COSMOS
Navigate the Compliance Universe

ethikos Volume 34, Number 3. March 01, 2020
An artificial intelligence code of ethics

By Marianne M. Jennings

Marianne M. Jennings (marianne.jennings@asu.edu) is Professor Emeritus, W.P. Carey School of Business, Arizona State University in Tempe, AZ.

Due to leg problems, compression stockings became my way of life. The sheer expense of these opaque stockings necessitated internet searches for the best prices, sites, selection, and colors. After ordering that first pair of 30–40 mg hose, ads for cremation, walk-in tubs, scooters, testosterone supplements, and canes rolled in, popped up, and obnoxiously entered into what was once a peaceful life. Someone somewhere assumed that the one singular purchase of support stockings meant that end-of-life and dotage products were just the ticket.

The companies and tech folks sold their data mining results on a woman who purchased Sigvaris compression stockings. Their analytics told them that they had hit pay dirt on a buyer seeking the comforts and treatments for old age and beyond. The precision, the ability to target, and the accessed contact information brought pride and, in other cases, perhaps, the sales of tubs and Poligrip, as well as prepayments for cremation. They were wrong in this case. Their analytics were not 100% on the money. In short, they got the wrong person. As the Monty Python folks would say, "I'm not quite dead yet, sir."

The use of private purchasing data for whatever purpose without consent is ethically problematic. Incorrect assumptions about purchasers raise additional ethical issues, including everything from profiling to privacy concerns. Facebook folks either did not see these ethical issues or were perfectly comfortable with their solution—gather it and sell it. The uses of artificial intelligence (AI) are varied but consistently involve ethical questions. The ethical risks of facial recognition also involve forms of profiling. The complexity of ethical issues with driverless cars is boundless, taking us back to one of the age-old philosophical dilemmas: "Do I swerve and hit one person to avoid hitting five?" Or even the more frequent, "Do I swerve to avoid a deer, or do I hit the deer and put myself at risk of injury or death?" Who makes those decisions in developing the technology for driverless cars?

An AI code of ethics is no small task. Perhaps the easier task is to evaluate what has already been done and offer insights into what is missing and how to improve upon the efforts to date.

What's out there in AI codes of ethics?

When thinking about ethics, AI, and codes coming together, the tech folks who mine, group, and sell the data believe that they have AI covered. There are issues with these AI codes of ethics, and those issues are described below.

The generic and lofty approach

The Asilomar AI principles[1] have directives for AI research, but even these approaches are problematic. Take this research goal:The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.

Akin to Google's former code of conduct clause, "Don't be evil," generic principles are not clear to everyone, and "evil" doesn't carry a universal definition. One person's evil is another's "not so bad." This AI ethical principle carries a noble and lofty generic quality. Loftiness in codes opens doors for, well, evil. One person's benefit is another person's nightmare. Driverless cars do have their benefits, but, as noted earlier, there are decisions programmed into the cars that make ethical choices in situations the human mind and centuries of philosophical debates have not yet resolved. These well-intentioned principles-based codes of ethics tend to offer little guidance but wide latitude on AI. An information technology executive once explained to me that it was perfectly ethical to collect health and medical information on patients because their goal was to provide those patients with suggestions, information, and care opportunities that would benefit them.

The lists of AI ethical principles and discussion documents

There are a few professional organizations that detail lofty principles. For example, the Institute of Electronics and Electrical Engineers (IEEE) has developed what it calls a discussion and recognition document "for the purposes of furthering public understanding of the importance of addressing ethical considerations in the design of autonomous and intelligent systems."[2]

In other words, as long as you help the public understand and consider the ethical issues, you can sally forth. The overall thoughtful reflection is then translated to these principles, followed by a brief analysis of each principle because, as the IEEE provides, it is looking for feedback and discussion.

1. **"Human Rights**. [Autonomous and Intelligent Systems] A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights." Currently, we live in a world where human rights aren't always provided to everyone. If the US is close, there are outliers, such as China. Does this mean that even the most beneficial AI that could help those in China will not have that benefit because it would be operating in a country that, most people agree, violates human rights?

2. **"Well-being**. A/IS creators shall adopt increased human well-being as a primary success criterion for development." There is the beneficial term again, phrased as "human well-being."

3. **"Data Agency.** AI/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity." The language is unclear; is it control over "their identity" or over the use of information about them that becomes part of the public internet world?

4. **"Effectiveness.** A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS." The driverless car has yet to establish its safety and fitness.

5. **"Transparency**. The basis of a particular A/IS decision should always be discoverable." Where will we be able to obtain that information? How will we know who is doing what with our information?

6. **"Accountability**. A/IS shall be created and operated to provide an unambiguous rationale for all decisions made." Whether there is an unambiguous rationale for all decisions is not as important as knowing what is being developed, why, for whom, and who will be affected by the decisions. Face recognition remains an ongoing debate because of the use of the technology by law enforcement and the decisions on who committed what crimes that have come from the mistaken identities.

7. **"Awareness of Misuse**. A/IS creators shall guard against all potential misuses and risks of A/IS in operation." Target, Equifax, and a host of other competent companies have not been able to curb the discovery of the data they hold on their customers.

8. **"Competence**. A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation." Perhaps a certification program could help with this principle, but there are 14-year-olds developing AI systems at home.

A code of ethics should provide boundaries, something more than the discussion of what can happen with AI. That issue in AI is clear. What is not clear are the lines that carry some sort of a "Thou shalt not."

The more detailed AI codes

IBM has made an effort with its AI code of ethics.[3] However, even IBM begins its code of ethics with this caveat:

"This document represents the beginning of a conversation defining Everyday Ethics for AI. Ethics must be embedded in the design and development process from the very beginning of AI creation.

"Rather than strive for perfection first, we're releasing this to allow all who read and use this to comment, critique and participate in all future iterations. So please experiment, play, use, and break what you find here and send us your feedback.

"Designers and developers of AI systems are encouraged to be aware of these concepts

and seize opportunities to intentionally put these ideas into practice. As you work with your team and others, please share this guide with them."

IBM, the company with Watson (an AI being that can tell you when an elevator needs repairs or the course of treatment for a patient), is still struggling with gritty details. IBM also proposed principles. However, IBM has added examples to make the principles clear. For example, one of the IBM principles is accountability, which declares, "Every person involved in the creation of AI at any step is accountable for considering the system's impact in the world, as are the companies invested in its development."[4] The example IBM uses is the placement of virtual assistants in hotel rooms—assistants that could control temperatures; reach the front desk; or turn on lights, radios, and televisions. They provide no benefit to the guest, but, oh, the data the virtual assistant can gather. Couple that data with the front desk's knowledge of the guest's physical address, email address, phone number, and status in any frequent flyer/hotel club program. What a marketing treasure trove!

The only thing this principle requires is that AI developers should be encouraged to think about the impact of their creation at all steps and be sure the companies investing in that development are also accountable. A promise is all it takes.

IBM does have some best practices for AI developers that are listed following the accountability principle, such as keeping detailed records of the development of the AI system, and recognizing that the developer may not have control over how the company uses what the system collects. Then there is advice to check with linguists, sociologists, and other professionals to further discuss the ethical issues.

Another IBM principle is value alignment, which is good advice. Don't do work for companies or on projects for others if their values do not align with your own. A simple principle that makes sense until we read the explanation:

"AI works alongside diverse, human interests. People make decisions based on any number of contextual factors, including their experiences, memories, upbringing, and cultural norms. These factors allow us to have a fundamental understand of 'right and wrong' in a wide range of contexts, at home, in the office, and elsewhere."[5]

That grant of broad discretion along with quotes around the phrase "right and wrong" send chills down an ethics professor's spine. This license provides AI developers and users with latitude under the umbrella of beneficial nobility.