Using Science Fiction in Teaching Artificial Intelligence

Milind Tambe
Computer Science Dept.
University of Southern California
Los Angeles, CA 90089
{tambe}@usc.edu

Anne Balsamo
Interactive Media Division
School of Cinematic Arts
University of Southern California
Los Angeles, CA 90089
abalsamo@cinema.usc.edu

Emma Bowring
University of the Pacific
Stockton
CA 95211
bowring@usc.edu

Abstract

Many factors are blamed for the decreasing enrollments in computer science and engineering programs in the U.S., including the dot-com economic bust and the increase in the use of "off-shore" programming labor. One major factor is also the lack of bold new vision and excitement about computer science, which thus results in a view of computer science as a field wedded to routine programming. To address this concern, we have focused on science fiction as a means to generate excitement about Artificial Intelligence, and thus in turn in Computer Science and Engineering. In particular, since the Fall of 2006, we have used science fiction in teaching Artificial Intelligence to undergraduate students at the University of Southern California (USC), in teaching activities ranging from an undergraduate upper division class in computer science to a semester-long freshman seminar for non-engineering students to *micro-seminars* during the welcome week. As an interdisciplinary team of scholar/instructors, our goal has been to use science fiction not only in motivating students to learn about AI, but also to use science fiction in understanding fundamental issues that arise at the intersection of technology and culture, as well as to provide students with a more creative and well-rounded course that provided a big picture view of computer science. This paper outlines the courses taught using this theme, provides an overview of our classroom teaching techniques in using science fiction, and discusses some of the lectures in more detail as exemplars. We conclude with feedback received, lessons learned and impact on both the computer science students and non-computer-science (and non-engineering) students.

"Science fiction like Star Trek is not only good fun, but serves a serious purpose, that of expanding human imagination" Physicist Stephen Hawking (from (Krauss 1995))

Introduction

Since 2000, many universities in the U.S. have witnessed decreasing enrollments in computer science programs(Vegso 2005; Zweben 2005). This trend has been explained as a consequence of a decrease in job opportunities due to the dot-com bust and/or as a result of the increasing reliance on off-shore programming labor. While the perception of decrease in job opportunities may certainly contribute to the lack of interest among students, it is also likely that the popular view of computer science — as a dreary form of non-creative labor that focuses purely on arcane programming skills — doesn't help create excitement about the field(Roboprof 2006). To combat this perception among students at the University of Southern California, we created three courses that use science fiction materials as a way to provide an imaginative and creative context for the discussion of topics in Artificial Intelligence (AI). As an interdisciplinary team of instructors/scholars, we developed a pedagogical framework on "Artificial Intelligence and Science Fiction" to teach a upper-division course in computer science, a freshmen seminar for non-engineering students, and a series of lectures for incoming freshmen students. This framework weaves the bold, fictional, futurist visions of computing into educational efforts to teach students fundamental concepts in AI. The aim is to expand their imagination about the importance, relevance, and creative potential of AI more specifically and computing in general.

Thus, our mission was not only to teach students fundamental concepts in AI, but also to excite them about AI and computing in general. In addition we wished to provide students with a bigger picture view of what they were doing and prepare them to deal with ethical issues relating to their work. It is in this context that science fiction provides four key benefits. First and foremost is that it instills a sense of wonder. The second is that it appeals to students' cultural sensibilities: many of them (but not all) are fans of science fiction films, book, and other forms of popular culture. Thirdly, science fiction stories provide a narrative context for discussing the social importance and significance of AI theories and research. Finally, in using historical works of science fiction, instructors can discuss progress in AI research made over the past forty years to explore paradigmatic changes that have shaped the research agenda of the field of AI.

The interdisciplinary team teaching these classes included a faculty member from the school of Engineering with expertise in AI (Tambe), a faculty member from the school of Cinema with focus on technology and culture (Balsamo) and a PhD student from Computer Science with keen interest and focus on undergraduate teaching (Bowring). Interdiscipli-

nary teaching techniques have been discussed in detail in the literature(Hanyes 2002). Our goal was to face the challenges and realize the benefits of interdisciplinarity in practice. In terms of (Hanyes 2002), our undergraduate upper-division course was *cross-disciplinary* in the sense of being dominated by one discipline (Computer Science or AI) but supplemented by other disciplines; whereas our freshman seminar was intended to be a interdisciplinary course(Hanyes 2002) that explored the theme of interaction of technology and culture, focusing on AI — and in this sense science fiction allowed us to investigate this interaction.

In the rest of this paper, we discuss first our upper division course in computer science (called "Intelligent Agents and Science Fiction"), followed by a discussion of the freshman seminar (called "Science, Technology and Culture: Artificial Intelligence and Science Fiction") and finally a discussion of our micro-seminars. We outline our syllabi, teaching techniques and discuss student evaluations and lessons learned.

Intelligent Agents and Science Fiction

This upper division course focuses on introducing concepts in "Intelligent agents and multiagent systems" using science-fiction stories and films. When first introduced, the course had three goals. The first was to introduce a new subfield of AI, that of inteligent agents and multiagent systems to undergraduate students. Intelligent agents and multiagent systems is a growing subfield of AI that focuses on the development of complete individual agents and the interaction among multiple agents. Our intent was to stimulate student interest in a subfield of AI that takes a more holistic view of AI. The second goal focused on using science fiction as an educational resource to provide a social and fictional context for the discussion of basic elements of intelligent agent design. Providing students with a more creative and well-rounded course that touched on philosophy, ethical concerns, etc was our third goal — to attract a broader student audience that was keen on understanding the social context of computing.

There are five key concepts in teaching this course. First, each lecture or group of lectures in this course builds on a science-fiction short story, a movie clip or a television episode. The science fiction texts were selected based on the following key characteristics:

- It had to exhibit some topic of interest in the agents and multiagent systems arena, e.g. agent modeling, emotions, teamwork, agent interactions under uncertainty, etc.
- The story or film clip had to feature the robot or AI as an active participant in the plot or story.
- The story needed to present the robot/agent in a positive light. There is already a lot of popular science-fiction presenting a negative view of robots — combating this perspective with a positive view of robots/agents was considered important in order to motivate our students. (Unfortunately, meeting requirement for all of our topics of interest proved to be quite difficult, and in one or two cases, we had to relax this requirement.)
- The story had to be short enough (maximum 30 pages); if a movie clip was to be used, it had to be short enough to be shown in 5-10 minutes at the beginning of class.

We chose several short stories by Isaac Asimov from his collection "Robot Vision" (Asimov 1991). In addition to the short stories, the book also includes several essays by Asimov that provided useful reading material for the third key concept described below. We also chose episodes from "Star Trek: The Next Generation" that focused on the issue of agents, robots and intelligence.

Course lectures highlighted key aspects of agent behavior and functioning. For example, in the Asimov story, "Little Lost Robot" a robot (Nestor-10) intentionally tries to hide among a group of similar robots, while a human tries to run tests to isolate it from other robots. Nestor-10 is just different from other robots in the rules it follows. In the story, Nestor-10 considers what the human believes the other robots will do in order to behave like all other robots — so it can blend in. Meanwhile, the human must devise tests that prevent Nestor-10 from blending in. In particular, she must infer the plan the robot is executing to pass the test; by observing the robot's actions, she can infer the plan the robot is executing, which in turn reveals whether the robot is Nestor-10 or not. Unfortunately, initially the human fails to recognize Nestor-10 because it anticipates the human's intentions, and foils those by continuing to successfully blend in. This story provides a fictional context for introducing basic concepts in agent modeling. In particular, the story provided three specific settings to investigate three particular aspects of agent modeling.

- The basic idea of human trying to infer the plan a robot is executing by observing its actions provides an initial introduction to plan recognition.
- Nestor 10 must predict what other robots would do, and immitate their actions. This allows us to discuss how agents predict other agents' behaviors.
- Nestor 10 must recursively model what the human believes other (non-NESTOR-10) robots would do, enabling a discussion about recursive agent modeling.

Similarly, the Asimov story, "RUNAROUND" is based on a robot facing conflicting directives, which in effect brings up the notion of intention reconsideration and conflicting commitments. We can explain the behavior in terms of Belief-desire-intention (BDI) concepts(Wooldridge, Muller, & Tambe 1996), and understand how to avoid such conflicts.

We outline in Table 1 some of the science fiction materials used and the concepts that were introduced using them. In all of these instances, students were either asked to read the story before class, or shown a short clip from the episode or film during class. In this way the story or film provided a context for a discussion about the key concept: i.e., why is agent modeling important, what are the difficulties in the problem; what are the key concepts offered in the filmic or textual "solution".

Our second key concept in teaching this course was to introduce cutting edge topics based on science fiction in the

Science Fiction Story or Movie	Brief Synopsis	Intelligent Agents and Multiagent Systems Concepts
Runaround by Isaac Asimov	Robot is stuck running around in circles because of internal rule conflict	Agents based on "Beliefs, desires, intentions" (BDI)
The Enemy (episode from "Star Trek: The next generation")	Humans and Romulans enter a dangerous game of who blinks first	Introduction to Game Theory
Descent Part I (episode from "Star Trek: The next generation")	The robot "Commander Data" shows emotions	Agent Emotions
Little Lost Robot by Isaac Asimov	Robot must reason what human trying to find it thinks about it	Agent modeling or plan recognition
2001	HAL	Adjustable autonomy and safety
Minority Report	Clip showing robotics spiders	Multiagent teamwork
Fast times at Fairmont high by Vernor Vinge	Small sensors create a network	Distributed constraint reasoning
The Swarm by Bruce Sterling	Insect species	Multiagent Swarms
The Offspring (episode from "Star Trek: The next generation")	Commander data creates an artificial offspring and must teach it.	Machine learning
Who watches the watchers (episode from "Star Trek: The Next generation")	Human colonists must be transported via shuttles	Coalition formation

Figure 1: Intelligent agents and science fiction: Science Fiction Material used and concepts introduced

Figure 2: Commander Data, the robot from Star Trek, displays anger in attacking the "Borg". This episode was used in teaching the topic of emotions in Agents.

Figure 3: A team of robotic spiders from the movie "Minority Report" was used in introducing the concept of teamwork and distributed POMDPs.

course, to instill that sense of wonder about AI and computing. We therefore intentionally introduced topics such as agent emotions in our class(Elliott 1992; Marsella & Gratch 2002). Figure 4 shows a picture of Commander Data from the popular science fiction series "Star Trek: The next generation" getting angry, i.e. showing emotions. This episode was used in teaching students about emotions in agents. The famous scene of the spiders from the movie "Minority Report" was used to introduce students to the concept of teamwork in general, and in particular, to the more cutting edge topic of distributed POMDPs(Nair *et al.* 2003; Becker *et al.* 2003).

Our focus on "Intelligent agents and Multiagent systems" had already restricted our choice of textbooks. Given that we were covering topics not covered in these textbooks, and that we also wished to cover cutting edge topics led us to the conclusion that we could not rely on any textbook; yet providing undergraduate students detailed mathematical research papers to read appeared to be an unappealing option. To compensate for the textbook we wrote a detailed set of class notes in Intelligent agents and multiagent systems. In their first incarnation, these notes unfortunately did not include many examples; this is a shortcoming that we are actively attempting to address as we revise these notes.

A third key concept in this course was to introduce students to social, ethical issues surrounding AI, as a means of going beyond basic computer science aspects. Our goal was also to engage students in active learning(Bonwell & Eison 1991) and teach them about challenges of defining AI as a field and what it means to be intelligent. The aim was also to tie the material in class to a broader context and (hopefully) make connections to other classes/areas of interest, e.g. philosophy. Here we showed students the film clip of the "The trial of Commander Data" from "Star Trek: The next generation." Commander Data is an Android who is put on trial to determine whether or not it has rights. In particular, if the ruling of the trial is that Commander Data has no rights, then he would be immediately dismantled for futher investigation — with a slim chance of being put back together.

The trial took place in two separate lectures. In the first lecture, students were shown the first half of the episode that chronicles the events leading up to Data's trial. We then

divided the class into four teams and gave the teams time to strategize/coordinate. Each team was assigned to argue either the pro or con of one of these two issues:

- Is commander Data intelligent? self-aware? sentient?
- Does Commander Data have rights? If Data creates art, who owns it? If Data kills someone, who is responsible?

We provided the students with a list of readings, supporting both pro and con positions. On one side were Turing's "Computing Machinery and Intelligence" (Turing 1950) and Asimov's essays (Asimov 1991) and on the other were readings such as "Can a computer have a mind" (Penrose 2002) and Searle(Searle 1980). To prepare for the trial, students had to do the following homework:

> *As homework, you need to read 2 readings from the list provided below, at least one must be non-Asimov. Then, write a short (roughly half a page) list of three supporting arguments for the position you have been assigned. Your arguments can either be direct support or refutations of likely counter-arguments that the opposing team will make. For each of your points, please provide some support based on a reading. 1 point of extra credit can be earned on this assignment if you bring in an additional credible source to support your argument and provide a citation.*

In the second lecture on this topic, students re-enacted the trial in class. During the trial, students had to speak up in support of their position based on their writeup submitted as part of their assignment. As instructors, we were not completely sure how this trial would work out. Our observation was that students were really passionate in support of their positions and had researched many extra sources. Several new and innovative arguments were brought to the floor, e.g. one student brought up the National historic preservation act to argue that Commmander Data could not be dismantled because it was a historic Engineering artifact.

A fourth key concept in this class was to introduce students to themes in the history of science fiction. The goal was not only to introduce students to this popular cultural genre, but also to explain its significance in the development of the technological imagination. In this section, students were given a presentation on the history of science fiction that covered topics such as its genre characteristics, its historical phases, and its relation to the development of technological insights. The point was to elaborate how many technological advances were first proposed and imagined by fiction writers. The broader point was to explore the relationship between technology and culture as the context for the development of intelligent agents.

The fifth and final concept in this course, a major decision contrasting it with other AI courses, was to ensure that students did not concentrate on programming. As a result, the course as a whole focused on key concepts, and our assignments reflected this choice. In fact, there were intentionally no programming assignments; instead the focus was on the mathematical concepts. For example, it was possible to ask students to obtain mixed-strategy nash equilibria in an assignment on game theory; or to provide a specification of a problem within a distributed POMDP framework.

Freshman Seminar

Using science fiction as its core component, this interdisciplinary seminar introduced non-engineering Freshmen students to the relationship between science, technology and culture through an investigation of the development of the science/engineering of Artificial Intelligence (AI). The freshman seminar runs for 11-weeks, and meets once a week. Our goal here was to reach students early on in their stay at USC, possibly before they get turned off of computer science. Our challenge was to make materials that were traditionally available in computer science — and that too after spending a number of years programming — to a different student audience at a different level. Yet to motivate and excite these students we wished to engage them in interesting open questions so that they could think big picture, and get inspired to take on computing and computer science.

To this end, the course had three key sections. The first was scientific material that included a discussion of AI's goals, and and introduction to fundamental concepts and techniques (which include basic planning, learning, multiagent interactions via basic game theory). Second, the course materials also included a selection of science fiction novels, films and short stories that demonstrate: 1) the concept of the technological imagination; 2) the process whereby this imagination is shaped by culture (i.e., fiction and film), and 3) the results of the technological imagination-in-action: the formation of a scientific/engineering research program on AI and the design of AI experiments and demonstrations. Third, the course materials engaged the students in active learning, once again by having them engage in the trial of Commander Data, as well as several classroom games and exercises.

Note that in contrast with the upper-division computer science course, the purpose of this course was to give equal coverage to science fiction. For example, midway through the course, students watched the science fiction film, Bladerunner (1985) as a "text" to use to discuss related course topics. The film is set in a future when artificial humans have become a corporate commodity. These AI humanoid androids are used for various purposes: off-world colonization efforts, hazardous condition mining, military pleasure. The plot of the film centers on the confusion and anxiety that results when a new version of android, the Nexus Six, is developed that not only mimics human behaviors, but is augmented with a set of human "memories". The result is that the Nexus Six android can effectively simulate realistic human emotions and believes, because it accesses human memories, that it is indeed human. The film thus presents a story world in which the science of artificial intelligence and bio-engineering have advanced to the point where humanoid robots (i.e., androids, or replicants) are indistinguishable from humans. The film stimulated discussion about the characteristics that differentiate humans from intelligent agents, where the more we discussed the qualities of an "agent" the clearer it was that the discussion was really about what it means to be "human" in a technological age.

Figure 4: Micro-seminars on AI and Science Fiction have been well attended.

Microseminar: Welcome Week

The concept of *microseminars* is to introduce incoming freshmen to academic discourse at USC, and to encourage faculty-student ineractions. Microseminars involve two 90 minute lectures, and these seminars are held during the welcome week before classes start. Our microseminars in 2006 and 2007 have focused on "Artificial Intelligence and Science Fiction", and gave students a very basic introduction to the field of Artificial Intelligence; these seminars have been very well attended.

A key question was to select the material to teach at the seminars. To this end, we selected materials that introduced students to AI, provided an overview of how one would program an agent, as well as some of the cutting edge topics to instill a sense of wonder and leave the students at the end of the lecture with some intriguing ideas. We used science fiction to aid in the following:

- Lecture 1 (part I): Using robots and AI programs from popular science fiction literature (e.g. Commander Data, HAL, R2D2), and contrasting them with programs such as "Deep Blue", focus on the concept of an intelligent agent.
- Lecture 1 (part II): Use Isaac Asimov's "Runaround" (mentioned above) to introduce basic concept of a Belief-desire-intention architecture for an agent.
- Lecture 2: Use clips from the movie "I, Robot" to introduce the notion of robots with emotions. This lecture also introduces game theory, and then impact of human emotion on their approach to reasoning in standard games.

In Lecture 1, and right through the middle of Lecture 2, students typically reject the utility of emotional robots. Lecture 2 introduces students to rational reasoning in game theory, and then to how people play in games such as prisoner's dilemma. By the end of the lecture, the students see the impact and potential benefit of having agents and robots with emotions.

Feedback and Lessons Learned

Results from student feedback on the courses have been encouraging. Students provided overwhelmingly positive response, not only about the course material but also about the teaching technique of using science fiction. With respect to the "Intelligent agents and science fiction" course, there were two sets of feedback. The first was prior to teaching this course. Here, the the feedback was from the faculty themselves. The key concern expressed was whether science fiction was diluting the content of the course. Once this concern was addressed, the faculty enthusiastically endorsed the course.

The response at the end of all of our courses was overwhelmingly positive. More specifically:

- We provided students with additional questions to evaluate the role of science fiction: 16/20 students who provided feedback thought that the science fiction really added value to the material taught. Students commented that without the science fiction they would not have taken the course.
- Some of the overwhelmingly positive comments that declared the teachers to be *the best professors at USC* clearly illustrated the effect of using Asimov and Star Trek in the classroom.
- Our computer science class and the microseminar lectures to incoming freshmen were covered separately in campus newspapers. At least one student interviewed for the campus newspaper right after the microseminar suggested that he would change his major to computer science; students from the microseminar have continued to correspond via email querying about AI.
- At least two students from the upper division computer science class have joined AI research labs to pursue research in Artificial Intelligence; one of whom is now all set to enter a PhD program in computer science.
- A few of the students also immediately followed up the courses to take more in-depth courses in computer science.

Given the success of our courses, it is useful to step back and understand some of the key factors that in our view contributed to this success. These include:

- We were teaching material for which there was no clear textbook, bringing in cutting edge research material into our courses, and simultaneously trying for a cross-disciplinary or interdisciplinary approaches. This required significant amounts of planning and we were fortunate enough to have been forced into this planning a year in advance partly by our faculty colleagues who engaged us in lively discussions.
- Figuring out the right science fiction material to use given our criteria proved to be extremely difficult. If the trend of using science fiction in AI were to be carried forward, it would help to build up a collaborative database of science fiction materials utilized.
- Having an interdisciplinary team was a significant advantage.
- Finally, colleagues in AI and in particular agents and multiagent systems from around the world were extremly sup-

portive of this effort, providing both encouragement and pointers to relevant materials.

Summary and Related Courseware

At USC, we have used science fiction as a means to generate excitement about Artificial Intelligence, and thus in turn stimulate student interest in pursuing educational programs in Computer Science and Engineering. Our mission was not only to teach students fundamental concepts in AI and excite them about AI and computing in general, but also to provide students with a broader view of the social and cultural context of the development of intelligent agents, including a discussion about the ethical issues relating to this work. To that end, since the Fall of 2006, we have taught an undergraduate upper division class in computer science, a semester-long freshman seminar for non-engineering students and microseminars during the welcome week. Detailed week to week syllabus and other details on these courses are available from the course web sites:

- http://teamcore.usc.edu/tambe/CS499 for the course "Intelligent agents and science fiction".
- http://teamcore.usc.edu/tambe/freshman-seminar.htm for our freshman seminar.

This paper outlined the courses taught using this framework, provided an overview of our classroom teaching techniques in using science fiction, and discussed some of the lectures in more detail as exemplars. We discussed the overwhelmingly positive student response and concrete examples of students turning to computer science and computer science research as a result.

There are other courses offered in other universities that use science fiction as a way of introducing science in general, of which in general AI might be a very small part. Prof. Barry Luokkala of Carnegie Mellon University, Teaching Professor in the Physics department, uses science fiction for introducing science. Another similarity is NIH's program called "Science in the Cinema" http://science.education.nih.gov/cinema and the American Chemical Society's similar program: http://www.scalacs.org/ScienceCinema/. In particular, those programs also use film and world leading scientists commenting on the films to educate students and the general public. However, to the best of our knowledge, ours is the first course in AI, and in general in Computer Science to employ science fiction.

References

Asimov, I. 1991. *Robot Vision*. Penguin publishers.

Becker, R.; Zilberstein, S.; Lesser, V.; and Goldman, C. V. 2003. Transition-independent decentralized Markov decision processes. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-03)*, 41–48.

Bonwell, C., and Eison, J. 1991. Active learning — creating excitement in the classroom. *ASHE-ERIC Higher Education Report* 1.

Elliott, C. 1992. *The affective reasoner: A process model of emotions in a multi-agent system*. PhD dissertation 32, Northwestern, IL, Northwestern University Institute for the Learning Sciences.

Hanyes, C. 2002. *Innovations in interdisciplinary teaching*. Oryx Press.

Krauss, L. 1995. *The physics of Star Trek*. Harper Collins.

Marsella, S., and Gratch, J. 2002. A step toward irrationality: Usign emotion to change belief. In *Proceedings of First International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS-02)*.

Nair, R.; Pynadath, D.; Yokoo, M.; Tambe, M.; and Marsella, S. 2003. Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, 705–711.

Penrose, R. 2002. *The emporer's new mind: Concerning computers, minds and the laws of physics*. Oxford University Press.

Roboprof. 2006. *http://insidehighered.com/news/2006/07/12/robot*. Insiderhighered.com.

Searle, J. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3:417–457.

Turing, A. 1950. Computing machinery and intelligence. *Mind* 49:433–460.

Vegso, J. 2005. Interest in cs as a major drops among incoming freshmen. *Computing Research News* 17:6–1.

Wooldridge, M.; Muller, J.; and Tambe, M., eds. 1996. *Intelligent Agents*, volume 2 of *Lecture Notes in Artificial Intelligence 1037*. Springer-Verlag, Heidelberg, Germany.

Zweben, S. 2005. 2004-2005 taulbee survey: Production at an all-time high with more new graduates going abroad; undergraduate enrollments again drop significantly. *Computing Research News* 18:7–17.