in order to get the work done. When someone points out that the procedures or laws don't match the reality, that issue usually winds up on a "to do someday" list.

Into this vacuum, new rules and procedures have been put into place — often by default. Unlike the older standards, they are put together and implemented by people who ordinarily wouldn't be considered policy setters. Ask yourself these questions:

- Does your payroll system properly account for the situation where an employee planning a vacation gets an advance on pay to cover an anticipated missed day? Does your payroll system handle that? (Chances are it does.)
- If the employee dies while on vacation, what happens to that payment? Is the procedure built into the payroll system or is it an ad hoc decision?
- Many organizations might formally or informally forgo an attempt to claw back the payment from the employee's estate. (Chances are this is an ad hoc decision.)
- What happens if the employee's death occurred while wearing a suicide vest in a terrorist attack?

Particularly in a case such as the last scenario, these decisions are generally made in a one-off and ad hoc manner (and that's probably correct). But there are many more intricate cases in which the decision making has been delegated by default to programmers.

There are ethical risks in most decisions, but they are magnified when corporate policies and procedures as well as laws no longer reflect reality. When that reality is defined by the engineers implementing systems, a significant change in management and governance has occurred, possibly without serious discussion or even awareness of the fact. Whether or not software engineers are better at decision making than managers — thanks to their comfort with high degrees of complexity — is a separate issue.

Like it or not, ethical considerations (formal or informal) are already embodied in corporate policies and procedures and are being implemented at the operational — not the policy-making — level. Operations must continue even while policies are being made, and ad hoc policies proliferate and become permanent in the absence of serious planning for the new operational environments.

BUILDING ETHICAL SYSTEMS

When most people think about ethics in system development, they think about new systems — the emphasis is on "development." Although ethical thinking and choices are present throughout our activities, explicit ethical choices often need to be confronted in the development of new systems.

To Build or Not to Build

The overriding ethical concern in building a new system is whether or not it should be built at all. (This question was at the heart of the recent discussion about compelling Apple to build a back door to encrypted user data. The discussion ended when authorities found someone to hack into the iPhone in question, but the issue remains unresolved.) Then, if the system should be built, its creator needs to ask the question again for each functional enhancement. Building systems that are intended to perform or facilitate illegal, immoral, or unethical activities may not be illegal in and of itself, but people may view doing so as immoral or unethical (including, in some cases, the people who are deliberately engaged in the activity and intend to profit by it!).

There are ethical risks in most decisions, but they are magnified when corporate policies and procedures as well as laws no longer reflect reality.

Dealing with Data

Systems, system developers, and system users have access to a wide variety of data that should be kept private. There are legitimate reasons for such access, and it is expected in many cases. However, in developing systems, attention has to be paid to identifying and protecting the data that should be kept private. It is remarkably easy to capture information surreptitiously and to store it without considering the security issues involved. (This is especially critical in cases where audit trails and debug statements are essential to the development process.) Privacy is becoming more and more problematic as laws are catching up with technology. Today, many developers consider the privacy laws and regulations in the EU to be stricter than those in the US. Where there are explicit laws, the situation is a bit clearer than in the not-legislated environment where only personal ethics rule.

A common ethical problem arises when the people who are implementing systems only gradually come to realize the true nature of the situation they are dealing with. Software engineers who are tasked with integrating data, for example, may only come to learn over time that the data in question should not be shared. For many, there is an ambiguous period of time when it is not clear exactly what is happening. Some developers of apps and websites have found themselves chafing at restrictions and being constrained by clients' refusal to divulge details that the developers feel are critical to completion of the system. The developers' requests for more information are interpreted as prying rather than a normal part of system development. Many developers have tales to tell of wasted hours (and weeks) struggling to implement a system with only sketchy details.

Not safeguarding an individual's Social Security number may seem like a minor infraction to a manager, but to a developer who understands that it really doesn't matter if it's one number or one million numbers, this is an ethical quandary.

In addition to keeping necessary information from developers, people also sometimes ignore laws and best practices, considering them nuisances. The developer is then placed in the middle of a messy situation. A common one cited by Web designers and app developers is the proper management of personally identifiable information (PII), particularly credit card and Social Security numbers. A generally accepted best practice is to never store this information in an encrypted manner unless absolutely necessary. This can involve writing a bit of either programming code or Web code to store just the critical data in an encrypted form, while storing part of the data (e.g., the last four digits of a credit card number) separately in an unencrypted form. This entails more work and generally requires implementation of a security mechanism so that users are identified and the appropriate version of the sensitive data (the full number or only the last four digits) is shown to each. Just to make things a little more complicated for the developer and manager, the encryption mechanism may add a bit more complexity and cost.

Not safeguarding an individual's Social Security number or credit card number may seem like a minor infraction to a manager, but to a developer who understands that it really doesn't matter if it's one number or one million numbers, this is an ethical quandary. After all, in the digital world, what can be done once can often be done millions of times. Because many managers are much less familiar with technology than system developers, it can be hard to find a way to talk about the issues at hand. The fact that many of the laws that might protect this data are in a state of flux doesn't make things any easier for anyone.

In the scope of a project, the protection of PII data is quite likely to play a minor part: so does the developer walk out or stay? As is the case with many ethical issues, there isn't a clear-cut solution that can be pulled out of a hat like a rabbit. To paraphrase a political saying, ultimately all ethics is personal (even if the "person" is a business).

Open Source Software and Frameworks

Once upon a time, people wrote programs that were totally self-contained. On occasion, they would call built-in (or "intrinsic") routines from the operating system or in the language to perform functions such as getting today's date, but most of the code in a program was *in* the program. Today, code is built using components from a variety of places. Some are shared components within an organization, and other are shared among a number of organizations or the public (i.e., open source software). Managers and engineers have to come to grips with the reality that developing software today doesn't necessarily mean writing line after line of code: the software engineer assembles code — writing some of it — into a composite bundle of code.

This can become an ethical question because "developing an app" means different things to developers and managers. Although you may consider this a simple misunderstanding, it becomes an ethical issue as different views of a project under development become developed in actual software. When the project includes open source software that is being reused from legitimate open source projects alongside newly written code, managers and even developers may not be used to the different management, development, and testing protocols that are necessary. People can easily look at the software as just code, but these distinctions in source and licensing influence how the software can be used. What might have looked at first like a practical issue can turn into an ethical one when the code used is not totally under the project manager's control.

MAINTAINING ETHICAL SYSTEMS

Managers and developers are frequently confronted with a serious ethical issue in maintaining existing systems. Many of these systems were written years or decades ago when computers and the Internet were very different from what they are today. We demand more from our software now, and some common coding techniques from the past (particularly in the areas of privacy and security) are no longer considered appropriate. For example, our understandings of privacy have evolved over time, and, as noted previously, our understanding of the basic use of software has changed from the enter-process-report model to a more dynamic and complex model, likely involving the Internet and multiple mobile devices.

It is an ethical decision to determine what to do when existing software (often mission-critical software) is discovered to have such flaws. Perhaps these flaws don't cause hard crashes, but they may be wrong by today's standards and best practices. It is easy to just let these things go and to patch what is necessary to patch, but it is the responsibility of the manager and developer (and perhaps even higher-level people) to decide when and how to revise software that is running smoothly but inappropriately. This is a difficult — and frequently expensive — decision to make.

There is a flip side to this issue, and that is making certain that ethically developed systems are not accidentally (or even deliberately) stripped of critical software. In many cases, such as credit card number encryption and unencrypted storage of the last four digits, the ethically valid code is more expensive to create and to maintain. Creating it is not the end of the issue, because many engineers without that expertise may touch it. Many times maintenance is delegated to the most junior people on staff (or sometimes to senior people who are not comfortable learning new things), and this makes the problem worse, as people without the training in ethical software development and maintenance may modify ethically developed software.

As is the case with many of the ethical issues described here, confronting them and understanding them can be a major step in building and maintaining ethically sensitive software. In general, when management and supervisors turn their backs on problems (ethical or technical), that is not the end of the story. There is usually a sequel, and it is rarely pleasant.

LIVING IN A MULTI-ETHICAL WORLD

The last point to consider in looking at ethics and system development is that although there is agreement across the globe on many things, ethics systems are not uniform worldwide. Many people think that is part of what makes life so interesting!

This mosaic of ethical systems across peoples, countries, religious creeds, and other groupings and distinctions runs into the reality of the global technology world. The most obvious manifestation of this is the Internet because, by default, even the simplest app posted by an individual developer on an app store is available to people around the world. Although language localization and other customizations and adaptations are often available, app and website developers (including corporations that develop apps and websites) increasingly understand that they are developing for a global audience. This can affect the interface of software products enormously in terms of the text and images that are part of the project. Something as simple as the image of an animal can have localization issues. One (possibly apocryphal) story relates that the image of a printed page in horizontal and vertical orientations was developed using a dog as the image. Because some cultures consider dogs unclean, the dog became another animal such as a horse or cow in those localizations.

In general, when management and supervisors turn their backs on problems (ethical or technical), that is not the end of the story. There is usually a sequel, and it is rarely pleasant.

Even if a company is focused on a local business opportunity, as soon as it sets up shop in any way on the Internet, it is a global business and should expect to deal with people from around the globe with all sorts of different viewpoints. The Web is open in a very basic sense, but it also allows people to find very small groups of like-minded individuals, be they professionals, hobbyists, co-religionists, terrorists, sexual predators, scholars of medieval church architecture, or any other self-selecting group. Within such a global but self-selected group, there may be a shared ethical system, but outside of such groups, the users of software development projects are likely to bring multiple viewpoints and ethical systems to the software.

How a given project approaches this depends on the project, but the fact that not every user of a software project shares the viewpoints and background of the project's developers and sponsors is a reality to be considered. In addition to considering the connotations of animals, images representing groups of people sometimes come into question as gender roles evolve in different ways in different societies. Depicting a business meeting with a woman standing and leading a discussion with a table of men is shocking in some cultures and contexts.

PLANNING AHEAD

Interest in ethics has risen and fallen over the centuries; the meaning of life and what matters in life are things that people in every era have pondered. The word "ethics" seems to be getting a bit more interest in some places today, as we pay attention to ethically sourced food, environmental ethics, and business ethics. (The first two areas of concern have only been identified and talked about in the last decade or two.)

It seems safe to say that with concern for ethics on the upswing (at least temporarily), and with interest in the development of software also increasing, more and more people will start to think about ethics in software development. Software development incorporates ethical choices, even if they have not been identified as such. It's time to think about ethics in software development so that as this discussion unfolds, it is not a surprise.

Jesse Feiler is a developer, consultant, and author specializing in database technologies and location-based apps. He has worked with databases and data management on computers from mainframes to iPhone, iPad, and Apple TV using data management tools from DB2 (IBM) and DMSII (Burroughs) to Enterprise Objects Framework and Core Data, MySQL, Oracle, and MySQL. Mr. Feiler is the creator of Trails & Places, a guide and authoring tool for trails and tours that includes location-based updates along with social media tools, and is also the creator of Minutes Machine, the meeting management app. As a consultant, he has worked with small businesses and nonprofits on projects such as production control, publishing, and project management. Mr. Feiler's books include iPad for Seniors for Dummies, Introducing SQLite for Mobile Developers, *and* iOS Programming with Swift for Dummies. *A native of Washington, DC, he has lived in New York City and currently resides in Plattsburgh, New York. He can be reached at champlainarts.com.*

About Cutter Consortium

Cutter Consortium is a truly unique IT advisory firm, comprising a group of more than 100 internationally recognized experts who have come together to offer content, consulting, and training to our clients. These experts are committed to delivering top-level, critical, and objective advice. They have done, and are doing, groundbreaking work in organizations worldwide, helping companies deal with issues in the core areas of software development and Agile project management, enterprise architecture, business technology trends and strategies, enterprise risk management, metrics, and sourcing.

Cutter offers a different value proposition than other IT research firms: We give you Access to the Experts. You get practitioners' points of view, derived from hands-on experience with the same critical issues you are facing, not the perspective of a desk-bound analyst who can only make predictions and observations on what's happening in the marketplace. With Cutter Consortium, you get the best practices and lessons learned from the world's leading experts, experts who are implementing these techniques at companies like yours right now.

Cutter's clients are able to tap into its expertise in a variety of formats, including content via online advisory services and journals, mentoring, workshops, training, and consulting. And by customizing our information products and training/consulting services, you get the solutions you need, while staying within your budget.

Cutter Consortium's philosophy is that there is no single right solution for all enterprises, or all departments within one enterprise, or even all projects within a department. Cutter believes that the complexity of the business technology issues confronting corporations today demands multiple detailed perspectives from which a company can view its opportunities and risks in order to make the right strategic and tactical decisions. The simplistic pronouncements other analyst firms make do not take into account the unique situation of each organization. This is another reason to present the several sides to each issue: to enable clients to determine the course of action that best fits their unique situation.

For more information, contact Cutter Consortium at +1 781 648 8700 or sales@cutter.com.

The Cutter Business Technology Council

The Cutter Business Technology Council was established by Cutter Consortium to help spot emerging trends in IT, digital technology, and the marketplace. Its members are IT specialists whose ideas have become important building blocks of today's wide-band, digitally connected, global economy. This brain trust includes:

- Rob Austin
- Ron Blitstein
- Tom DeMarco
- Lynne Ellyn
- Vince Kellen
- Tim Lister
- Lou Mazzucchelli
- Ken Orr
- Robert D. Scott