SANDIA REPORT
SAND2006-0608
Unlimited Release
Printed May 2006

Ethical Principles and Guidelines for the Development of Cognitive Systems

Wendy L. Shaneyfelt

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy Sandia Corporation.

Ethical Principles and Guidelines for the Development of Cognitive Systems

Wendy L. Shaneyfelt
Cognitive and Exploratory Systems and Simulation
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1188

Abstract

As cognitive systems technologies emerge, so too do the ethical issues surrounding their development and use. To develop cognitive systems technologies responsibly, Sandia National Laboratories is establishing a framework to proactively address both real and potential ethical issues. This report contains the principles and guidelines developers can use to guide them as they are confronted with ethical issues related to developing cognitive systems technologies as they apply to U.S. national security. A process to apply these principles offers a practical way to transfer these principles from paper to a working strategy. Case studies are presented to reflect upon potential scenarios and to consider resolution strategies.

Acknowledgements

Special thanks are given to the following for their insightful comments and suggestions by way of review of this report:

Preface

Cognitive systems technologies can augment the human process to acquire, retain, and apply knowledge at an unprecedented level of speed and effectiveness. Applying cognitive systems technologies, we can increase our proficiency to comprehend and interact with the world, as well as recall experiences to solve future problems. The benefits encompass numerous aspects of our lives including national security, manufacturing, education, and medicine.

Foreseeing the potential positive and negative impacts of the application of cognitive systems is essential to developing these new technologies responsibly. Sandia National Laboratories is committed to addressing the ethical and legal considerations in concert with the development of cognitive systems technologies. The *Ethical Principles and Guidelines for the Development of Cognitive Systems* establishes the foundation for responsibly developing cognitive systems with a focus on ethics.

Contents

Introduction

As new technologies emerge, the introduction of these innovations also brings potential unforeseeable impacts on our lives. The ethical implications of these new technologies require us to carefully consider their potential uses and misuses, and our resulting roles and ethical responsibilities as developers.

Advances in cognitive psychology, neurophysiology, and computer science offer a breakthrough in our ability to model human cognition. Cognitive systems are a new technology that has minimal precedent within the history of science and technology. They are inherently complex, dynamical, and non-linear; therefore, clear-cut answers and simple solutions to complex issues are not apparent nor are they appropriate. While traditional software development has ethical guidelines established, cognitive systems introduce an array of new issues related to individual privacy, security, public awareness, and societal affects, to name a few. Recognizing the potential for unintended consequences as well as abuse by users compels us to address the ethical concerns as developers.

This paper presents principles and guidelines that developers of cognitive systems technologies can use when confronted with related ethical issues now and in the future. Sandia National Laboratories (Sandia) strives to ensure that a mindset on ethical considerations is core to the development of cognitive systems technologies, recognizing a responsibility that technology advancements must work in harmony with society. Development of these ethical principles and guidelines has been a grassroots effort by the technology developers themselves as they contemplate the implications of this capability.

Focusing on practical principles, the guidelines provided are intended to help the developer of cognitive systems formulate ethical decisions supported by professional codes already established. This community of developers will be responsible for refining and reshaping these principles as the technological development progresses by engaging in conversations and exchanges of ideas. Addressing the ethical dimension of this work is expected to be an ongoing, iterative process as the technology expands.

This report is organized by first presenting a specific definition for Cognitive Systems, followed by the principles and guidelines for developers. Next, we present a process whereby these principles can be applied to particular development efforts. Finally, case studies are presented for the purpose of allowing reflection upon the practicality of applying the principles. They reveal some of the ethical dilemmas that might be incited throughout the development of cognitive systems technologies. Some of these examples are based on actual events occuring in our society today, and some serve as credible hypothetical situations.

What is a Cognitive System?

In recent years, the term "cognitive systems" has been used to identify a variety of programs, initiatives, and technologies leading to ambiguity with regard to its meaning. Sandia has taken a leadership role in forming a community of researchers who share a common definition:

> *Cognitive systems consists of technologies that utilize as an essential component(s) one or more plausible computational models of human cognitive processes.*

This definition distinguishes cognitive systems from mainstream artificial intelligence, and related fields, in that it establishes human-based plausibility requirements. While many researchers from other fields might draw inspiration from human cognition, plausibility requirements demand that the design details of a computational system be traceable to established scientific understanding of human cognition. Sandia is unique in its approach to building technologies that are applicable to complex, real-world problems that use human plausibility as a core requirement.

The focus of Sandia's cognitive systems work today is on the development of computer models of human cognition that are applied to create unique technology solutions. By creating machines that have cognitive characteristics of humans, we can take advantage of the basic strengths of humans and machines while mitigating the basic weaknesses of each.

The decision to pursue the development of cognitive systems technologies is based on the belief that there are numerous positive impacts cognitive systems technologies could have on our national security. A model describing how a human acquires knowledge through the process of reasoning, intuition, or perception can be customized to reflect an individual's knowledge and disposition toward various topics, tasks, technology, people, etc. Areas of application include:

- Detection, recognition, analysis, and forecasting of human behavior and performance
- Machine representation and application of human knowledge and experience (synthetic subject matter expert)
- System adaptation to the knowledge, skill, situation awareness, or intentions of individual operators or teams of operators
- Preservation and transfer of knowledge and experience
- Aides to human attention, memory, situation awareness, decision-making, and other cognitive functions
- Technologies in which human-machine interaction are vital to the performance, safety, and security of systems
- Training for jobs or tasks in which human interaction under unpredictable and stressful conditions is essential to success

Ethical Principles and Guidelines

This section presents the ethical principles specifically established for the responsible development of cognitive systems. Following the ethical principles are the guidelines which further elaborate on the principles. Due to the unforeseeable risks of this emerging science, these principles and guidelines will be refined and reshaped regularly to accommodate the new risks revealed as cognitive systems mature.

Ethical Principles for the Development of Cognitive Systems

1. Cognitive systems developers will remain knowledgeable about and apply their respective established professional/ethical codes and guidelines as appropriate.

2. Cognitive systems developers will proactively consider the intended uses and impacts of their specific technologies, as well as the potential for accidental use, misuse, and abuse.

3. Cognitive systems developers will provide inherent safety features to the extent considered reasonably possible to maximize the prevention of accidents, misuse, and abuse.

4. Cognitive systems developers will proactively initiate ethical discussions among themselves and support public engagements as these technologies develop.

5. Cognitive systems developers will provide human test subjects with a clear understanding of the personal information acquired and how it will be stored, analyzed, and applied.

6. Cognitive systems developers will responsibly handle any personal information obtained from test subjects.

7. Cognitive systems developers will respect the limitations of a cognitive model as an imperfect representation of a test subject.

Ethical Principles and Guidelines for the Development of Cognitive Systems

1. Cognitive systems developers will remain knowledgeable about and apply their respective established professional/ethical codes and guidelines as appropriate.

Cognitive systems developers belong to a variety of disciplines including engineering, computer science, psychology, sociology, neuroscience, etc. The ethical standards established for individual professions should be applied as appropriate to the development of cognitive systems.

2. Cognitive systems developers will proactively consider the intended uses and impacts of their specific technologies, as well as the potential for accidental use, misuse, and abuse.

While the applications of cognitive systems are typically beyond the control of the developer, those involved with the development should proactively determine the technology's capabilities and limitations. They should then systematically apply them to various scenarios of intended use, accidental use, misuse, and abuse. Potential positive and negative impacts to societal issues such as economic, health, and social or cultural changes should be considered. Diversity training which fosters a multicultural understanding should be included to ensure fairness in the technology development by minimizing personal biases, prejudices, stereotypes, and generalizations.

3. Cognitive systems developers will provide inherent safety features to the extent considered reasonably possible to maximize the prevention of accidents, misuse, and abuse.

Developers will identify areas where safety features are or might be necessary to maximize the prevention of accidental use, misuse, and abuse. The safety features will be included in the design and development to the extent considered possible.

4. Cognitive systems developers will proactively initiate ethical discussions among themselves and support public engagements as these technologies develop.

Developers will resolve ethical controversies openly and expediently. They will recognize the need to extend public awareness and the appreciation of the science via public discussions, community classes, forums, debates, articles, etc. Understanding cognitive systems technologies and the issues involved will enable people to make informed decisions and choices. Developers will clearly and carefully describe both the risks and the potential benefits associated with the technology and its applications with care taken to not over inflate the capabilities.

5. Cognitive systems developers will provide human test subjects with a clear understanding of the personal information acquired and how it will be stored, analyzed, and applied.

As part of the development of cognitive systems, human test subjects might be asked to provide developers access to personal information and behavior. Under the guidance of a Human Test Subjects Board, any human test subject involved in the development of the cognitive systems technologies for test purposes must be informed about the personal information acquired, modeled, and analyzed. This information, as well as completed cognitive models, are subject to review by the test subject and can be corrected or rejected, if desired. The completed cognitive model should be validated to ensure the model does represent the construct for which it was developed. Developers will respect the test subject's limitations and restrictions imposed for selective data gathering.

6. Cognitive systems developers will responsibly handle any personal information obtained from test subjects.

Under the guidance of a Human Studies Board, any information acquired, modeled, analyzed, etc. from a test subject will not be used outside the limits imposed by the developer and test subject without consent. This information is not available for others to view or use without prior consent from the test subject (e.g., demonstrations of applications containing a test subject's personal information from whom the model was developed). The identity of the person from whom the cognitive model was developed will remain confidential, if desired, and not be given to anyone without the test subject's permission. Procedures must be established to destroy data if the test subject changes his/her mind about participating.

7. Cognitive systems developers will respect the limitations of a cognitive model as an imperfect representation of a test subject.

A cognitive model of a test subject is limited to the amount, type, and quality of information collected and used in the model, which is limited in part by the technical capabilities for collecting such information. Additionally, since test subjects will be aware that data is being collected, they might intentionally or unintentionally further limit the completeness of the cognitive model. Therefore, the data collected will not give a complete representation of a person's knowledge and disposition. Because of these significant limitations, developers will resist using the cognitive model to judge mental health, intellect, personality, competence, or natural ability. The cognitive model will not be used to determine any absolutes about the individual.

Process for Applying the Ethical Principles and Guidelines

The principles and guidelines established herein address the developers of cognitive systems technologies. The developers involved comprise an interdisciplinary partnership including different types of scientists, engineers, technologists, program managers, business developers, ethicists, and academic advisors.

We expect the ethical principles and guidelines for the development of cognitive systems technologies to mature and expand as the scientific effort grows. The unforeseen risks of an emerging science demand continued attention to new risks as they transpire. Applying the ethical principles and guidelines using an iterative process-based approach requires refinements to be made as new ethical issues arise from the new technical capabilities that are developed. This recommended disciplined process to apply these principles and guidelines includes the following steps: articulation, approval, accountability, and audit.

A clear *articulation* of the ethical principles and guidelines to all the developers is imperative to establishing a common mindset of how to address ethical issues. Forums for discussing and debating actual and potential ethical issues and their resolutions include meetings, workshops, round-tables, videos, websites, and papers. The goal of these forums is to educate each developer regarding the principles and guidelines and how they can be applied to ethical issues.

Approval of these ethical principles and guidelines by Sandia's upper management, as well as management's active involvement in developing the guidelines, emphasizes the commitment to keep potential soeciatal impacts – negative and positive – in mind. As these technologies advance, we recognize that in addition to the significant positive impacts we expect them to have on our national security and society in general, we will also endeavor to recognize any negative impacts to society and address those issues responsibly. With the support of upper management and their approval of these principles, ethical issues encountered can be addressed with the same mindset at the development and management level.

Developers and managers are *accountable* for the effective implementation of the ethics policies when resolving ethical issues. How thoroughly is an issue considered? What "grey" areas are involved? Risk-benefit tradeoff studies can be performed, relying on the direction provided by the ethical principles and guidelines. The resolutions can be traced to a structured decision-making process and include such information as the state-of-the-art of the technologies, political factors, environmental factors, resources, current events, need for the technology, etc.

To sustain high standards of the ethical development of cognitive systems, independent *audits* can be performed for such purposes as the following:
- to examine, evaluate, and report on the effectiveness of the ethical resolutions
- to reveal ethical irregularities
- to evaluate the effectiveness of the established principles and guidelines and the extent to which they are adhered to by the development team

The audit assessment will recommend ideas to improve upon the effectiveness of the ethics policy for the developers, providing the information required to make the appropriate changes.

Case Studies

The case studies presented here are hypothetical examples based on real concerns in the cognitive systems community. These case studies serve to stimulate discussions concerning potential resolutions for potential ethical issues.

Case Study #1
A large automotive company is seeking ideas for ways to reduce the deaths and injuries due to automobile accidents caused by driver inattention. Researchers believe they can provide an effective and cost-efficient solution quickly using cognitive systems technologies. While the potential benefit appears obvious to the researchers, some of the team members identify a situation where the proposed technology could be abused if it fell into the hands of someone other than the intended customer. Should the researchers develop the capability?

Case Study #2
A customer is interested in using a cognitive systems technology application to collect and analyze data. The application has the potential to reduce the customer's collection and analysis time from years to weeks. However, the data collected could be of interest to other analysts with intentions different from this customer. Knowing the possibility exists for the data collected to be exploited by others, what can the developers do to circumvent the potential misuse or abuse of the data?

Case Study #3
A project is underway that involves developing a cognitive model of a human test subject. The intent of this project is to capture the test subject's expertise in a particular area. The Human Subject Board has approved this project and the test subject selected cooperates fully. Once the cognitive model is built, a review of it reveals the test subject's expertise, but it also subtly reveals some areas of weakness. The test subject was aware of these weaknesses himself, but is not interested in having others learn of it when they use the cognitive model. He is concerned how it might affect his job performance review and even his professional reputation. What responsibilites do the developers have to protect the human test subject?

Case Study #4
A project is underway to develop a cognitive systems technology that can be used to further an individual's predictive capabilities in order to improve decision-making skills. Some project team members believe this is a jump too far in the future and are concerned about the imbalance of power distribution this could cause in our society. They want nothing to do with developing such a capability and want off the project. Management needs these key team

members in order to successfully develop this technology. What options do these employees have? What options does the employer have? What structure could be put in place to mitigate this situation in the future?

<u>Case Study #5</u>
The medical community has made life better for thousands of people with devices such as pacemakers and cochlear implants. These implants are acceptable to society as a whole as it provides a functionality that is considered normal to the average person. Consider a breakthrough in the medical community that could restore memory functions to a person who has suffered a stroke. Such a development would also likely be positively embraced by society. What if that same technology could be implanted in a working memory to give someone super-human memory capabilities? Is there a benefit to our society in providing such a technology to super-enhance our cognitive capabilities?

Recommendations

The principles presented within this paper are intended to be used to guide the developers of cognitive systems technologies in resolving ethical issues introduced by the applications of cognitive sciences. However, stating the principles is only the first step. The process of applying the principles must be recognized by management for them to take hold. With Sandia management's approval of the principles and guidelines stated herein, clear articulation to the development team can be made with assurance that management concurs with these responsible development practices. A structure accessible to the cognitive development team must then be formed to provide a forum to address and resolve ethical issues.

As these principles and guidelines are applied to actual and potential ethical scenarios, it is recommended this SAND Report be amended to effectively address relevant updates and modifications from lessons learned. As the process of applying these principles is developed, improvements and details should be added to reflect a course of action that is reasonable to administer and act upon. Oversight or independent review committees are also strongly recommended to improve upon the effectiveness of these principles and their applications.

Bibliography

American Medical Association. Principles of medical ethics. July 06, 2005.
URL:http://www.ama-assn.org/ama/pub/category/2512.html.

American Psychological Association. Ethical Principles of Psychologists and Code of Conduct. June 1, 2003. URL:http://www.apa.org/ethics/code2002.html.

Forsythe, C., Bernard, M., Goldsmith, T. (2006). Cognitive Systems: Human Cognitive Models in Systems Design. Lawrence Erlbaum Associates, Inc.

Greely, Henry T. "Neuroethics: The Neuroscience Revolution, Ethics, and the Law. Markkula Center for Applied Ethics. 2005. Santa Clara University. Date of Access: June 2005.
URL:http://www.scu.edu/ethics/publications/submitted/greely/neuroscience_ethics_law.html.

Illinois Institute of Technology. Center for the Study of Ethics in the Professions, Codes of Ethics Online. 1996. URL:http://www.iit.edu/departments/csep/codes/.

Jacobstein, Neil and Reynolds, Glenn Harlan. "Foresight Guidelines on Molecular Nanotechnology, Version 4.0." October 2004.
URL:http://www.foresight.org/guidelines/current.html.

McLean, Margaret R. "A Framework for Thinking Ethically About Human Biotechnology." Markkula Center for Applied Ethics. 2005. Santa Clara University. Date of Access: June 2005.
URL:http://www.scu.edu/ethics/publications/submitted/mclean/biotechframework.html.

National Academy of Sciences. On Being a Scientist, Responsible Conduct in Research. 1995. National Academy Press, Washington, DC.

National Institutes of Health Office of Human Subjects Research. Regulation and Ethical Guidelines. Date Accessed: December 2005.
URL:http://ohsr.od.nih.gov/guidelines/guidelines.html.

Distribution

1		American Staff Augmentation Providers, LLC Attn: Mary McDaniel Clum 1300 Eubank Blvd. SE, Suite B Albuquerque, NM 87123
1	MS 1005	Russ Skocypec, 06640
1	MS 1188	Chris Forsythe, 06641
1	MS 1188	Wendy Shaneyfelt, 06641
1	MS 1188	John Wagner, 06641
1	MS 9158	Nina Berry, 08961
2	MS 9960	Central Technical Files, 8945-1
2	MS 0899	Technical Library, 4536