AC 2008-1665: TOWARDS AN UNDERSTANDING OF ARTIFICIAL INTELLIGENCE AND ITS APPLICATION TO ETHICS

William Birmingham, Grove City College

Towards an Understanding of Artificial Intelligence and Its Application to Ethics

1. Introduction

Artificial intelligence (AI) is a broadly defined discipline involving computer science, engineering, philosophy, psychology, political science, and a host of other disciplines. Because AI is so broad, it is hard to succinctly define; for the sake of brevity, we will use the handle of "thinking machines," without commitment to depths of this thinking.

The machines that AI researchers develop are unlike any machines ever built in history. Before AI, machines were constructed only as a mean to perform work. Machines were mostly built to save labor, entertain, or measure things. While classic literature has stories of machines (Pinocchio) or statues (Pygmalion) coming to life, the scientific and technical communities did not, until recently, believe there was any real possibility of such a thinking or acting machine. The technological breakthroughs of cheap, easy-to-use, large-scale artificial memory and computation radically changed the conception of what machines were capable of doing. In a landmark paper, Turing challenged the scientific and technical communities to create machines that could make humans believe they were interacting with another machine.[2] That is, Turing desired machines that could *think and act similar to a human being,* i.e., artificial, non-organic, non-evolved human-like machines. Thus, the idea that a machine could have the distinctly human abilities of thinking and self-reflection entered the scientific and engineering realms.

In some way, the AI enterprise can be considered a response to Turing's challenge, where engineers are developing ever more powerful thinking machines, eventually leading to machines that some might believe are indistinguishable from humans. The creation of more complex artificial agents inevitably leads to a question of what constitutes humanness, which in many AI circles is, by and large, rooted in a view that is materialistic and purely rationalistic.6 The nearly uniformly held position of AI researchers is that once we have created the proper rational superintelligent machine, scientists and engineers will have fulfilled the goals of AI.

The philosophical discussion in AI centers on the functionality (computation and memory) needed to get the right kind of rational thinking machine that will necessarily yield human (-like) machines. The position held by many is that it is simply a matter of time until we hit upon the right mix. By and large, the critics of AI do not dispute the prevailing view; rather, they argue against or for a particular technology or function that will lead to human-like behavior. Or, that it currently technology cannot possibly supply these functions.

The reduction of humanness to computing ability and memory is disconcerting to those trained in or strongly influence by humanist traditions that ascribe inherent dignity to the human person. We contend that, for example, those drawing from philosophical traditions influenced by Thomas Aquinas, would argue that while rational thinking is an important element of what humans do, defining humanness by "the right mix" of the rational functions we perform or even the way we think is problematic. Other philosophical traditions, such as personalism or phenomenology in the 20th century, would also argue for a richer conception of humanness than a view of humanness predicated on mere functionality. This paper proposes to present an account of humanness distinctive from that offered by many working in the AI field and an argument those working in AI should take this view seriously; moreover, this paper will argue that, while AI cannot re-create humanness, AI can be enthusiastically accepted and utilized by those committed to the main currents of the Western intellectual tradition, either religious or purely humanistic.

2. Machines Are Not Persons

We can claim without controversy that humans are not computers and computers are not humans. Aside from the obvious biological conundrum such equivalence would raise, there are no programs that have the range of performance over the vast cognitive and sensory functions of humans, even if these machines have "human-like" performance on some specific task.

So what? one might ask: the issue that we must address is will these machines become with future technology that for all practical purposes will yield unlimited memory and computational power 'human' in some meaningful sense or have the same rights and duties as humans.

We wholeheartedly agree that technology will come about in the not too distant future that will allow AI systems to have computational abilities, both in terms of speed of computation and access to vast amounts of information, that far outstrip human performance in at least a great many functions. We will even cede the point that the same will eventually hold for sensory input; that is, AI machines will likely have the ability to perceive the world in a way similar to human perception, including understanding art, music, and literature.[3] We will not address whether these machines can "enjoy" or "create" these sensory media, as it is not the point of this paper.

We argue that all this technology masks what is the core issue: does this clever, extraordinary simulation actually equal humanness? Or, more importantly, do machines that exhibit this sort of artificial humanness demand to be treated in the same fashion as humans. From a moral point of view, do these machines become equal to human beings. Do we have the right to turn them off, to destroy or create them at our whim, and so forth? While Pinocchio actually became a boy, do these machines, while not become human beings in the sense of organic embodiment, become boys, too?

We contend that no matter how sophisticated, a simulation does not become the thing[12]. For some reason, it seems that AI is given special exception to this principle that we would never afford to other very clever simulations. For example, no matter how good and knowledgeable the actor, we would never claim that person playing the character Ben Franklin is *Ben Franklin*. In fact, in a recent episode of television series *The Office,* an actor playing Ben Franklin passes the Turing Test when being interviewed by the Dwight Schrute, almost as if to emphasis the fact the facsimile is not the thing.[6]

Most importantly, we contend that no artificial thing can become the moral equivalent of a human being.

The question of what constitutes a person is not an uncontested question. Those AI proponents who see humanity as the capacity for computation and memory are heirs of the Platonic or Cartesians tradition which separate body and soul. There is a philosophical tradition which rejects Descartes' 'person as intellect' view and which understands humanity not as one entity or even as two entities (body and mind) which happen to co-exist, but instead understands human persons as body and soul which subsist together.

This approach to the human person is articulated in Thomas Aquinas, although he draws upon the 6th century philosopher Boethius. Aquinas understands human beings as an entity which combines the incomplete substances of body and soul. The human person is a dynamically cohesive being which is self-preserving. According to this view, persons are rational, but that rationality subsists in and through a particular form of the body. The body is not simply the vessel for the rationality. The body as a whole and the rationality of an individual person are essentially connected. Hart says that for Aquinas that "soul is life itself, of the flesh and of the mind"[4] and that the soul "encompasses every dimension of human existence." Aquinas, drawing upon Boethius, further argues that each person is not only an individual substance, but an individual substance that is connected to other similar substances. The person then exists both individually and in communion with other persons. (Summa Theologica I.29 art 1)

The high regard for the person as such is not limited to those who would consider themselves Thomists. Immanuel Kant argues the Grounding of the Metaphysics of Morals that persons are end in themselves and that human communities should be viewed as a "kingdom of ends'. Kant does not explicitly adopt or a reject the person as unity of body and soul, but he certainly sees human beings as using their bodies as autonomous agents.[7]

Kant's defense of the dignity of the person influences a tradition of philosophers and theologians who come to be known as personalism. This philosophical approach, which had advocates in the United States in the late 19th and early 20th centuries, gained a following in Europe after WWI and whose best known advocate was Emmanuel Mounier. Mounier asserts that a person was "a living activity of self-creation, communication and commitment that comes to know himself through actions by which he becomes a person." Mounier asserts that the acceptance of the reality of the person, with the body and intellect united, could be held by both theists and atheists. The body could be aided through modern science, but the body should certainly not be destroyed or overcome.

Complementing personalism in the 20th century was the phenomenological philosophical tradition that includes figures such as Edmund Husserl (1859-1938), Max Scheler (1874-1928), and Edith Stein (1891-1943). Husserl sees human acts as products of intentionality and as acts of consciousness in and through a body. Scheler's ethical vision draws upon Kant's conception of the autonomous person, but seeks to 'personalize' Kant by characterizing the person as a creature who engages in acts of love. Stein seeks to understand to understand the meaning of the human person through acts of sympathy. This phenomenological tradition then sees persons engaging in intellectual action in and through embodiment.[11]

Martin Luther King, Jr. was influenced by the personalist tradition, and perhaps the most recent prominent public figure influenced by this tradition is Pope John Paul II. In 1969, he authored his most dense philosophic work, *The Acting Person*, which is an exploration of the meaning of human embodiment. John Paul II continues this exploration of human embodiment in a series of address given in the early years of his pontificate. In these addresses he asserts that the body 'reveals' and 'expresses' the meaning of the person. The body is not just an outward sign of the person, but is essential to the very being of the person. John Paul II also argues that the person comes to know himself as a being with a body. The body is then essential to self-knowledge. While some of John Paul II's statements about embodiment are influenced by passages from the Christian Scriptutures, many comments about embodiment are based on observations of the experience of embodiment and could be regarded as compelling to non-Catholics or non-theists.5

This brief survey suggests there is a vibrant philosophic alternative, with proponents who are both Christian and non-Christian and theistic and non-theistic, and this approach regards persons as different from the capacity for computation and memory. Our experience of humanness confirms our regard for the body as not merely a vessel for the intellect, but to be an essential to our humanness.

3. The argument for AI

We stand against the idea that it is possible to create the moral equivalent humans or proto-humans from technology. AI system will forever remain machines, and thus we are free to do with these machines what we do with other machines: turn them on or off, create or destroy, modify or dismantle in any way that we will. We further argue that artificial agents, even if they are autonomous, interactive or adaptable, still lack the capacity for responsibility and therefore cannot held responsible for their actions in the same way that we hold human being responsible for their actions. In this respect we disagree with Floridi and Sanders who argue that some artificial agents, depending on their level of abstraction, can be held morally responsible.7 Certainly we can praise or blame the actions of artificial agents, but they cannot be imprisoned because they cannot understand the larger social consequences of actions.

While we assert that artificial agents cannot be moral agents, there still remains the question: does AI entail particular moral concerns? Does AI research have similar moral issues concomitant with biological research? Does creating ever more capable machines, for example, raise the possibility of creating something that will endanger humans? Do AI researchers have to be careful of what they create in the way that researchers in biology must be concerned with both what they make (e.g., clones) and stuff from which they make it (e.g., embryonic stem cells).

Is AI research more akin to development of nuclear weapons? The morality of nuclear weapons is debatable, with strong opinions on both sides. The Catholic Church, for example, has declared the use of these weapons to be morally unjustifiable in that they have no purpose other than annihilation of the human race. By extension, it is reasonable to conclude that working on these weapons is, if not morally untenable, at the very least highly suspect.

Or, does AI research require no special moral consideration separate from what we consider for other types of engineered systems, e.g., safety, reliability, usability, etc. If, in fact, AI systems are nothing more than machines that help us with work and thinking, they have no special moral concern.

We start with the proposition that the most an AI system can ever do is create a machine that is extremely "powerful." For some applications, the power might be in the ability to compute extremely fast and to access vast amounts of information (also very quickly). For some, the power might in a robotic machine that is both strong in sense of applying force and sensing its surroundings. In other applications, the AI system might be part of a computer game (computer controlled character). We do not consider AI systems that are hybrid human/computer, such as "cyborgs," as legitimate insofar as such systems would not simply assist the human person, but would radically alter the person, in the same way that genetically altering a human being by incorporating non-human elements. Our moral intuitions and our moral reasoning would be greatly troubled by human transformation.

3.1 Computing systems

AI systems are composed of the same elements that one might find in a payroll system or a radar system: hardware and software. The thing that differentiates AI system is the algorithms used to implement them.

Thus, those working in AI do not have to consider any special moral concerns about the “stuff” that they use for building their systems. Unless the elements of the computing system inherent moral concerns from other areas (e.g., the use of embryonic stem cells as system memory), the components themselves pose no problems. In this way, AI systems are different from biological systems. In the same way, the particular processes and implementation of software used in AI systems poses no special moral challenges. Software itself is morally neutral, in the same way that any other material is morally neutral.

The systems themselves, however, are where issues arise. If we consider AI systems to be like other software systems, then they fall under the same moral concerns as other computation systems. Developers need to ensure they are safe, reliable, secure, legal, protect privacy, do not harm persons or property, and do what they are supposed to do.[13]

This still leaves open a range of moral questions. For example, is it ethical to build an AI system that guide a nuclear weapon to its target so it can fulfill its purpose of annihilating an entire city? This question, however, is no different for AI system than it is for a non-AI computing system. Thus, AI itself raises no special considerations and can be considered “regular software.” The ethical frameworks used computing ethicists to analyze these issues are sufficient.

3.2 Crossing a sensitive line

Are there situations where AI system leave the realm of “regular software.” We believe that there are. As we start to understand the operation of the human genome, it is possible that a new type of AI system may emerge, where the *information* attributes of humanness are able to be integrated into a computer program or even simply be accessible as a “design example” for those writing AI programs.10

Consider the following scenario: imagine that we have decoded the human genome and the biological mechanisms controlled by it. Further suppose that we are able to write a program that very accurately simulates these mechanisms in such a way that they can be “programmed” by a person’s genome and that the program can perform actions in the natural world.

One might argue that such a system has become a very accurate simulation of a particular person. Given embodiment in a suitably realistic robot, it might even be mistaken for the person whose genetic code was used to program the system. In other words, we will have a replica of a person that was programmed by genetic code (e.g., the "clone" army of Star Wars).

While we contend that this robot, no matter how perfect the system is at simulation, is not a person, such a system raises serious moral issues. While the most obvious issue is bioinformatics, such as protecting personal privacy, appropriate use of information, etc., this is not the most important consideration.

This use of this type of information is intrinsically dehumanizing. It separates out an element of personhood from other integral elements; and, in fact, separates the manifest of a person's genetic code from the body for which it was intended. This directly leads to the view of a person simply as a commodity. It is not hard to see how certain genetic traits may be enhanced or suppressed to fit a particular utilitarian aim. Thus, the person is merely an instrument to some other end.

The ACM Code of Ethics (1.1)[1] asserts that one should 'contribute to society and human well-being." The above example would not serve to promote human well-being; in fact, for the person whose genetic code was copied into another, there would be the potential for significant harm to that person. AI would be acceptable to a wide range of ethical systems if it were directed for improving work, meeting human needs for improved health and safety, and for providing entertainment with socially beneficial results (or at least without social harm).

Thus, it is difficult to make an argument for the 'uploading of a persons' genome" while maintaining regard for the dignity of the person. We suggest that research in this area carries a high, negative, moral consequence.

4. Summary

AI systems are sophisticated and will become increasingly more sophisticated over time. From a humanistic view, they can never become the moral equivalent of humans. Thus, these systems deserve no special moral consideration; they do not get the same moral status as persons or as agents fully accountable for their actions.

While we view AI a worthy and legitimate intellectual activity and a way to truly improve the human condition, there are some moral issues that one must consider. The sort of research most commonly undertaken by researchers, developing hardware and software systems, that perform "human-like" tasks have no significant moral consideration aside from that which covers non-AI software. The problem comes when particular elements of a human person, genetic code for example, is used in the creation of an AI system. In this case, we believe that such work can have significant negative moral consequence. We would advise our students to not participate in such work.

Bibliography

1. Association for Computing Machinery. Code of Ethics. http://www.acm.org/about/code-of-ethics/.

2. Turing, A.M.,"Computational machinary and intelligence," 1950 Mind, Vol. 59, pp. 433-460.

3. Cope, David, *THe computers and musical style.* Madison, WI : A-R Editions, 1991.

4. Hart, David B. John Paul II and the Ethics of the Body. The New Atlantis (Summer 2005): 65-82.

5. Floridi, Luciano and J. W. Flanders. "On the Morality of Artificial Agents." Minds and Machine 14 (2004): 349-379.

6. Kaling, Mindy. *The Office "Ben Franklin", Episode Number 42.* [perf.] Randall Einhorn. February 1, 2007.

7. Kant, Immanuel. Grounding of the Metaphysics of Morals. Indianapolis: Hackett Publishing, 1993.

8. May, William. Pope John Paul II and Bioethic: Bodily Life and Integral to the Human Person. Undated Manuscript available at www.catholicsocialscientists.org.

9. Norvig, S. Russell and P, *Artificial intelligence: A modern introduction 2nd Edition.* : Prentice-Hall, 2003.

10. .Sanders, Luciano Floridi and J.W. *On the morality of artificial agents*.: Kluwer Academic Publishers, 2004, Minds and Machine, Vol. 14, pp. 349-379.

11. Schmiesing, Kevin. A History of Personalism. Undated Manuscript at www.Acton.org

12. Searle, John.*Minds, brains, and programs.* . 1980, The Behaviorial and brain sciences, pp. 417-457

13. Tavani, Herman T.. *Ethics and technology: Ethical issues in an age of information and communication technology.* s.l. : John Wiley and Sons, Inc, 2007.