Ethical guidelines for AI in education: Starting a conversation

Article *in* International Journal of Artificial Intelligence in Education · January 2000

CITATIONS
40

READS
6,886

2 authors:

Robert M. Aiken
Temple University
124 PUBLICATIONS **662** CITATIONS
SEE PROFILE

Richard G. Epstein
West Chester University
39 PUBLICATIONS **199** CITATIONS
SEE PROFILE

International Journal of Artificial Intelligence in Education (2000), 11,163-176

Ethical Guidelines for AI in Education: Starting a Conversation

Robert M. Aiken *CIS Department Temple University Philadelphia, PA 19122, email: aiken@joda.cis.temple.edu*

Richard G. Epstein *Department of Computer Science West Chester University of PA West Chester, PA 19383 email: epstein@wcupa.edu*

Dedication to Martial Vivet

The first author of this paper had the good fortune to interact with Martial Vivet over many years; in particular with Martial and his colleagues and students in Le Mans for five weeks during the Spring of 1999. His passing has left us without an important voice in the AI & Education community. In addition to his many professional contributions he was an inspiration to many students and colleagues. His warm personality, his adherence to rigorous scientific standards and his concern for the people with whom he interacted will always be a beacon for us to follow. He was concerned about ethics and the impact, both for good and potential harm, that AI research could have on education. It is in his memory and with his concern for students that we would like to dedicate this paper.

Abstract: This paper explores the human and ethical issues implicit in the use of AI in education. Our intention is to begin a discussion that will lead to a deeper understanding of the issues and eventually to a consensus within the research community concerning what is desirable and what is not in the use of AI in education.

INTRODUCTION

It is interesting to speculate where research in AI might lead us in the next ten years even as some people are predicting that we will be fortunate to survive the year 2000 and all of the associated Y2K problems. We prefer to take an optimistic view - at least with respect to our ability to cope with and solve the technical problems we might experience as we move into the new millennium. Indeed, there are reasons to believe that the twenty-first century will present breathtaking advances in science and technology, transforming the nature of human life as we know it. Authors such as Moravec (1996) and Kurzweil (1999) predict exciting developments in many spheres of human endeavor, although it can be argued that some of their views are not well-grounded philosophically. (We are referring, specifically, to the idea that human beings will be able to upload their consciousness into a computer medium, achieving a form of immortality.)

It is clear, however, that computer-based education will be a fact of life in the twenty-first century. Artificial intelligence will play a significant role in this emerging technology. For example, in his predictions concerning life in 2010, Andy Hines of the World Future Society states:

> "The teacher of 2010 will rarely spend a day lecturing, but will be primarily a facilitator and coach. ... The teacher will coach students through video lectures, educational television programs, and artificial intelligence-based programs. Only

> occasionally will teachers instruct classes themselves. Instead, they will be freed up to deliver the personalized instruction critical to educational achievement.
>
> "The artificial-intelligence tutor will become a valuable assistant, providing the individualized instruction that a teacher with 20 or more pupils does not have the time for. Learning can take place at the student's pace." (Hines, 1996, pp. 9-10).

In her predictions concerning future careers, Barbara Moses sees computer-based education as one of the most promising areas for career development in the coming decades (Moses, 1999). This applies both to traditional education and also to what she calls "edutainment", the integration of educational and entertainment technologies.

Just as we are optimistic about Y2K, so we are somewhat concerned about the introduction of AI technology into the classroom. While there is good reason for optimism about the technological dimension of AI in education, care must be taken that the introduction of AI into the classroom is not driven by the technology as much as by genuine human need. This paper will examine some of these issues that we believe are critical to the future development and deployment of such systems. We want to begin a conversation about the ethical principles that can and should guide the development of AI systems for education. What are the bedrock ethical concerns? What makes for a good educational technology, in terms of its social effect, and the kind of student that it produces? What is the risk of harm that this technology represents and how can harm be avoided? What new potentialities will this technology open, and how can they best be exploited?

One tool that can help us explore this subject is the use of stories about the future of technology. Stories can help us build hypothetical scenarios to explore and evaluate the possible social impacts of computer technologies. They constitute a starting point for a discussion of fundamental principles. The role of stories in developing ethical principles is discussed in Artz (1998). In conjunction with this paper the second author is creating a web site (http://www.cs.wcupa.edu/~epstein/AIStories.html) that contains nearly forty stories about AI. While many of these do not relate to AI in education specifically, all of them raise issues about the social implications of AI, wherever that technology is applied. The intent of this site is to provide access to the stories and to create a resource for persons interested in this topic. An introduction to this web site can be found at the end of this paper.

Another theme in this paper is that we are blessed with five senses and we must ensure that technology enhances and does not diminish any of them. Thus, a number of disparate issues are discussed that tie directly and indirectly back to this theme. We hope this paper will initiate a discussion and be more than a compendium of our concerns.

WHY DO WE NEED PRINCIPLES FOR THE USE OF AIED SYSTEMS?

We are at a turning point. Unless we seriously discuss our philosophical premises before AI moves in any significant way into the classroom, we will limit the scope, effectiveness, and positive contributions that AI can bring to learning. Computer-based education, including AI technology, has the potential to harm young people in various ways, including ethically, aesthetically, physically, psychologically, intellectually and socially.

Consider the manner in which computer technology can provide a means for unethical behavior. For example, there is a problem with students using information obtained from the web without giving proper credit to the original authors. This has been called plagiarism. In the "pre-web" culture plagiarism was considered a major intellectual sin. In the current culture many students no longer even think it is cheating if they do not provide appropriate attribution for what they find on the web. Other students undoubtedly realize that they are cheating, that they are getting away with the theft of information, and that they will probably not be caught. In other words, these students are not writing their term papers or essays in good faith. Unless attribution is absolutely demanded, we will develop a breed of students who think plagiarism is normal.

With the growth of AI technology, the ethical problems of intellectual property and honesty become much more subtle and complicated. One of the stories on our web is entitled "The New York Times Book Reviewer." This describes an AI system (circa 2028, when all of our stories take place) that is being marketed by the New York Times. This system will create a book review for any desired book in the style of the New York Times Book Review. So, the student who has access to this technology would not just steal information from the web. The student would cause an intelligent agent (the New York Times Book Reviewer) to generate a book report that the student could then hand in. From an ethical standpoint, this is slightly different from blatantly stealing information, but how different is it? Certainly, it would not be ethical for the student to claim that she wrote the book review. The availability of intelligent agents that create stories and other intellectual property will present a major ethical problem for educators in the coming century.

A student might be hurt aesthetically if her sense of beauty or sensitivity is harmed by the use of the technology or if her creativity is stunted. A common theme in the stories on our web relates to the fact that intelligent systems that surpass human capabilities might result in human beings becoming intellectually lazy. One such story describes an intelligent system that composes music in the style of Tchaikovsky. What if the existence of such a system were to convince an aspiring human composer that musical composition is no longer a viable career and that mimicking the style of Tchaikovsky is truly creative?

There is already considerable concern that computers in education are harming students physically, causing repetitive strain injuries, eye problems, obesity, and so forth (Gross, 1999). If computers become ubiquitous in the classroom, if students spend many hours in front of a computer screen each day, the physical harm is likely to be considerable. Problems with posture, repeated motion stress injuries, and other related physical ailments are directly linked to how people use a computer. A friend of one of the authors is a speech therapist, and she reports a dramatic increase in people with vocal chord damage due to the use of voice recognition systems. These people were accommodating to this new kind of computer interface by speaking in a monotone, thus straining their vocal chords - a new kind of repetitive strain injury.

Moreover, as we mentioned in the introduction we feel that more attention needs to be paid to the impact of technology on our other senses. How does technology impact hearing, touching and smelling? Are we missing opportunities for experiencing simple pleasures in life because we are relying too heavily on technology, i.e., taking care of a simulated pet rather than a live one?

Computer technology can harm a student intellectually in various ways. For example, an intelligent system might induce in a student intellectual laziness simply by showing itself to be far superior to the student in its problem-solving skills. We already observe that many people can not do simple mathematical operations in their head. Without a calculator they do not know how to perform simple calculations. Furthermore, intelligent systems will almost certainly embody a certain kind of limited or narrow intelligence that would be incapable of dealing with certain kinds of creativity that a student might exhibit, thus discouraging the student's own development.

Perhaps the greatest danger that computer technology poses for the student of the future is the social damage that comes from limiting the range of interaction with other human beings. There is clearly a profound danger for society if computers are introduced into the classroom in such a way as to discourage meaningful human interactions. Edward Cornish, President of the World Future Society, discusses this in his predictions for the year 2025 (Cornish, 1999). Speaking about people in general and not just about students, he makes the following predictions:

> "The new infomedia may make people increasingly egocentric and selfish. Since the infomedia cannot be controlled by any nation, religion, or community, individual consumers will dominate in shaping its content and character. ... Consumers will thus become more narcissistic - infatuated with themselves rather than caring for things beyond themselves. As televsion and other electronic entertainments absorb

> more and more time, people will feel ever less motivated to do things for anyone but themselves....
>
> "People may lose much of their ability to think rationally and to make wise decisions. ... The proliferation of information sources - more TV channels, specialized news services, and databanks - has overwhelmed people's ability to focus on particular issues and think logically about them. ...
>
> "Interpersonal relationships will likely be increasingly unstable. The rapid changes and heightened mobility encouraged by infotech will tend to break up human groups, not only in the worksite but in the family and community. Job shifts will separate colleagues and even family members." (Cornish, 1999, pp. 12-13)

These effects of information technology, which Cornish is applying to society as a whole, also apply to the use of technology in education. We need to be careful to protect our social fabric and our sense of community as well as to be careful not to diminish the variety of human endeavor. And at all cost we must preserve the human capacity to solve problems and think rationally.

The social costs of poor educational technology could be high. Negative impacts must be prevented. This can be done by the careful development of and adherence to fundamental principles when developing educational software. An open discussion of principles to be used among those who will be responsible for the creation and the eventual deployment of this software is urgently needed.

RESOURCES FOR DEVELOPING FUNDAMENTAL PRINCIPLES

A set of fundamental principles concerning the development and use of AIED systems requires a philosophical underpinning. Some premises or assumptions upon which our principles are based are discussed below. After presenting our philosophical premises, we shall present our list of fundamental principles.

Professional societies, such as the ACM, have codes of ethics that can be applied to the development of AI systems for education (Anderson, 1993). The following "general moral imperatives" from the ACM Code of Ethics, can all be applied to the development of computing systems in general and to AI systems for education in particular.

General Moral Imperatives (Excerpts from the ACM Code of Ethics)

1.1 Contribute to society and human well-being

1.2 Avoid harm to others

1.3 Be honest and trustworthy

1.4 Be fair to take action not to discriminate

1.5 Honor property rights including copyrights and patents

1.6 Give proper credit for intellectual property

1.7 Respect the privacy of others

1.8 Honor confidentiality

(Anderson et al., 1993, p. 101)

In addition, the "more specific professional responsibilities" listed in the ACM code are also relevant for developers of AI systems for education:

More Specific Professional Responsibilities (excerpted from ACM Code of Ethics)

2.1 Strive to achieve the highest quality, effectiveness and dignity in both the process and products of professional work.

2.2 Acquire and maintain professional competence.

2.3 Know and respect existing laws pertaining to professional work

2.4 Accept and provide appropriate professional review.

2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.

2.6 Honor contracts, agreements, and assigned responsibilities.

2.7 Improve public understanding of computing and its consequences

2.8 Access computing and communication resources when authorized to do so

A related approach to ethics, specifically oriented towards software engineering (and thus, applicable to the development of AI systems for education) is given in Gotterbarn, Miller, and Rogerson (1997).

Collins et al. (1994) presented an interesting approach to the issue of deciding whether it is appropriate to release a particular software system. Their analysis, based upon Rawlsian principles (Rawls, 1989), requires that we assess the obligations among the various parties involved, including the vendor, the client, the users, and the penumbra. Each party has specific obligations to the other parties. An important principle in such an analysis is to protect the least advantaged and the most vulnerable, those who might be negatively impacted by a poorly designed system.

In the case of an AI system in education, the vendor would be the company or institution that develops the AI software. The client would (usually) be the school or university or institution that buys the software. The users are the students. (In the future, clients and users for educational software might merge, as adults manage their own lifelong learning). The penumbra includes all people that are affected by the introduction of the new technology. In the case of educational software, the penumbra could ultimately encompass society at large. Thus, Collins et al. would have us ask, of an AIED system, whether the vendors have fully understood and complied with their obligations to the clients, users, and to the penumbra. They would apply this same kind of analysis to the other parties. For example, the penumbra would have the obligation to protect itself against harm and to influence the decisions of vendors so that vendors would not have the economic incentives to create harmful AIED software.

Roger Clarke (Clarke, 1993; Clarke, 1994) wrote two articles for IEEE Computer that analyzed and then modified Asimov's Laws of robotics to information technology. Clarke's discussion is also relevant to the present task, which is to create a framework for the safe and benevolent application of AI technology in education. Asimov's three laws for robots were first published in 1940 and appeared in several of his collections of short stories (Asimov, 1968; Asimov 1983).

Asimov's Laws of Robotics (Asimov, 1940)

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

We think that it is safe to say that since the publication of Clarke's discussion (beginning in December of 1993), progress has been made in artificial intelligence. (For example, see the

historical survey in the opening chapters of Moravec, 1996). Thus, as we enter a new century and a new millennium, Asimov's laws seem ever more relevant.

Clarke had the foresight to recognize this and to attempt to apply this line of thinking to information technology more generally. In effect, Clarke was trying to state ethical principles for information systems as ethical agents. This line of thinking will certainly become more and more relevant as AI progresses, and it certainly applies to the ethical analysis of AI systems for education. Clarke developed an extended set of laws and discussed the implications of these laws for future robotics technology and information technologies more generally:

An Extended set of the Laws of Robotics (Clarke, 1994)

- **The Meta-Law**: A robot may not act unless its actions are subject to the Law of Robotics
- **Law Zero**: A robot may not injure humanity, or, through inaction, allow humanity to come to harm.
- **Law One**: A robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate a higher order law.
- **Law Two**: (a) A robot must obey orders given it by human beings, except where such orders would conflict with a higher order law. (b) A robot must obey orders given it by superordinate robots, except where such orders would conflict with a higher order law.
- **Law Three**: (a) A robot must protect the existence of a superordinate robot as long as such protection does not conflict with a higher order law. (b) A robot must protect its own existence as long as such protection does not conflict with a higher order law.
- **Law Four**: A robot must perform the duties for which it has been programmed, except where that would conflict with a higher order law.
- **The Procreation Law**: A robot may not take any part in the design, manufacture, or maintenance of a robot unless the new or modified robot's actions are subject to the Laws of Robotics.

At the end of his second paper, Clarke states a position that is very close to the position that we are taking with respect to the need for ethical principles to guide the development of AIED systems:

> "The issues raised in this article suggest that existing codes of ethics need to be reexamined in the light of developing technology. Codes generally fail to reflect the potential effects of computer-enhanced machines and the inadequacy of existing managerial, institutional, and legal processes for coping with inherent risks. Information technology professionals need to stimulate and inform debate on the issues. Along with robotics, many other technologies deserve consideration. Such an endeavor would mean reassessing professionalism in the light of fundamental works on ethical aspects of technology." (Clarke, 1994, p. 65)

As soon as we discuss intelligent systems as ethical agents, we might want to consider fundamental ethical frameworks for human beings. Edgar (1997) has a good discussion on philosophical approaches to ethics in the introductory chapters of her book. Johnson and Nissenbaum (1995) and Baase (1997) are also highly respected authors in this arena (i.e., computer ethics). Tavani (1999) presents a comprehensive review of textbooks in computer ethics. For example, Edgar discusses Kant's categorical imperative. One method that an ethical agent might use in assessing a proposed course of action is to ask what would happen if everyone behaved in this manner. Implicit in Clarke's critique of codes of ethics is the fact that a developer of an AIED system might pass a Kantian critique even as she releases a harmful technology into the environment. This would be the case if that developer bases her analysis on

professional codes of ethics without giving sufficient attention to the social and human impact of technology. The relevant question is no longer "What if everyone were to produce AIED software the way that I do?", but "What if all AIED systems behave like my AIED system?" Thus, as Clarke suggests, and as we are suggesting, the ethical principles need to be applied to the systems themselves, *as if* they were ethical agents.

James Moor of Dartmouth is a lucid and provocative author in the field of computer ethics (Moor, 1997; Moor, 1998a; Moor, 1998b). In his paper, "If Aristotle were a Computing Professional" (Moor, 1998b) Moor presents some compelling arguments for the venture that we are attempting: that is, the attempt to formulate fundamental ethical principles for AIED. He compares the manner in which computers are encroaching into ever more aspects of human life to urban sprawl, calling it "computer sprawl". He notes that there is often an ethics gap in which we find our ethical principles and understanding lagging behind the latest technological advance:

> "Computer sprawl is worldwide and culturally transforming. Computer sprawl is not necessarily rational or harmless, but it is an undeniable force in the world that will affect not only the lives of all of us in technological societies but quite possibly everyone on the planet and their descendants for centuries to come. The ethics gap that is generated because we massively computerize without taking time to consider the ethical ramifications is therefore quite wide and deep." (Moor, 1998b, p. 14)

Appealing to Aristotle, Moor contends that ethics must be grounded in virtue. Virtue itself leads to happiness. This then gives us a fundamental philosophical underpinning for discussing ethics, rooted in an understanding of human virtue.

> "A courageous software programmer is not one who acts rashly and puts herself and others at risk by not thoroughly testing her program but neither is she one who never releases software until she can prove with absolute certainty that it contains no problems. The virtuous programmer is one who balances the risks and benefits properly. ... It is the ability to find the mean, find the balance point, that is the mark of practical wisdom in a person for Aristotle." (Moor, 1998b, p. 16)

Consequently, according to Moor, our discussion of ethical principles for AIED systems cannot ignore the character traits of the actual developers of these systems. But, according to Aristotle, the development of positive personality traits is a matter of developing the correct ethical habits.

In "Reason, Relativity, and Responsibility in Computer Ethics" Moor emphasizes the need to base computer ethics on universal core values (Moor, 1998a). Moor lists these core values in a paper on privacy (Moor, 1997). These are life, happiness, freedom, knowledge, ability, resources and security. Moor then shows how a theory of privacy can be developed around the core value of security. This is consistent with the approach that we are taking in this paper. However, instead of discussing "core values" we will look for basic dimensions of human beings (e.g., the ethical dimension or the physical dimension) and we will demand that AIED systems not damage human beings along any of these fundamental dimensions.

Another approach to discussing computer ethics was presented in Shneiderman's keynote address at the ACM CQL'90 Conference (reprinted in Shneiderman, 1999). Clearly, many of the fundamental principles of user interface design are relevant to the design of AIED systems. Shneiderman attempts to create a philosophical foundation for analyzing information technologies by starting with fundamental goals, such as world peace, freedom of expression, privacy protection, and so forth. In his address, Shneiderman proposed a "Declaration of Responsibility" that would include the following statements:

> "1) We, the researchers, designers, managers, implementers, testers, and trainers of user interfaces and information systems, recognize the powerful influence of our science and technology. Therefore, we commit ourselves to studying ways to enable

> users to accomplish their personal and organizational goals while pursuing higher societal goals and serving human needs.
>
> 2) We agree to preparing a Social Impact Statement (patterned on the Environmental Impact Statement) at the start of every human-computer interaction project. The Social Impact Statement will identify user communities, establish training requirements, specify potential negative side-effects (health, safety, privacy, financial, etc.), and indicate monitoring procedures for the project's lifetime." (Shneiderman,1999, p. 6)

In his address, Shneiderman lists ten questions for designers. These questions are relevant to this discussion. Here is a sampling of the questions that Shneiderman proposes as a "useful checklist for designers":

> "2) *Alienation*: Can we build user interfaces that encourage constructive human social interaction?
>
> "4) *Impotence of the individual*: While large complex systems may overwhelm individual initiative, it seems clear that computers have the potential of dramatically empowering individuals. How best to ensure that this happens?
>
> "9) *Lack of professional responsibility*: Complex and confusing systems enable users and designers to blame the machine, but with improved designs responsibility and credit will be properly given and accepted by the users and designers.
>
> "10) *Deteriorating image of ourselves*: Rather than be impressed by smart machines, accept the misguided pursuit of the Turing test, or focus on computational skills in people, I believe that designs that empower users will increase their appreciation of the richness and diversity of unique human abilities." (Shneiderman, 1999, p. 8)

The fundamental question that underpins Shneiderman's checklist is presented in the form of a quote from Lewis Mumford:

> "The real question before us lies here: do these instruments further life and enhance its values, or not?" (Shneiderman, 1999, p. 8; this quote is from Mumford, 1934).

This leads us to two meta-principles that will ground our discussion of our list of principles for the design of AIED systems.

TWO META-PRINCIPLES

The fundamental meta-principles that we propose as a basic philosophical underpinning for any discussion of AIED systems are the following:

- **The Negative Meta-Principle for AIED** AIED technology should not diminish the student along any of the fundamental dimensions of human being.
- **The Positive Meta-Principle for AIED** AIED technology should augment the student along at least one of the fundamental dimensions of human being.

This is the basic philosophical approach that was taken in the stories "Is Your Computer Stealing From You?" in Epstein (1997a) and "The Great Brain Robbery" in Epstein (1997b). These stories take the form of interviews with Professor Lowe-Tignoff who is concerned that computer technology might diminish human beings. The second of these stories is available on the web site.

Another way of expressing the two meta-principles that we have presented is the following:

The Golden Rule for Computers in Education:

Teach others as you would like to be taught.

This Golden Rule for the use of computers in education (Aiken and Aditya, 1997; Aiken, 1989) is closely related to our negative and positive meta-principles. We would like to be taught in such a manner that our personality is expanded and augmented. Certainly none of us would want to be diminished aesthetically, ethically, or physically by a computer-based educational system.

The bottom line is that we do not want the new technologies to damage the students in any way. Yet, the dangers with any new technology are great. For example, Burke and Ornstein (1997) offer a fascinating view of the unexpected impacts of new technologies. Particular caution needs to be exercised when applying new technologies to young children, who are especially vulnerable because their brains are still developing. A significant cautionary tale that we might consider is currently being played out in our schools and libraries. A preliminary version of this tale is told in Kimberly Young's book, Caught in the Net (1998). Dr. Young, a psychologist at the University of Pittsburgh, is documenting many of the negative impacts of internet addiction on young people. Certainly, the Internet was not developed with much concern for the impact it would have on young people. However, AIED technology needs to be fundamentally concerned with this. The actual situation when we introduce many new technologies into the classroom, as some futurists are predicting for the next decade, may become quite complicated, and we need to carefully track the negative and positive impacts of individual technologies and the subtle interactions between technologies.

The two meta-principles refer to fundamental dimensions of human being. We propose that these are the following:

Fundamental Dimensions of Human Being

1. **Ethical:** actions and behaviors insofar as they might have an impact upon other human beings, creatures, and the environment. This dimension relates to an understanding of basic ethical principles and a willingness to act in accordance with that understanding.
2. **Aesthetic:** having an appreciation of beauty in all of its manifestations. This includes beauty in nature, the arts, mathematics, science and technology.
3. **Social:** an individual's concept of self and his/her relation to others. This dimension has to do with the values of community, family, and friendship.
4. **Intellectual:** the human intellect and its manifest and manifold powers. These include the ability to understand existing knowledge and to create new knowledge.
5. **Physical:** basic physical health, including all aspects of physical well-being, including exercise and avoiding harmful substances and habits.
6. **Psychological:** individual's ability to lead a happy and fulfilling life. This dimension is also related to the social, intellectual, aesthetic and ethical dimensions

PRINCIPLES FOR AIED SYSTEMS

The following ten principles are derived from the two meta-principles presented in a previous section. Our goal is not to preach, rather, we hope that we can raise the awareness of researchers in designing educational systems.

1. Design systems that encourage and do not demoralize the user.

It has been well understood that AIED systems need to be responsive and adaptive to individual learning styles. They have improved immeasurably over CAI programs that were often very condescending in their response to students. Moreover, AIED researchers have also realized that the systems do not need to understand and assess everything that a student does. In most cases it is best for the student or the teacher to assess the student's performance.

2. Encourage collaborative learning and the building of healthy human interactions.

The proliferation of technology-driven long distance learning environments has raised numerous possibilities for expanding AIED-based collaboration research. In fact, six of the twenty-one sessions at the recent AIED Biennial Conference were devoted to aspects of collaboration (Lajoie and Vivet, 1999). The key here is to consider the human aspects of collaboration and not to simply focus on system components.

3. Support the development of positive character traits.

By positive character traits, we especially mean ethical behavior. We want to help students learn to be considerate of others, to be helpful and creative, and to thrive in the workplace of the future. Some of these character traits can be strengthened in multi-user domains, or in multi-player games based upon virtual reality.

4. Avoid information overload.

Systems should provide students with "bite-size" pieces of information that they can assimilate and understand. Shenk (1997) writes convincingly that we have a problem with information overload in our society, generally. Cornish (1996) suggests that deciding which information is important for students will be the significant challenge for educators in the coming decades.

A student who confuses masses of information with knowledge has been damaged both intellectually and aesthetically. This issue is discussed in the story "Toxic Knowledge" (Epstein, 1998), which is also available on our website. Information overload, if it is associated with sitting long hours at a computer screen, can also lead to health problems. In his predictions concerning the year 2025, Edward Cornish states:

> "Infotech is encouraging a physically inactive lifestyle that endangers people's health. The number of seriously overweight children and adolescents in the United States has more than doubled over the past three decades. the National Center for Health Statistics reported in 1995. ... Research by William Dietz of Tufts University and others points the finger at physical inactivity, induced largely by TV, video games, and PCs, plus too much munching on high-calorie foods. Public health officials fear that today's overweight children will be tomorrow's overweight adults, at risk for premature heart attacks, strokes, and diabetes. If the obesity trend continues, college-age youths may start needing heart transplants and bypass operations." (Cornish, 1999, p. 14)

5. Build environments that promote inquisitiveness and curiosity and that encourage students to learn and explore.

Students should be able to discover new interests and talents within themselves. Technology is opening up incredible new opportunities. For example, interactive video brings students to situations previously unimaginable (Smith and Reiser, 1997). A good overview of some of the current and exciting work with computational tools, based on hand-held devices is reported in Soloway et al. (1999). In the future, virtual reality is likely to prove a significant tool for

students who want a computer created experience. Yet it must be balanced by the realization that it can also limit their imagination.

6. Consider ergonomic features to avoid injuries such as eyestrain, repetitive strain injuries, back problems, etc.

Young students are especially vulnerable to harm due to poorly designed computer systems. Because of the intrinsic limits of technology as it exists today, one must tailor his or her interaction with the computer into a few, constrained, muscular patterns. The vocal chord injuries that are being reported due to the use of voice recognition software is an injury due to new uses of technology. Others may follow. There needs to be a kind of advocacy for student health in this sphere.

7. Develop systems that give teachers new and creative roles that might not have been possible before the use of technology. Systems should not attempt to replace the teacher.

We need to carefully assess the impact of computer technology on the teaching profession. Teaching can be stressful and technology might be able to improve the experience for many teachers who now face severe classroom management and discipline problems. A number of studies emphasize how the role of the teacher changes from a "talking head" to a facilitator. The teacher now has more time to work with students individually or in small groups. Yet while this has happened in some cases and while computers have been successfully deployed in the classroom (Koedinger et al., 1995; Schofield, 1995) we still do not have a good understanding of the relationship between the way teachers teach and what makes interactive computing in the classroom a success.

8. Respect differences in cultural values; avoid "cultural imperialism".

The development of innovative and effective educational software is a time-consuming and expensive undertaking. Most of it is developed in English for a number of reasons. The question whether school systems in other countries utilize this software or design their own will have to be answered. And if they do not elect to develop software in their own language how can, or will, they incorporate culture specific content into existing educational software? Moreover, what is the impact of computer technology on human language? Today there are approximately 6,000 languages in existence. Experts are predicting that more than half of these will disappear by year 2050 (Ostler, 1999). Other experts are predicting that English will become the native language of almost all human beings during the next century. Is this a desirable outcome? (Clearly, not, in the opinion of these authors.) What role can the developers of educational software play in the preservation of linguistic and cultural diversity?

9. Accommodate diversity and acknowledge that students might have different learning styles and skill levels.

Clearly this has been a major goal in many of the education systems that have been developed. But has this goal been met? In our view not successfully because it is a VERY hard problem. Perhaps we should step back and look at this as an expertise of the teacher. Several researchers are studying this problem from different perspectives. (See, for example, du Boulay, Luckin and del Soldato, 1999; Barnard and Sandberg, 1996; Lepper et al., 1993).

The objective of influencing humans for the better (the positive meta-principle) without acknowledging diversity and different learning styles is not possible. Diverse teaching styles are required to stimulate maximum learning and creativity.

10. Avoid glorifying the use of computer systems thereby diminishing the human role and the human potential for learning and growth.

In game two of his match with Deep Blue, Kasparov reports that he experienced profound fear and anxiety when he came to see his opponent as embodying a form of genuine intelligence. It may well be that this chess match is a turning point in the history of AI. As more and more computer systems surpass human levels of performance, how many of us will experience similar attacks of anxiety and fear? How many students will come to see their intelligent computer-based tutors as omnipotent beings beyond human reckoning? Clearly, every effort must be made to avoid any scenario that leads to people being diminished by computers.

Perhaps, in the long run, the greatest threat posed by artificial intelligence is the prospect of increasing the intellectual laziness of human beings. Several stories on our web treat this issue. We need to initiate a discussion about what is important in education and how we should teach it. Moreover, we have to consciously decide what role computers should play in this process and what must remain in the human domain.

In his recent book, Slaves of the Machine, Gregory Rawlins states "All [humans] play [chess] in their own unique style, by doing so, the best of them don't simply play chess, they create art on the chessboard. Today, chess machines don't do that. They play with no heart." (Rawlins, 1998, p. 107) We need to understand the human heart and the manner in which education engages the human heart.

A BRIEF INTRODUCTION TO OUR WEB RESOURCE

The second author intends to create a web resource that will provide access to over thirty short stories about artificial intelligence. Indeed, these stories are already available at www.cs.wcupa.edu/~epstein/may2028.htm, but a new web is being created specifically for the purpose of promoting the agenda behind this article, which is to stimulate discussion about the proper role of AI systems in general and AIED systems in particular. The new web site will be located at http://www.cs.wcupa.edu/~epstein/AIStories.html.

These stories use various formats to present the world of technology circa 2028. They include newspaper stories, television and newspaper interviews, college lectures, a commencement address, book reviews, and infomercials. Some of them explore the social implications of computer technology in great depth. A few of the stories are rather short newspaper accounts of specific technologies.

It is hoped that they become a basis for generating classroom discussions concerning AI technologies and their social implications. In addition, we intend to add the following information:

- Introductory materials describing the purpose of the web and how it might be used.
- An index to the stories based upon the six dimensions implicit in our negative and positive meta-principles.
- An index to the stories based upon the ten principles for AIED that we have enunciated.
- Discussion questions for each story.
- Suggested writing assignments and research papers that relate to the stories.
- Contributions from professors and students who might want to discuss a particular story or aspect of the web.
- Additional links to other web resources that relate to the subject matter of this web.

CONCLUSIONS

There is little doubt that human beings are capable of astonishing creativity in the realm of technology. Ongoing research in artificial intelligence will eventually lead to new applications of computers in the classroom, in schools, in colleges, in universities, and other environments where learning takes place. We are at a critical juncture. We must clearly formulate for ourselves the fundamental goals and principles for the development of technology in the classroom. We need to articulate humane, compassionate, and wise principles for the use of artificial intelligence in education. The potential for harm is too great for us to ignore.

Acknowledgments

The authors would like to thank the editor, John Self, and three anonymous reviewers for comments on an earlier version of this paper.

References

Aiken, R. M. (1989), "The Impact of Artificial Intelligence on Education: Opening New Windows", in *Proceedings of CEPES-UNESCO International Symposium on Artificial Intelligence in Higher Education*, Prague, 1989, Springer-Verlag, pp. 1-14.

Aiken, R. M. and Aditya, J. N. (1997), "The Golden Rule and the Ten Commandments of Teleteaching: Harnessing the Power of Technology in Education", *Education and Information Technologies*, 2(1), pp. 5-15.

Anderson, R. E., Johnson, D. G., Gotterbarn, D., and Perrolle, J. (1993), "Using the New ACM Code of Ethics in Decision Making", *Communications of the ACM*, February 1993, pp. 98-107.

Artz, J. M. (1998) "The Role of Stories in Computer Ethics", *Computers and Society*, March 1998, pp. 11-13.

Asimov, I. (1968) I, Robot, Grafton Books, London.

Asimov, I., Warrick, P. S., and Greenberg, M. H. (1983), *Machines That Think*, Holt, Rinehart, and Wilson, London.

Baase, S. (1997), *The Gift of Fire: Social, Legal, and Ethical Issues in Computing*, Prentice-Hall, Upper Saddle River, New Jersey.

Barnard, Y. F. and Sandberg, J. A. C. (1996) "Self-explanations: Do We Get Them from Our Students?" in P. Brna, A. Paiva, and J. Self, editors, *Euroaied: European Conference on Artificial Intelligence in Education*, pp. 115-121, Edicoes Colibri, Lisbon.

Burke, J. and Ornstein, R. (1997), *The Axemaker's Gift: Technology's Capture and Control of Our Minds and Culture*, G. P. Putnam's Sons, New York.

Clarke, R. (1993) "Asimov's Laws of Robotics: Implications for Information Technology: Part I", *IEEE Computer*, December 1993, pp. 53-61.

Clarke, R. (1994) "Asimov's Laws of Robotics: Implications for Information Technology: Part II", *IEEE Computer*, January 1994, pp. 57-66.

Collins, R. W., Miller, K. W., Spielman, B. J., and Wherry, P. (1994) "How Good is Good Enough?: An Ethical Analysis of Software Construction and Use", *CACM*, 17(1), pp. 81-91.

Cornish, E. (1999) *The Cyberfuture*, The World Future Society, Bethesda, MD.

Du Boulay, B., Luckin, R. and del Soldato, T. (1999), "The Plausibility Problem: Human Teaching Tactics in the 'Hands' of a Machine", *Proceedings of AIED 99 World Conference on Artificial Intelligence in Education*, pp. 225-232, IOS Press, Amsterdam.

Edgar, S. (1997), *Morality and Machines: Perspectives on Computer Ethics*, Jones and Bartlett, Sudbury, Massachusetts.

Epstein, R. (1997a), *The Case of the Killer Robot*, John Wiley and Sons, New York. (Specifically, the story "Is Your Computer Stealing from You?")

Epstein, R. (1997b), "The Great Brain Robbery", *Computers and Society*, December 1997, pp. 35-40.

Epstein, R. (1998), "Toxic Knowledge", *Computers and Society*, June 1998, pp. 86-91.

Gotterbarn, D., Miller, K., and Rogerson, S. (1997), "Software Engineering Code of Ethics", *Communications of the ACM*, 40(11), pp. 110-116.

Gross, J. (1999) "Missing Lesson in Computer Class: Avoiding Injury", *New York Times*, March 15, 1999.

Hines, A. (1996), "Jobs and Infotech: Work in the Information Society", in *Exploring Your Future: Living, Learning, and Working in the Information Age*, pp. 7-11, Edward Cornish (editor), World Future Society, Bethesda, MD.

Johnson, D. and Nissenbaum, H. (1995) *Computers, Ethics and Social Values*, Prentice-Hall, Englewood Cliffs, New Jersey.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. (1995), "Intelligent Tutoring Goes to School in the Big City". In J. Greer (editor), *Proceedings of AIED 95 World Conference on Artificial Intelligence in Education*, pp. 421-428. Charlottesville, VA. Association for the Advancement of Computing in Education.

Kurzweil, R. (1999), *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Viking Penguin, New York.

Lajoie, S., and Vivet, M. (1999) (Editors) *Proceedings of AIED 99 World Conference on Artificial Intelligence in Education*, IOS Press, Amsterdam.

Lepper, M. R., Woolverton, D. L., Mumme, and Gurtner, J. -L. (1993), "Motivational Techniques of Expert Human Tutors: Lessons for the Design of Computer-Based Tutors", in S. Lajoie and S. J. Derry (editors), *Computers as Cognitive Tools*, pp. 750105. Lawrence Erlbaum, Hillsdale, New Jersey.

Moor, J. H. (1997) "Toward a Theory of Privacy in the Information Age", *Computers and Society*, September 1997, pp. 27-32.

Moor, J. H. (1998a), "Reason, Relativity, and Responsibility in Computer Ethics", *Computers and Society*, March 1998, pp. 14-21.

Moor, J. H. (1998b), "If Aristotle were a Computing Professional", *Computers and Society*, September 1998, pp. 13-16.

Moravec, H. (1996), *Robot: Mere Machine to Transcendent Mind*, Oxford University Press, New York.

Moses, B. (1999), "Career Intelligence: The 12 New Rules for Success", *The Futurist*, 33(7), pp. 28-35.

Mumford, L (1934) *Technics and Civilization*, Harcourt Brace and World, Inc., New York.

Ostler, R. (1999), "Disappearing Languages", *The Futurist*, 33(7), August-September 1999, pp. 16-22.

Rawlins, G. J. E., (1998), *Slaves of the Machine*, Cambridge, MA, MIT Press.

Rawls, J. (1989). *A Theory of Justice*, Cambridge, MA, Harvard University Press.

Schofield, J. W. (1995), *Computers and Classroom Culture*, Cambridge University Press, New York.

Shenk, D. (1997), *Data Smog: Surviving the Information Glut*, HarperEdge, San Francisco.

Shneiderman, B. (1999), "Human Values and the Future of Technology: a Declaration of Responsibility", *Computers and Society*, September 1999, pp. 5-9. (This paper was originally given as the keynote address at the CQL Conference in 1990.)

Smith, B. K. and Reiser, B. J. (1997) "What Should a Wildebeest Say? Interactive Nature Films for High School Classrooms" in *ACM Multimedia 97 Proceedings* (pp. 193-201). ACM Press, New York.

Soloway, E., Grant, G., Tinker, R., Roschelle, J., Mills, M., Resnick, M., Berg, R., and Eisenberg, M. (1999), "Log on Education: Science in the Palms of their Hands", *Communications of the ACM*, 42(8), pp. 21-26.

Tavani, H. (1999) "Computer Ethics Textbooks: a Thirty-Year Retrospective", Computers and Society, September 1999, pp. 26-31.

Young, K. (1998), Caught in the Net: How to Recognize the Signs of Internet Addiction - And a Winning Strategy for Recovery, John Wiley and Sons, New York.