Technology at the crossroads

Engineers pride ourselves on solving problems. But what happens when we don't know the questions? We are great at minimising one cost function, and maximising another. But what if those are the wrong functions to measure?

As our technologies change society, these dilemmas make us consider our values, ethical frameworks, and the design methodologies that can genuinely decrease harm and increase wellbeing.

ARTICLE BY: **PROF RAFAEL A. CALVO AND JOHN C. HAVENS**

The Good Life At Risk

The board room was already full when I walked in. Someone walked out to get an extra chair while I heard the round of introductions. I knew most of the academics in the room, but did not know much about their interest in the fourth industrial revolution or in engineering ethics, the theme of our meeting.

Technologies were reshaping the concept of what constitutes a good life. They were redefining work, basic emotions, and even volition.

The specific goal of the meeting was to prepare a response to Australia's Human Rights Commissioner, who was seeking comments on their human rights and technology issues paper.[1] Technologies, like the ones we were about to discuss, impinge on our perceived rights[2], for example around privacy, free speech, workplace technologies and government surveillance.

Not everyone talked. Some were probably asked to be there by their managers, but the majority felt a sense of urgency uncommon in academics. Most of us explained what brought us there. There was a manufacturing researcher who explained how 3D printing was being taken to an industrial level and they could not predict what manufacturing would look like in 10 years.

Then there was a biomedical engineer who described how brain stimulation could be used to supress the excretion of adrenaline and therefore supress the fear response that stops soldiers from going into a battlefield (the same emotion that many of us were feeling).

Then it was my turn to describe how social networks have been used by companies and governments to manipulate the attention of people like us, both pushing their commercial and political agendas. How companies like Cambridge Analytica were risking the core democratic values of the enlightenment.

In all these cases, technologies were reshaping the concept of what constitutes a good life. They were redefining work, basic emotions, and even volition. These are not 'academic' questions, they have daily consequences to citizens, and to engineers like us.

Technologies are designed to shape what we spend time on, and the outcome was that we spent one billion human hours watching Youtube videos in 2017, millions of hours sharing selfies and making lip-sync videos on musical.ly. Technologies are transforming, or altogether eliminating jobs (according to some, 800M jobs will disappear by 2030[3]).

No human activity is safe from disruption: even caring for others could

How far should we go?
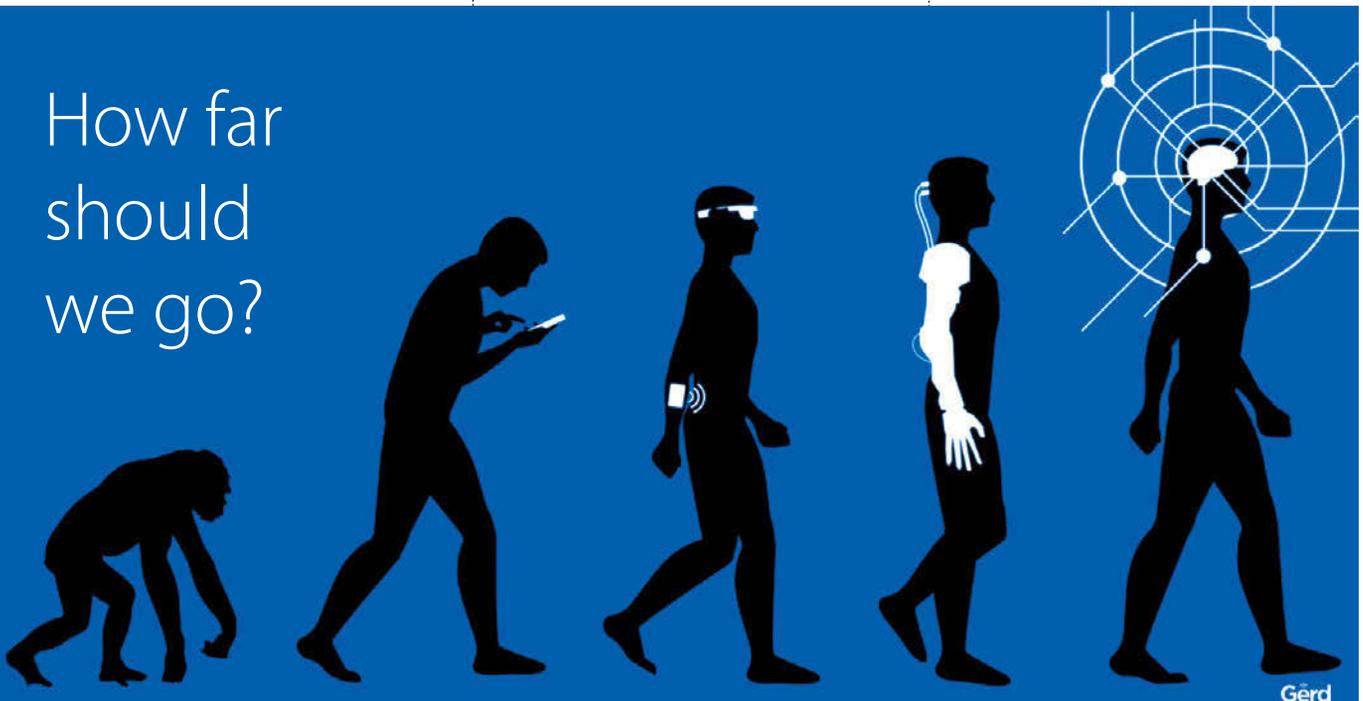
be the purview of robots (consider that in Japan 80% of the elderly will be taken care of by robots – in 2020[4]).

As reflective professionals and engineers we should be asking, is this the sort of world we want?

I think the academic who called the meeting, a world leader in robotics, did not expect the general sense of unease many of us had. He told the story of a public panel where he discussed his vision for the future, together with a journalist and a philosopher. At the beginning of the event the host asked the audience how many thought that AI would make the world a better place. Roughly ½ raised their hands. An hour into the event, when the three speakers had presented their arguments they asked again. This time only twenty percent raised their hands.

My colleague thought it had been his fault – a poor presentation of the benefits. But I was there and knew he had done an excellent job presenting his views. Could it be that it was not him, but us, (engineers in general) who do not have a compelling description of 'the good life'? Could it be that our vision of what technology can provide did no longer satisfy people's ideas of what they want?

As technologies become part of our lives, their implicit values (including that of their designers) raise moral dilemmas. Should technology be designed to drive us away from pain and towards the hedonic pleasures offered by the market economy? Or should they be designed to support what Aristotle called the *eudaimonic* life, where the good life consists of achieving potentials?

When I (Rafael) wrote *Positive computing: Technologies for wellbeing and human flourishing*, we could already see technologies having positive and negative effects on the human psyche. We discussed a new discipline of how technologies could (and should) be designed to support psychological wellbeing. That is still the case.

Design engineers cannot leave to chance the unintended impact of technologies. In the same way that we design products to be safe and respect our physical health, we need to design them so they respect our psychological health, and that of our environment.

Engineers are often stereotyped as naïve technophiles. Of course some are, but the actual perspectives of most engineers do not reflect this attitude. Colleagues from around the world are expressing their concerns, questioning long-standing ideas, and raising alarms.

In such situations it may seem easy to despair, but we can also find strength in numbers with a growing awareness that emerging technologies require a new level of due diligence to protect and increase the wellbeing of humanity at large.

What has become increasingly clear is that the definitions of 'good' and 'bad' are more complicated than we could expect. Anyone only looking at today's media would think that moral norms vary amongst age groups, political parties, even genders. So what are the universal values that should drive our technology design?

Reframing Design in the Age of The Algorithm

One way to answer this question is through open, multidisciplinary and multistakeholder debate. This is what the Institute of Electrical and Electronic Engineers (IEEE) have been doing since 2016 in relation to the ethics of autonomous and intelligent systems.

IEEE is the largest professional organisation of its kind. Over 420,000 engineers from 160 countries are members who go to IEEE conferences, participate in forums, read (and write) articles in its many journals.

The IEEE Global Initiative on Ethics of

> Anyone only looking at today's media would think that moral norms vary amongst age groups, political parties, even genders. So what are the universal values that should drive our technology design?
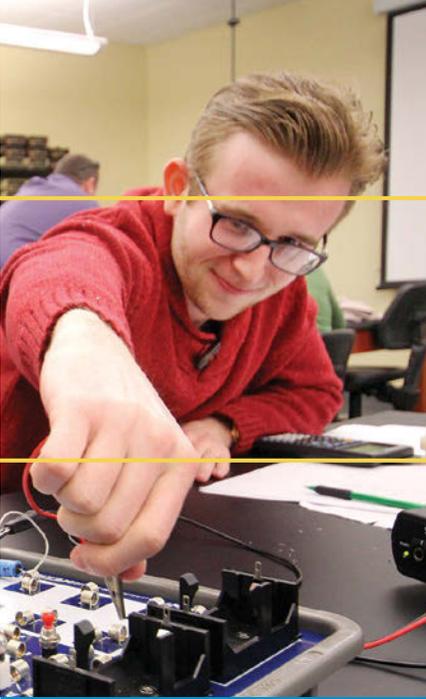
Should guidelines be narrow and strict, bordering regulations, **or should they be broad and an instrument for critical thinking?**

Ethically Aligned Design Version 1

You can download Ethically Aligned Design version 1 at
HTTP://BIT.LY/EADV1[7]

You can also see an overview of the document at
HTTP://BIT.LY/EADV1-SUMMARY[8]

Ethically Aligned Design Version 2

You can download Ethically Aligned Design Version 2 at
HTTP://BIT.LY/EADV2[9]

You can also see an overview of the document at
HTTP://BIT.LY/EADV2-SUMMARY[10]

Autonomous and Intelligent Systems[5] (A/IS) was launched in April of 2016 to move beyond paranoia, or the uncritical admiration, of autonomous and intelligent technologies. Its goal is to align technology development and use with ethical values, advancing innovation in a way that truly serves humanity while diminishing fear in the process.

The IEEE Global Initiative also aims to incorporate ethical aspects and values relating to human well-being in ways that may not automatically be considered in the current design and manufacture of A/IS technologies.

This is why the Mission Statement of The IEEE Global Initiative is, *"to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity."*

By training all engineers or manufacturers to utilise applied ethical frameworks (or what is often called, "values-based design") before projects are sent to be developed, progress will not be measured only in terms of materialistic criteria but can include the intentional prioritisation of individual, community, and societal flourishing as measured by both subjective and objective criteria.

The discussions of the IEEE Global Initiative and the IEEE P7000™ Standards Working Groups that it inspired,[6] are open to the public or any experts who

wish to join. In this process entrepreneurs, psychologists, sociologists and philosophers are sharing their expertise with other engineers, data scientists and engineering stakeholders moving the overall work from 'principles to practice'.

The IEEE Global Initiative has several outputs including the creation and iteration of a body of work known as *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* and the identification of multidimensional indicators of wellbeing. *Ethically Aligned Design* (EAD) is a Creative Commons document so any organisation can utilise it as an immediate and pragmatic resource.

Version 1 was released in 2016; Version 2 was released in 2017; both versions were released as Requests for Input and received over five hundred pages of aggregate feedback. *Ethically Aligned Design, First Edition* will be released in early 2019 and will feature over one hundred top A/IS Issues and pragmatic Recommendations.

It was created by over three hundred global experts in A/IS and another seven hundred Initiative members were able to review it so that EAD can be the 'go-to' resource to help technologists and policy makers prioritise ethical considerations in A/IS.

Of course, the IEEE is not alone. Other initiatives include the Ada Lovelace

IMAGE: © Cdbrice00-wiki

AUTHOR:

Rafael Calvo is Professor at the University of Sydney, and Future Fellow of the Australian Research Council. He is Director of the Wellbeing Technology Lab that focuses on the design of systems that support wellbeing in areas of mental health, medicine and education. Rafael is co-author of over 200 publications, including "Positive Computing" (MIT Press, 2014) and co-Editor of the *Oxford Handbook of Affective Computing*.

🌐 **WWW.RAFAEL-CALVO.COM** 🐦 **@rafael_a_calvo**

AUTHOR:

John C. Havens is Executive Director of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and The Council on Extended Intelligence (CXI). He is the author of the books, *Heartificial Intelligence: Embracing Our Humanity To Maximize Machines* and *Hacking Happiness: Why Your Personal Data Counts* and *How Tracking it Can Change the World*.

🌐 **WWW.JOHNCHAVENS.COM** 🐦 **@johnchavens**

Institute, The Leverhulme Centre for the Future of Intelligence, The Future of Humanity Institute, the AI Now Research Institute, and many more.

Companies are also acknowledging some of the issues: John Giannandrea (who leads AI at Apple), Mustafa Suleyman (co-founder of DeepMind), Satya Nadella CEO of Microsoft have, for example, acknowledged the risk of bias in AI systems.

Achieving consensus about what are the important issues is hard. Achieving consensus on how to address them would seem impossible. Even the form that guidelines should take is controversial. Our work in IEEE has seen many heated discussions. Should guidelines be narrow and strict, bordering regulations, or should they be broad and an instrument for critical thinking?

Multiple approaches are probably needed. For example, the IEEE has also established the Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)[11] dealing with transparency, accountability and reduction of algorithmic bias.

Some groups are top down, with a small group of 'experts' (who may consult the public) producing their own reports and research. Some are industry based, others more academic. Consumer or civil society groups are also joining the fray, including The European Consumer Organisation (BEUC) and AlgorithmWatch.

The Good Life, Reengineered

Engineers, and those designing technologies, need to move beyond the traditional method of design to deal with the new realities facing humans in the algorithmic age. And we are beginning that critical work. We are working with others across disciplinary, cultural and political boundaries.

The engineering community, through IEEE and other professional organisations, will continue to develop these guidelines in consultation with the broader community. The development of such guidelines will have impact in at least two areas: education and professional standards.

It is expected that these guidelines will be used in engineering curricula worldwide, and by other professional organisations. This will affect what we expect from future graduates and the way they go about their work. Another way of influencing professional practices is through the certification of AI systems (similar to the ISO quality standards accreditation), an initiative that IEEE will be developing in 2019.

And we are not naïve: these discussions may lead to regulation that limits what the major tech companies, those that shape our lives, can do. But we are at a crossroads, and such transformation will need a new social contract.

Society is struggling to decide which way to go, and with a renewed focus on defining risk as the elements that keep us from realising our values or full potential, engineers can help minimise risk and increase wellbeing for the future. **AQ**
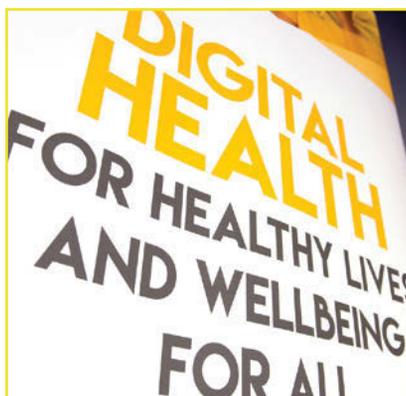
IMAGE: © ITU Pictures-Flickr