Reaching and Teaching through "The Matrix"; Robosapiens, Transhumanism, and the Formidable in Engineering Ethics

Rosalyn W. Berne
University of Virginia

Abstract

Much of what is taught in engineering ethics is a matter of practical decision-making and professional responsibility—the avoidance of harm and the doing of good in industrial and technical settings. But what about the questions of metaethics; those issues about which there is deep concern, but also great difficulty in formulating a system of ethics which can be applied to them?

How might students consider, for example, the writing of Ray Kurzweil, and others, who put forth arguments that the evolving intelligent, 'spiritual' machine will soon be nearly indistinguishable from its human creator? How do we teach the ethics of such a daunting, futuristic prediction? How might classroom instruction provide for probing, insightful thinking which would equip our students with the tools they need to navigate such unknown and unfamiliar moral terrain? One helpful pedagogical approach is the use of film.

This paper relates the use the film The Matrix as a teaching tool in Engineering Ethics. It suggests that the pedagogical power of using film for Engineering Ethics lies in its ability to uncover otherwise elusive ideas and feelings, while providing a vehicle for constructive thinking, writing, and discussion about them.

1. Portraying a Technological Apocalypse

In the book entitled "Virtual Faith", Tom Beaudoin writes that today's young adults use technology to form "a shared generational culture amid a world of tension and ambiguity."[i] The truth of this claim became apparent to me when my students began excitedly referring to the film "The Matrix" during our classroom discussions. Something about this film was speaking to these young people, and so I decided to bring it into the classroom for formal study and review. As a classroom tool, the Matrix could be fodder for stimulation of the moral imagination Mark Johnson refers to in his book of the same title. It was my hope that through consideration of the film's dominate narrative constructs, and an understanding of its metaphors, the students could approach the moral dimensions of the development of super intelligent machines, with deep and enlightening imagination. Johnson asserts that "...it is in sustained narratives that we come closest to observing and participating in the reality of life."[ii] It is of utmost importance, I believe,

that engineering students learn to observe and participate in the potential realities of a future we are creating today. And how else might they develop this ability, but through the exercise of moral imagination? I have found that to do so can be especially fruitful through the genre of science fiction.

The Matrix is a science fictional account of a future where highly intelligent, human-like machines take over and control the human race. In the film, humans have lost the war against our own technological, super intelligent creations. These machines have become the dominant presence on the planet earth and are in need of energy to survive. The massive destruction of the war between humans and machines created a cloud mass that blocks out the sun's rays, and so the most immediate source of energy is humans themselves. The super intelligent machines farm humans through genetic engineering, and maintain their bodies for the heat they produce. But because the human body needs its consciousness to stay alive, the machines must sedate and pacify the human minds. Thus, the creation of a virtual reality termed, "the matrix." The human bodies are wired at birth to a virtual reality program that simulates life on earth before the war. Humans perceive that their state of mind is actual conscious reality, while in fact they live in isolated, confined pods and are unaware of their true condition as energy producing slaves. Loaded into a virtual reality, the human mind has no consciousness of what is real, of self-knowledge or knowledge of the truth. Only a small group of human rebels have gained "enlightenment" to their true condition. A hero among them has the capacity to fight the technological enemy, and to restore self-knowledge to the human race.

2. Predicting a Transhuman Evolution

Ray Kurzweil, author of The Age of Spiritual Machines, internationally renowned scientist, and computer engineer, is predicting and proclaiming a future in which the human bodily form will become more technological then biological. He says that two and a half million years of DNA based evolution may soon be irreversibly altered by the dominance of technological creations. Through artificial intelligence, quantum computing, nanotechnology and robotics, our brains and bodies
will be recreated, reshaped and enhanced in a transformation being fueled in laboratories worldwide.

Kurzweil is a transhumanist; one who studies the "ramifications, promises and potential dangers of the use of science, technology, creativity and other means to overcome fundamental human limitations." Transhumanism is an "intellectual and cultural movement that affirms the possibility and desirability of fundamentally altering the human condition through applied reason, especially by using technology to eliminate aging and greatly enhance human intellectual, physical, and psychological capacities."iii

Kurzweil believes that the transhumanist vision is a matter of technological inevitability: an evolution that is rapidly approaching near time reality. His argument states that because of the Law of Accelerating Returns, the computational powers of our machines will soon far exceed that of the human brain. And, that this artificial intelligence will solve for us the myriad of problems that have plagued the human existence on earth. Kurzweil says that the 21st century will bring radical changes not only to our computational powers, but also to the very nature of human biological existence. What

were once dreams and fantasies of human immortality and bodily perfection, limited to our science fiction, arts and religions, are soon to take their place in the domain of technological probability.

"In the second decade of this century," writes Kurzweil, "it will be increasingly difficult to draw any clear distinction between the capabilities of human and machine intelligence...Over the next several decades, machine competence will rival -and ultimately surpass- any particular human skill one cares to cite, including our marvelous ability to place our ideas in a broad diversity of contexts."iv The problem, as Kurzweil defines it, is that because of its extraordinarily slow computational speed, the human brain is severely limited, and flawed. Through technology, we will create something Kurzweil attests to be superior to the human brain.

"DNA based evolution will eventually have to be abandoned."v Such a statement might seem like the ravings of one megalomaniac, but Ray Kurzweil speaks of widely pursued technological advancements. Research and development all over the world is moving in just the direction he cites. He points, for example, to the Advanced Telecommunications Research Lab in Kyoto Japan that is building an artificial brain with a billion electronic neurons. Although that represents only one percent of the number of neurons of the human brain, these neurons run at a computational speed a million times faster than the human brain. The computational speed of this artificial brain will be thousands of times greater than the human brain. By the year 2010, he says, supercomputers will reach the 20 million billion calculations per second capacity of the human brain. By the year 2060, a personal computer will be able to simulate the brain- power of a trillion human brains.

Kurzweil forecasts the development of molecular computers that use DNA as the basis for quantum computing. This possibility opens the prospect of computers that will compete with and surpass the full range of human capabilities. And at that point, which according to Kurzweil is only a few decades away, we will be able to use "reverse engineering" to examine and replicate the brain layer by layer, synapse by synapse, neuron by neuron.

The implications for scientific research are tremendous. But for Kurzweil, and other transhumanists, the implications suggest transcendent possibilities; the possible replacement of the 'flawed, inadequate, DNA based brain we've been stuck with for two and a half million years'. Kurzweil claims, "An ultimately feasible scenario will be to scan someone's brain to map the locations, interconnections and contents of somas, axons, dendrites, presynaptic vesicles, and other neural components. Its entire organization could then be re-created on a neural computer of sufficient capacity including the contents of its memory."vi

Kurzweil is speaking here of the use of scanning technology for the purpose of eventually downloading the brain's contents into another receptacle. He continues, "Objectively, when we scan someone's brain and reinstate their personal mind file into a suitable computing medium, the newly emergent "person" will appear to other observers to have very much the same personality, history, and memory as the person originally scanned. Interacting with the newly instantiated person will feel like interacting with the original person. The new person will claim to be that same old person and will have memory of

having been that person, having grown up in Brooklyn, having walked into a scanner here, and woken up in the machine there. He'll say, 'Hey, this technology really works.'"

And what kind of container will this reinstantiated personal mind file have? Perhaps the original human body, thawed. Or, a body built from the engineering of nanotechnology, self- replicating machines built at the atomic level. Or, perhaps robotic bodies will be used. Kurzweil suggests that maybe our brains will find a home in a virtual reality. As a result, Kurzweil claims, "there won't be mortality by the end of the twenty first century. Not in the sense we have known it."vii

3. Moral Challenges to Technological Realities

Perhaps it is true that human destiny will bring a merging with super intelligent machines, for the evolution of a transhuman being. For the transhumanists, that could mean the indefinite extension of individual human consciousness and life. For humanists, it means the potential loss of much of what makes us human, like suffering, love, imagination, sorrow, joy and mortality. Inside of the film "The Matrix" it means loss of human freedom and disconnection from reality. While the film is fantastic, it never the less speaks to scientific possibility. The pedagogical move to questions about an ethic of developing technology is easily made reading Kurzweil and viewing the Matrix. Is living immortality morally permissible? Should we protect the integrity of the biological body and mind? Under what circumstances ought we pursue potentially self-destructive technologies? What moral claims can be made on the human consciousness?

In the life sciences, current research is mapping out detailed knowledge of the human body. The human genome project is making rapid progress. Meanwhile, in the technological and engineering sciences machines are being designed and built to function like the human organism. Robots are being built and designed not just to perform dirty, repetitive tasks humans prefer not to do, but for complex, reasoning tasks requiring cognition, perception and sensitivity. In Japan, at the Science University of Tokyo, a robot has shape-memory actuators that move like muscles creating facial expressions beneath the robots silicon skin.' Tatsuya Matsui, creator of RoboCup, the robot soccer playing team, is now working on the next evolution of his robo SIG. Tatsuya hopes to endow SIG with "sufficient eyesight, hearing, and processing power to follow instructions given by several people in a crowd."viii And, not long ago, two researchers at Brandeis University created the first self-replicating robotic. ix The realization super-intelligence in robots awaits only the development of quantum computing. And that way has just been shortened, with the work of Harvard scientist Lene Hau. She has demonstrated, for the first time in history, the ability to manipulate, and halt, the speed of light.

In his own labs at MIT, Ray Kurzweil has taught a computer to read, interpret and write poetry. And another has learned to compose complex musical arrangements. Professor Kevin Warwick, head of the Cybernetics department at the University of Reading, plans this summer to implant a silicon chip to communicate with his brain. Warwick believes that intelligent robots will quickly realize their superiority to flesh and blood beings. Ultimately, he says, they will wipe us out.x If he is correct, then the science fiction of "The Matrix" may one day be our reality.

While Warwick anticipates our demise, other robotics experts predict a transformation that will mean the end of human beings, as we know ourselves to be. In the book Robo Sapiens, Hans Moravek of Carnegie Mellon is quoted as saying that humans will literally become robots, 'uploading their consciousness and memories into their computers.'xi Kurzweil agrees writing, "As we cross the divide to instantiate ourselves into our computational technology, our identity will be based on our evolving mind file. We will be software, not hardware... Today, our software cannot grow. It is stuck in a brain of a mere 100 trillion connections and synapses. But when the hardware is trillions of times more capable, there is no reason for our minds to stay so small... As software, our mortality will no longer be dependent on the survival of computing circuitry. (By which he means our biological bodies.) Computers will be deeply embedded in our bodies, brains, and environment."xii

Photographer Peter Menzel and journalist Faith D'Aluisio spoke with and visited roboticists in labs all over the world.xiii What they found was widespread agreement upon the prediction that robots will radically transform our future, but they disagree as to how that will happen. Some predict that robots will simply be integrated into our society as companions, servants, and workers. Other engineers anticipate that robots will overtake, dominate and compete with human society. Some, such as Ray Kurzweil, are clear about the co-evolution and fusion of humans with technology.

Questions of whether and how to proceed with the development and design of these technological advancements are fundamental to engineering ethics. Bill Joy, Chief Scientist for Sun Microsystems, responds to Kurzweil with a call for moral deliberation. Joy agrees with Kurzweil that coming advances in computing power seem to make it possible that an intelligent being will be built by the year 2030. And, he also agrees that shortly thereafter, we will gradually replace ourselves with our robotic technology, "achieving near immortality by downloading our consciousness." But, Joy asks, "If we are downloaded into our technology, what are the chances that we will thereafter be ourselves or even human?" xiv

Joy feels that a major redefinition of what it means to be alive has already begun. He points to the promises of genetic engineering to "revolutionize agriculture to increase crop yields while reducing use of pesticides; to create tens of thousands of novel species of bacteria, plants, viruses, and animals; to replace reproduction, or supplement it, with cloning; to create cures for many diseases, increasing our life span and our quality of life; and much more." Joy explains, "We know with certainty that these profound changes in the biological sciences are imminent, and will challenge all of our notions of what life is." Notions about the meaning of our human and earthly lives, and how we want to live together, are at the core of ethical reflection.

4. Reaching through the Matrix

To bring into the engineering classroom the moral questions of these developing new technologies requires moving beyond the parameters of ethical principles and rules and into the realm of intuition, emotion, and imagination. Traditional deontology simply cannot be applied to futuristic technologies because life then may in no way resemble

what we now know life to be. To approach futuristic ethics in the engineering classroom, we must abandon our rationalistic methodologies and open students to their feelings and even fears, excitements and dreams about the role they will play in the design of the future. Only then can they learn to articulate and define an ethic about the development of technologies that promise such profound changes to the characteristics of human life.

On their own, through popular culture films like the Matrix, our students engage in fantasy about the role technological development might play in the alteration of human meaning and purpose. When the film comes into the formal classroom as subject for papers and dialogue, coupled with theoretical readings on society, ethics and technology, they extend that fantasy into moral imagination. The profundity and richness of the metaphors in the film call forth emotional and analytic responses from the students that are reflective and impassioned. Many of my students feel optimistic and even excited about a technological evolution that could mean stronger, healthier bodies and higher intelligence inside a newly evolved human being. One student, expressing the hope of technology's role in human salvation wrote, *Humans are destructive. We pillage the land for natural resources leaving irrevocable scars. We destroy the ozone with "good" we manufacture and use. We destroy ourselves motivated by some human-conceived cause... The complexity and vastness of the problems we have created for ourselves are well beyond our reach of our current efforts to solve, at least directly. I think the most likely, (and quickest) solution is to create better humans*.xv

Other students express horror and ask what it is that motivates us to continue in our insatiable quest to dominant and control nature, even to the extent of self-obliteration? One such essay states: *The problem is that humans lack the ability to see beyond the immediate future: logically, a cell phone or DVD player is not going to result in our extinction. However, our quickness to accept and implement new technologies in our every day lives will be our demise because we are not looking at the future implications of their merging. It is like when you are floating on an inner tube at the edge of the ocean. You close your eyes for what seems like a brief moment, drifting in peaceful unconsciousness, while the tide drags you far away from shore... You only have two choices: paddle back or drown. We, as humans, are floating and what will we do when we realize it? I fear the worse- I am afraid we have already forgotten how to swim*.xvi

Like the worried students, Bill Joy has deep concerns. Joy writes, "We each seek to be happy, but it would seem worthwhile to question whether we need to take such a high risk of total destruction to gain yet more knowledge and yet more things; common sense says that there is a limit to our material needs, and that certain knowledge is too dangerous and is best foregone... Neither should we pursue near immortality without considering the costs, without considering the commensurate increase in the risk of extinction. Immortality, while perhaps the original, is certainly not the only possible utopian dream."xvii

Joy wants us to pause over the power of destructive self-replication in genetics, nanotechnology, and robotics."xviii He wants us to look at the frailties and inefficiencies of the human made systems already built and to have humility. He wants us to pull back for cautious reflection, and the development of shared values, ethics and goals. This is what is hoped for as we reach towards these challenges through the narratives of "The

Matrix", which portray the human masses as ignorant of their true state. It seems hopeful that bringing the film into the classroom helps engineering students to become more fully conscious and morally aware of the possibilities that come from the technological creations we make. In identification with varied characters, stimulation from the special effects and fantastical images, and conjecture about the symbolism and metaphor, students find a way to express and understand beyond the confinements of rigid analysis, and into an imaginative realm of possibilities. But when those possibilities become formidable, we can shift away from so called reality into the protection of the matrix. There, the perplexities, tensions and ambiguities have a safe and imaginative haven.

Science fiction films such as The Matrix provide a worthy venue for consideration of the unknown in technology, a vehicle for the expression of our horrors and elations about what life may become through our engineering hands. With moral imagination as a primary analytical tool, engineering students can take hold of the enterprise of meta-ethics, to pursue moral judgments in light of emotion, attitudes, and preferences, through the creative and illuminative power of the human intuition.

Endnotes

i. Beaudoin, Tom, Mark, Virtual Faith, Jossey Bass, San Francisco, 1998, pg. 11
ii. Johnson, Mark, Moral Imagination, University of Chicago Press, 1993. Pg. 196
iii. Definition of transhumanism, by Nick Bostrom et al, The Transhumanist FAQ, http://www.transhumanist.org/
iv. Kurzweil, Ray, The Spiritual Machine, 1999, pg. 4
v. ibid. Pg. 102
vi. Kurzweil 1999, pg. 124
vii. Kurzweil, 1999 pg. 128
viii. Menzel and D'Aluisio, Robo Sapiens, MIT Press, 2000 pp. 76, 83
ix. Hop Lipson and Jordon B. Pollack, Computer Science Department, Volen Center for Complex Systems, Brandeis University, Waltham Massachusetts.
x. "Professor to Wire Computer Chip into Nervous System," Dec 7, 2000, www.CNN.com
xi. Menzal, pp 32
xii. Kurzweil, 1999, pg. 129
xiii. For a full explanation of their travels and studies, see Menzel, Peter and D'Alusio, Faith, Robo Sapiens; Evolution of a New Species. Cambridge, MIT Press, 2000.
xiv. Joy, Bill, The Future Doesn't Need Us, Wired, April 2000, pg. 244.
xv. Schweitzer, Dan, "Humans Creating a Superhuman Species: A Necessary Step for Human Survival?" unpublished essay for Engineering TCC 401, University of Virginia, fall semester, 2000
xvi. VanDoren, Beth, "Society Two," unpublished essay for Engineering TCC 401, University of Virginia, fall semester, 2000.
xvii. ibid. pp. 249
xviii. ibid. pp. 248.

Bibliography

1. Beaudoin, Tom. Virtual Faith. San Fransicso: Jossey Bass Publishers, 1998
2. Johnson, Mark. Moral Imagination. Chicago: University of Chicago Press, 1993
3. Joy, Bill. "The Future Doesn't Need Us." Wired, April 2000, pg. 244
4. Kurzweil, Ray, The Spiritual Machine.New York: Penguine Books, 1999
5. Menzel, Peter and D'Alusio, Faith, Robo Sapiens; Evolution of a New Species. Cambridge: MIT Press, 2000.

Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition

ROSALYN W. BERNE
Rosalyn W. Berne is Assistant Professor of Technology, Culture and Communication at the University of Virginia. Her research interests are ethics in future studies, religion and technology, and the development of meaning in young adults. She has been Executive Director of the Olsson Center for Applied Ethics, Director of Darden Admissions, and Head of Tandem Friends School.

Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition