

# Addressing the Technical, Philosophical, and Ethical Issues of Artificial Intelligence through Active Learning Class Assignments

**Pamela K. Fink**

St. Mary's University, One Camino Santa Maria, San Antonio, Texas, USA  
pfink@stmarytx.edu

## Abstract

Artificial intelligence (AI) is an extremely large and complex field technically, while at the same time it captures our imagination and prompts us to explore major philosophical and ethical questions concerning humanity and human intelligence. Teaching a course that does justice to all these aspects of the field is a big challenge. However, due to the increase in computational capability with a commensurate decrease in cost, a wealth of products and materials are available that can be used to provide students with rich, meaningful, and memorable experiences within the context of a primarily technical course in AI. Toys, articles, and movies can all be used to foster student exploration of key questions in the technical, philosophical, and ethical issues of AI.

## Introduction

The field of artificial intelligence is challenging to teach for a variety of reasons. As stated by Russell and Norvig in the Preface of their third edition of *Artificial Intelligence: A Modern Approach*, “Artificial Intelligence (AI) is a big field, and this is a big book.” (Russell and Norvig, 2010). As a matter of fact, this leading textbook used in the teaching of AI for over 20 years is over 1000 pages. Indeed, it is difficult to fit all that is needed into an introductory AI course. In order to be successful, a balance must be found in covering the theory and the practice of AI, as well as the philosophy and ethics. The field covers a very large set of tools and techniques including search strategies, logic, and various other knowledge representation schemes, as well as a wide range of problem solving areas such as game playing, natural language processing, planning, perception, and learning. In a computer science version of the course, one wants to do justice to the technical components of the field. However, at the same time, it is important to place AI into its philosophical and ethical context regarding intelligence and humanity.

The intelligent agent construct utilized in most of the leading textbooks since the mid-1990's (see, for example, Russel and Norvig's first edition of their textbook, from 1995, and Nilsson's *Artificial Intelligence: A New Synthesis* from 1998) provides a coherent means of assimilating several goals of an AI course, including most of the technical issues of how such behavior can be implemented, as well as an initial look at the philosophical question of what constitutes intelligent behavior, and the ethical implications that arise as a result of creating a non-human intelligent entity.

## Example Course Assignments

The following describes three different assignments utilized in teaching concepts in each of the three areas of technical, philosophical, and ethical issues, that exploit the concept of an intelligent agent, as well as the readily available multimedia materials to provide hands-on, active learning experiences for students of AI. These assignments are just three examples among many, much more technical assignments given in a primarily technical introductory course on AI.

### Technical Assignment: Adopt an Intelligent Agent

The author has had experience teaching AI since the mid-1980's and has had, therefore, to teach it when it was difficult to have a machine powerful enough to implement much intelligent behavior in a reasonable period of time and cost. With the advent of smaller, faster, cheaper computers came the development of commercial products displaying intelligent behavior at a price inexpensive enough to market them as toys and, thus, to use in a classroom setting. One of the first assignments in the introductory AI course is to take an intelligent agent toy home for a week. Example intelligent agents used have included Tomadachi/Pocket Pet, Furby, the Sims computer game, Meow-chi/Poo-Chi, RoboBaby,

and Dimrock Don. The students spend a week exploring their assigned intelligent agent, analyzing its behavior, and then answering a number of questions about it, ultimately classifying the type of agent it represents (e.g. simple reflex, model-based reflex, goal-based, utility-based, or learning). To report on their work to the class, students generate a presentation on their intelligent agent that involves a physical description, the environment in which it was designed to operate, whether or not the agent progresses through different stages, and if/how randomness was used in its behavior. Students also must provide a targeted demonstration of their intelligent agent's capabilities and a diagram that illustrates a behavior map that elucidates the decision making behind the selected observed behaviors and reflects what category of agent the toy represents. The final piece of the assignment is to discuss whether or not they think that their agent displays intelligent behavior, and why or why not.

### Philosophical Assignment – Exploring Intelligence through Sphexishness

Countless articles, and even books, have been written over the years addressing what intelligence is and how it relates to the possibility of machine intelligence. To address the question, students are asked to explore the ideas put forth in three seminal articles on the subject: Douglas Hofstadter's "On the Seeming Paradox of Mechanizing Creativity" (Hofstadter, 1985), Alan Turing's "Computing Machinery and Intelligence" (Turing, 1950), and Herbert Dreyfus' *What Computers Can't Do* (Dreyfus, 1972). Hofstadter's article, in particular, provides an interesting and insightful take on what intelligence is and how to identify it, coining the term "Sphexishness" to describe what appears, at least at first glance, to be intelligent behavior. In the assignment students are asked to explore Hofstadter's use of the Sphex wasp and the concept of loops and loop identification, and apply this idea to their intelligent agent analysis. In addition, they are asked to address several points made by each of Turing and Dreyfus from Hofstadter's perspective of what intelligence is. Finally, the students are asked to summarize the various arguments for and against the ability to generate an intelligent machine, and discuss where they stand on the question.

### Ethical Assignment – Exploring the Three Laws of Robotics in the Movie *I, Robot*

Numerous movies made over the years have addressed the potential ethical implications to humanity of having non-human intelligent entities living and working alongside humans. Examples of such movies include *The Bicentennial Man* (1999), *A.I. Artificial Intelligence* (2001), and *I, Robot* (2004). An assignment that has students explore the ethical implications of artificial intelligence centers on the movie *I, Robot* because the theme revolves around the classic three

laws of robotics (originally introduced in Asimov's short story "Runaround" in 1942), and what can happen when these laws are applied strictly with logic, without consideration of standard human sensitivities. The assignment asks students to synthesize and compare ideas and themes found commonly in several of the activities in which they have participated during the semester, comparing and contrasting VIKI from *I, Robot* and HAL-9000 from *2001 – A Space Odyssey*, vs. Sonny from *I, Robot* and Number 5 from *Short Circuit*. The final question that the student must address is the role that pure logic vs. emotion plays in ethical behavior, a key theme in the success or failure of an intelligent machine to truly make decisions that are acceptable from the human perspective.

### Conclusions

The potential for machine intelligence has captured the imagination of humans since the creation of the first computer. The theme has shown up in countless books and movies that explore its promise as well as its drawbacks. It causes us to explore what it really is to be human as well as what could happen if we should succeed in this endeavor. Due to the decrease in cost and increase in power of modern computational devices, it is possible to purchase toys that represent basic intelligent behaviors that can be readily explored and evaluated relative to the concept of an intelligent agent. This rich and extensive set of resources allows for the development of interesting and meaningful assignments for students in an introductory computer science AI course that effectively address this complex field from the technical, philosophical, and ethical viewpoints.

### References

- Asimov, I. 1950. Runaround. In *I, Robot*, 40. New York, NY: Doubleday.
- Dreyfus, H. 1972. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York, NY: Harper & Row.
- Hofstadter, D. 1985. On the Seeming Paradox of Mechanizing Creativity. In *Metamagical Themas: Questing for the Essence of Mind and Pattern*, 526-546. New York, NY: Bantam.
- I, Robot*. 2004. Proyas, A. (director). 20<sup>th</sup> Century Fox.
- Nilsson, N. 1998. *Artificial Intelligence: A New Synthesis*. San Francisco, CA: Morgan Kaufmann.
- Russell, S. and Norvig, P. 2010. *Artificial Intelligence A Modern Approach*. Third Edition. Upper Saddle River, New Jersey: Pearson Education.
- Turing, A. 1950. Computing Machinery and Intelligence. *Mind* 236: 433-460.