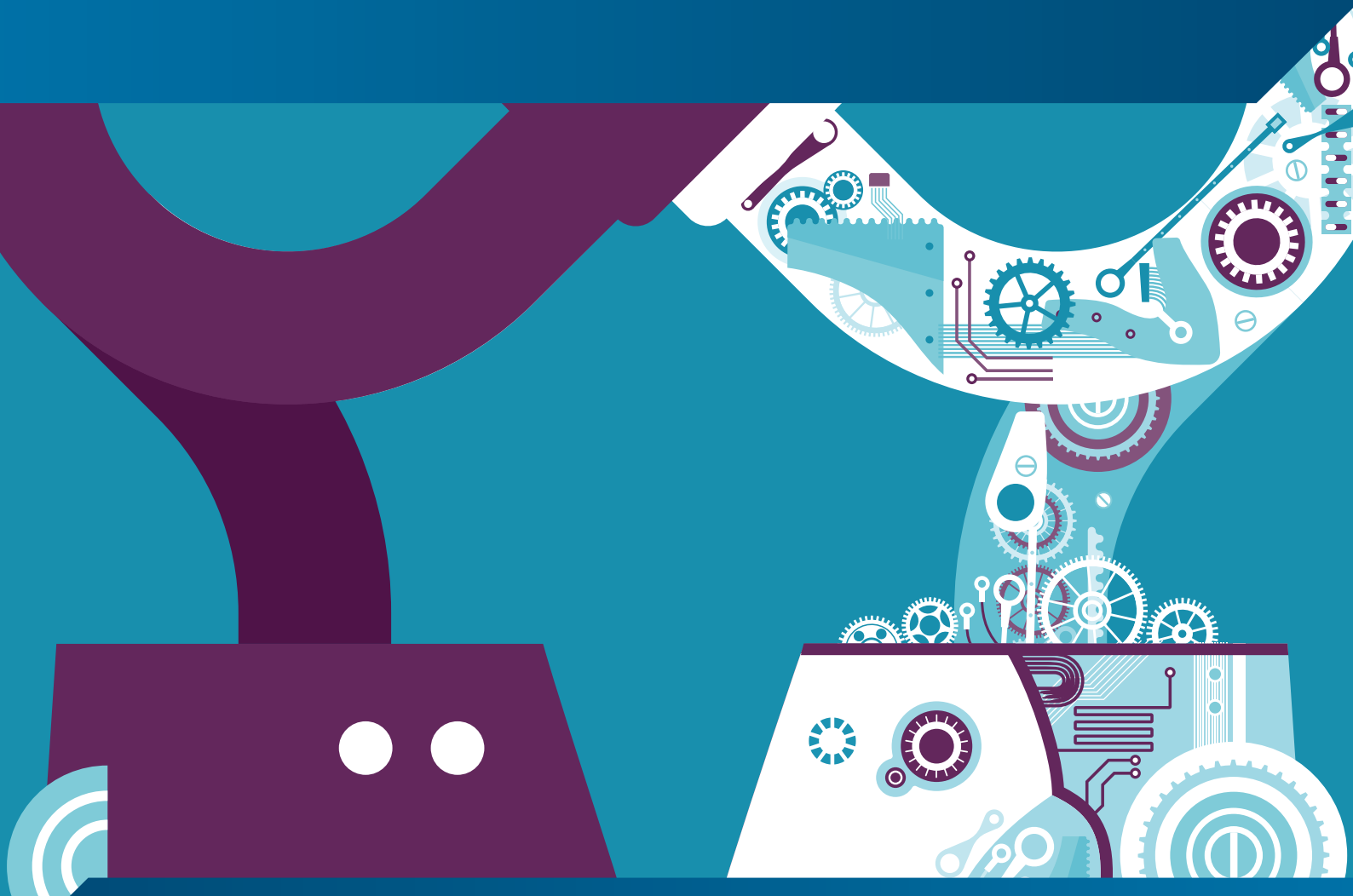
IEEE AI & ETHICS SUMMIT 2016

Artificial Intelligence & Ethics – Who Does the Thinking?

Tuesday 15 November 2016
The Sofitel Brussels Europe
Place Jourdan 1, Brussels, 1040, Belgium

Summit Report



Join the conversation at:
#IEEEAISUMMIT

www.ieee.org
© IEEE 2016

IEEE
Advancing Technology for Humanity

Artificial Intelligence can be programmed to perform as human. **But is it the smartest thing to do?**

Artificial Intelligence is quickly finding its way into the lives of people across the world thanks to the imagination and hard work of technologists and innovators, many of whom are IEEE members. As AI becomes a greater part of our everyday lives so does the discussion about managing its risks and rewards.

IEEE encourages innovators to follow their ideas — from the first insightful thought to assessing and discussing how that implemented idea affects the future of humankind. As more applications of AI are developed, IEEE members are leading the discussions on how to integrate ethical considerations into the design of AI innovations and on open-ended questions that we need to keep in mind as we work to achieve AI's potential benefits to humanity.

IEEE: Fostering technological innovation for the benefit of humanity.

www.ieee.org

IEEE
Advancing Technology for Humanity

Table of contents


IEEE AI & ETHICS SUMMIT 2016
Artificial Intelligence & Ethics
Who Does the Thinking?
Tuesday 15 November 2016 – The Sofitel Brussels Europe

Welcome: Marko Delimar – Chair, IEEE European Public Policy Initiative

The chair of the IEEE European Public Policy Initiative welcomed the audience to the "highly relevant" summit on artificial intelligence and ethics. "We rely more and more on artificial intelligence in many applications," said Delimar, further explaining that IEEE is bringing together thought leaders in ethics and artificial intelligence in order to contribute to "an ongoing dialogue regarding the issues that must be considered and addressed prior to the widespread adoption of artificial intelligence centric technologies."

Delimar said that recent innovation and development in artificial intelligence herald its fully-fledged arrival in everything from automobiles to cognitive computing. But he added, the question remains: "Will all of this benefit humanity?"

"A simple answer is 'yes of course.' The full answer is much more complex than that and that is one of our conversations here today," he continued.

Delimar then outlined some of the numerous ethical concerns that have to be addressed: What safeguards should be put in place to protect the massive amount of personal data needed to power artificial intelligence? When can we expect shifts in responsibilities in the workplace? What will those shifts look like? And will there be impacts on the workers? How do we ensure compliance with safety standards? Who takes responsibility if artificial intelligence malfunctions?

"It's up to us to build a world where we want artificial intelligence to operate within. We have all seen the *Terminator* movies and I'm trying very hard not to end this speech with 'I'll be back'," he joked.

"We are all too familiar with cinematic and literary storylines in which artificial intelligence runs amok and humanity suffers. But this is fiction. In the real world, decisions and deliberations must come before widespread adoption. From computers to communication technology, that model has been true for more than a century," said Delimar, citing electricity's roll out more than 125 years ago, and the Internet that has been around for more than 25 years.

"The best approach to building an ethical future for artificial intelligence is based within the global community with as many concerned entities as possible participating in the discussion. Only then will foundationally ethical artificial intelligence deliver its promise," he concluded. ■

Keynote Speech: Wojciech Wiewiórowski – Assistant Supervisor, European Data Protection Supervisor (EDPS)

Wiewiórowski opened his address with a parable from Nick Bostrom's book Superintelligence: It is the story of some sparrows that decide they need an owl to help them build nests faster — how easy life would be! — despite the fact that they know nothing about owls. There is only one sparrow who is a bit skeptical about it as owls are big and difficult to manage and he fears the sparrows don't know what they're getting into. He is told he has a fretful temperament.

"The role of the European Data Protection Supervisor is not to be the prophet of the apocalypse," said Wiewiórowski, "but we want to keep the position of asking questions. Sometimes difficult questions...even sometimes the questions [people] do not want to hear."

The assistant EDPS also drew on the world of fiction to illustrate his points: "There were of course a lot of hopes and expectations around artificial intelligence that were created in the last half of the century. And of course there were fears as well. When we watched the movies that we already heard about, we think about Hal 9000, we think about Skynet, we think about the replicants from Blade Runner," said Wiewiórowski, noting that, according to the story, some of the fictional androids were assembled in 2016.

"But now the artificial intelligence is very practical and artificial intelligence is the reality that has already woven its way into our everyday life with navigation systems with spam filters and weather forecasts just to name a few," he continued.

Wiewiórowski mentioned IBM's Watson supercomputer and Ross, an AI lawyer described by Richard Suskind in his book The End of the Lawyers. "This amount of attention shows that it's not too early to talk about artificial intelligence and its implications to everyday life," said Wiewiórowski. It was for this reason that the global data protection authorities' conference in Marrakesh chose the challenges of artificial intelligence as the main topic of discussion, he explained.

The Universal Declaration of Human Rights takes human dignity as its starting point said Wiewiórowski. "In early 21st century individuals are increasingly disclosing information about themselves and digital profiles can be created thanks to artificial intelligence techniques and shared in microseconds. The use of artificial intelligence to predict people's behaviors risks stigmatization, reinforcing existing stereotypes, social and cultural segregation, and exclusion, subverting individual choice and equal opportunities."

And Wiewiórowski does see some risks. Since artificial intelligence systems learn from the information provided and have no way of seeing "the bigger picture," whatever bias is introduced in training will be reinforced and will influence the predictions made, explained Wiewiórowski. "If those predictions are used to take decisions, a vicious circle starts to work and self-fulfilling prophecies can be created," he warned.

"Machine learning is one of the most researched subjects on artificial intelligence and it involves the construction of algorithms that can learn and make predictions using data. Machine learning algorithms represent knowledge and structures which sometimes cannot be translated into a form which is intelligible for us or without sacrificing its meaning. And this is serious. It has serious implications for data protection, as it means that we may not have appropriate information about how our personal data is used, and importantly how decisions concerning us are taken therefore making it impossible to meaningfully consent to use our data," he said.

"Data protection framework in Europe requires organizations and controllers of data to be transparent. As far as the data is concerned, but also as far as the algorithms that are in use are concerned, this is especially demanding in the world of machine learning where algorithms which are in use may be unknown and unpredictable even for those who have developed them."

"In the near future data protection authorities will deal with the cases where machine learning has been used for challenging or supporting a decision," Wiewiórowski predicted. "We need to adopt a realistic approach to artificial intelligence. We now have a window of opportunity to build the right values in today's technology."

But how does the story of the sparrows and the owl end? We don't know, especially in the absence of actual owl to practice on.

"The role of the European Data Protection Supervisor is not to be the prophet of the apocalypse,"

"I came to this conference looking for an owl to practice on as well," concluded Wiewiórowski, "and I have also come here to meet those who know more about owls than I do." ■

Q&A

When asked whether new regulatory schemes and new laws for AI should be developed or whether it fits into the existing body of law and regulation, Wiewiórowski replied: "One of the worst things that we can do is to regulate fast. It's not a good idea. We have our data protection regulation, which in my field tries to be the general solution for all the problems. Of course it will not answer all the questions, but at the moment that's what we have. I would not be the supporter of strong regulation before we really find out what we are regulating."

"We have examples in the history of European law, as well as other situations, that went the other way around. And the best example for me was the directive on electronic signature. The idea in this directive was actually great. The only problem was that the market didn't want to follow this idea. So even if it was great in the beginning, it was not used in practice. And after fifteen years we found ourselves more or less in the same position we were at the moment when the regulation was created," he said. ■

"After a car accident we used to cross-examine the driver; now it will be robots"

Session 1: Autonomous Systems – Managing Risk and Reward

The moderator, John C. Havens, author and executive director of The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, opened the panel by pointing out that "the advent and increased sophistication of autonomous systems offers significant potential benefits in diverse application domains including manufacturing and transportation, healthcare and financial services, exploration, maintenance and repair. As well as cost and risk reduction, potential benefits include enhanced productivity, precision and accuracy, better health outcomes, lower mortality and injury rates due to human error, as well as opportunities for greater human creativity."

However, he continued, "these are counterbalanced by a broad range of ethical, social, philosophical and legal concerns, including further dehumanizing warfare, creating existential threats and damaging the fabric of human society."

The main question for the panel, from the perspective of reducing the likelihood of negative as well as unintended consequences, is what is the best way to manage risk and reward?

Jérôme Perrin, VP Scientific Director of Groupe Renault, who comes from an academic background in physics now focuses on ethical questions of machine learning and the automated car. "Automated vehicles are a special kind of robot because they are operated by non-expert people," he said. They also interact with many social elements in a complex environment and must be managed in real-time for obvious safety reasons.

The question of safety was the overriding theme of the panel, and unsurprisingly many participants referred to seminal science fiction writer, Isaac Asimov's Laws of Robotics — three rules to which his fictional robots must adhere to in order to protect human life.

"If you remember Asimov's rules, one of them was basically that eventually the robot should self-sacrifice or commit suicide if would harm a human being," said Perrin. "In the case of automated cars, where you have a human being outside and inside the vehicle, there is a debate. It is about managing risk."

Engineers may design the algorithms, but if a user can set these parameters then is it the driver's responsibility or liability if something goes wrong, he added. We must define the levels of car automation. "Suppose I could tune my car to say I don't want a scratch at any cost," said Perrin.

Kay Firth-Butterfield, a distinguished scholar, barrister and judge, co-founder of AI Austin and co-founder of the Consortium on Law and Ethics of AI and Robotics, at the Robert Strauss Center for International Security and Law, University of Texas, Austin said: "In common law it could be argued that if you set your car up so that it only kills other people not you, it could be determined to be premeditated murder."

"We, at AI Austin, are looking at AI in healthcare and education," explained Firth-Butterfield, but she believes that ethical principles can apply to any company. "We need to design in ethical procedures and we need to do that throughout the product cycle and even into the sales cycles. One of the problems we have in the United States, is that an American company is designed only to serve its shareholders, maximizing profit is the only motivation, in Europe you have different motivations. I think there is room for something called a Chief Values Officer in companies," she said.

Raja Chatila, Director of the Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, Paris discussed whether or not it makes sense to have preemptive bans and where to draw the lines. "A preemptive ban on what?" he asked. "My point is you have to be clear what you are banning. Are you banning, for example, Mary Scott of Scott Curie from discovering radium that is used every day in hospitals to cure people? Or are you banning Oppenheimer from leading the Manhattan Project? Even with the Manhattan Project and the decision to use the atom bomb, debates are still ongoing was it right or wrong many years after the event. So banning is really something that you have to reflect upon for a long time before making a decision."

"In fact," Chatila continued, "there have been discussions on autonomous weapons for the past three years in the United Nations on the Convention on unconventional weapons to decide if a ban should be pronounced or →

→ not, and the ban, if it is decided would be on very specific definitions of autonomous weapons. So I don't believe that banning preemptively is really useful. If it's applied, it will stop us — I mean humanity — from doing research or from understanding phenomenon or developing useful applications."

The next question is then how to enforce a ban if one is agreed: "There are bans, for example there is an international convention about the proliferation of nuclear weapons and if you know the story, it's quite difficult to really apply." said Chatila. "Banning is an interesting idea if we are sure we know what we are banning. Take arms. They are not banned in the United States and probably won't be for a long time, but they are banned unless specific conditions apply in most European countries. We know what that banning means: That any person cannot just buy a weapon and walk around with it. We have been living with weapons for millennia, so we know what we are talking about and we know the consequences. Therefore, banning has practical sense," he explained.

"But when we are trying to ban something that we don't understand yet – that's quite challenging. I would say it's probably impossible and certainly not desirable at the stage where we are trying to discover what we are doing," he continued.

"It's also kind of a sad irony in terms of ethics. If you ban something too fast, obviously you're making a judgment about it and your point of view may not be the same around the world from different countries," commented Havens.

"The question is always how much green washing is pushing what's actually going on in the company," said

Firth-Butterfield, prompting Havens to remark that "ethics is the new green," a comment that resonated with many at the summit.

"That's why I feel we could see ethics panels within companies, we could see chief values officers, we could see chief AI officers where all of these would help validate and to help brands say 'we are behaving ethically, and we have bought into that,' and it is good for profit margins," explained Firth-Butterfield.

"I think that it's probably the way to go and I'm encouraged by, for example, the AI partnership where you have five big AI companies saying we recognize that AI is something that we need to think about self-regulating. That's an important need.

"But I want to come back to something we mentioned earlier and that's vulnerable people. In the law, we think about vulnerable people as people with disabilities, people with children and I think it's actually really important that we do you continue on with this conversation about ethics and regulation and think particularly about our children. When I was leading the ethics advisory panel at Lucid AI. My first pick was actually someone from UNICEF and everybody asked why. But actually if this is this is about our children and the future, we should be thinking about protecting our vulnerable kids. AIs in Barbie's listening all the time, for example. We need to think more structurally about regulation on that," she said.

Juha Heikkilä, Head of Unit, Robotics and Artificial Intelligence in the European Commission's DG Connect department which has been funding robotics for approximately 12 years said that he sees huge potential in robotics and automation to do away with those "dull, dirty and dangerous jobs."

"We have conducted a Eurobarometer survey on the public attitude on robots and autonomous systems and these would of course obviously involve AI, AI-based systems and AI components. We did that twice in fact and made some interesting discoveries on people's attitudes on how they respond to robots. So, for example, eighty-nine percent of people think AI is a technology that requires careful management. Seventy percent of people think that these kinds of systems could steal their jobs. On the other hand, a very high percentage — more than seventy percent — thought robots can be very useful. Because they help people and they can be doing dull and dangerous jobs," he said.

According to Heikkilä a very mixed picture is emerging of the public mood and perceptions in this area.

"We see on the one hand, trust, and on the other hand also all a lot of skepticism about the technology. So it's not all negative, not all positive. Perhaps the most interesting finding is that, at least in some respects, familiarity tends to alleviate fears, so those who are familiar with these systems tend to have fewer fears about it — they don't have a kneejerk reaction if you like." ■

> "The question is always how much green washing is pushing what's actually going on in the company"

Session 2: Programming Human Ethics: Cui Bono?

There have been proposals to program ethical algorithms into machines, such as cars or robots. But are machines even capable of making what humans consider ethical or moral decisions? That was the question tackled by the second panel.

According to Joanna Bryson from the Department of Computer Science at University of Bath and the Center for Information Technology Policy at Princeton University there are at least three ways we can program ethical algorithms.

"One possibility is that humans explicitly program in the instructions. We just say 'in this condition, do that,' 'in this other condition, do this.' So we can set the priorities. There are still ethical concerns about that. For example, if all the people involved are male, will they overlook considerations of diversity?" she said.

> "Ethics is the new green"

"Another possibility is for systems to learn rules automatically. Even then, you can train systems to learn rules or you could do sort of statistical learning, or non-symbolic learning like the deep neural networks where we have less explicit access afterwards to see what's been learned. I don't think that's the biggest concern. We're not that good at explaining why humans do things either. I think the biggest problem there is — and this is something actually the White House has been saying a great deal — is if we learn about society, and how we do things now, we may entrench bad things as well as good things," stated Bryson.

"So machine learning off of our flawed society will replicate its flaws," she warned.

But she pointed out that is what human learning does too, and it's only okay "if it doesn't get

completely codified, and set down and used to perpetuate the errors we've already made. It's incredibly important to distinguish between what I would call normative versus descriptive."

Science is descriptive, policy isn't. "Science is about explaining and predicting. Policy governance is about saying what should happen," she explained.

"When we consider our ethics we should never think of humans strictly as means. Some people think that says we shouldn't think of humans as means [at all], but of course we're the means. We're the ones who do everything. So humans are the means. But we're not only the means, we must always be the ends. And I think that very much informs my ethics that humanity is of ultimate importance and that we have to be supported," said Bryson.

Event moderator, Havens, added to Bryson's comments by mentioning the morphology of robots: "The majority of embodied robots, 90 percent, meaning ones that are designed to look like people, look like attractive, young, white women. What a shocker!" he joked before making a serious point. "I have nothing against young, attractive, white women, but why would you build robots that only look like one type of person versus like the communities where they're going to be placed?"

Sarah Spiekermann from the Institute for Management Information Systems, Vienna University of Economics and Business, said that language is part of the problem. "The IT world has used terminology from the humanities for decades and misused those terms by promising things and values that the systems have never lived up to. Let's say one of the values in a system should be the system's transparency or the degree to which the system is maintaining a person's privacy. The first thing is that you actually need to understand by the value itself, is what is privacy or what is transparency," she said. "Half of my book is about making these distinctions clear to bring value definitions and understanding of values to computer science."

"We need to break down the values that we want to have in the system. We need to model and document the systems we are building and we have to run systems through risk assessments. But, risk assessments where we can plug in any value into a system and not just security as we already have security risk assessment types of privacy risk assessments."

"Ethical theory is not just about preventing harm or being moral. It's also about making good decisions — simple as that. And for a company to make a good decision it has to maximize the value proposition of the system. The value proposition is at the core of the business model and an ethical development lifecycle has that goal as well. It's not just preventing harm. It's also fostering the value of assistance for the service of humans," continued Spiekermann.

Asked whether machines are capable of making ethical or moral decisions or if humans always need to be in the loop, the third speaker, Corrine Cath, a PhD student at the University of Oxford and an Alan Turing Institute doctoral student replied: "The question is not whether humans should always be in the loop, because they always ARE to some degree."

"Even if you work with algorithms that come from neural networks where humans are for a large part taken out, whatever that algorithm is going to be applied to will most likely affect people," she said. "But I think thus far, we have seen that when we let algorithms just run amok, we get very bad outcomes especially for vulnerable people: Look at algorithms for sentencing in the courtroom, look at algorithms for predictive policing, even look at trying to hire new people and you have an algorithm looking at their CVs."

An algorithm is not going to flag up all the biases that are in society in its results, said Cath arguing that you need a human to recognize subtle racism or sexism or any of the other -isms at play. "I would definitely say you need a human being in the loop to be able to catch those things and call them out. And we haven't even started speaking about autonomous weapons yet! There, if you take the human out, you also get into a very grey zone of international humanitarian law."

"I see a lot is people saying 'machines can do ethics' or 'we can program this.' I'm not sure they can. I think they can approximate ethics, but humans have choice. We have agency. There are a lot of 'squishy' concepts that go into →

→making ethical decisions including our understanding of human dignity and our understanding of compassion. These are very difficult things to teach an algorithm to do and as such I would be very hesitant to outsource ethical decision-making to a machine," concluded Cath.

Mady Delvaux, Member of the European Parliament, who has taken the lead in AI and robotics issues said that she represents the normal citizen who is interested in the development of robotics and trying to understand what is going on.

"How can you explain an algorithm to a normal person? I am not so optimistic when I follow the discussions in my group. Our concern as policy makers is to make it accessible to the majority. We want transparency on the algorithms — how to get it is another matter. I don't think that technology makes people more equal, because access to technology is an enormous consideration," she said before focusing on the issues of privacy, transparency and trust.

"In the Parliament we have very passionate debates on privacy and data protection and I think, for example, care robots who have access to a very private sphere should have mechanisms that you can switch off from time to time. We want more privacy. My question to the scientific community would be, is it possible to deliver that?" said Delvaux.

"The other concern is about transparency and sometimes I'm afraid when I talk to the specialists, because how can you explain it to a normal consumer? It should be explained in a way so that someone who's not a scientist can understand what is happening with these objects."

With androids, Delvaux wants it to be clear that while some devices appear to show empathy they can never feel empathy. "We will not prevent [AI and robotics] from being on the market, but what can we do, is to make sure that at least humans have the option of saying yes we want this or we don't want this," she said.

Bryson later took up the question of moral decisions, referring to the first panel's discussion about self-driving cars. "We need to discriminate. There are two different things about applying ethical rules or having the robot

make decisions that we consider to be moral. You may have to embody moral decisions in the car, but that doesn't mean that the cars are responsible for the decisions. This is the key distinction. It's incredibly ill-advised to say that the machine is responsible for that decision. Someone programmed that machine, someone owns that machine and someone has made the decision to use that machine," she said highlighting the difficulty in ascribing liability.

In her report for the European Parliament, Delvaux suggested ascribing a legal personality for robots, at least for those robots with a higher degree of artificial intelligence and with self -learning capabilities as a possible way forward. The MEP explained that it was an idea worth exploring, but debunked the idea that she was trying to "humanize" robots.

"Post-digital age is about the demystification of technologies"

"This is far from my intention," she said. "But we asked the Commission to study and to evaluate the pros and cons of such a solution. I have to confess that I don't know. And there are so many things we don't know yet, but I believe that in the development of different scenarios it would be better if we studied the different possibilities before it is too late."

"I'm really enjoying this discussion because I think it's the first real discussion that I've been to where I feel that people are really getting to some of the issues," said Cath. "But I want people to take away AI and robots for a minute, and just replace them with the word property and look at them as property. I want people to ask questions like why are why are we as a community asking people if they can develop empathy for a robot or AI property. Why are we concerned whether people can have attachments? Whether they can feel supported? Through my research, I began to realize that these are exactly the same kinds of questions marketing executives ask people in focus groups — they want to know if you buy a product how it is going to make you feel."

Spiekermann argued that the "post-digital age is about the demystification of technologies." "A time where we know about the limits of technology, but where we work in tandem with power technologies. But a very important part of that demystification process is this idea of 'better.'

That everything that is digital is automatically better is an important fallacy. I do believe in many respects technology helps us to make better decisions. I'm not opposed to technology, but what I deeply resist is the idea that anything that is new and digital is automatically better."

Bryson was more nuanced: "There's a slippery slope tradeoff argument. I know psychologically we already are there. People already think that I'm racist because I don't think that AI is human. I mean, I get it — we're already aware that people are mistaking the contemporary AI for something that is human-like. Everything is changing so fast now that we're having trouble keeping up with it," she concluded. ■

Session 3: Social Implications – Perils & Promises of AI

"There are clearly social implications and ethical challenges arising from the use of AI in rapidly evolving sectors despite the technical capabilities that may exist," said Havens introducing the third and final panel of the day. "There are often practical challenges associated with balancing competing demands on finite capacity and resources in a context where decisions related to the suitability or prioritization of individuals to access services. For instance, health insurance and technological innovations may potentially be based on increasingly automated assessments of perceived risk factors. How do we ensure that critical decision making continues to incorporate a strong ethical dimension aligned with human values in an increasingly complex black box decision making environment?"

In other words, *quis custodiet ipsos custodes* — who watches the watchmen?

Nikolaos Mavridis, graduate of MIT and founder director of the Interactive Robots and Media Lab (IRML) asked attendees to "imagine a world where robots and AI are part of our everyday life and not only as helpers, but also as companions and friends. And then to imagine a world in which we might all be able to take part in giant intelligent entities that are made up of potentially thousands of humans and machines."

Even if such a thing was possible only for five or ten seconds a day, Mavridis says such entities might end up by far surpassing the current limits both of human as well as artificial intelligence. But he also looked at the labor implications of automation: "So for every American worker with an hourly wage of less than $20 there is an 83 percent probability of automation of his actual activity in the future. Of course all of this goes very well to the huge discussion that is taking place at the moment regarding what the real effect of automation will be on the labor force, how government will deal with this and whether we can change it."

"One point I want to touch upon is the fact that all of these questions regarding ethics have various stakeholders that are playing an important role and we don't yet have either standardized nor, I think, effective processes for being able to bring these stakeholders together. Take for example the question of robotics in warfare: You have academia, you have governments, you have international organizations, you have the military industry and you have mass media which all play a very important role. And in the middle of all this you have the citizens of the world. I would like to ask how much are the citizens of the world really involved in the kinds of possible futures that we might have regarding these technologies," said Mavridis.

"The big question is how we present ourselves, how can we decide what kind of architects we want to be, and what kind of world we want to create for the future. Remembering that this future will not only be our future but also the future of our children," he concluded.

"The challenge we are facing when we discuss artificial intelligence is the classic issue we're facing with technology impact assessments," said Paul Nemitz, Director for Fundamental Rights and Union Citizenship in DG Justice at the European Commission, before going on to elaborate on the "precautionary principle."

The problem comes when "artificial intelligence is able to do things which we as humans can't foresee," he said. Using the example of environmental law, he explained: "The precautionary principle says if two conditions are fulfilled namely, one, a chain of causality is put in motion and we don't know where it's going to lead. Second, if the condition/situation can have a huge impact on human beings — maybe even extinction, but maybe also the huge changes of the way we live — then we have the duty to invest in technology impact assessment."

"Also [we must] put ourselves into the political and emotional state to make decisions today, to make sure that these long-term negative impacts on humanity do not occur," said Nemitz.

"This is normally something where ethics is very important. But even more important is, of course, the law, which through a democratic process creates obligations — including obligations that limit the ability to roll out and use such technology. So I think that's the tipping point."

"For me as a lawyer and fundamental rights policy-maker the key question is: have we already reached the point where we have to admit and live with the fact that artificial intelligence will create its own new causalities? →

"There are clearly social implications and ethical challenges arising from the use of AI in rapidly evolving sectors despite the technical capabilities that may exist"

→ That raises a lot of issues which I'm convinced, in a democracy, legislators will eventually have to address," Nemitz said. "It's on the horizon."

Greg Adamson, Chair of IEEE's Technical Activities, Ethics, Society and Technology Initiative said we shouldn't trust a company on the simple basis that 'they know what they're doing,' but rather on the precautions they're taking around creating those technologies.

Elaborating on the idea that the outputs of machine learning are not always discernable to their creators, Adamson used the example of the steam engine. "The steam engine was commercialized around the last decade of the 18th century and at the time it was built, it was unclear how it actually worked. It was only with the development of thermodynamics around the 1850s or 1860s that scientists started to understand how the steam engine actually worked," he said.

"That may sound strange — how could you build something when you don't know how it would work? It means initially you build it using rules of thumb. So you try it out and you see if it blows up. If it blows up, you need to make the metal a bit thicker. You just keep on trying it out, but until you've actually got a theory of thermodynamics, you can't figure out how to maximize the design of a particular device, in this case the steam engine."

"So we now come to the into the area of artificial intelligence and if we look at the promises of the artificial intelligence in the 1960s, I think things haven't moved as quickly as expected. I'm not talking here about *Terminator*, or sentient artificial intelligence," said Adamson, be he added it was reasonable to expect some sort of breakthrough in the next 12 months to 40 years.

"When that breakthrough occurs, by definition, the company or that organization that achieves it won't understand why they made the breakthrough. They won't understand because if they understood they'd go and make the breakthrough now. Basically lots of different communities are trying out different things," he continued.

"So they cannot say 'trust us with this, we know how it works,' because they don't know how it works and may not know how it works for decades," said Adamson. "The suggestion here is that instead of saying we trust this large organization, we ask what precautions are being taken today by this organization as it's developing this work? Obviously commercial organizations will strongly resist the release of their algorithms. That's a truism. However, to ask them to release that evidence of the precautions is a different sort of discussion."

Aurélie Pols, data governance and privacy advocate, who is a member of the European Data Protection Supervisor's (EDPS) Ethics Advisory Group, highlighted human-machine hybrid systems and examined where the human being sits in such a framework.

"My main question, taking into account this possible future and all of the other configurations of human machine systems, is what are the roles that humans are playing in relation to AI and to robotics?"

"First of all we have the creators of artificial systems and they might be the visionaries, or designers or engineers. Then you have the approvers of such systems — people who work on funding, people who are doing planning with the ministries, etc. After that, the next stage is the interaction partners of systems, i.e. people that are interacting with a robot or with a machine and this could be either in a collaborative role or in a master-slave role or all the other possible combinations within a team," she said.

"Beyond that, you slowly start to have the trainers of artificial systems. But from the moment you have a machine that relies on machine learning to form its behavior, the past experiences that the machine has had from the person who trains it, or past datasets, will explicitly change how the machine will behave in the future," said Pols.

Nemitz, too, tackled the master-slave question in relation to humans and AI systems. "I would say the simple rule here is that a human being can never become an object. The human being is a subject, which has its own rights that we have to respect. I think in Europe it would probably be impossible to have a worker who has a sensor on his belt that sees how much the worker is moving. And if the worker is not moving enough, and therefore not working enough, the machine algorithm automatically decides you're fired. This type of model in Europe would be

illegal and, I would think, everywhere that values human rights."

"But I think the principle must be that it's human autonomy which is the precondition for human responsibility and only if we maintain this autonomy that it is that the human being decides and not the machine," he said. Although technology moves much faster than legislation, Nemitz says this isn't necessarily a bad thing — we don't want human rights or democracy principles to change as fast as technology because we need some stable constants. ■

"We have now a window of opportunity to build the right values in today's technology"

Final thoughts — John C. Havens

How will machines know what we value if we don't know ourselves?

As moderator for The IEEE AI & Ethics Summit, this is the question I asked audience members to consider as we closed the event. This is because while it's easy to assume other people (academics, scientists, manufacturers) are creating Artificial Intelligence technology we are in fact all contributing to its creation via our data and daily interactions.

And in this sense, the ethical questions that can sometimes seem esoteric become very pragmatic – how will you want your autonomous car to respond when it confronts a small child in the middle of the road? Should the vehicle kill the child to spare you as the driver if the accident won't be your fault?

Whatever your specific answer, what was clear from The IEEE AI & Ethics Summit is we must take the discussions around these issues into a realm where we are thinking of solutions to potential issues before they happen. For instance, we cannot program algorithms without ensuring they have both transparency and a level of accountability to demonstrate their lack of negative bias. And because of legislation like the GDPR we cannot allow the arbitrary tracking of personal data to continue without organizations demonstrating they're utilizing customer information in a way that accelerates trust, safety, and more accurate data.

While navigating ethical considerations is challenging, it's certainly no more difficult than evolving machines that may someday attain human level sentience. And while it may be inappropriate to mandate a particular ethical viewpoint for individuals, measuring and building to their self-identified values is a must in our algorithmic era. In this regard, "ethics is the new green" and the companies that identify and provably align their products and services with end use values will demonstrate their dedication to honoring people's ethical choices for themselves, their families, and their communities.

So how will machines know what we value if we don't know ourselves? They won't. Unless we tell them, together. ■

There's no such thing as science fiction. Just stuff we haven't made yet.

From 20,000 feet under the sea to the dark side of the moon, and everywhere in between, you'll find the work of IEEE members. It's in a thousand things our founders could only dream of in 1884.

In fact, IEEE members have taken the stuff of science fiction and made it part of every day life. That's why, whether you need to draw on the knowledge of technology's past pioneers or today's innovators, one fact is clear—you'll see that IEEE members aren't just waiting for the future, they're engineering it—and the best chapters are yet to come.

IEEE: Fostering technological innovation for the benefit of humanity.

www.ieee.org

IEEE
Advancing Technology for Humanity

IEEE

Advancing Technology for Humanity

Operations Center:
445 Hoes Lane
Piscataway, NJ 08854 USA
Phone: +1 732 981 0060
Fax: +1 732 981 966
www.ieee.org

Global Offices:
Bangalore, India
Beijing, China
Los Alamitos, CA, USA
New York, NY, USA
Solaris, Singapore
Tokyo, Japan
Vienna, Austria (opening soon)
Washington, DC, USA

Disclaimers:

i. The thoughts shared are those of the individual speakers and do not necessarily represent the opinions of IEEE. The presentations and sessions should not be considered legal or business advice or recommendations

ii. IEEE prohibits discrimination, harassment and bullying. For more information visit www.ieee.org/nondiscrimination