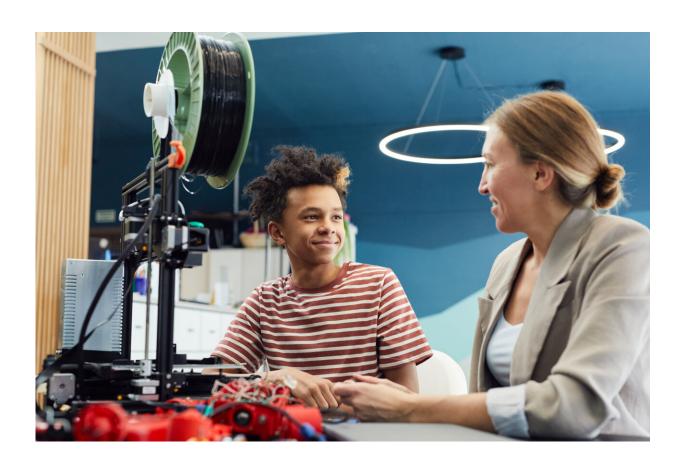# Engineers, philosophers and sociologists release ethical design guidelines for future technology

December 13 2017, by Rafael A Calvo And Dorian Peters

Credit: Vanessa Loring from Pexels

If kids spend hours a day speaking to digital personal assistant Alexa, how will this affect the way they connect to real people? When a self-

driving car runs over a pedestrian, who do you take to court? Is it okay to manipulate people's emotions if it's making them happier?

Together with an international team of researchers in fields as diverse as philosophy, engineering and anthropology, we set out to tackle these questions. The result is a new set of guidelines focused on the ethical and social implications of autonomous and intelligent systems. That includes everything from big data and social media algorithms to autonomous weapons.

The report, Ethically Aligned Design, was released today by the Institute of Electrical and Electronics Engineers (IEEE). It is the culmination of a year's work by 250 world leaders in technology, law, social science, business and government spanning six continents.

IEEE is the world's largest technical professional organisation. With over 420,000 members in 160 countries, it's the global authority for professional standards related to technology. The latest report proposes a set of recommendations (suggestions) that are open to public feedback.

Once adopted, the guidelines in the report will be implemented by professional organisations, accreditation boards and educational institutions to ensure the next generation of engineers incorporate ethical considerations into their work.

Guiding principles

The big questions posed by our digital future sit at the intersection of technology and ethics. This is complex territory that requires input from experts in many different fields if we are to navigate it successfully.

To prepare the report, economists and sociologists researched the effect of technology on disempowered groups. Lawyers considered the future

of privacy and justice. Doctors and psychologists examined impacts on physical and mental health. Philosophers unpacked hidden biases and moral questions.

The report suggests all technologies should be guided by five general principles:

- protecting human rights
- prioritising and employing established metrics for measuring wellbeing
- ensuring designers and operators of new technologies are accountable
- making processes transparent
- minimising the risks of misuse.

Sticky questions

The report runs the spectrum from practical to more abstract concerns, touching on personal data ownership, autonomous weapons, job displacement and questions like "can decisions made by amoral systems have moral consequences?"

One section deals with a "lack of ownership or responsibility from the tech community". It points to a divide between how the technology community sees its ethical responsibilities and the broader social concerns raised by public, legal, and professional communities.

Each issue tackled includes background discussion and a set of candidate recommendations. For example, the section on autonomous weapons recommends measures to ensure meaningful human control. The section on employment recommends the creation of an independent body to track the impact of robotics on jobs and economic growth.

A section on affective computing – an area that studies how computers can detect, express and even "feel" emotions – raises concerns about how long-term interaction with computers could change the way people interact with each other.

This brings us back to our question: if kids spend hours a day speaking to Siri or Alexa how will these interactions change them?

The report makes two recommendations on this point:

1. To acknowledge how much we don't know (we need to learn much more before these systems become widely used);
2. That humans who witness negative impacts – parents, social workers, governments – learn to detect them and have ways to address them, or even shut technologies down. Experience shows this is not always easy – try forbidding your child from watching YouTube and see how well that flies.

Clearly affective computing is an area in which we are at a particular loss for evidence of its human impact.

Consultation and feedback

IEEE standards are developed iteratively and the organisation will use the findings in this report to build a definitive set of guidelines over time.

Early feedback on an earlier version of the report highlighted its Western-centric bias. As a result, a larger and more diverse panel was recruited. A number of new sections were added, including the section on affective computing, along with policy, classical ethics, mixed reality (including augmented reality technologies like Google Glass) and wellbeing.

Over the next year, the final version will be released as a handbook with recommendations that technologists and policy makers can turn to, and be held accountable for, as our technological future unfolds.

This is an important step toward breaking the protective wall of specialisation that allows technologists to separate themselves from the impact of their work on society at large. It will demand that future tech leaders take responsibility for ensuring that the technology we build as humans genuinely benefits us and our planet.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation