Leveraging the Singularity:  
Introducing AI to Liberal Arts Students

James B. Marshall

Department of Computer Science  
Sarah Lawrence College  
One Mead Way  
Bronxville, NY 10708  
jmarshall@slc.edu

Abstract
In recent years, the notion that computers and robots will attain superhuman levels of intelligence in the next few decades, ushering in a new “posthuman” era in evolutionary history, has gained widespread attention among technology enthusiasts, thanks in part to books such as Ray Kurzweil’s The Singularity Is Near. This paper describes an introductory-level AI course designed to examine this idea in an objective way by exploring the field of AI as it currently is, in addition to what it might become in the future. An important goal of the course is to place these ideas within the broader context of human and cosmic evolution. The course is aimed at undergraduate liberal arts students with no prior background in science or engineering.

Introduction
The field of Artificial Intelligence is now over a half-century old, having established itself as a viable intellectual discipline in the early 1950s, with the advent of the first programmable electronic computers. Since then, the dream of creating intelligent machines has lost none of its original excitement and fascination, and impressive advances have been made in almost every area of the field. Today, machines can recognize faces, process speech, learn from experience, drive themselves autonomously for hundreds of miles, explore the surfaces of distant planets, beat human chess masters with ease, diagnose some diseases better than human experts, compose music, handle customer service calls interactively over the phone, and even improve their performance through simulated evolution, to mention just a few of the many accomplishments that AI researchers have achieved over the years.

Yet for all of its successes, AI still has far to go. Progress in many areas has been painstakingly slow, and we still lack robust general-purpose systems that exhibit intelligence across a wide range of different problem domains, beyond those for which they were specifically designed. Most of the successful real-world AI systems developed to date operate only within narrowly defined boundaries. Master chess programs typically can’t play checkers or poker, even poorly; a robot that can navigate mazes or drive itself across town probably isn’t very good at diagnosing diseases or composing music, even at a very basic level. Efforts to develop natural language understanding systems and automated translators, an area of intense research for the past several decades, have yielded only a modest amount of progress. Developing truly flexible and robust systems with a broad spectrum of intelligence and common sense, what some researchers have begun calling Artificial General Intelligence (AGI), remains a distant and formidable goal.

Nevertheless, despite the many challenges still facing AI, there has been a remarkable increase over the past few years in the level of speculation and excitement about the prospects of achieving human-level intelligence in computers and robots within the next few decades. A new vision of humanity’s future has taken hold among many technology enthusiasts, and has been gradually making its way into the wider public consciousness. According to this vision, the continuing acceleration of computer power and other information-based technologies, combined with advances in molecular biology, genetic engineering, and the nascent field of nanotechnology, will soon lead to the development of machines that will match, and then quickly surpass, the intellectual powers of humans. Once machines attain this level of sophistication, they will be able to analyze, understand, and enhance their own designs, setting in motion an ever-accelerating cycle of recursive self-improvement. This will utterly transform our civilization, in ways that are almost impossible to foresee, and will mark the beginning of a new “posthuman” era in evolutionary history. This scenario is often referred to as the Singularity, by way of analogy with physics. Just as a physical singularity such as a black hole has an event horizon beyond which no information about it can be known, the technological Singularity represents a discontinuity in the flow of human history, beyond whose “event horizon”
nothing can be known or predicted.

The term “Singularity” seems to have first been suggested by the mathematician and science fiction author Vernor Vinge in the 1980s, but these ideas can be traced back much further. In 1965, for example, the statistician I. J. Good wrote of a coming “intelligence explosion”:

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an “intelligence explosion”, and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make.

More recently, the well-known AI pioneer and technologist Ray Kurzweil, and the robotics pioneer Hans Moravec of CMU, have introduced these ideas to a much wider audience through several popular science books (Moravec, 1988, 1999; Kurzweil 1999, 2005). In particular, Kurzweil’s book The Singularity Is Near has inspired much discussion and controversy within academic circles since its publication in 2005 (Singularity Summit 2006, 2007), and there is even a movie version of the book in the works, slated for release later this year. Kurzweil confidently and enthusiastically predicts that by 2029 computers will have fully passed the Turing Test, as a result of progress in reverse engineering the human brain made possible by the continued development of faster computational hardware according to Moore’s Law. Moravec predicts that cheap computer hardware will match the computational capacity of the brain by 2020. Others, however, including the respected computer scientist Bill Joy, are profoundly worried about the dangers posed to humanity by the development of intelligent robots and other powerful technologies. They argue that we should impose strict limits on current research in AI, genetics, and nanotechnology, or even abandon work in some areas entirely, before it is too late (Joy, 2000).

What can one make of all this? On the one hand, it is tempting to dismiss these ideas as simply the fanciful speculations of technology buffs prone to overindulging in science fiction. On the other hand, if one examines the arguments more carefully and objectively, they become harder to dismiss outright, especially if one takes a longer-range view of the evolution of life on Earth. There can be no doubt that something profound has been happening on our planet over the past few millennia. Around thousand years ago, the technology of agriculture first appeared on the Earth, invented by itinerant tribes of humans who were already quite adept at making tools for hunting and gathering their food. By a thousand years ago, things had changed dramatically. There were towns and cities and empires scattered across the Earth. The saddle and the stirrup had revolutionized travel on land, and great wooden sailing ships plied the seas. Within a few more centuries, the printing press and the steam engine and the harnessing of electricity had transformed human society as thoroughly as the invention of farming and the plough had several millennia earlier. And the past one hundred years have witnessed the most radical technological advances of all. Think of the world of 1908, as compared to the world of 2008. Back then, automobiles, telephones, phonographs, and flying machines were curiosities only recently invented. Television, digital computers, hydrogen bombs, space travel, mobile phones, and the Internet were as yet unimaginable, though they lay only a few decades away in the future.

What of the next one hundred years? Viewed from a larger historical perspective, it is hard to imagine that the technological advances of the coming decades will be any less profound or far-ranging than those of the past century. In all likelihood their impact will be far greater. Barring some unprecedented catastrophe, the arc of the trajectory seems clear: ever upward, at an ever faster pace. Our most advanced current technologies will seem far cruder and more antiquated a hundred years from now than the slide rules, flying machines, and Model T’s of 1908 seem to us today. Think of the implications.

But does this mean that we will have succeeded by then in creating conscious robots with superhuman intelligence, poised to supersede or even supplant us in the relentless ongoing march of evolutionary development? Could we homo sapiens really be on the verge of creating our own successors? The prospect of an imminent transition from carbon-based human intelligence to silicon-based machine intelligence does in fact have its precedents in the history of life on Earth: the origin of life itself was a “phase transition” from inanimate to animate matter, which took place billions of years ago on the early Earth; much later, life changed substrate again with the colonization of the land by early amphibians. Such things are possible. Or is it more likely, as Joy fears, that the development of ever more powerful technologies will, one way or another, unleash a catastrophe on a vast scale that destroys our technological society, or even renders humanity extinct? Perhaps all emerging intelligent civilizations in the Universe (if indeed any others exist) go through a similar dangerous period of technological adolescence, a kind of evolutionary “test” to determine if they are mature enough and wise enough to manage their newfound powers, and thus worthy enough to survive and flourish. Perhaps we are entering our own time of testing.

An Idea for a Course

Regardless of how much credence one gives to the arguments of Kurzweil and Moravec, the issues they raise are broad and fascinating. The concept of the Singularity, though centrally concerned with the future of AI and robot-
ics, encompasses a remarkably diverse range of topics in science and technology. This seemed to me like the ideal subject matter for a course aimed at liberal arts students that would introduce them to AI from a much wider perspective than is normally possible in a typical AI survey course. I decided to use Kurzweil’s book The Singularity Is Near as the focal point, and entitled the course First-Year Studies: Is the Singularity Near?. The interrogative title was chosen to emphasize that rather than simply taking Kurzweil’s arguments at face value, a central goal would be to examine these arguments as critically and objectively as possible. First-Year Studies designates a type of course required of all first-year students at Sarah Lawrence College. These are year-long seminars, typically having no more than about 13 students, which center around class discussions and the development of students’ writing skills. The rest of this paper describes the structure and content of this two-semester course. However, adapting it to a one-semester format by choosing a representative subset of topics and readings would not be difficult.

The overarching goal of the course, in accordance with the mission of First-Year Studies, is to teach students the writing, discussion, research, and presentation skills they will need in order to be successful during their four years at Sarah Lawrence, and beyond. The Singularity Is Near serves as an excellent framework for achieving this goal. Within this broader framework, the course focuses on the following more specific objectives:

Give students a realistic overview of AI by examining the successes and failures of AI research over the past fifty years and the challenges still facing the field. Kurzweil describes many current projects of interest, and discusses at length several important AI paradigms such as neural networks, genetic algorithms, expert systems, and heuristic search. However, I decided to supplement TSIN with Stan Franklin’s book Artificial Minds (1995) during the second half of the course, for deeper coverage of these topics. The book provides a balanced, detailed, and very accessible account of contemporary AI and its history, while limiting speculation about the future, which complements TSIN well.

Teach the basics of quantitative reasoning. The course was designed for students with non-technical backgrounds who may lack a strong preparation in science or mathematics. Since Kurzweil uses scientific notation and order-of-magnitude estimates extensively throughout TSIN to support his arguments, it is crucial for students to understand these concepts. For instance, the idea of exponential growth is central to many of Kurzweil’s claims. He often refers to the “knee of the curve” to describe the apparent transition from slow to explosive growth on a linear-scale exponential graph. But, as Douglas Hofstadter has pointed out, this is misleading. The “knee” is strictly an artifact of the particular scale chosen for the axes, and is not an inherent property of the curve itself or the underlying process it describes. This can serve as an excellent jumping-off point for a review and discussion of the graphical presentation of data, and the mathematical meaning of linear, exponential, and logarithmic growth.

Teach critical thinking skills. In the last chapter of TSIN, Kurzweil identifies fourteen specific criticisms that others have put forward in an effort to show that his projections about the future are either wrong or implausible. He then presents a detailed and carefully argued set of responses to his critics. This chapter provides an excellent opportunity for students to learn to analyze the strengths and weaknesses of the arguments on both sides of the debate, and could serve as the starting point for discussions of how to evaluate and interpret evidence, and how to recognize logically sound or fallacious reasoning.

Get students excited about the big questions of science, and get them to think about humanity’s place within the larger context of evolutionary history. TSIN ranges across the full spectrum of modern scientific ideas, touching on, in varying levels of detail, the Big Bang and cosmic evolution; the evolution of life on Earth; the possibility of extraterrestrial intelligence; the anthropic principle; molecular biology, DNA, and genetics; neuroscience and the brain; the chemistry and physics of nanotechnology; complex systems and cellular automata; Turing machines and undecidability; Moore’s Law and computer engineering; digital logic; quantum mechanics and quantum computing; consciousness and the mind-body problem; and many other topics as well, including of course, AI and robotics. While this presents many opportunities for individual students or the class as a whole to delve more deeply into particular areas of interest, it also assumes a fairly extensive amount of general scientific knowledge on the part of the reader. For this reason, I decided to devote the first few weeks of the course to reading and discussing Carl Sagan’s book Cosmos (1980), by way of preparation for TSIN. Sagan provides a marvelous account of the history and scope of science that sets the stage perfectly for the ideas to come.

Teach basic research skills. TSIN includes over one hundred pages of extraordinarily detailed footnotes and references, which constitute one of the most useful features of the book. Many of these references are to cutting-edge information available on the Web, which makes it easy for students to dig deeper. Each semester, students are required to write an in-depth research paper and to give an oral presentation on some topic relevant to the course, broadly construed, and the notes section serves as an invaluable resource for their research. Some examples of research topics chosen by students include gene therapy and its ethical implications; the current state of nanotechnology; brain-machine interfaces such as silicon retinas, cochlear implants, and artificial limbs; human-computer interaction and user interface design; telepresence and virtual reality; the impact of social networking sites such as Facebook and MySpace on society; the impact of the Internet on the music industry; near-term plans for the human colonization of space; genetically modified foods; online
virtual worlds and massively multiplayer games; 21st century military technologies; and the threat of genetically engineered pathogens and bioterrorism.

Make students more aware of the dangers posed by emerging technologies. If we are to survive, the public will need to have a much deeper understanding of science and technology than it currently has, in order for society to make informed choices in the coming decades. Thus it is important for students to understand the potential risks of future technological advances. TSIN discusses these issues in depth, from Kurzweil’s point of view. For a different perspective, I also included on the reading list Our Final Hour by Sir Martin Rees (2003), the distinguished Cambridge astrophysicist. Rees gives a balanced and sober assessment of the risks, which serves as a useful counterpoint to Kurzweil’s perhaps overly optimistic account.

Course Synopsis

An outline of the major course sections and topics is given below:

Numbers and Scale
• Scientific notation
• Orders of magnitude
• Exponential processes

Evolution and the Cosmic Perspective
• Our place in the Universe
• Biological evolution
• Are we alone?

The Acceleration of Technology
• Moore’s law and the increasing power of computation
• Reverse engineering the human brain
• Biotechnology and genetic engineering
• Nanotechnology
• AI and robotics

Prospects For Survival
• The promise and peril of advanced technologies
• Existential risks

Artificial Intelligence: A Closer Look
• Symbolic AI
• The Turing Test and the Chinese Room
• Connectionist AI
• Robots and embodied AI
• Machine creativity

As mentioned earlier, a fundamental grasp of quantitative reasoning is essential for understanding Kurzweil’s arguments, so these concepts are covered first. Students read the first two chapters of Carl Sagan’s book Billions and Billions (1997), which explains the basics of scientific notation and exponential growth, and also On Number Numbness by Douglas Hofstadter, an entertaining discussion of techniques for estimating large and small quantities, taken from his book Metamagical Themas (1985). In the spirit of Hofstadter’s article, a follow-up homework assignment asks students to calculate or estimate various quantities such as how many nanoseconds there are in a year, how many music CDs could fit on an iPod with one petabyte of storage, how many tons of garbage New York City produces each week, how fast one’s hair grows in miles per hour, and so on. This is a fun way for students to become more comfortable with scientific notation and back-of-the-envelope calculations. It is also an opportunity to review the differences between million, billion, trillion, and other such big numbers, which many non-technically-inclined students tend to find confusing. Students also watch the famous short video Powers of Ten by Charles and Ray Eames (1977), which illustrates scientific notation and the idea of exponential and logarithmic scaling in a vivid, compelling way. This sets the stage for the next unit of the course, on evolution and the Universe at large.

To fully appreciate the concept of the Singularity and to grasp its true significance, one must gain an understanding of what Carl Sagan used to call the Cosmic Perspective—a view of humanity’s place within the vast expanses of space and time of our Universe as revealed by modern science, especially through astronomy, biology, and anthropology. Sagan’s book Cosmos is extremely effective at conveying this view, which rests on an understanding of biological evolution in general, and the evolutionary history of human beings in particular. Students spend several weeks reading and discussing Cosmos, and watching selected episodes from the accompanying TV series. Sagan also considers the possibility of the evolution of intelligent life elsewhere in the Universe, as does Kurzweil in TSIN, although interestingly they come to quite opposite conclusions on this matter. An additional reading selection by Hofstadter on “viral” sentences and self-replicating structures (1985, Chapters 1-3) helps to explain and clarify the evolutionary mechanisms underlying self-replication, a key concept that plays an important role in Kurzweil’s later discussions of the power and potential dangers of nanotechnology.

At this point, having finished Cosmos, the students are prepared to tackle The Singularity Is Near, which serves as the main reading for the next unit of the course, on the acceleration of technology. At slightly over six hundred pages, TSIN is a rich feast of ideas and information about current research in many branches of science, and at least several weeks are needed to digest all of it. Some sections of the book are quite dense and technically detailed, so class time is often spent discussing and clarifying these more difficult passages. Other articles may occasionally be assigned to supplement the information from TSIN. For example, Richard Feynman’s classic lecture There’s Plenty of Room at the Bottom (1959) is still a
wonderfully lucid introduction to the main ideas behind nanotechnology. Students also regularly post written responses to the assigned readings on a class webboard system, which further helps to facilitate discussion and understanding.

After finishing TSIN, we turn our attention to the risks associated with newly emerging technologies. Kurzweil discusses these at length from his own perspective, but it is important to consider other viewpoints as well. In addition to Rees’ Our Final Hour, students read Joy’s famous article Why The Future Doesn’t Need Us (2000), which received widespread attention when it was first published. This article was written as a direct response to the scenarios of the future being publicized by Kurzweil, Moravec, and others, and argued strongly in favor of relinquishing certain technologies that Joy saw as being too dangerous to pursue. Soon afterward, Jaron Lanier, another well-known computer scientist and a pioneer in virtual reality, published One-Half of a Manifesto (2000), an article that criticized the assumptions underlying the arguments of both Kurzweil and Joy, on the basis of software complexity. Having read TSIN in detail, students are in a position to evaluate the arguments of Lanier and Joy for themselves, and to make up their own minds about where the balance point currently lies between the promise and the peril of future technologies.

The last section of the course focuses on AI as it currently is, as opposed to what it might or might not become in the future. I strongly believe that a deeper understanding of the issues and technical challenges facing real AI, right now, is essential in order to be able to effectively evaluate claims made about the future of AI. Thus the goal of this final unit, which occupies the majority of the second semester, is to convey these key ideas in a sufficiently detailed but still comprehensible way. To this end, I chose as the main text Artificial Minds by Stan Franklin, supplemented by numerous other articles.

Our survey of AI is organized into five components: symbolic approaches, philosophical questions centering on the Turing Test and the Chinese Room, connectionist approaches, embodied approaches, and machine creativity. An excerpt from Luger’s AI textbook (2001, Chapter 1) provides a concise summary of the philosophical foundations and major paradigms of AI. Hubert Dreyfus’ article From Micro-Worlds to Knowledge Representation: AI at an Impasse (1979) gives a good overview of early work in symbolic AI (such as Winograd’s program SHRDLU) while criticizing its fundamental tenets. This leads naturally to the Turing Test, and to Searle’s famous Chinese Room thought experiment, in which he attempts to show that the Turing Test is unsound. Students read and discuss the original papers (Turing, 1950; Searle, 1980), as well as Hofstadter’s piece A Coffeehouse Conversation on the Turing Test (1985, Chapter 22). They also read Hofstadter’s dialogue A Conversation With Einstein’s Brain (1981), which provides some useful and amusing perspective on the debate, and on Searle’s arguments in particular.

From there we move to neural networks and connectionism. The Appeal of Parallel Distributed Processing (McClelland, Rumelhart, and Hinton, 1986) provides an excellent introduction. Although some students may find this article challenging at first, a computer demo of the Jets & Sharks network discussed in the article helps to make the ideas more concrete and understandable.

The next section on embodied cognition and robots focuses on Braitenberg vehicles and Rodney Brooks’ subsumption architecture (Brooks, 1991). The concept of emergent behavior and the importance of sensorimotor interactions with the environment are the key ideas here. Students invariably find the first few chapters of Braitenberg’s seminal book Vehicles: Experiments in Synthetic Psychology (1984) to be as fascinating and thought-provoking as they are accessible and entertaining. The article by French (2000) connects the idea of embodiment back to earlier discussions of the Chinese Room and the Turing Test in a particularly effective way.

Discussions at this point can be supplemented with in-class videos and demos of actual robots. An excellent choice is The Great Robot Race: The DARPA Grand Challenge (NOVA, 2006), which describes the 130-mile race across the Mojave Desert in 2005 between 23 autonomous vehicles, ultimately won by Stanford’s entry. See the course’s home page (Marshall, 2008) for links to other good videos available on the Web.

Finally, we spend some time discussing the idea of machine creativity, focusing on AARON, a program that creates drawings and paintings (McCorduck, 1991); EMI, a program that composes music in the style of Bach and Mozart (Cope, 2005); and Copycat, a program that solves analogy problems (Hofstadter, 1994, 1995).

Conclusion

In designing this course, my main goal has been to convey as wide a perspective of science as possible, using TSIN as the centerpiece. Although the course incorporates AI and computation as major themes, in its present form it may fit more naturally within an interdisciplinary Science, Technology, and Society program, rather than Computer Science. By omitting some or all of the material on cosmic evolution, biotechnology, nanotechnology, or our future survival, it could be tailored to place greater emphasis on core CS topics such as algorithms and computability, and could go more deeply into neural networks, genetic algorithms, or other AI topics. TSIN could then serve as a supplementary text, useful for its wealth of information on these topics and on recent AI research. Either way, the concept of the Singularity provides a lofty vantage point from which to explore the heady ideas and implications of AI, and to ponder its ultimate impact on the future.
Acknowledgments

I would like to express my thanks to Doug Hofstadter for bringing this fascinating topic to my attention in a memorable conversation one evening in his home, and to the anonymous reviewers for their comments and suggestions.

References


